

Instituto de Ciencias del Seguro

MÉTODOS ESTOCÁSTICOS DE ESTIMACIÓN DE LAS PROVISIONES TÉCNICAS EN EL MARCO DE SOLVENCIA II

Irene Albarrán Lozano
Pablo Alonso González

FUNDACIÓN MAPFRE

© FUNDACIÓN MAPFRE

Prohibida la reproducción total o parcial de esta obra sin el permiso escrito del autor o de FUNDACIÓN MAPFRE

FUNDACIÓN MAPFRE no se hace responsable del contenido de esta obra, ni el hecho de publicarla implica conformidad o identificación con la opinión del autor o autores.

Prohibida la reproducción total o parcial de esta obra sin el permiso escrito del autor o del editor.

© 2010, FUNDACIÓN MAPFRE
Paseo de Recoletos 23
28004 Madrid (España)

www.fundacionmapfre.com/cienciasdelseguro
publicaciones.ics@mapfre.com

ISBN: 978-84-9844-218-2
Depósito Legal: SE-7765-2010

*A nuestra hija Míriam,
por quien todo esfuerzo es poco*

... el valor de las provisiones técnicas debe corresponder al importe que otra empresa de seguros o de reaseguros (la empresa de referencia) previsiblemente necesitaría para poder asumir y cumplir las obligaciones subyacentes de seguro y reaseguro. La cuantía de las provisiones técnicas ha de reflejar las características de la cartera de seguros subyacente.

Punto (55) del Preámbulo de la Directiva 2009/138/CE del Parlamento Europeo y del Consejo, de 25 de noviembre de 2009, sobre el seguro de vida, el acceso a la actividad de seguro y de reaseguro y su ejercicio (Solvencia II)

PRESENTACIÓN

Desde 1975, FUNDACIÓN MAPFRE desarrolla actividades de interés general para la sociedad en distintos ámbitos profesionales y culturales, así como acciones destinadas a la mejora de las condiciones económicas y sociales de las personas y sectores menos favorecidos de la sociedad. En este marco, el Instituto de Ciencias del Seguro de FUNDACIÓN MAPFRE promueve y desarrolla actividades educativas y de investigación en los campos del seguro y de la gerencia de riesgos.

En el área educativa, su actuación abarca la formación académica de postgrado y especialización, desarrollada en colaboración con la Universidad Pontificia de Salamanca, así como cursos y seminarios para profesionales, impartidos en España e Iberoamérica. Estas tareas se extienden hacia otros ámbitos geográficos mediante la colaboración con instituciones españolas e internacionales, así como a través de un programa de formación a través de Internet.

El Instituto promueve ayudas a la investigación en las áreas científicas del riesgo y del seguro y mantiene un Centro de Documentación especializado en seguros y gerencia de riesgos, que da soporte a sus actividades.

Asimismo, el Instituto promueve y elabora informes periódicos y publica libros sobre el seguro y la gerencia de riesgos, con objeto de contribuir a un mejor conocimiento de dichas materias. En algunos casos estas obras sirven como referencia para quienes se inician en el estudio o la práctica del seguro, y en otros, como fuentes de información para profundizar en materias específicas.

Dentro de estas actividades se encuadra la publicación de este libro, resultado de la “Ayuda a la investigación Riesgo y Seguro” que FUNDACIÓN MAPFRE concedió a sus autores en la convocatoria de 2008. El trabajo fue tutorizado por Jose Antonio Aventín Arroyo (MAPFRE EMPRESAS, S.A.).

Desde hace unos años, Internet es el medio por el que se desarrollan mayoritariamente nuestras actividades, ofreciendo a los usuarios de todo el mundo la posibilidad de acceder a las mismas de una manera rápida y eficaz mediante soportes Web de última generación a través de www.fundacionmapfre.com\cienciasdelseguro.

Irene Albarrán Lozano: Actuario de seguros y licenciada en Ciencias Económicas y Empresariales por la Universidad Complutense de Madrid. Doctora en Ciencias Económicas y Empresariales por la misma universidad. Profesora titular de Estadística Actuarial, actualmente vinculada al Departamento de Estadística de la Universidad Carlos III de Madrid. Ha dirigido y participado en numerosos proyectos de investigación. Es autora de varios libros y artículos científicos centrados en temas relacionados con el sector asegurador. Su actividad investigadora se centra en los aspectos cuantitativos ligados a la puesta en marcha de Solvencia II y al estudio de los efectos económicos y actuariales del Seguro de Dependencia.

Pablo Alonso González: Actuario de seguros y Economista, licenciado por la Universidad Complutense de Madrid. Es Doctor en Ciencias Económicas y Empresariales por la misma universidad. Ha prestado servicio en diversas entidades financieras en áreas de Tesorería y Control de Gestión. Autor de varios libros y artículos científicos sobre aspectos financieros y aseguradores. Actualmente es profesor del Departamento de Estadística, Estructura Económica y O.E.I. de la Universidad de Alcalá. Su actividad investigadora se centra en los aspectos cuantitativos ligados a la puesta en marcha de Solvencia II y al estudio de los efectos económicos y actuariales del Seguro de Dependencia.

ÍNDICE

PREFACIO	1
CAPITULO 1. ANTECEDENTES NORMATIVOS	5
1. Normativa previa a Solvencia II	5
1.1. La normativa anterior: Solvencia I	7
2. Elementos básicos para el cálculo de las provisiones técnicas en Solvencia II	12
3. Resultados empíricos previos: los Q/S	15
3.1. El Q/S 1	15
3.2. El Q/S 2	17
3.3. El Q/S 3	22
3.4. El Q/S 4	24
4. La directiva de Solvencia II	32
CAPITULO 2. MÉTODOS CLÁSICOS PARA EL CÁLCULO DE LAS PROVISIONES TÉCNICAS	35
1. Introducción	35
2. El método <i>Grossing-up</i>	38
2.1. Incorporación de la información de años previos	41
2.2. Estimación del peor caso posible	42
2.3. Utilización de la información del resto de filas	44
3. El método <i>Link ratio</i>	49
3.1. Estimación pesimista	50
3.2. Media simple de las <i>link ratios</i>	52
3.3. Media ponderada de las <i>link ratios</i> de años diferentes	54
4. Métodos basados en la ratio de siniestralidad	56
4.1. <i>Loss ratio</i> ingenuo	56
4.2. Proyección paso a paso del <i>Loss ratio</i> (<i>Step by step</i>)	58
5. El método <i>Chain ladder</i>	62
6. Métodos que incorporan cierta información previa	64
6.1. Método Bornhuetter-Ferguson (<i>B/F</i>)	64
6.2. Método Benktander-Hovinen (<i>BH</i>)	71
6.3. Método Cape-Cod (<i>CC</i>)	73

CAPÍTULO 3: LA DISTRIBUCIÓN LIBRE DE MACK	77
1. Fundamentos teóricos de la propuesta	77
2. Ejemplo	81
CAPÍTULO 4. INTRODUCCIÓN A LOS MODELOS LINEALES GENERALES	87
1. Concepto y finalidad	87
1.1. Representación en este formato de las distribuciones más habituales	90
1.2. Utilización del modelo anterior	94
1.3. Enlaces y complementos añadidos (<i>links y offsets</i>)	95
2. Estimación del modelo	96
2.1. Estimación por máxima verosimilitud	97
2.2. Estimación mediante el uso de métodos numéricos	100
3. Inferencia con <i>GLM</i>	104
3.1. Fundamentos teóricos	104
3.2. La ratio entre neperianos de verosimilitud	107
4. Diagnóstico del modelo	111
4.1. Estudio de los residuos	111
4.2. El fenómeno de la sobredispersión (<i>overdispersion</i>)	113
4.3. Estadísticos del modelo	114
CAPÍTULO 5. MÉTODOS BASADOS EN DISTRIBUCIONES DE PROBABILIDAD	119
1. Introducción	119
2. Modelo lognormal para datos acumulados	121
2.1. Ejemplo	127
3. Distribución para datos en incrementos	134
3.1. Modelo de Poisson	134
3.2. Modelo de Poisson con sobredispersión	137
3.3. Modelo basado en la distribución Gamma	142
3.4. Modelo basado en la distribución lognormal	147

CAPÍTULO 6. MÉTODOS DE REMUESTREO	153
1. Introducción al remuestreo o bootstrap	153
1.1. <i>Bootstrap</i> no paramétrico	153
1.2. <i>Bootstrap</i> paramétrico	154
2. Enfoques más habituales	155
2.1. England y Verrall (1999): fundamentos de la propuesta	155
2.2. England y Verrall (1999): ejemplo	159
2.3. England (2002): fundamentos de la propuesta	166
2.4. England (2002): ejemplo	167
CAPÍTULO 7. OTROS MÉTODOS ESTOCÁSTICOS PARA EL CÁLCULO DE PROVISIONES TÉCNICAS	173
1. Métodos basados en credibilidad	173
2. Modelo de Gogol (1993)	178
3. Modelo de Poisson sobredispersa con Gamma	186
BIBLIOGRAFÍA	195
NORMATIVA CONSULTADA	203
Colección “Cuadernos de la Fundación” Instituto de Ciencias del Seguro	205

PREFACIO

Hasta ahora la tarea de estimar el nivel de provisiones técnicas -reservas en la literatura científica actuarial- era un trabajo algo tedioso pero dotado de un elevado componente mecánico. En esencia, se trataba de repetir un mismo cálculo sin cuestionarse ningún aspecto más. Se utilizaba uno de los múltiples métodos disponibles para tratar de obtener el nivel esperado de provisiones que fuera capaz de hacer frente a los pagos futuros asociados a siniestros que se hubieran producido durante el ejercicio del cálculo y en los anteriores. Básicamente, se trataba de calcular un valor esperado, es decir, una media. Sin embargo, con la aparición de los nuevos esquemas de evaluación del nivel de capital basado en los riesgos soportados -Solvencia II es un claro ejemplo de ellos-, el interés se amplía hacia otros aspectos tales como intentar conocer qué nivel de posibles siniestros se pueden producir en el futuro y cómo afectarían estas situaciones sobre el nivel de reservas previamente estimado. Por estas razones, parece obligado pasar a un planteamiento en el que se incorporen elementos estocásticos que, por una parte, justifiquen el mecanismo de cálculo del nivel de reservas y que además, permitan cuantificar la incertidumbre asociada a estos mecanismos.

A pesar de la ganancia que ello supone, es necesario recordar que los algoritmos estocásticos no proporcionan soluciones en aquellos casos en los que los métodos deterministas fracasaron sino que permiten obtener una estimación de la incertidumbre ligada a los citados métodos deterministas mediante el uso de un modelo de probabilidad adecuado a la realidad que se desea medir.

En este contexto es en el que se enmarca este trabajo. El objetivo del mismo es mostrar una colección de métodos de cálculo de las provisiones técnicas en los que la aplicación de modelos basados en distintas funciones estadísticas permite obtener tanto el nivel medio esperado -lo que se conoce como *Best Estimate*- como el nivel de incertidumbre ligado a ese cálculo.

Para ello, el trabajo se estructura en dos partes muy diferenciadas. La primera de ellas recoge lo que podrían denominarse los antecedentes del cálculo estocástico de provisiones técnicas. Se aborda esta cuestión desde una doble perspectiva legal y técnica, recogida en los capítulos 1 y 2 respectivamente. El primero de ellos se centra en describir el contexto normativo en el cual se va a desarrollar la aplicación de estos métodos. La norma central que marcará su aplicación es la Directiva 2009/138/CE del Parlamento Europeo y del Consejo, de 25 de noviembre de 2009, sobre el seguro de vida, el acceso a la actividad

del seguro y de reaseguro y su ejercicio (Solvencia II). En este capítulo se describe cómo era el sistema anterior y cómo será el nuevo, pasando por la descripción de los pasos que se han dado hasta llegar a la nueva norma.

El capítulo 2 se centra en los aspectos técnicos de la estimación tradicional de las provisiones técnicas. Se describe lo que podríamos denominar métodos clásicos de cálculo, tales como el *Grossing-up*, *Link ratio* o *Chain Ladder*, entre otros.

El resto de la obra se centra en el estudio de técnicas y métodos de naturaleza estocástica que, además de estimar el nivel medio previsto de reservas, permiten evaluar el nivel de incertidumbre ligado a la estimación del volumen de provisiones. Así, en el capítulo 3 se recoge el método conocido como de la distribución libre de Mack para calcular el error asociado al cálculo. Este método puede considerarse como de transición entre los tradicionales y los basados en algún tipo de distribución pues lo que hace es calcular el error asociado al cálculo cuando se utiliza el *Chain Ladder* y no impone ningún tipo de restricción en la naturaleza estocástica de los datos.

El capítulo 4 se centra en exponer los fundamentos de los modelos lineales generalizados. Esta técnica es ampliamente utilizada tanto en las tareas de estimación de provisiones como en las de tarificación de primas. En lo que concierne a este trabajo, la presencia de este capítulo se justifica porque la estimación a realizar se basa en la suposición de un determinado comportamiento estadístico en la variable de respuesta.

El capítulo 5 aborda la estimación del nivel de reservas mediante la utilización de un cierto número de modelos, cada uno de los cuales supone que los datos siguen una determinada distribución de probabilidad. A partir de aquí y mediante el empleo de los modelos lineales generalizados se procede a obtener la cifra prevista de reservas junto con el nivel de incertidumbre ligado a ella. Se distingue entre modelos para datos acumulados y aquéllos para datos en incrementos.

El capítulo 6 está dedicado al cálculo de reservas mediante la técnica del remuestreo o *bootstrapping*. Tras pasar revista a los principales aspectos teóricos de esta técnica, se aborda el estudio de dos modelos de amplia difusión como son los de England y Verrall y la modificación del anterior, planteada por England pocos años después. A diferencia de los modelos anteriores, en este caso se hace un uso intensivo de las técnicas de simulación.

Finalmente, el capítulo 7 trata de recoger otros modelos susceptibles de ser utilizados. Son los basados en la teoría de la credibilidad, la cual goza de un amplio predicamento y una extensa utilización en el ámbito actuarial.

La exposición de todos los modelos que se presentan en esta obra se ven acompañados del correspondiente ejemplo numérico que permita facilitar la comprensión del mismo. En este sentido, y para hacer más visibles las diferencias entre unos y otros modelos, se ha utilizado siempre el mismo conjunto de datos. El esquema seguido en la exposición y ejemplos también es idéntico en todos los casos. Se estima el nivel esperado de reservas, se calcula el error asociado a dicho cálculo y por último, se desglosa el volumen previsto de reservas por ejercicios futuros, los cuales se presentan tanto en valor nominal como descontado, tal y como establece la Directiva de Solvencia II.

Para concluir, nos gustaría agradecer a D. José Antonio Aventín, tutor de este trabajo, tanto el tiempo que ha dedicado como sus comentarios e indicaciones, que han permitido acercar más la siempre densa y compleja teoría al prosaico y directo mundo de la práctica diaria. Obviamente, todos los errores que pueda contener este trabajo son responsabilidad única y exclusiva de los autores.

CAPÍTULO 1

ANTECEDENTES NORMATIVOS

1. NORMATIVA PREVIA A SOLVENCIA II

En los tiempos que corren, con tremendos sobresaltos en los mercados financieros y con desagradables sorpresas en cuanto al cumplimiento de los compromisos adquiridos por parte de las entidades financieras -entiéndase tanto bancos y cajas como aseguradoras-, la capacidad para pervivir y, por tanto, respetar las obligaciones pactadas, ha devenido en crucial. Indudablemente, para lograr este objetivo es básico evaluar los peligros, técnicamente riesgos, a los que las entidades se enfrentan.

Para cumplir con este fin, las autoridades económicas impulsaron en el pasado una serie de medidas que trataban de cuantificar los recursos que habrían de destinarse para garantizar la viabilidad de este tipo de empresas. Para el sector asegurador que opera en la Unión Europea, las primeras normas surgieron en fechas ya tan lejanas como 1973 y 1979¹, en las que se impuso la creación de un fondo de capital suficiente para absorber los resultados debidos a cambios imprevistos. A comienzos de este siglo, la normativa sobre solvencia fue modificada. Es lo que se ha denominado Solvencia I. De forma muy general, los cambios se han centrado en lo siguiente:

- Publicación de dos directivas, una para vida y otra para no vida, en las que se modifican los requisitos del margen de solvencia²,
- Aparición de una Directiva que regula la actividad aseguradora dentro de los conglomerados financieros³, la cual se añade a la de 1998 sobre grupos aseguradores⁴,

¹ Para los seguros de no vida, la Primera Directiva del Consejo 73/239/CEE y para los seguros de vida, la Primera Directiva del Consejo 79/267/CEE.

² Son la Directiva 2002/13/CE del Parlamento Europeo y del Consejo de 5 de marzo de 2002 por la que se modifica la Directiva 73/239/CEE del Consejo en lo que respecta a los requisitos del margen de solvencia de las empresas de seguros distintos del seguro de vida y la Directiva 2002/83/CE del Parlamento Europeo y del Consejo, de 5 de noviembre de 2002, sobre el seguro de vida.

³ Directiva 2005/1/CE, del Parlamento Europeo y del Consejo, de 9 de marzo de 2005, por la que se modifican las Directivas 73/239/CE, 85/611/CEE, 91/675/CEE, 92/49/CEE y 93/6/CEE del Consejo y las Directivas 94/19/CE, 98/78/CE, 2000/12/CE, 2001/34/CE, 2002/83/CE y 2002/87/CE, a fin de establecer una nueva estructura organizativa de los comités de servicios financieros.

- Publicación de la Directiva que regula el reaseguro⁵ y
- Directivas que establecen los requisitos para la reorganización y quiebras de aseguradoras -el llamado *winding-up*-⁶.

A pesar de los cambios y de ofrecer una oportuna protección a los asegurados, esta normativa supone cargas para las aseguradoras, lo cual introduce un elemento distorsionador en el funcionamiento del mercado. Dicho de otra forma: el capital exigido, o legal, no se ajusta adecuadamente al que deberían tener de acuerdo a los riesgos soportados, o capital económico. En concreto:

- a) El margen de solvencia de los aseguradores no vida depende del número de contratos y no del riesgo inherente a esos contratos.
- b) La normativa exige que las provisiones técnicas se calculen con prudencia, concepto vago y difuso que lleva a que las provisiones se doten de forma arbitraria.
- c) Los beneficios de la agrupación y diversificación de riesgos se reconocen de manera muy limitada. En concreto, los requisitos de capital sólo recogen de forma muy parcial el reaseguro y otras formas de reducción del riesgo.
- d) Los grupos de compañías no se pueden beneficiar de la diversificación que existe entre los riesgos en sus filiales.
- e) Para el negocio de vida, la interacción que existe entre las provisiones técnicas y el margen de solvencia puede ocasionar efectos perversos en los casos en los que altas dosis de prudencia en las provisiones técnicas conducen a un aumento en el margen de solvencia.
- f) El riesgo de los activos no está recogido en los requisitos de capital. En su lugar se recogen restricciones cuantitativas las cuales pueden distorsionar la formación de carteras.

Con el fin de corregir estas deficiencias, se puso en marcha el proceso de cambio que ha culminado en la Directiva 2009/138/CE del Parlamento europeo y del Consejo, de 25 de noviembre de 2009, sobre el seguro de vida, el acceso a la actividad de seguro y de reaseguro y su ejercicio (Solvencia II). Antes de exponer las principales características de la misma, se procede a hacer un repaso de la normativa y estudios previos en el ámbito objeto de este trabajo, que no es otro que el de la estimación y cálculo del nivel de las provisiones técnicas mediante la aplicación de modelos estocásticos.

⁴ Directiva 98/78/CE, del Parlamento Europeo y del Consejo, de 27 de octubre de 1998, relativa a la supervisión adicional de las empresas de seguros y de reaseguros que formen parte de grupos de seguros o de reaseguros (a partir de ahora se la denominará por sus siglas en inglés, *IGD*).

⁵ Directiva 2005/68/CE del Parlamento Europeo y del Consejo, de 16 de noviembre de 2005 sobre el reaseguro.

⁶ Directiva 2001/24/CE del Parlamento Europeo y del Consejo, de 4 de abril de 2001, relativa al saneamiento y a la liquidación de las entidades de crédito.

1.1. La normativa anterior: Solvencia I

En esencia, se trataba de un sistema de cálculo del capital exigido a partir de una serie de ratios que relacionan dichas exigencias con algún aspecto del volumen de negocio, tales como primas o siniestros, eligiéndose el caso más desfavorable. Su regulación se encuentra recogida en dos Directivas:

- Para vida, la 2002/83/EC, del Parlamento Europeo y del Consejo, de 5 de marzo, en su artículo 28 y
- Para no vida, la 73/239/EEC, en su artículo 16, modificado parcialmente por la Directiva 2002/13/CE.

Por tanto, se diferencia en su forma de cálculo según el negocio propio de cada aseguradora. De forma muy general, para cada tipo de seguro se establecen normas para calcular las cifras correspondientes al margen de solvencia obligatorio y para el fondo de garantía mínima, así como unas normas para la actualización de cantidades. En líneas generales, se trata de un sistema sencillo de poner en práctica que permite la comparación de resultados entre entidades. Sin embargo, tiene ciertas limitaciones, tales como:

- Los activos y pasivos no se valoran de acuerdo con el mercado.
- No se consideran riesgos distintos de los técnicos (primas, siniestros, reservas en no vida y CER-capital en riesgo- en vida). Por ejemplo, el riesgo asociado a la inversión no se considera a la hora de evaluar las exigencias de capital, y únicamente se aborda como una restricción a la hora de invertir los activos.
- Aspectos tales como la diversificación del riesgo o su transferencia no implican reducciones en la cantidad exigida de capital.
- La incidencia del reaseguro es muy reducida -sólo a través de la tasa de retención- y no se consideran aspectos como la calidad crediticia de los reaseguradores.

▪ Seguros de no vida

Por lo que se refiere al margen de solvencia obligatorio -a partir de ahora, MSO_{NV} - se calcula a partir de la siguiente expresión:

$$MSO_{NV} = \max\{I'_P; I'_S\}$$

en donde $I'_P = k \cdot I_P$ y $I'_S = k \cdot I_S$ siendo k una constante con valor 1,5 para los seguros de responsabilidad civil -excepto autos-, marítimos y de aviación, y con valor 1 para el resto de ramos. Por su parte I_P es el índice de primas, el cual se calcula como:

$$I_P = \begin{cases} 0,18 \cdot x \cdot r & x \leq 50 \\ 9r + 0,16(x - 50)r & x > 50 \end{cases}$$

en donde x es el volumen de primas, expresado en millones de euros, y r es la tasa de retención de primas, la cual se calcula como:

$$r = \max \left\{ 0,5; \frac{SN_t}{\frac{1}{m} \sum_{t=0}^{m-1} SB_{t-i}} \right\}$$

en donde m vale 7 para las aseguradoras que suscriban negocio de impago de crédito, tempestades, granizo o helada, y 3 en el resto de los ramos. SN_t y SB_t son los siniestros netos y brutos de reaseguro respectivamente, en el ejercicio t .

Por su parte, I_s es el índice de siniestros, que se calcula como:

$$I_s = \begin{cases} 0,26 \cdot s \cdot r & s \leq 35 \\ 9,1r + 0,23(s - 35)r & s > 35 \end{cases}$$

siendo s el importe de los siniestros, expresado en millones de euros y r la tasa de retención de siniestros, que se calcula como:

$$r = \max \left\{ 0,5; \frac{SN_t}{\frac{1}{m} \sum_{t=0}^{m-1} SB_{t-i}} \right\}$$

en donde m tiene igual interpretación que en caso de las primas.

En cuanto al Fondo de Garantía Mínima, se obtiene como el máximo entre una cantidad fija -Z- y un tercio del MSO_{NV} , es decir:

$$FGM_{NV} = \max \left\{ Z; \frac{MSO_{NV}}{3} \right\}$$

en donde Z vale 3 millones de euros si la empresa cubre todos o parte de los riesgos incluidos en los ramos de responsabilidad civil, crédito o caución, y 2 millones de euros en el resto de los casos.

- Seguros de vida

Por lo que se refiere al margen de solvencia obligatorio - MSO_V -, dependerá de cuál sea la naturaleza del seguro. Así, para los seguros distintos de los ligados

a fondos de inversión, la forma de cálculo se puede recoger en la siguiente expresión:

$$MSO_V = 0,04 \cdot PMB \cdot r_p + k \cdot CER \cdot r_C$$

PMB son las provisiones matemáticas brutas y r_p representa la tasa de retención de las citadas provisiones, la cual se calcula como:

$$r_p = \max \left\{ 0,85; \frac{PMN_t}{PMB_t} \right\}$$

en donde PMN_t son las provisiones matemáticas netas de reaseguro del ejercicio t y CER es el capital en riesgo o importe de riesgo máximo para el asegurador, que puede ser:

- En el seguro mixto, la diferencia entre la prestación por muerte y la reserva matemática ya constituida
- En seguros temporales y de accidentes, es la suma asegurada.

Por su parte, k es distinto según la duración del seguro:

- 0,001 si se trata de un seguro temporal con duración de hasta tres años,
- 0,015 si se trata de un seguro temporal con duración de entre tres y cinco años y
- 0,030 para el resto de los casos.

En cuanto a r_C , es la tasa de retención de los CER y se calcula como:

$$r_C = \max \left\{ 0,50; \frac{CERN_t}{CER_t} \right\}$$

siendo $CERN$ y CER los capitales en riesgo netos y brutos de reaseguro respectivamente.

Para las operaciones de capitalización, el MSO_V se obtiene únicamente a partir de las provisiones matemáticas, es decir:

$$MSO_V = 0,04 \cdot PMB \cdot r_p$$

El caso de los seguros ligados a los fondos de inversión presenta la siguiente casuística:

- Si la empresa asume el riesgo de inversión, entonces:

$$MSO_V = 0,04 \cdot PMB \cdot r_p$$

- Si la empresa no asume el riesgo de inversión pero el importe destinado a cubrir los gastos de gestión se fijan para un periodo superior a los cinco años, entonces:

$$MSO_V = 0,01 \cdot PMB \cdot r_p$$

- Si la empresa no asume el riesgo de inversión pero el importe destinado a cubrir los gastos de gestión se fijan para un periodo no superior a los cinco años, entonces será el 25% de los gastos de administración netos de dicha actividad correspondientes al último ejercicio presupuestario $-GAN_r$, es decir:

$$MSO_V = 0,25 \cdot GAN_t$$

- Si asume el riesgo de mortalidad, entonces:

$$MSO_V = k \cdot CER \cdot r_C$$

con k definido como en el caso de los seguros no ligados a fondos de inversión.

Finalmente, en el caso de las operaciones tontinas, el MSO_V se fijará como el 1% de sus activos $-A_r$:

$$MSO_V = 0,01 \cdot A_t$$

En el caso de este tipo de seguros vida, el Fondo de Garantía Mínima se obtiene como el máximo entre un fijo de 3 millones de euros y un tercio de MSO_V , es decir:

$$FGM_V = \max \left\{ 3; \frac{MSO_V}{3} \right\}$$

- Regla de actualización de las cantidades de FGM y umbrales

Tanto los umbrales de primas -50 millones- como de siniestros -35 millones- y la cantidad fija del FGM , tanto en vida como en no vida, están sujetos a revisión anual a partir del 20 de septiembre de 2003, de acuerdo a las modificaciones experimentadas por el índice de precios europeo en los Estados Miembros con arreglo a lo que publique Eurostat. Las cantidades se adaptarán automáticamente aumentando su importe inicial en igual cambio porcentual que dicho índice, redondeado hasta un múltiplo de 100.000 €. Si el cambio porcentual desde la última actualización es inferior al 5% no se efectuará actualización alguna.

- Ejemplo de cálculo del *MSO* y *FGM* en no vida

Supongamos una compañía que opera en alguno de los ramos de no vida, por ejemplo, en Responsabilidad civil en autos, cuyas cifras de actividad se pueden resumir en los siguientes datos:

- Nivel de primas de 120 millones de euros
- Siniestros por valor de 100 millones de euros, de los cuales, el 40% es por cargo del reaseguro.
- La siniestralidad bruta de los dos ejercicios anteriores fue de 95 y 90 millones de euros, respectivamente.

Además, se sabe que desde 2003 hasta la fecha el índice de precios europeos ha aumentado en un 10%.

Con toda esta información, se procede a calcular tanto el *MSO* como el *FGM*. En primer lugar, se calcula el valor de la tasa r . Sustituyendo en su expresión, se obtiene que:

$$r = \max \left\{ 0,5; \frac{100 \times 0,6}{\frac{1}{3}(100 + 95 + 90)} \right\} = \max \left\{ 0,5; \frac{60}{95} \right\} = \frac{12}{19} = 0,63$$

A partir de aquí, se calculan tanto el índice de primas como el de siniestros. Comenzaremos por el de primas. Dado que el índice de precios ha crecido más del 5%, se modifica al alza el umbral de 50 millones, quedando en 55, por lo que el índice I_P se calculará en el segundo tramo, obteniéndose:

$$I_P = 9,9r + 0,16(x - 55)r = 9,9 \cdot \frac{12}{19} + 0,16(120 - 55) \frac{12}{19} = 12,82$$

A continuación se calcula I_S . Igual que en el caso anterior, primeramente se actualiza el umbral, pasando de 35 a 38,5. Como la siniestralidad supera este límite, habrá que situarse en el segundo tramo obteniéndose:

$$I_S = 10,01r + 0,23(s - 38,5)r = 10,01 \frac{12}{19} + 0,23(100 - 38,5) \frac{12}{19} = 15,25$$

Como el *MSO* es el máximo de ambos índices, entonces su valor será de 15,25. Por lo que se refiere al *FSM*, habrá de ser el máximo entre una cantidad fija, actualizada de acuerdo al aumento del índice de precios y la tercera parte del *MSO*, es decir:

$$FGM_V = \max \left\{ 3,3; \frac{MSO_V}{3} \right\} = \max \left\{ 3,3; \frac{15,25}{3} \right\} = 5,08$$

2. ELEMENTOS BÁSICOS PARA EL CÁLCULO DE LAS PROVISIONES TÉCNICAS EN SOLVENCIA II

Los contratos de seguros suponen la generación de obligaciones para la aseguradora basadas en la posible ocurrencia de siniestros en el futuro. Estos compromisos tienen un elevado grado de incertidumbre tanto en su cuantía como en el instante de tiempo en que se producirán.

Hasta ahora, el enfoque tradicional para valorar las obligaciones de una compañía de seguros buscaba acotar el riesgo asociado a desviaciones adversas mediante la realización de estimaciones conservadoras. Las Directivas actuales exigen una valoración prudente pero aportan una escasa guía sobre cuánta y cómo debe ser esa prudencia. Los Estados Miembros generalmente confían en las valoraciones usadas en los informes financieros referidos a temas de prudencia, encontrándose diferentes enfoques a lo ancho de la UE y diferentes niveles de prudencia⁷ que, según un estudio sobre el sector de seguros de vida en la UE⁸, es casi imposible hacer comparaciones de solvencia con un cierto sentido entre aseguradoras de distintos países.

El enfoque tradicional a menudo fracasa al reflejar cambios en la incertidumbre subyacente asociada al compromiso objeto de estudio porque el margen requerido fluctúa en función de otras variables tales como la tasa de descuento, o simplemente porque no está cuantificado y por lo tanto, se puede usar para ocultar los impactos de las pérdidas inesperadas. Como resultado, pueden darse cambios en la fortaleza del capital de una aseguradora sin que sea evidente para el regulador y, a veces, incluso para la propia dirección de la empresa. De esta forma, se retrasa tanto la respuesta adecuada de la dirección como la intervención del regulador aumentando el riesgo de insolvencia.

Este tipo de enfoques de valoración prudente tuvo sentido en un contexto de mercados nacionales cerrados a la competencia exterior, con una regulación detallada de los productos. Pero en un entorno como el actual, más dinámico, competitivo, y con fronteras abiertas y libre circulación de capitales, las consecuencias de una respuesta tardía por la dirección es probable que se materialicen más rápidamente. Además, la ausencia de transparencia en las normas de valoración y, por tanto, en la fortaleza del capital de las aseguradoras, puede elevar sus costes de capital.

Por todo ello, es crucial contar con un enfoque que permita obtener la mejor estimación posible de sus compromisos, además de contar con un margen adicional que refleje las características reales del riesgo subyacente en los contratos de seguro. Es decir, dos son las piezas necesarias para lograr un

⁷ KPMG / European Commission (2002): *Study into the methodologies to assess the overall financial position of an insurance undertaking from the perspective of prudential supervision*, May 2002.

⁸ Mercer Oliver Wyman (2004): *Life at the end of the Tunnel*, June 2004.

valor adecuado de las provisiones técnicas. Por lo que se refiere al mejor estimador de las provisiones, -en inglés *Best Estimate Liabilities* y a partir de ahora, *MES*- es la media o valor esperado de la distribución estadística de pagos por siniestros que también:

- tenga en cuenta una tasa de descuento apropiada para los pagos futuros,
- calcule los compromisos netos de reembolsos por reaseguro,
- evite la aplicación inadecuada de valores mínimos de rescate -*surrender values floors*-,
- evalúe opciones y garantías incorporadas con su valor correcto -*fair value*- e
- incluya los compromisos contractuales en los que el asegurador tiene cierto poder discrecional sobre los beneficios. Es decir, los beneficios que un asegurado espera razonablemente obtener y todos aquéllos que un asegurado podría legalmente obtener.

Estos requisitos suponen un cambio realmente significativo respecto de la situación actual. Por ejemplo, las directivas actuales prohíben el descuento de futuros compromisos en los contratos de no vida en algunos casos y en otros deshace el efecto de cualquier descuento. Pero, cualquier estimación del valor de un flujo futuro que no tenga en cuenta el valor temporal del dinero está sesgada.

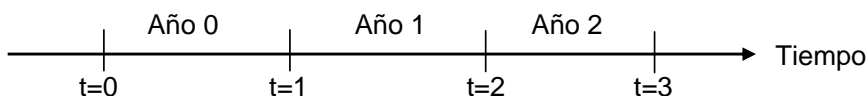
Una vez definido lo que se entiende por el *MES*, el siguiente paso consiste en la definición del margen que cuantifique la incertidumbre asociada al compromiso que se conoce como Margen de Riesgo -en inglés, *Risk Margin*, y a partir de ahora, *MAR*-. Para su cuantificación existen dos posibilidades: bien usar un margen basado en un intervalo de confianza, bien calcularlo basándose en valoraciones consistentes de mercado. La cuestión se centra en saber cómo valorar dicho margen. En este punto es crucial la diferencia entre riesgos susceptibles de cobertura y los que no lo son. Los primeros son aquéllos que es posible negociar y por tanto, valorar a precios de mercado, mientras que en el caso de los segundos, es necesario utilizar métodos indirectos de valoración. Uno de ellos, que además es el recogido en la Directiva de Solvencia II -artículo 76.5- es el conocido como enfoque del coste del capital. Este sistema fue propuesto en el Test Suizo de Solvencia -*TSS*-⁹. Es similar al propuesto por el Comité Europeo de Seguros¹⁰ y por el Foro de Supervisores del Riesgo¹¹.

⁹ Swiss Federal Office of Private Insurance (2004): *White Paper of the Swiss Solvency Test*.

¹⁰ Comité Européen des Assurances (2006): *CEA Document on Cost of Capital*, Brussels, 21 April.

¹¹ Comité Européen des Assurances and Chief Risk Officers Forum (2005): *Solutions to major issues for Solvency II*.

Para entender cómo se calcula el *MAR* en el *TSS* es necesario tener en cuenta que dicha magnitud es el coste asociado del valor actual de los futuros *SCR* -acrónimo de *Solvency Capital Requirement*- que son necesarios dotar durante el periodo en el que se va a llevar a cabo la liquidación de la cartera. Es decir, es un importe que se calcula teniendo en cuenta las circunstancias vigentes en mercado en el momento de la valoración. De ahí que se diga que es un margen consistente con el mercado o, en sus siglas en inglés, *MVM -Market Value Margin*-. Sea $t = 0$ el instante actual y $t = 1$ el comienzo del próximo ejercicio. Al intervalo de tiempo entre ambos instantes se le designará como Año 0 y así sucesivamente. Gráficamente:



El *SCR* absorberá pérdidas durante el año cero hasta un cierto nivel. Si aquéllas superasen ese nivel, la empresa pasaría a ser insolvente. Se supone que en $t = 1$ la empresa es absorbida por otra compañía, la cual deberá ser compensada por el capital adicional que deberá dotar para llevar a cabo la liquidación. Es decir, de forma muy resumida, se puede decir que en $t = 0$, cuando la empresa es solvente, el valor de los activos era igual al valor de mercado de los pasivos (*MES + MAR*)- más el *SCR*, mientras que en $t = 1$, el valor de los activos era solamente igual al valor de los pasivos.

Para poder calcular el importe asociado a *MVM* es necesario realizar algunas hipótesis:

- la empresa insolvente no sabe quién le comprará,
- a efectos del *TSS*, el *SCR* adicional que deberá dotar la empresa adquirente se calcula única y exclusivamente en función de las partidas de la empresa adquirida. Es decir, su cálculo se realiza al margen del necesario para evaluar el *SCR* de la cartera del adquirente,
- el riesgo de los activos en $t = 1$ es el mismo que en $t = 0$

Para proceder al cálculo del *MAR*, se requieren tres pasos:

- 1) Determinación del *SCR* necesario para los próximos años. Para su cálculo sólo se considera el riesgo asociado al proceso de liquidación y no a las nuevas actividades pues se supone que no se realizarán. Se puede hacer por los siguientes caminos:
 - calcular la cifra precisa de cada año usando las proyecciones de activos y pasivos o
 - suponer que el riesgo de liquidación es proporcional a la mejor estimación de las provisiones técnicas y a partir de ahí, calcular los *SCR* asociados a los próximos años.

- 2) Descuento de las futuras cifras de *SCR* usando la curva de tipos libres de riesgo para así obtener su valor actual,
- 3) Obtención del coste, sin más que multiplicar ese valor actual por el coste del capital, que en el caso del *TSS* se fijó en el 6%¹². El resultado es el *MAR*.

3. RESULTADOS EMPÍRICOS PREVIOS: LOS QIS

El camino hasta llegar al texto final de la Directiva ha sido largo y a lo largo de él se han realizado diversas pruebas de campo, los denominados *QIS* (siglas de *Quantitive Impact Study*) cuya finalidad ha sido la formulación y calibración de un modelo general de estimación del nivel de capital adecuado a los riesgos asumidos por las aseguradoras. En las primeras fases se ha partido de principios generales para, paso a paso, ir perfilando las expresiones analíticas y el valor de los parámetros incluidos en ellas. A continuación, se va a realizar un breve repaso de lo que ha supuesto cada una de las cuatro pruebas realizadas en lo que a valoración de las provisiones técnicas se refiere.

3.1. EL QIS 1

En este primer trabajo de campo se realizó una descripción muy general de cómo deberían ser las estimaciones tanto del *MES* como del *MAR* sin entrar en expresiones de tipo cuantitativo. Así, por lo que respecta al *MES*, se trataba de evaluar el valor más probable que tendrán las provisiones técnicas para lo que es necesario estimar cuál será el valor presente de los flujos asociados a todas y cada una de las pólizas. Para alcanzar estos valores esperados se supuso que se utilizarían hipótesis actuariales realistas y basadas en la experiencia. No obstante, también se deberían incorporar evoluciones previsibles de ciertas variables. Para alcanzar el valor del mejor estimador existen factores que son específicos de cada ramo y otros que son generales a todos. Dentro de estos últimos cabe señalar los siguientes:

- *Gastos*: se deberían considerar todos aquéllos en los que se pudiera incurrir en el futuro para poder hacer frente a los compromisos adquiridos.
- *Inflación*: se deberían considerar hipótesis adecuadas sobre este fenómeno. Por ejemplo, para algunos flujos la inflación relevante será la asociada al *IPC*, mientras que para otros puede ser la relativa a la evolución de los salarios.
- *Impuestos*: se deberían considerar todos los pagos relativos a aspectos fiscales necesarios para hacer frente a los compromisos adquiridos.
- *Descuento de flujos*: se usará el tipo libre de riesgo.

¹² Es el coste que estimó el *FOPI* para una compañía con un *rating* de *BBB* en la escala de Standard & Poor's.

- *Reaseguro*: se supuso que las reaseguradoras no quiebraban, por lo que se podría hacer frente a los compromisos adquiridos aún cuando existiese un desfase temporal entre el momento de la indemnización al beneficiario por parte de la aseguradora y el momento de la recepción de los fondos por ésta procedentes de la reaseguradora.
- *Calidad crediticia propia*: no se permitían reducciones en los compromisos adquiridos por este motivo.

Por lo que se refiere a los aspectos privativos de cada ramo de actividad, es necesario distinguir entre negocio de vida y el de no vida. Por lo que se refiere al negocio de vida, se tendrían en cuenta los siguientes aspectos:

- Las hipótesis de mortalidad, longevidad y enfermedad se formularían para distintos grupos de riesgo. En principio, se supone que esos riesgos eran independientes, pero si no fuera así, deberían señalarse las hipótesis y los motivos para no considerarlos de esta forma.
- Se debería tener en cuenta las opciones que los tomadores tengan para cambiar las condiciones del contrato.
- De igual forma, se debería hacer el cálculo suponiendo que los tomadores podrían rescatar las pólizas o transferirlas a otro asegurador siempre que las situaciones financieras sean tales que encuentren ventajas en ello.
- Para cada póliza se tomaría el mayor valor entre el mejor estimador y el valor actual del rescate.
- Se tendrían en cuenta acciones tomadas por la dirección y que puedan afectar a la proyección de flujos, tales como cambios en la colocación de activos o en la tasa de reparto de los beneficios.
- Por lo que se refiere a las tasas futuras de reparto de beneficios, se supuso que se utilizarían las que se apliquen en el momento de la simulación. No obstante, un aspecto a considerar es el del nivel de reservas dedicado a tal fin en el momento del cálculo.
- En cuanto a los contratos del tipo *unit-linked*, se aplicaría el mismo esquema de proyección de flujos, debiéndose considerar todas las posibles contingencias incorporadas en sus cláusulas.

Por lo que se refiere al negocio no vida, se tuvieron en cuenta los siguientes aspectos:

- Las provisiones técnicas que se incluyeran en los cálculos serían las referidas a siniestros pendientes y las asociadas a primas -no cobradas y riesgos no vencidos-. Se sugirió que, siempre que fuera posible, se deberían calcular por separado.
- Se deberían usar al menos dos métodos estadísticos, lo cual sugiere la utilización de triángulos de *run off* y se debería suministrar información sobre:

1. origen y tipo de datos usados,
2. impacto de la inflación sobre el coste total de los siniestros,

3. impacto de los grandes siniestros sobre el coste total, haciendo mención expresa sobre su posible tratamiento separado,
4. indicar cómo se han obtenido las estimaciones en el caso en el que no se usen los métodos estadísticos.

Por lo que se refiere al *MAR*, en este primer *QIS* se definió como la diferencia entre el valor obtenido para un cierto nivel de confianza y el valor medio o mejor estimador. Para obtener el margen de riesgo se deberían realizar simulaciones de tipo estocástico de las posibles sendas que puedan seguir los flujos previstos. Los márgenes requeridos eran los asociados a los percentiles 75 y 90. En todos los casos, las garantías y opciones incorporadas en los contratos se deberían valorar a precios de mercado.

3.2. El *QIS 2*

En este segundo estudio de campo el tema de la estimación de las provisiones técnicas se abordó centrándose en los siguientes aspectos:

- criterios de valoración
- clasificación de riesgos por ramos
- aspectos específicos del negocio de vida
- aspectos específicos del negocio de no vida
- requisitos especiales

3.2.1. Criterios de valoración

Los requisitos exigidos para este estudio de campo se pueden resumir en dos:

- a) valores consistentes con los de mercado para aquellos riesgos cuya cobertura está disponible en mercado -por ejemplo, para riesgos financieros- y
- b) utilización, para cualquier otro tipo de riesgos -por ejemplo, algunos riesgos actuariales- de un enfoque basado en la suma del mejor estimador que se pueda obtener más un margen de riesgo basado en el percentil 75 de la distribución asociada a la liquidación de siniestros¹³.

En cualquiera de los dos casos, las provisiones técnicas se deberían expresar tanto en términos brutos como netos de reaseguro.

El enfoque asociado al requisito b) implica la utilización del *MES* y del *MAR*. Por lo que se refiere al *MES*, hace referencia a la obtención del valor más creíble a partir de la información disponible. Para su obtención, era necesario tener en cuenta los siguientes aspectos sobre los flujos de caja que se esperan generar:

¹³ Este planteamiento se modificó en el *QIS 3*.

- a) Se debería usar el valor actual esperado de los flujos de caja futuros. En principio, se debería basar en un estudio póliza a póliza pero se pueden utilizar aproximaciones y métodos actuariales adecuados.
- b) Los flujos de caja esperados deberían basarse en hipótesis que sean realistas, las cuales deberían formularse de acuerdo con la experiencia previa pero sin olvidar la información del sector o la del mercado en los casos en los que la experiencia propia sea limitada o nula.
- c) Las proyecciones de flujos de caja deberían recoger aspectos demográficos, legales, médicos, tecnológicos, sociales o económicos. Por ejemplo, una previsible tendencia en la esperanza de vida debería recogerse en los cálculos.
- d) Los flujos de caja deberían descontarse al tipo libre de riesgo adecuado al periodo de tiempo de generación del flujo. Si ese tipo no estuviese disponible, entonces se podría interpolar o extrapolar, según el caso.
- e) Las proyecciones de flujos de caja también deberían tener en cuenta hipótesis sobre la inflación, teniendo especial cuidado en identificar el tipo de inflación al que los flujos se verán sometidos -por ejemplo, distinguiendo entre inflación salarial o de precios de consumo-.

Por tanto, para llevar a cabo una valoración correcta de activos y pasivos, se requería tener perfectamente identificados todos los posibles flujos de caja que pueden aparecer para atender los compromisos con los asegurados. Así:

- Todos aquellos gastos en los que sea necesario incurrir en el futuro deberían estar provisionados.
- Todos los costes administrativos futuros, incluyendo la gestión de inversiones, comisiones, gastos asociados a siniestros y gastos generales se debían tener en cuenta.
- También deberían considerarse todos los gastos asociados a depósitos y primas.
- Todos aquellos flujos asociados a pagos de impuestos que no han sido dotados previamente, también deberían ser tenidos en cuenta.
- Ocurre que, a veces, en el reaseguro existe una profunda diferencia entre el tiempo que se tarda en recuperar y el tiempo que se tarda en pagar, lo cual tiene su efecto en la tesorería. Sin embargo, cuando se calcule la provisión neta de reaseguro se supondrá que la reaseguradora no va a quebrar.
- No se harían reducciones de la cifra de compromisos por causas de solvencia de la compañía.

En cuanto a la estimación del *MAR*, se asoció al cálculo del percentil 75, lo cual supuso obtener un *MAR* lo suficientemente amplio como para cubrir la contingencia en el caso de que no existan instrumentos de cobertura. Ese margen se definió como la diferencia entre el valor esperado y el valor necesario para alcanzar un nivel dado de confianza que se debería calcular usando técnicas adecuadas de cálculo actuarial, como por ejemplo, simulación estocástica de cambios en los flujos de caja.

Además de los cálculos referidos al estimador y al percentil, se debería estimar el valor asociado a 0,5 veces la desviación típica de la distribución de daños para cada uno de los grupos de ramos que se indicarán en el siguiente epígrafe. Este cálculo sería coherente con lo exigido en el *Framework for Consultation* de la Comisión Europea.

Por otra parte, las garantías financieras y las opciones deberían valorarse de forma consistente con los datos de mercado. Las provisiones técnicas destinadas a cubrir ambos instrumentos se debían calcular usando tasas de descuento neutrales al riesgo sobre los datos del balance.

3.2.2. Clasificación de los riesgos por ramos

Los valores que se obtengan se agruparon según los grupos de ramos que recoge el artículo 63 de la Directiva del Consejo sobre cuentas anuales y cuentas consolidadas de compañías de seguro¹⁴, estos son:

- accidentes y enfermedad
- responsabilidad civil en autos
- otros ramos de autos
- marítimo, aéreo y transporte
- incendio y otros daños a la propiedad
- responsabilidad civil en general
- crédito y caución
- defensa jurídica
- asistencia
- varios de no vida
- reaseguro

Para el caso de vida, las agrupaciones a realizar son:

- contratos con cláusulas de participación en beneficios
- contratos en los que el asegurado soporta el riesgo de la inversión
- otros contratos sin cláusula de participación en beneficios -excepto enfermedad-
- reaseguro

Los contratos para seguros de enfermedad con características similares al negocio de vida deberían clasificarse separadamente. Obviamente, los ramos expresados no son mutuamente excluyentes, por lo que los contratos se asignarían al ramo más afín.

¹⁴ Directiva del Consejo 91/674/CEE, de 19 de diciembre, relativa a las cuentas anuales y a las cuentas consolidadas de las empresas de seguros.

Además de las cantidades calculadas para cada ramo, se calculó una cifra para el total de vida, enfermedad y no vida. Esta cifra global pudo incorporar alguna deducción debida a la diversificación, basada en el uso de técnicas actuariales correctas.

3.2.3. Aspectos específicos del negocio de vida

Los factores relevantes de riesgo que se consideraron fueron, al menos:

- tasa de mortalidad
- tasa de morbilidad
- longevidad
- tasa de suspensión *-lapse rate-*
- tasas de cambio de condiciones contractuales *-take-up option rate-*
- hipótesis sobre gastos

Las hipótesis sobre mortalidad, morbilidad y longevidad para distintos grupos se debían hacer por separado. Si se asume correlación entre los riesgos de distintos grupos, se debían hacer constar, al igual que las hipótesis que se hagan sobre la volatilidad de la mortalidad, morbilidad y longevidad.

Era importante no perder de vista las posibilidades de que el tomador cambiase las condiciones del contrato, por eso, las proyecciones de flujos de caja debían tener en cuenta la proporción de asegurados que se espera realicen estas modificaciones *-take up options-*, lo cual dependerá de las condiciones financieras vigentes en el momento en que expire esa posibilidad. También se debía considerar el impacto del cambio en las condiciones no financieras. Por ejemplo, cabía esperar que un deterioro de la salud de los asegurados tuviera un impacto en las tasas de cambio de las opciones sobre seguridad garantizada *-guaranteed insurability-*.

También se debía tener en cuenta la experiencia sobre discontinuidad en los cobros de primas. Siempre que un rescate se hubiera pagado en una de esas fases de discontinuidad, el estimador debería considerar el pago que el asegurador haría bajo el escenario que se está analizando.

Las acciones futuras de la dirección también deberían tenerse en cuenta. Las hipótesis deberían reflejar las acciones que la dirección espera llevar a cabo en las circunstancias de cada escenario, tales como cambios en la asignación de activos, cambios en los productos o en la política de bonus.

Las provisiones técnicas deberían incluir cantidades en relación con los beneficios, como los pagos garantizados, recogidos en los estatutos o de naturaleza discrecional. En concreto:

- si los beneficios difieren de un grupo de pólizas a otro, se debe tener en cuenta a la hora de hacer hipótesis sobre tasas futuras de bonus y

- se deberían tener en cuenta las restricciones emanadas de la normativa legal o de las cláusulas de reparto de beneficios en el articulado de las pólizas.

Para el caso de las cantidades en las que exista reparto de beneficios, se solicitaba que se diferenciara entre las siguientes cifras:

- total de provisiones técnicas,
- la cantidad de provisiones técnicas relacionadas con beneficios garantizados y estatutarios y
- la cantidad de provisiones técnicas relacionadas con beneficios de naturaleza discrecional.

El mismo enfoque de proyecciones de flujos de caja debería utilizarse para los *unit-linked* y similares. En este caso, se exigía considerar todos los flujos de caja que pueda generar el producto, incluidos los gastos, indemnizaciones por muerte y otros flujos que pueda recibir el asegurado. Si el asegurador tenía el derecho de aumentar las cargas, las hipótesis que sustentasen esta subida debían ser consistentes con los principios generales de gestión.

3.2.4. Aspectos específicos del negocio de no vida

En este caso, las provisiones técnicas incluirían:

- provisiones para siniestros vivos y
- provisiones para primas: provisión por primas no cobradas y por riesgos en curso.

La valoración de la provisión para siniestros vivos y para primas se debía realizar de forma separada.

Las compañías deberían usar métodos estadísticos compatibles con la mejor práctica actuarial y considerar todos los factores que pudiesen tener un impacto en la experiencia futura esperada sobre siniestros. Generalmente esto supone usar datos de siniestros tanto en el año de ocurrencia como en el año de desarrollo -triángulos *run off*-.

3.2.5. Requisitos adicionales

Se pedía que se tuviera en cuenta los siguientes aspectos:

- descuento: En los negocios de no vida, se debería estimar el valor del cambio en las provisiones técnicas –dotación- utilizando una tasa de descuento del 0%.

- rescates: Se debería calcular el impacto a pagar si todas las pólizas se rescatasen de inmediato.

3.3. El QIS 3

En este caso debería distinguirse entre riesgos susceptibles de cobertura de los que no lo son. En caso de duda, se seguiría el esquema previsto para los segundos, es decir, la suma del *MES* más el *MAR*. Como ocurriera con el *QIS 2* no se aceptan reducciones en las provisiones técnicas por motivos de calidad crediticia del asegurador. Se entendía por riesgos con cobertura aquéllos cuyos flujos pudieran ser replicados mediante instrumentos financieros negociados en mercado suficientemente profundos. Para el resto de los casos, la valoración se realizaría sumando al mejor estimador el margen de riesgo calculado según el enfoque del coste del capital.

El *MES* se definió como el valor presente esperado de los potenciales flujos de caja futuros obtenidos a partir de la información disponible y de las hipótesis utilizadas. Debería evaluarse usando, al menos, dos métodos diferentes que sean considerados apropiados, eligiéndose el más adecuado de ellos. Se entiende que el más apropiado era el que mejor recogiese la naturaleza del pasivo. Para su cálculo se debería tener en cuenta todos los flujos futuros potenciales que se pudieran producir para hacer frente a los compromisos con los tomadores. Se debería utilizar un horizonte temporal que capturase todos los flujos de caja del contrato o grupo de contratos. Si no fuera así, la aseguradora debería garantizar que este hecho no pudiera afectar significativamente a los resultados. Para la obtención del *MES* se deberían tener en cuenta los siguientes elementos:

a) Hipótesis para el cálculo:

tenían que ser realistas y basadas en la experiencia de la empresa respecto de la distribución de probabilidad de cada factor de riesgo, si bien se debía tener en cuenta datos del entorno si la experiencia propia era escasa o poco creíble. Los flujos de caja deberían recoger los desarrollos esperados de tipo legal, médico, tecnológico, social o económico. Se debería tener en cuenta la evolución prevista de la inflación.

b) Descuento de flujos:

los flujos de caja se descontarían a la tasa libre de riesgo disponible en el momento de la valoración correspondiente al plazo al cual se espera que se produzca ese flujo. Para *QIS 3*, *CEIOPS* facilitó información sobre tipos de interés del euro, dólar americano, yen japonés y franco suizo. Si se necesitara información de otras monedas se tomaría siempre el tipo de interés libre de riesgo correspondiente al vencimiento considerado. A veces se podrían utilizar los tipos de la curva swap.

c) Gastos:

se deberían incorporar en el cálculo todos los gastos en que estuviera previsto incurrir para hacer frente a los compromisos pactados con los tomadores. En concreto:

- gastos de administración futuros, comisiones, gastos por siniestros y una parte de los gastos generales. Se debería considerar la inflación que puedan sufrir estas partidas
- gastos relacionados con futuros depósitos o primas.

d) Impuestos:

todos aquéllos necesarios para cumplir con las obligaciones con los tomadores. Los cálculos también habrían de incluir aquellas situaciones en las que está aprobado pero no puesto en marcha un cambio fiscal.

e) Reaseguro:

MES se debería calcular bruto y neto de reaseguro -en este caso, se suponía que la reaseguradora no iba a quebrar-. Se debería tener en cuenta la divergencia entre los momentos del pago del siniestro y el de la recuperación de fondos procedentes del reasegurador.

f) Primas futuras de contratos vigentes:

todos aquéllos que excedían del nivel necesario para afrontar las obligaciones de un contrato no se tendrían en cuenta. Si un contrato incluyera opciones o garantías que otorgasen derechos al tomador, deberían ser tenidas en cuenta.

Por lo que se refiere al *MAR*, se propuso que se utilizara la metodología del coste de capital, si bien se aceptaban metodologías distintas para los ramos de no vida con distribuciones de probabilidad que presentasen colas largas. Se asumió un coste del capital equivalente al tipo libre de riesgo más un margen de 600 puntos básicos. Para llevar a cabo la evaluación del *MAR* se distinguiría entre el cálculo del *SCR* del primer año del correspondiente a años posteriores. Los riesgos incluidos en su cálculo eran:

- Para el año 1 los riesgos de mercado, de crédito y el de primas, si bien este último se excluiría en los sucesivos *SCR*.
- Para los años 2 y siguientes, únicamente los riesgos técnicos y operativos más el riesgo de impago por la contrapartida asociada al reaseguro.

Con respecto al SCR del primer año, el cálculo se hizo diferenciando entre distintos ramos o entre distintos grupos homogéneos de riesgo¹⁵. Así, cada SCR iba acompañado de dos subíndices: el primero hacía referencia al año y el segundo al ramo o grupo de riesgo. Las entidades pudieron usar el procedimiento que juzgaron más oportuno para dividir la cantidad obtenida por el margen de riesgo entre los distintos SCR_{ij} . No obstante, si el mecanismo era distinto al propuesto por el CEIOPS deberían indicar por qué lo encuentra más adecuado. Los sistemas propuestos por CEIOPS fueron los siguientes:

- Sistema 1: para el año 1, $SCR_{1j} / j = 1, 2, \dots, k$ se obtenía proporcionalmente a la cantidad que se calcularía debida al riesgo técnico distinto al de catástrofe - SCR_{rtdc1j} . Se usaría igual sistema para años siguientes salvo que se utilicen aproximaciones al SCR.
- Sistema 2: el SCR se calcula como la suma de las distintas cantidades asociadas a cada uno de los riesgos analizados, para lo cual se supuso que cada elemento del activo estaba asociado a un determinado riesgo.

3.4. EI QIS 4

3.4.1. Principios generales

Las provisiones técnicas debían valorarse a un nivel al cual pudieran ser transferidas a un tercero. Su valor estaría basado en lo que se denomina valor actual de salida (*current exit value*). Para llevar a cabo el proceso de valoración se debería tener en cuenta la información vigente en los mercados financieros y todos los compromisos de la compañía para con los asegurados y beneficiarios de las pólizas contratadas.

Como es preceptivo en Solvencia II, el valor de las provisiones técnicas se estimaría como la suma del MES más el MAR, los cuales deberían calcularse por separado, excepto en los riesgos que sean susceptibles de ser cubiertos con instrumentos negociados en mercados (los denominados *hedgeable risks*). En caso contrario, esos flujos se deberían estimar inter o extrapolando a partir de la información disponible en el mercado. El MES es el valor medio de los flujos de caja futuros ponderados por su probabilidad de ocurrencia y descontados al tipo libre de riesgo. El MES debería calcularse de forma bruta, es decir, sin deducir las cantidades que se pudieran recuperar bien mediante el reaseguro, bien mediante contratos con algún SPV (siglas de *Special Purpose Vehicle*). Por lo que se refiere al MAR, debía ser tal que el valor de las provisiones técnicas fuera equivalente al que debía pagar un tercero que se hiciera cargo de esos compromisos y su importe sería equivalente al SCR necesario para hacer frente a las obligaciones contraídas en la vida residual que le quede a los contratos implicados.

¹⁵ A los efectos de QIS 3, se entiende por tal al conjunto de contratos con iguales o similares características de riesgo, tales como, duración del seguro, coberturas por enfermedad, rentas, etc.

En el caso de los riesgos no susceptibles de cobertura por instrumentos negociables en mercado, el *MAR* se determinaría usando el método del coste del capital.

Una de las diferencias fundamentales del *QIS 4* es que permitía el uso de simplificaciones para obtener los valores estimados de las provisiones técnicas. Obviamente, estos métodos abreviados tanto estadísticos como actuariales, serían adecuados si satisfacían determinadas condiciones. En concreto, se estableció que tales métodos podían usarse si los contratos implicados eran de naturaleza básica o si los importes no eran elevados. En este sentido, se admitía este enfoque si, alternativamente, se cumplía lo siguiente:

- El resultado obtenido era inferior a los 50 y a los 10 millones de euros en vida y no vida respectivamente.
- Por cada ramo o grupo homogéneo de riesgo en donde se utilizase el valor del *MES* así obtenido no se debía superar el 10% del total bruto de los *MES* de cada uno por separado ni el 30% del total bruto para el conjunto de todos ellos.

Otro de los cálculos simplificados que se permitieron en el *QIS 4* era lo que se conoce como *proxies*. Estos sistemas se podían utilizar cuando existía insuficiencia en el volumen de datos disponibles para obtener el *MES* mediante algún método estadístico fiable. Esto ocurriría, por ejemplo, en situaciones en las que se suscriba algún tipo de contrato por primera vez.

3.4.2. El *MES* en el *QIS 4*

Principios generales

Como se ha indicado anteriormente, el *MES* es el valor actual esperado de los flujos futuros, ponderándose cada uno de ellos por su probabilidad de ocurrencia. Es necesario resaltar que se han de considerar todos los flujos potenciales que pudieran surgir.

Otro aspecto a tener en cuenta es que, como el *MES* debería calcularse considerando el valor presente de salida de acuerdo con las condiciones imperantes en mercado en el momento de la valoración, es perfectamente posible que la cifra obtenida fuera diferente a la que se obtendría utilizando las normas contables vigentes en cada país. Independientemente de la forma en la que había de estimarse el valor presente de los flujos futuros era necesario tener claro que el horizonte de proyección de flujos que se utilizara habría de cubrir la totalidad de la vida de la cartera objeto de estudio.

Respecto a las proyecciones de flujos de caja, deberían reflejar todos los factores previstos de índole demográfica, legal, médica, tecnológica, social o económica. En este último aspecto, un elemento crucial es el referido a las perspectivas de inflación, ya que era perfectamente posible que se deban manejar diferentes sendas en función de los contratos implicados en el cálculo.

Una vez obtenidos los flujos previstos, éstos se deberían descontar al tipo libre de riesgo correspondiente al plazo al que se prevé que se reciba el mismo. Para los flujos ligados a contratos en euros, se utilizarían los tipos de la curva *swap*. Para las monedas del *EEE* y del resto del mundo, se aplicaría una metodología semejante¹⁶. Se trataba de usar la curva *swap* (intercambio de tipo fijo contra EURIBOR 6 meses) para, a partir de ella, obtener la curva de tipos cupón cero. Dado que la mayoría de las operaciones *swap* se concentran en los plazos ligados a años completos, el problema surgía al estimar tipos bien entre plazos, bien al final de la curva. Para ello, el procedimiento seguido era el de suponer que los tipos a plazo o *forward* entre dos puntos se mantenían constantes y, a partir de ahí, obtener el tipo de contado o *spot* correspondiente a dicho plazo.

Los datos que se utilizan para obtener los tipos buscados se toman de *Bloomberg*, código CMPL, en donde se muestran tipos para plazos 1 a 10 años de forma ininterrumpida, 12, 15, 20, 25, 30, 40 y 50 años. Para obtener los tipos cupón cero se aplicó la metodología de *bootstrapping* o recurrencia, es decir, que para obtener el tipo a un determinado plazo, es necesario antes haber obtenido los anteriores. Así, si se denominan r_t y z_t a los tipos publicados y al tipo de cupón cero respectivamente del plazo t -ésimo, y asumiendo valoración a la par, entonces se tiene que, por ejemplo, si $t = 1$:

$$1 = \frac{1+r_1}{1+z_1} \Rightarrow z_1 = r_1$$

Si $t = 2$, se haría:

$$1 = \frac{r_2}{1+z_1} + \frac{1+r_2}{(1+z_2)^2} \Rightarrow z_2 = \left(\frac{1+r_2}{1 - \frac{r_2}{1+z_1}} \right)^{1/2} - 1$$

¹⁶ Dicha metodología se haya recogida en el Anexo TP1: Adoption of interest rate term structure methodology, dentro del Q/S 4 Technical Specifications (MARKT/2505/08)

Para un plazo más largo, por ejemplo, si $t = 10^{17}$:

$$1 = \frac{r_{10}}{1+z_1} + \frac{r_{10}}{(1+z_2)^2} + \dots + \frac{1+r_{10}}{(1+z_{10})^{10}} \Rightarrow z_{10} = \left(\frac{1+r_{10}}{1 - \sum_{t=1}^9 \frac{r_{10}}{(1+z_t)^t}} \right)^{1/10} - 1$$

Para los casos en los que no existe tipo swap a plazo intermedio, por ejemplo a 11 años, se aplica el supuesto de constancia de los tipos forward, es decir, suponiendo que $f_{10,11} = f_{11,12} = f_{10,12}$, entonces:

$$1 = \sum_{t=1}^{10} \frac{r_{11}}{(1+z_t)^t} + \frac{1+r_{11}}{(1+z_{11})^{11}} \Rightarrow z_{11} = \left(\frac{1+r_{11}}{1 - \sum_{t=1}^{10} \frac{r_{11}}{(1+z_t)^t}} \right)^{1/11} - 1 = \left[(1+z_{10})^{10} (1+f_{10,12}) \right]^{1/11} - 1$$

Siendo $f_{10,12} = \left[\frac{(1+z_{12})^{12}}{(1+z_{10})^{10}} \right]^{1/2} - 1$

¹⁷ No obstante lo anterior, es posible calcular todos los tipos de una sola vez mediante un planteamiento matricial del problema. Así, asumiendo valoración a la par, todo el conjunto anterior de ecuaciones se puede representar como:

$$P = C \cdot F$$

es decir:

$$\underbrace{\begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{bmatrix}}_P = \underbrace{\begin{bmatrix} 1+r_1 & 0 & \cdot & \cdot & 0 \\ r_2 & 1+r_2 & 0 & \cdot & 0 \\ \cdot & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_N & \cdot & \cdot & \cdot & 1+r_N \end{bmatrix}}_C \underbrace{\begin{bmatrix} FD_1 \\ FD_2 \\ \cdot \\ \cdot \\ FD_N \end{bmatrix}}_F$$

siendo FD_j igual a:

$$FD_j = \frac{1}{(1+r_{0j})^j}$$

Premultiplicado por C^{-1} se obtiene los factores de descuento buscados. A partir de ellos se obtienen los tipos cupón cero despejando r_{0j} en la expresión anterior:

$$r_{0j} = \left(\frac{1}{FD_j} \right)^{1/j} - 1$$

Gastos

Se debían considerar todos los gastos en los que se pudiera incurrir en el futuro para hacer frente a los compromisos adquiridos. No obstante lo anterior, se deberían tener en cuenta ciertas reglas para su correcta valoración, tales como:

- Considerar todos los gastos administrativos, incluyéndose gastos de gestión, comisiones, gastos ligados a siniestros y la parte correspondiente de gastos generales, debiéndose tener en cuenta el posible incremento que tales partidas tuvieran en el futuro.
- Los gastos ligados a la renovación de pólizas futuras: serían excluidos si tales pólizas se excluyen del cálculo del *MES*.
- Se deberían tener en cuenta los análisis de gastos elaborados por las propias empresas, futuros planes de negocio y cualquier otro dato de mercado que resulte relevante. Sin embargo, no se deberían considerar los recortes en gastos que puedan surgir por la existencia de economías de escala si aún no se hubieran materializado.

Pago de impuestos cargados a los asegurados

En algunos estados miembros se produce esta circunstancia. En ese caso, la estimación de los flujos de caja esperados a partir de los que se obtuvo el nivel de las provisiones técnicas debían incluir las obligaciones fiscales que serían traspasadas a los asegurados. En este caso, las obligaciones fiscales de la compañía irían reflejadas en otra rúbrica distinta del pasivo. Obviamente, el reconocimiento de estos derechos y obligaciones fiscales debía ser coherente con la cantidad y calendario previsto de pérdidas y ganancias gravables que se esperen obtener en el futuro.

Si al hacer el cálculo se sabía que iba a producirse alguna modificación fiscal en el futuro, debería incorporarse al cálculo aunque no se hubiera llevado aún a la práctica.

Recuperación de fondos a partir del reaseguro y de SPV

Cuando se hubieran firmado contratos de reaseguro o con algún *SPV*, la estimación de los compromisos futuros se hizo en términos brutos. Es decir, en el pasivo se estimarían los importes como si tales contratos no estuvieran firmados y, simultáneamente, aparecería en el activo de la compañía una partida que reflejaba lo que se esperaba recuperar con ellos.

Los ajustes debidos a la quiebra de la contraparte debían basarse tanto en la probabilidad de quiebra de la misma como en la pérdida media resultante de esa quiebra, conocidas como *probability of default (PD)* y *loss given default (LGD)* respectivamente.

En caso de no tener información suficiente para estimar esta última, se supondría una pérdida del 50%. También se deberían tener en cuenta la duración de los contratos.

La valoración de la probabilidad de quiebra debería tener en cuenta la posible ocurrencia de diversos escenarios. Además, si dicha probabilidad dependía del importe a pagar, entonces se debería usar la probabilidad media de quiebra, la cual se ponderaría con el producto del importe a pagar y con la probabilidad de que ese importe se pague¹⁸.

Finalmente, siempre que los importes susceptibles de recuperación estuvieran cubiertos por algún tipo de garantía o línea de crédito, entonces tanto la *PD* como la *LGD* de estas garantías sustituirían a las de la contraparte en el cálculo de la pérdida esperada.

Primas futuras ligadas a los contratos vigentes

Los flujos de caja incluidos en el *MES* sólo incorporarían los asociados a los contratos existentes y cualquier obligación de pago hacia los asegurados que de ellos se pudiera derivar. Por tanto, no se incluirían renovaciones futuras no incluidas en los contratos actualmente en vigor.

Asimismo, si un contrato incorporase opciones y garantías que otorgasen ciertos privilegios a los asegurados en el futuro, dichos elementos deberían incluirse en los cálculos.

3.4.3. El *MAR* en el Q/S 4

Se indicaba que su cálculo se realizaría utilizando la metodología del Coste del capital. Según este sistema, el *MAR* se calcularía determinando el coste de conseguir un cierto nivel de recursos propios igual al *SCR* necesario para soportar las obligaciones de pago a lo largo de la vida residual de la compañía.

¹⁸ Sea el siguiente ejemplo: supongamos que hubiera que pagarse bien 100 €, bien 10.000 €, con probabilidades del 99% y del 1% respectivamente. Entonces, el *MES* será la media ponderada de ambos importes por sus probabilidades de ocurrencia, es decir:

$$BE = 0,99 \times 100 + 0,01 \times 10.000 = 199$$

Si suponemos que el importe menor no tiene ningún problema en pagarse (*PD* = 0%), pero con el superior la reaseguradora quebrará (*PD* = 100%), entonces la probabilidad media de quiebra, *PD*, será:

$$\begin{aligned} PD &= \frac{\text{pérdida posible}}{BE} = \frac{0,99 \times 100 \times 0\% + 0,01 \times 10.000 \times 100\%}{199} \\ &= \frac{100}{199} = 0,5025 = 50,25\% \end{aligned}$$

Como, a falta de otra información, se supone que la pérdida prevista por la quiebra de la contraparte, o *LGD*, es del 50%, entonces la pérdida esperada será de:

$$\text{Pérdida} = (BE \times LGD) \times PD = (199 \times 0,50) \times 0,5025 = 49,99$$

Para ello, era necesario realizar una proyección de las obligaciones de pago hasta la extinción de la compañía y el SCR que cada año haría frente a dichas obligaciones.

Si el *MAR* se calculaba a partir de la fórmula estándar, se debería estimar como cifra neta de reaseguro. Sin embargo, si se usaba un modelo interno podía hacerse tanto en términos netos como brutos, bien de forma separada, estimando por un lado el *MAR* asociado a las provisiones técnicas y por otro, el *MAR* ligado a los importes recuperables vía reaseguro o *SPV*.

Para calcular el *MAR* se debían considerar los riesgos operacionales, técnicos respecto de los negocios existentes y de quiebra de la contraparte respecto al reaseguro cedido. Se asumiría un coste del capital del 6% y los pasos a seguir para su cálculo eran los siguientes:

- *Paso 1:*

obtención para $t = 0$ y años sucesivos del SCR asociado a cada segmento de negocio considerando únicamente los riesgos señalados anteriormente

- *Paso 2:*

multiplicación de cada SCR futuro por la tasa de coste del capital -6%- para obtener el coste por mantener esos SCR

- *Paso 3:*

descuento de las cantidades obtenidas en el paso 2 usando la curva de tipos libres de riesgo en $t = 0$. La suma de los valores descontados se corresponde con el *MAR* que se debe añadir al *MES* en $t = 0$

- *Paso 4:*

el *MAR* total será la suma de todos y cada uno de los *MAR* obtenidos en cada segmento de negocio, siguiendo en cada uno de ellos los pasos 1 a 3.

Ejemplo:

Sea una compañía que quiebra en $t = 0$. Los datos relevantes para el cálculo del *MAR* son:

- *MES* de 250

- *SCR* de 250 que se descompone en:

1. asociado al riesgo de mercado y crédito = 125
2. asociado al riesgo de operaciones del año actual = 75
3. asociado al riesgo de liquidación en un año = 50

- activos en forma de títulos públicos y acciones
- liquidación paulatina de las provisiones técnicas, de forma que el saldo vivo sigue el siguiente calendario:

t	Saldo
1	250
2	125
3	75
4	40
≥ 5	0

- Se supone una cartera tal que los flujos generados por los nuevos negocios coinciden con los de las actividades que desaparecen, es decir, es lo que se conoce como cartera estable.

Como para el cálculo a realizar no se consideran las nuevas actividades, entonces $SCR_1 = 175$, es decir, el correspondiente al instante actual menos el asociado al riesgo de operaciones del año actual. Además, se supondrá que en $t = 2$ todas las acciones se habrán canjeado por títulos públicos, por lo que la cartera de títulos se habrá convertido en lo que se conoce como cartera replicante óptima¹⁹.

Por tanto, los pasos a dar serán los siguientes:

Paso 1: cálculo de los SCR en $t = 1$ y años sucesivos

Se supone que las cantidades asociadas al riesgo de liquidación serán proporcionales al MES de las provisiones técnicas en $t = 0$. Es decir, si en el instante inicial el SCR por tal concepto es de 50 y las provisiones son de 250, se supondrá que la proporción de 1/5 se mantendrá hasta el final, por lo que resultan las siguientes cifras:

t	Provisiones técnicas	SCR
1	250	50
2	125	25
3	75	15
4	40	8
≥ 5	0	0

¹⁹ Es un concepto asociado al de casación de flujos *-cash flow matching-*. Se entiende por cartera replicante aquella que reproduce los flujos de caja esperados de los pasivos bajo escenarios de riesgo en los mercados financieros. Se compone de activos líquidos. Es aquella que mejor reproduce tales flujos, es decir, que minimiza los riesgos de mercado y de crédito. Por tanto, esta cartera inmuniza los flujos del pasivo contra cambios en los riesgos subyacentes de mercado. Si una cartera replica perfectamente los flujos de caja de otra, entonces el valor de mercado de los pasivos coincidirá exactamente con el de la cartera replicante.

SCR_1 se completa con la parte correspondiente al riesgo de mercado y crédito, es decir, agregándole 125, quedando un total de 175. A partir de $t = 2$ no es necesario agregar importe alguno por este concepto ya que se supone que los activos se han convertido a títulos públicos.

Paso 2: cálculo del coste de los SCR futuros

Suponiendo un coste del 6% resultan las siguientes cifras:

t	SCR	Coste
1	175	10,50
2	25	1,50
3	15	0,90
4	8	0,48
≥ 5	0	0

Paso 3: obtención de los valores actuales asociados a los costes

Supondremos que la curva de tipos cupón cero y los valores descontados a esos tipos son las columnas Z_t y VA_t respectivamente de la siguiente tabla:

t	Z_t	VA_t
1	1,50%	10,34
2	2,00%	1,44
3	2,25%	0,84
4	3,00%	0,43

Paso 4: obtención de la cifra final de MAR

Ello se logra sin más que sumar los valores actuales calculados en el paso 3. Para este caso, el resultado final es de 13,05.

4. LA DIRECTIVA DE SOLVENCIA II

El 17 de diciembre de 2009 apareció publicada en el Diario Oficial de la Unión Europea como Directiva 2009/138/CE del Parlamento Europeo y del Consejo, de 25 de noviembre de 2009, sobre el seguro de vida, el acceso a la actividad del seguro y de reaseguro y su ejercicio (Solvencia II).

La Directiva consta de 312 artículos y siete anexos. La parte dedicada al cálculo de las Provisiones técnicas se encuentra en la sección 2 del capítulo VI, concretamente en los artículos 76 a 86, ambos inclusive. El artículo 76 recoge las disposiciones generales. En concreto, se establece que el valor de las provisiones será el valor actual que las empresas de seguros o reaseguros tendrían que pagar si transfirieran sus obligaciones de pago a otra empresa de seguros o reaseguros (artículo 76.2). Su cálculo se realizará utilizando la información facilitada por los mercados financieros y con los datos disponibles

sobre riesgos de suscripción. De manera más precisa, se establece que las provisiones técnicas se calcularán como la suma de la mejor estimación de los pagos futuros más un margen de riesgo, calculándose cada uno de ellos por separado (artículo 77.4). Se entiende por mejor estimación la media de los flujos futuros ponderados por su probabilidad de ocurrencia, teniendo en cuenta el valor temporal del dinero (valor actual esperado de los flujos de caja futuros) mediante la aplicación de la pertinente estructura temporal de tipos de interés sin riesgo (artículo 77.2). Por su parte, el margen de riesgo será aquel importe que garantice que el valor de las provisiones técnicas sea equivalente al importe que las empresas de seguros y reaseguros previsiblemente necesitarían para poder asumir y cumplir las obligaciones de seguro y reaseguro (artículo 77.3). Será igual al coste de financiación de un volumen de fondos propios admisibles igual al capital de solvencia obligatorio necesario para asumir las obligaciones de seguro y reaseguro durante su periodo de vigencia. La tasa utilizada para determinar el coste financiero del citado importe de fondos propios admisibles será la misma para todas las entidades y será igual al tipo adicional, por encima del tipo libre de riesgo que tendría que satisfacer una aseguradora o reaseguradora por mantener un importe de fondos propios admisibles igual al capital de solvencia obligatorio necesario para asumir las obligaciones de seguro y reaseguro durante el periodo de vigencia de las mismas (artículo 77.5)

Para llevar a cabo los cálculos se habrá de tener en cuenta todos los gastos en que se incurra para hacer frente a las obligaciones, la inflación y los pagos que esté previsto realizar a los tomadores y beneficiarios de seguros, incluyendo las futuras participaciones discrecionales en beneficios (artículo 78), así como el valor de las garantías financieras y de las posibles opciones contractuales incluidas en los contratos de seguros y reaseguro (artículo 79).

Todos estos cálculos deberán hacerse para grupos homogéneos de riesgo, siendo el ramo²⁰ la unidad máxima admisible (artículo 80).

El resto de los artículos ligados al cálculo de las provisiones técnicas se refieren a cuestiones no directamente vinculadas a la forma de cálculo, aunque sí con incidencia en los resultados que se obtengan. Así:

- El artículo 81 recoge cómo tratar los importes recuperables por aseguradoras y reaseguradoras y cómo ajustar el resultado de ese cálculo para tener en cuenta las pérdidas esperadas por incumplimientos de la contraparte.
- El artículo 82 impone la exigencia de velar por la calidad de los datos que se empleen en el cálculo y de cómo actuar cuando no se disponga de un volumen suficiente de información estadística.
- El artículo 83 recoge la necesidad de contrastar periódicamente las estimaciones realizadas con los datos reales y cómo actuar ante la

²⁰ Realmente la Directiva habla de líneas de negocio

presencia continuada de desviaciones significativas entre la realidad y las estimaciones.

- El artículo 84 impone la necesidad de demostrar ante los reguladores la idoneidad de los datos y métodos empleados para el cálculo de provisiones técnicas así como la adecuación del nivel de dicha partida.
- El artículo 85 refleja la potestad de los reguladores de exigir el aumento de provisiones técnicas si las cifras calculadas no se atienen a lo exigido en los artículos 76 a 83.
- Finalmente, el artículo 86 expone las medidas de ejecución que adoptará la Comisión en temas tales como:
 - a) metodologías estadísticas y actuariales para calcular la mejor estimación
 - b) estructura temporal de tipos de interés aplicable al cálculo anterior
 - c) casos en los que las provisiones técnicas se pueden calcular como un todo o como la suma de margen de riesgo y mejor estimación
 - d) cómo calcular el margen de riesgo
 - e) ramos o agrupación de ellos en que puede dividirse la cartera para calcular el nivel de provisiones técnicas
 - f) requisitos que han de cumplir los datos que se utilicen en los cálculos
 - g) circunstancias en las que podría recurrirse a aproximaciones
 - h) técnicas utilizables para ajustar el cálculo por incumplimiento de la contrapartida, tal y como se definió en el artículo 81
 - i) uso de posibles métodos y técnicas simplificadoras para calcular las provisiones técnicas que garanticen que las metodologías actuariales y estadísticas a utilizar sean las adecuadas a la naturaleza, volumen y complejidad de los riesgos soportados por los aseguradores y reaseguradores.

CAPÍTULO 2: MÉTODOS CLÁSICOS PARA EL CÁLCULO DE LAS PROVISIONES TÉCNICAS

1. INTRODUCCIÓN

En este capítulo se pasa revista a los principales métodos tradicionales para el cálculo del valor de las provisiones técnicas para siniestros pendientes. Todos ellos se caracterizan por ser sistemas en los que, mediante un cálculo más o menos complejo basado en la información histórica recogida en forma de triángulos, se trata de estimar el valor que deberían tener las provisiones técnicas para hacer frente al valor de los siniestros futuros.

Dado que en este tipo de métodos no se realizan hipótesis alguna sobre la naturaleza estadística de la distribución subyacente, lo que se pretende es simplemente obtener el valor previsible que deberían tener las reservas, siempre en función solamente de los datos pasados de liquidación. Ésta es una gran diferencia con respecto a los métodos estocásticos objeto de estudio en este trabajo. En estos últimos se parte de una determinada distribución y, mediante diferentes métodos -los cuales se abordarán en capítulos sucesivos- se trata de estimar el valor medio junto con un intervalo de confianza en el cual se sitúa dicho valor.

Todos estos métodos se pueden clasificar en dos grandes grupos:

1) Métodos sencillos:

Los cuales, a su vez, se pueden dividir en dos bloques:

a) métodos caso a caso:

en ellos se realiza una estimación individual de la cuantía de cada siniestro en tramitación en el momento del cierre. Suelen estar basados en el juicio subjetivo de un experto y son adecuados en aquellos casos en los que se tienen pocos siniestros pero cada uno de ellos de muy elevada cuantía

b) métodos estadísticos tradicionales:

en ellos no se estudian las variaciones aleatorias presentes en los datos. En este caso, los siniestros analizados se agrupan en un

cierto número de clases homogéneas. Para aplicar este sistema es necesario estimar previamente los siguientes datos:

- el número de siniestros en el año i de la clase j , que representaremos por $n_{i,j}$
- el coste medio del siniestro ocurrido en el año i en la clase j , que representaremos por $\bar{C}_{i,j}$ y que se obtiene como:

$$\bar{C}_{i,j} = \frac{\sum_{z=1}^{n_{i,j}} C_{i,j,z}}{n_{i,j}}$$

siendo $C_{i,j,z}$ el coste individual del siniestro z -ésimo ocurrido el año i dentro de la clase j .

- la tasa de inflación acumulada, representada por α_{ij}

Con estos elementos, el valor de la provisión para siniestros pendientes -PT- se puede expresar como:

$$PT = \sum_{i=1}^k \sum_{j=1}^h \bar{C}_{i,j} n_{i,j} (1 + \alpha_{i,j})$$

El uso de esta clase de métodos suele venir justificado en los casos en los que se tenga un gran número de siniestros donde las desviaciones de unos compensen las de otros.

2) Métodos globales:

Caracterizados por lo siguiente:

- utilizan algoritmos para eliminar las influencias aleatorias en los datos
- se basan en una misma presentación de los datos: la efectuada mediante el triángulo de siniestros (también conocido como *run-off triangle*) y que puede expresarse de la siguiente manera:

	0	1	2	$k-1$	k
0	$C_{0,0}$	$C_{0,1}$	$C_{0,2}$	$C_{0,k-1}$	$C_{0,k}$
1	$C_{1,0}$	$C_{1,1}$	$C_{1,2}$	$C_{1,k-1}$	
2	$C_{2,0}$	$C_{2,1}$	$C_{2,2}$...	$C_{2,k-2}$		
...			
...				
$k-1$	$C_{k-1,0}$	$C_{k-1,1}$					
k	$C_{k,0}$						

En donde las filas hacen referencia al año de ocurrencia, origen o notificación del siniestro y las columnas al año del pago o desarrollo. En el caso del triángulo anterior, el número máximo de años de los cuales se tiene información es k y, dado que las filas comienzan en cero²¹, se debe verificar que, para todos los datos conocidos, $i + j \leq k$. En este trabajo, los números máximos de años de origen y de desarrollo se representarán por I y J respectivamente.

- pueden aplicarse sobre distintas magnitudes, siempre que utilicen el mismo formato de presentación de los datos, es decir, mediante triángulos de liquidación de siniestros. Desde este punto de vista, pueden aplicarse tanto a los importes como al número de siniestros. A su vez, el triángulo puede recoger incrementos anuales o cantidades acumuladas hasta un cierto año. Para distinguir entre una u otra, en este trabajo se representarán los incrementos por $X_{i,j}$, mientras que las cifras acumuladas se escribirán como $C_{i,j}$, por lo que
$$C_{i,h} = \sum_{j=0}^h X_{i,j} \quad \forall h = 0, 1, \dots, k$$
- pueden tener en cuenta factores exógenos muy influyentes, como es el caso de la inflación en los siniestros

Existe una amplia colección de métodos dentro de esta familia. De todos ellos, los más conocidos son, seguramente, el Chain Ladder (CL) y el *Bornhuetter-Ferguson*. En los siguientes apartados, se explicarán tanto los anteriores como otros métodos de cálculo de amplia difusión.

A partir de ahora, y en todos los casos que se analicen en este capítulo y siguientes, se van a usar los mismos datos, que son los que aparecen a continuación:

	0	1	2	3	4	5	6	7	8	9
0	242	775	1.217	1.584	1.996	2.459	2.616	2.774	3.006	3.121
1	238	852	1.522	2.395	2.762	3.051	3.507	3.780	4.187	
2	196	891	1.556	2.310	2.898	3.061	3.493	3.776		
3	210	978	1.540	2.683	2.930	3.245	3.462			
4	300	783	1.493	2.069	2.474	2.869				
5	268	919	1.530	2.132	2.684					
6	298	888	1.698	2.487						
7	243	980	2.010							
8	255	940								
9	233									

²¹ Existe literatura en la que el origen temporal se sitúa en 1. En ese caso, lo que se debe cumplir es que $i + j \leq k + 1$.

A su vez, dado que la Directiva de Solvencia II establece que las provisiones técnicas se calcularán a valor actual (artículo 77.2), se hace preciso conocer la estructura temporal de tipos de interés. En concreto, y a efectos de los cálculos que se realizan, se supondrá la siguiente curva de tipos cupón cero (el plazo se mide en años):

Plazo	Tipo	Plazo	Tipo
1	1,25%	6	3,15%
2	1,37%	7	3,67%
3	1,50%	8	3,90%
4	2,25%	9	3,95%
5	2,49%	10	4,05%

De forma general, y para simplificar tanto los cálculos como la exposición, en todos aquellos casos en los que se calcule $C_{i,\infty}$, se supondrá que dicho instante de tiempo se corresponde con un año más al máximo de años de desarrollo, es decir, si J es igual a 9, entonces el límite se alcanza en 10.

2. EL MÉTODO GROSSING-UP

Este método consiste en calcular el porcentaje del total de siniestros pagados en cada año de desarrollo. Un dato fundamental para realizar los cálculos es el de la estimación de la siniestralidad total del primer año considerado, es decir, el importe total que se deberá pagar por parte de la compañía correspondiente al primer año del cual se tengan datos. Este importe se divide en dos:

- por un lado, lo que ya se conoce por el transcurso del tiempo, que son bien los $C_{0,j}$, bien los $X_{0,j}$, $\forall j = 0, \dots, J$
- por otro, la cifra estimada del total de gasto correspondientes a los siniestros de ese año, que será un importe al menos igual a $C_{0,j}$. A esta cifra se la representa por $C_{0,\infty}$.

El método consiste en calcular la proporción que representa la cantidad acumulada hasta el año j de desarrollo, $C_{0,j}$, en relación al total del primer año de ocurrencia, $C_{0,\infty}$. Es decir, para este primer año que se toma como referencia se trata de calcular los cocientes p_j , definidos como $p_j = C_{0,j}/C_{0,\infty}$.

El supuesto implícito es que estas proporciones se mantendrán constantes sea cual sea el año de desarrollo. Dichas proporciones se aplicarán al resto de los años de ocurrencia con el objetivo de obtener las $\hat{C}_{i,\infty}$, las cuales se pueden calcular como:

$$\hat{C}_{i,\infty} = \frac{\text{último } C_{i,j} \text{ disponible}}{p_j \text{ correspondiente}}$$

Finalmente, la cifra estimada de reservas será la suma de las diferencias entre ese nivel $\hat{C}_{i,\infty}$ y la última cifra conocida de reservas, $C_{i,J-i}$, es decir:

$$PT = \sum_{i=0}^I (\hat{C}_{i,\infty} - C_{i,I-i}) \quad I = \text{máximo de filas}$$

Ejemplo:

Supongamos el triángulo de datos anteriormente expuesto y que $C_{0,\infty} = 3.320$

Como primer paso, se calculan los p_j correspondientes a cada año de desarrollo, a partir de las cifras del primer año de ocurrencia, obteniéndose que:

	año de desarrollo									
	0	1	2	3	4	5	6	7	8	9
$C_{0,j}$	242	775	1.217	1.584	1.996	2.459	2.616	2.774	3.006	3.121
$p_i(\%)$	7,29	23,34	36,66	47,71	60,12	74,07	78,80	83,55	90,54	94,01

A partir de estos p_j , se calculan los $\hat{C}_{i,\infty}$ correspondientes, cuyos valores estimados son:

año de origen	$\hat{C}_{i,\infty}$
0	3.320
1	4.624
2	4.519
3	4.394
4	3.874
5	4.464
6	5.213
7	5.483
8	4.027
9	3.197

En valor estimado de las provisiones será igual a la suma de los anteriores menos lo ya dotado (suma de las cantidades que aparecen en la última diagonal del triángulo de liquidaciones), obteniéndose una cifra de 17.346, es decir:

$$\begin{aligned}
 PT &= \sum_{i=0}^I (\hat{C}_{i,\infty} - C_{i,I-i}) = \\
 &= (3.320 - 3.121) + (4.624 - 4.187) + (4.519 - 3.776) \\
 &\quad + (4.394 - 3.462) + (3.874 - 2.869) + (4.464 - 2.684) \\
 &\quad + (5.213 - 2.487) + (5.483 - 2.010) + (4.027 - 940) \\
 &\quad + (3.197 - 233) \\
 &= 17.346
 \end{aligned}$$

Una vez estimadas las cifras finales $\hat{C}_{i,\infty}$ y los porcentajes de paso, p_j , la parte inferior derecha del triángulo se puede rellenar, obteniéndose los siguientes resultados:

	0	1	2	3	4	5	6	7	8	9	∞
0											3.320
1										4.347	4.624
2									4.092	4.248	4.519
3								3.671	3.978	4.130	4.394
4							3.052	3.237	3.507	3.641	3.874
5						3.307	3.518	3.730	4.042	4.197	4.464
6				3.134	3.861	4.107	4.355	4.720	4.900	5.213	
7			2.616	3.297	4.061	4.321	4.582	4.965	5.155	5.483	
8		1.476	1.921	2.421	2.983	3.173	3.365	3.646	3.785	4.027	
9	746	1.172	1.525	1.922	2.368	2.519	2.671	2.894	3.005	3.197	

y las dotaciones anuales se obtienen sin más que restar a cada $\hat{C}_{i,j}$ el inmediatamente anterior en el mismo año de origen. Obviamente, en el caso de la primera cifra estimada $\hat{C}_{i,i-1}$, la diferencia ha de calcularse con respecto al último dato conocido de ese año, es decir, $C_{i,i}$.

El reparto de las cifras a dotar por ejercicios se obtiene simplemente al sumar una diagonal y restarle la inmediatamente anterior. Las cifras resultantes, tanto en valor nominal como descontadas, se recogen en la siguiente tabla:

año	dotación anual	
	valor nominal	valor actual
1	3.992	3.943
2	3.414	3.323
3	2.770	2.649
4	2.175	1.990
5	1.648	1.458
6	1.174	975
7	936	727
8	692	509
9	352	248
10	192	129
a dotar	17.346	15.951
Saldo total	43.115	41.720

Este método admite diversas variantes en las cuales se trata de incorporar un mayor volumen de información. A continuación, se recogen tres de ellas.

2.1. Incorporación de la información de años previos

En este caso las p_j utilizadas no son las correspondientes al año 0, sino al resultado de una media aritmética entre las p_j de ese año y las de varios años anteriores. Una vez obtenida esta nueva proporción, el resto del método es igual que en caso general.

Ejemplo:

Supongamos que para el caso anterior, se tuviera la siguiente información sobre proporciones (expresadas en porcentajes):

	0	1	2	3	4	5	6	7	8	9
Año -1	9,50	21,12	38,00	52,24	59,15	77,68	81,04	86,57	91,05	96,88
Año -2	8,76	20,38	37,26	51,50	58,41	76,94	80,30	85,83	90,31	96,14
Año -3	8,03	19,65	36,53	50,77	57,68	76,21	79,57	85,10	89,58	95,41

Cada uno de los p_j se obtendrá como la media aritmética de esos tres más el correspondiente al ejercicio 0, es decir:

$$\bar{p}_j = \frac{1}{4} \sum_{t=0}^3 C_{-t,j} = \frac{C_{0,j} + C_{-1,j} + C_{-2,j} + C_{-3,j}}{4}$$

obteniéndose los siguientes resultados (cifras en porcentajes):

	0	1	2	3	4	5	6	7	8	9
\bar{p}_j	8,39	21,12	37,11	50,56	58,84	76,22	79,93	85,26	90,37	95,61

con lo que los $\hat{C}_{i,\infty}$ resultantes son ahora:

año de origen	$\hat{C}_{i,\infty}$
0	3.320
1	4.633
2	4.429
3	4.332
4	3.764
5	4.562
6	4.919
7	5.416
8	4.450
9	2.776

Ahora, con los nuevos $\hat{C}_{i,\infty}$ y los porcentajes p_j obtenidos para este caso, los valores correspondientes a la parte inferior derecha del triángulo son:

	0	1	2	3	4	5	6	7	8	9	∞
0											3.320
1									4.430	4.633	
2								4.002	4.234	4.429	
3							3.693	3.914	4.141	4.332	
4						3.008	3.209	3.401	3.599	3.764	
5					3.477	3.646	3.889	4.122	4.361	4.562	
6			2.895	3.187	3.750	3.932	4.194	4.446	4.703	4.919	
7		2.738	3.187	4.128	4.329	4.618	4.895	5.178	5.416		
8	1.651	2.250	2.618	3.392	3.557	3.794	4.022	4.255	4.450		
9	586	1.030	1.403	1.633	2.116	2.218	2.367	2.508	2.654	2.776	

con lo que, el valor de la reserva será igual a la suma de las cantidades así estimadas, 42.600, menos lo ya dotado, 25.769, es decir, 16.831. Operando como en el caso anterior, las dotaciones por ejercicio, tanto en valor nominal como descontado, son las recogidas en la siguiente tabla:

año	dotación anual	
	valor nominal	valor actual
1	4.032	3.982
2	3.372	3.282
3	2.723	2.604
4	2.087	1.909
5	1.592	1.408
6	1.075	892
7	875	680
8	613	451
9	341	240
10	122	82
a dotar	16.831	15.530
Saldo total	42.600	41.299

2.2. Estimación del peor caso posible

Como en el caso anterior, se utilizan las p_j del año cero y de los años anteriores. Sin embargo, la diferencia estriba en que ahora no se hace la media aritmética, sino que se escoge para cada año aquella p_j que suponga una mayor dotación de reservas, es decir, la menor posible. Por tanto, la proporción seleccionada será:

$$p_j = \min\{p_j(0), p_j(-1), p_j(-2), \dots, p_j(-m)\}$$

siendo m el número de ejercicios anteriores disponibles para el cálculo. Una vez escogida la proporción adecuada a este criterio, el resto del cálculo es exactamente igual.

Ejemplo:

Aplicando este criterio, las proporciones a aplicar serán:

	0	1	2	3	4	5	6	7	8	9
\bar{p}_j	7,29	19,65	36,53	47,71	57,68	74,07	78,80	83,55	89,58	94,01

con lo que los $\hat{C}_{i,\infty}$ resultantes son ahora:

año de origen	$\hat{C}_{i,\infty}$
0	3.320
1	4.674
2	4.519
3	4.394
4	3.874
5	4.654
6	5.213
7	5.503
8	4.785
9	3.197

Ahora, con los nuevos $\hat{C}_{i,\infty}$ y los porcentajes p_j obtenidos para este caso, los valores correspondientes a la parte inferior derecha del triángulo son:

	0	1	2	3	4	5	6	7	8	9	∞
0											3.320
1										4.394	4.674
2									4.048	4.248	4.519
3								3.671	3.936	4.130	4.394
4							3.052	3.237	3.470	3.641	3.874
5						3.447	3.667	3.888	4.168	4.375	4.654
6					3.006	3.861	4.107	4.355	4.669	4.900	5.213
7				2.625	3.174	4.076	4.336	4.598	4.929	5.173	5.503
8			1.748	2.283	2.760	3.544	3.770	3.998	4.286	4.498	4.785
9		628	1.168	1.525	1.844	2.368	2.519	2.671	2.863	3.005	3.197

con lo que, el valor de la reserva será igual a la suma de las cantidades así estimadas, 44.131, menos lo ya dotado, 25.769, es decir, 18.362. Operando como en el caso anterior, las dotaciones por ejercicio, tanto en valor nominal como descontado, son las recogidas en la siguiente tabla:

año	dotación anual	
	valor nominal	valor actual
1	4.171	4.119
2	3.627	3.529
3	2.903	2.776
4	2.326	2.128
5	1.764	1.560
6	1.220	1.013
7	996	774
8	734	541
9	428	302
10	192	129
a dotar	18.362	16.872
Saldo total	44.131	42.641

2.3. Utilización de la información del resto de filas

En este caso, se trata de ir obteniendo los $\hat{C}_{i,\infty}$, no sólo a partir de los datos de la fila del ejercicio 0, sino también considerando todas las demás filas. El mecanismo es sencillo: a partir de los p_j del año 0, a los que denominaremos $p_j(0)$ y que se obtienen como se ha visto anteriormente, se calcula el valor de $\hat{C}_{1,\infty}$. Usando esta estimación y apoyándose en los valores de C_{1j} , se obtienen los valores para $p_j(1)$. Mediante algún tipo de cálculo que implique la utilización $p_j(0)$ y $p_j(1)$ -se verán dos posibilidades- se calcula $\hat{C}_{2,\infty}$. Como antes, a partir de este valor estimado y usando los datos de C_{2j} , se obtiene los correspondientes a $p_j(2)$. A continuación, combinando $p_j(0)$, $p_j(1)$ y $p_j(2)$, se calcula $\hat{C}_{3,\infty}$ y el proceso sigue como se ha descrito anteriormente.

Existen tantas variantes como mecanismos de combinación de los $p_j(t)$ se utilicen. A continuación, se van a describir dos posibilidades: la primera en la que se obtienen los $p_j(t)$ a partir de medias aritméticas de los ya calculados, mientras que en la segunda, se toma el valor más desfavorable de los ya calculados, es decir, aquél que genere un mayor valor de la reserva.

2.3.1. Utilización de la media aritmética

Sabemos que para $t = 0$ y con un $C_{0,\infty}$ de 3.320, los p_j , a los que denotaremos por $p_j(0)$, eran:

	año de desarrollo									
	0	1	2	3	4	5	6	7	8	9
$p_j(\%)$	7,29	23,34	36,66	47,71	60,12	74,07	78,80	83,55	90,54	94,01

con lo que el valor de $\hat{C}_{1,\infty}$ se obtiene como:

$$\hat{C}_{1,\infty} = \frac{C_{1,8}}{p_8(0)} = \frac{4.187}{0,90542} = 4.624$$

A partir de este valor de $\hat{C}_{1,\infty}$, se obtienen los valores de $p_j(1)$ como:

$$p_j(1) = \frac{C_{1,j}}{\hat{C}_{1,\infty}}$$

resultando ser los siguientes:

	año de desarrollo								
	0	1	2	3	4	5	6	7	8
$p_i(\%)$	5,147	18,424	32,913	51,791	59,727	65,977	75,837	81,741	90,542

Obsérvese que ahora se estima un valor menos que antes. La razón hay que buscarla en que el proceso se inicia con el valor límite que tendrían las liquidaciones correspondientes al ejercicio 1, en vez del 0.

Como ahora se tiene información de $p_j(0)$ y $p_j(1)$, para obtener el valor de $\hat{C}_{2,\infty}$ se calcula la media aritmética entre $p_7(0)$ y $p_7(1)$:

$$p_7(2) = \frac{0,83554 + 0,81741}{2} = 0,82648$$

con lo que el valor de $\hat{C}_{2,\infty}$ será:

$$\hat{C}_{2,\infty} = \frac{C_{2,7}}{p_7(2)} = \frac{3.776}{0,82648} = 4.569$$

y, a partir de este valor, se calculan los $p_j(2)$ como

$$p_j(2) = \frac{C_{2,j}}{\hat{C}_{2,\infty}}$$

resultando ser los siguientes:

	año de desarrollo							
	0	1	2	3	4	5	6	7
$p_i(\%)$	4,290	19,502	34,057	50,560	63,430	66,998	76,453	82,648

Repitiendo este procedimiento para cada caso, se obtienen todos los p_j . La relación entera de ellos se recoge en el siguiente triángulo (cifras en porcentajes):

	0	1	2	3	4	5	6	7	8	9
0	7,29	23,34	36,66	47,71	60,12	74,07	78,80	83,55	90,54	94,01
1	5,15	18,42	32,91	51,79	59,73	65,98	75,84	81,74	90,54	
2	4,29	19,50	34,06	50,56	63,43	67,00	76,45	82,65		
3	4,67	21,76	34,26	59,70	65,19	72,20	77,03			
4	7,30	19,05	36,33	50,34	60,20	69,81				
5	6,16	21,14	35,19	49,04	61,73					
6	6,17	18,40	35,18	51,52						
7	4,22	17,04	34,94							
8	5,38	19,83								
9	5,63									

y los promedios resultantes son:

	año de desarrollo									
	0	1	2	3	4	5	6	7	8	9
$p_i(\%)$	5,63	19,83	34,94	51,52	61,73	69,81	77,03	82,65	90,54	94,01

Con lo que los $\hat{C}_{i,\infty}$ que se obtienen son:

año de origen	$\hat{C}_{i,\infty}$
0	3.320
1	4.624
2	4.569
3	4.494
4	4.110
5	4.348
6	4.827
7	5.753
8	4.740
9	4.141

La evolución de los pagos acumulados es la recogida en el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9	∞
0											3.320
1										4.347	4.624
2									4.137	4.295	4.569
3								3.715	4.069	4.225	4.494
4							3.166	3.397	3.721	3.863	4.110
5						3.035	3.349	3.593	3.937	4.087	4.348
6					2.980	3.370	3.718	3.989	4.370	4.538	4.827
7				2.964	3.551	4.016	4.431	4.754	5.208	5.408	5.753
8			1.656	2.442	2.926	3.309	3.651	3.917	4.292	4.456	4.740
9		821	1.447	2.134	2.556	2.891	3.190	3.422	3.749	3.893	4.141

Con lo que los valores asociados a las dotaciones anuales, tanto en valor nominal como descontado, son los siguientes:

año	dotación anual	
	valor nominal	valor actual
1	4.371	4.317
2	3.724	3.624
3	2.982	2.852
4	2.247	2.056
5	1.778	1.572
6	1.447	1.201
7	1.095	851
8	836	616
9	428	302
10	248	167
a dotar	19.156	17.558
Saldo total	44.925	43.327

2.3.2. Estimación a partir del caso más desfavorable

Como en el caso anterior, se parte de los valores de $C_{0,\infty}$ y de los $p_j(0)$ e igual que antes, a partir de ellos, se obtiene el valor de $\hat{C}_{1,\infty}$, el cual sirve de base para calcular los valores de $p_j(1)$.

Las diferencias con la variante anterior comienzan aquí, al intentar estimar el valor de $\hat{C}_{2,\infty}$. Partiendo de la información disponible, $p_j(0)$ y $p_j(1)$, el valor de $p_7(2)$ será aquél que genere un mayor valor en la reserva, es decir:

$$p_7(2) = \min\{p_7(0); p_7(1)\}$$

que en este caso será:

$$p_7(2) = \min\{0,8355; 0,8174\} = 0,81741$$

con lo que el valor de $\hat{C}_{2,\infty}$ será:

$$\hat{C}_{2,\infty} = \frac{C_{2,2}}{p_7(2)} = \frac{3.776}{0,8174} = 4.619$$

Partiendo de este último valor, se calculan todos los $p_j(2)$ como $p_j(2) = \frac{C_{2,j}}{\hat{C}_{2,\infty}}$, obteniéndose:

	año de desarrollo							
	0	1	2	3	4	5	6	7
$p_i(\%)$	4,243	19,288	33,684	50,006	62,734	66,263	75,615	81,741

Para calcular $\hat{C}_{3,\infty}$, se dispone de información sobre $p_6(0)$, $p_6(1)$ y $p_6(2)$. En este caso $p_6(3)$ será:

$$p_6(3) = \min\{0,78795; 0,75783; 0,75615\} = 0,75615$$

con lo que $\hat{C}_{3,\infty}$ será:

$$\hat{C}_{3,\infty} = \frac{C_{3,6}}{p_6(3)} = \frac{3.462}{0,75615} = 4.578$$

Repitiendo este procedimiento para cada caso, se obtienen todos los p_j . La relación entera de ellos se recoge en el siguiente triángulo (cifras en porcentajes):

	0	1	2	3	4	5	6	7	8	9
0	7,29	23,34	36,66	47,71	60,12	74,07	78,80	83,55	90,54	94,01
1	5,15	18,42	32,91	51,79	59,73	65,98	75,84	81,74	90,54	
2	4,24	19,29	33,68	50,01	62,73	66,26	75,61	81,74		
3	4,59	21,36	33,64	58,60	64,00	70,88	75,61			
4	6,90	18,01	34,33	47,58	56,89	65,98				
5	5,68	19,48	32,43	45,19	56,89					
6	5,42	16,14	30,86	45,19						
7	3,73	15,04	30,86							
8	4,08	15,04								
9	3,73									

y los mínimos resultantes son:

	año de desarrollo									
	0	1	2	3	4	5	6	7	8	9
$p_i(\%)$	3,73	15,04	30,86	45,19	56,89	65,98	75,61	81,74	90,54	94,01

con lo que los $\hat{C}_{i,\infty}$ resultantes son:

año de origen	$\hat{C}_{i,\infty}$
0	3.320
1	4.624
2	4.619
3	4.578
4	4.349
5	4.718
6	5.503
7	6.514
8	6.248
9	6.246

La evolución de los pagos acumulados es la recogida en el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9	∞
0											3.320
1									4.347	4.624	
2								4.183	4.343	4.619	
3							3.742	4.145	4.304	4.578	
4						3.288	3.555	3.937	4.088	4.349	
5					3.113	3.567	3.856	4.271	4.435	4.718	
6				3.131	3.631	4.161	4.498	4.983	5.173	5.503	
7			2.944	3.706	4.298	4.926	5.325	5.898	6.124	6.514	
8		1.928	2.824	3.555	4.123	4.725	5.108	5.657	5.874	6.248	
9	940	1.927	2.823	3.554	4.121	4.723	5.106	5.655	5.872	6.246	

con lo que los valores asociados a las dotaciones anuales, tanto en valor nominal como descontado son los siguientes:

año	dotación anual	
	valor nominal	valor actual
1	5.166	5.103
2	4.707	4.580
3	3.856	3.688
4	3.104	2.839
5	2.477	2.190
6	2.032	1.687
7	1.488	1.156
8	1.157	852
9	591	417
10	374	252
a dotar	24.952	22.764
Saldo total	50.721	48.533

3. EL MÉTODO *LINK RATIO*

Este método se basa en las tasas de variación de la siniestralidad en un ejercicio de ocurrencia entre un ejercicio de desarrollo y el siguiente. La citada tasa (*Link ratio*) es 1 más la anterior proporción y se expresa como:

$$R_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$$

Una vez calculadas cada una de las tasas, el elemento básico del cálculo lo constituyen los factores de proyección, que no son más que los productos de las ratios previamente calculadas. Como su nombre indica, no es más que una estimación de cuál será la siniestralidad de un ejercicio suponiendo que en el paso de un año a otro, dicha variable crecerá al ritmo fijado por la *link ratio* o el

producto de éstas, si se tratase de un periodo superior a un ejercicio. Formalmente, se puede expresar como:

$$F_k = \prod_{j=k}^h R_j$$

siendo k el ejercicio considerado, h el máximo posible de ejercicios y R_j la *link ratio* aplicada en el cálculo.

Como ocurriera en el *grossing-up*, existen diferentes modalidades de aplicación de este método. A continuación, se explicarán las más frecuentes. Para ello, se utilizarán los datos que sirven de base en todos los ejemplos, con los cuales se calculan los R_{ij} , obteniéndose los siguientes resultados:

	0	1	2	3	4	5	6	7	8	9
0	3,2025	1,5703	1,3016	1,2601	1,2320	1,0638	1,0604	1,0836	1,0383	1,0638
1	3,5798	1,7864	1,5736	1,1532	1,1046	1,1495	1,0778	1,1077		
2	4,5459	1,7464	1,4846	1,2545	1,0562	1,1411	1,0810			
3	4,6571	1,5746	1,7422	1,0921	1,1075	1,0669				
4	2,6100	1,9068	1,3858	1,1957	1,1597					
5	3,4291	1,6649	1,3935	1,2589						
6	2,9799	1,9122	1,4647							
7	4,0329	2,0510								
8	3,6863									
9	---									

En la tabla anterior, $R_{0,9}$ se obtiene como la proporción entre la siniestralidad límite del ejercicio 0 y la acumulada hasta el noveno año de desarrollo.

3.1. Estimación pesimista

En este caso, se trata de elegir la mayor ratio de cada columna, es decir, de cada año de desarrollo, para así estimar la siniestralidad futura. Por tanto:

$$R_j = \max\{R_{ij}\} \quad i = 0,1,\dots,l-j-1 \quad j = 0,1,\dots,J-1$$

En nuestro ejemplo, los valores de R_j son los siguientes:

	0	1	2	3	4	5	6	7	8	9
R_j	4,657	2,051	1,742	1,260	1,232	1,149	1,081	1,108	1,038	1,064

Una vez obtenidos los R_j , el siguiente paso consiste en calcular los factores de proyección. Aplicando la expresión de los F_k se obtienen los siguientes valores:

	0	1	2	3	4	5	6	7	8	9
F_k	39,272	8,433	4,111	2,360	1,873	1,520	1,322	1,223	1,104	1,064

Con estos factores, el valor de los $\hat{C}_{i,\infty}$ se obtiene sin más que multiplicar el último valor conocido en cada ejercicio de ocurrencia por su correspondiente factor de proyección, obteniéndose la siguiente tabla:

año de origen	$\hat{C}_{i,\infty}$
0	3.320
1	4.624
2	4.619
3	4.578
4	4.361
5	5.027
6	5.869
7	8.264
8	7.927
9	9.150

Siendo la suma de estas cantidades igual a 57.740 y la provisión total, por tanto, igual a la diferencia entre este importe y lo ya dotado, obteniéndose un valor de 31.971. La evolución en el tiempo de los gastos ligados a los siniestros se recoge en el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9	∞
0											3.320
1									4.347		4.624
2								4.183	4.343		4.619
3							3.742	4.145	4.304		4.578
4						3.298	3.565	3.949	4.100		4.361
5					3.307	3.801	4.109	4.551	4.725		5.027
6				3.134	3.861	4.438	4.797	5.314	5.517		5.869
7			3.502	4.413	5.436	6.249	6.755	7.482	7.769		8.264
8		1.928	3.359	4.233	5.214	5.994	6.479	7.177	7.452		7.927
9	1.085	2.226	3.877	4.886	6.019	6.919	7.480	8.285	8.602		9.150

A partir de estas cifras, las dotaciones anuales, tanto en valor nominal como descontado son los siguientes:

año	dotación anual	
	valor nominal	valor actual
1	6.076	6.001
2	5.811	5.655
3	5.253	5.024
4	4.030	3.687
5	3.371	2.981
6	2.617	2.173
7	1.896	1.473
8	1.575	1.160
9	792	559
10	548	369
a dotar	31.971	29.082
Saldo total	57.740	54.851

3.2. Media simple de las *link ratios*

La diferencia con el caso anterior se centra en la forma de calcular los R_j aplicables a cada ejercicio. En esta situación, cada R_j es la media aritmética de los R_{ij} de cada columna. Así, por ejemplo, R_0 es igual a:

$$R_0 = \sum_{j=0}^9 \frac{R_{0,j}}{10}$$

mientras que R_1 se calcula como:

$$R_1 = \sum_{j=0}^8 \frac{R_{1,j}}{9}$$

Por tanto, los valores numéricos de R_j en este caso son:

	0	1	2	3	4	5	6	7	8	9
R_j	3,636	1,777	1,478	1,202	1,132	1,105	1,073	1,096	1,038	1,064

Una vez obtenidos los R_j , los F_k asociados a ellos son los siguientes valores:

	0	1	2	3	4	5	6	7	8	9
F_k	18,652	5,130	2,888	1,954	1,625	1,435	1,299	1,210	1,104	1,064

Con estos factores, los valores de los $\hat{C}_{i,\infty}$ son los que se recogen en la siguiente tabla:

año de origen	$\hat{C}_{i,\infty}$
0	3.320
1	4.624
2	4.569
3	4.496
4	4.118
5	4.361
6	4.859
7	5.804
8	4.822
9	4.346

siendo la suma de estas cantidades igual a 45.319 y la provisión total es, por tanto, igual a la diferencia entre este importe y lo ya dotado, obteniéndose un valor de 19.550. La evolución en el tiempo de los gastos ligados a los siniestros se recoge en el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9	∞
0											3.320
1									4.347	4.624	
2								4.137	4.295	4.569	
3							3.715	4.070	4.226	4.496	
4						3.171	3.403	3.728	3.871	4.118	
5					3.038	3.358	3.604	3.948	4.100	4.361	
6				2.990	3.385	3.742	4.015	4.399	4.568	4.859	
7			2.971	3.572	4.044	4.470	4.796	5.255	5.456	5.804	
8		1.670	2.468	2.968	3.360	3.713	3.985	4.366	4.533	4.822	
9	847	1.505	2.224	2.675	3.028	3.347	3.591	3.935	4.085	4.346	

A partir de estas cifras, las dotaciones anuales, tanto en valor nominal como descontado son los siguientes:

año	dotación anual	
	valor nominal	valor actual
1	4.438	4.383
2	3.795	3.693
3	3.048	2.915
4	2.298	2.103
5	1.816	1.605
6	1.479	1.228
7	1.118	869
8	858	632
9	440	310
10	260	175
a dotar	19.550	17.913
Saldo total	45.319	43.682

3.3. Media ponderada de las *link ratios* de años diferentes

En este caso, el cálculo a realizar consiste en ponderar los R_{ij} de diferentes ejercicios con un determinado peso. Por ejemplo, si se considerasen únicamente el último y el penúltimo ejercicios, con pesos de 2/3 y 1/3 respectivamente, entonces tendríamos los siguientes valores para R_j :

	0	1	2	3	4	5	6	7	8	9
R_j	3,802	2,005	1,441	1,238	1,142	1,092	1,080	1,100	1,038	1,064

Una vez obtenidos los R_j , los F_k asociados a ellos son los siguientes valores:

	0	1	2	3	4	5	6	7	8	9
F_k	22,234	5,848	2,917	2,025	1,636	1,432	1,312	1,215	1,104	1,064

Con estos factores, los valores de los $\hat{C}_{i,\infty}$ son los que se recogen en la siguiente tabla:

año de origen	$\hat{C}_{i,\infty}$
0	3.320
1	4.624
2	4.586
3	4.541
4	4.108
5	4.390
6	5.035
7	5.864
8	5.497
9	5.181

siendo la suma de estas cantidades igual a 47.146 y la provisión total, por tanto, igual a la diferencia entre este importe y lo ya dotado, con lo que se obtiene un valor de 21.377. La evolución en el tiempo de los gastos ligados a los siniestros se recoge en el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9	∞
0											3.320
1										4.347	4.624
2									4.152	4.311	4.586
3								3.739	4.111	4.269	4.541
4							3.132	3.382	3.719	3.862	4.108
5						3.066	3.347	3.614	3.975	4.127	4.390
6				3.079	3.517	3.839	4.146	4.559	4.733	5.035	
7			2.896	3.585	4.095	4.470	4.828	5.309	5.512	5.864	
8		1.884	2.715	3.361	3.839	4.191	4.526	4.977	5.168	5.497	
9	886	1.776	2.559	3.168	3.618	3.950	4.266	4.691	4.870	5.181	

A partir de estas cifras, las dotaciones anuales, tanto en valor nominal como descontado son los siguientes:

año	dotación anual	
	valor nominal	valor actual
1	4.732	4.674
2	4.188	4.075
3	3.298	3.154
4	2.544	2.327
5	1.971	1.743
6	1.585	1.316
7	1.272	988
8	967	712
9	509	359
10	311	209
a dotar	21.377	19.558
Saldo total	47.146	45.327

4. MÉTODOS BASADOS EN LA RATIO DE SINIESTRALIDAD

Son aquéllos en los que el cálculo del nivel de reservas se efectúa a partir de algún mecanismo en el que intervenga la citada ratio, la cual se define como el cociente entre la siniestralidad y el nivel de primas -a partir de ahora, S y P respectivamente-. Dicha proporción también se la conoce en la literatura como *loss ratio*. Conviene aclarar que por siniestralidad se entiende la correspondiente a un cierto año de origen que se encuentre totalmente desarrollada, es decir, que se conozca el valor total que tendrá la citada variable a lo largo de todos los posibles ejercicios en los que se manifieste. Por lo que a las primas se refiere, es necesario tomar un indicativo de las mismas, por ejemplo, las devengadas en un ejercicio. Entonces, en ese caso, el nivel de las mismas vendrá formado por la agregación de las primas emitidas en un determinado ejercicio con las primas no consumidas emitidas en el ejercicio anterior, restándose finalmente el importe correspondiente a las primas no consumidas emitidas en el ejercicio en cuestión.

Existen diversos mecanismos para obtener el nivel de provisiones a partir de esta ratio. En este apartado se van a exponer los más habituales.

4.1. *Loss ratio* ingenuo

Este método se basa en suponer que la citada ratio entre S y P tiene un valor constante a lo largo del tiempo, por lo que, conociendo el nivel de primas y de la citada ratio, es posible calcular la siniestralidad asociada a cada ejercicio de origen. Por tanto, el nivel de reservas se obtendrá, año a año, sin más que restar a la cifra así calculada de siniestralidad, el volumen de gastos ya

registrados, que viene recogido en la última diagonal del triángulo de liquidaciones y que representaremos por $C_{i,t-i}$. Finalmente, el volumen total de reservas será la suma de todas y cada una de las diferencias anuales previamente calculadas.

Ejemplo:

Además de los datos recogidos en el triángulo de siniestralidad aflorada, se conoce el nivel de primas durante estos ejercicios, los cuales se recogen en la siguiente tabla:

0	1	2	3	4	5	6	7	8	9
3.905	5.640	5.416	5.069	4.657	5.352	6.332	6.250	4.837	3.660

Otro dato que se conoce es que la ratio entre siniestralidad $-S-$ y primas $-P-$ es del 85%. A partir de los valores de P y de la ratio, se estima S , como el producto de $0,85 \times P$, es decir:

0	1	2	3	4	5	6	7	8	9
3.319	4.794	4.604	4.309	3.958	4.549	5.382	5.313	4.111	3.111

Dado que en este método no se dice nada de cómo evolucionan estas cifras hasta llegar al final y que solamente se conocen las cantidades iniciales y finales, una forma de rellenar la parte inferior del triángulo consiste en calcular la tasa a la cual deberían incrementarse las cantidades de la última diagonal para llegar a $\hat{C}_{i,\infty}$. Dichas tasas se recogen en la siguiente tabla:

Ejercicio	Tasa
0	6,4%
1	7,0%
2	6,8%
3	5,6%
4	6,6%
5	9,2%
6	11,7%
7	12,9%
8	17,8%
9	29,6%

Aplicando estas tasas, la evolución en el tiempo de los gastos ligados a los siniestros se recoge en el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9	∞
0											3.319
1										4.480	4.794
2									4.034	4.309	4.604
3								3.657	3.862	4.079	4.309
4							3.060	3.263	3.480	3.712	3.958
5						2.931	3.200	3.494	3.815	4.166	4.549
6					2.777	3.101	3.462	3.866	4.317	4.820	5.382
7			2.270	2.563	2.894	3.268	3.690	4.167	4.705	5.313	
8		1.107	1.305	1.537	1.811	2.134	2.514	2.962	3.490	4.111	
9	302	391	507	657	851	1.103	1.430	1.853	2.401	3.111	

A partir de estas cifras, las dotaciones anuales, tanto en valor nominal como descontado son los siguientes:

año	dotación anual	
	valor nominal	valor actual
1	2.168	2.141
2	2.171	2.113
3	2.063	1.973
4	1.983	1.814
5	1.988	1.758
6	1.995	1.656
7	1.875	1.457
8	1.558	1.148
9	1.170	825
10	710	478
a dotar	17.681	15.362
Saldo total	43.450	41.131

4.2. Proyección paso a paso del *Loss Ratio* (Step by step)

La principal diferencia con el método anterior es que en este caso la proporción entre siniestralidad y primas no es constante, sino que se va modificando a lo largo de los años de desarrollo. Es decir, para cada uno de esos años, se calcula la citada proporción, la cual se denotará por la letra griega Λ , como el cociente entre la siniestralidad aflorada hasta ese ejercicio y la prima para un cierto año. A partir de los porcentajes que se obtengan, se trata de rellenar la parte inferior derecha del triángulo formado por todos y cada uno de los valores de Λ . Para ello, se utiliza algún método estadístico de estimación de tendencia, como por ejemplo, la regresión mínimo cuadrática.

A partir de los valores de Λ que se obtengan, se estima el valor de la siniestralidad esperada, con lo que el valor de la provisión técnica será la diferencia entre la cifra así obtenida y lo ya afluado.

A continuación, se va a exponer un ejemplo que ilustre la aplicación de este método. Se utilizarán los mismos datos que en el apartado anterior.

Paso 1: Cálculo de los Λ_{ij}

Sabiendo que por definición son:

$$\Lambda_{ij} = \frac{C_{i,j}}{P_i} \times 100$$

entonces, los valores para este caso son:

	0	1	2	3	4	5	6	7	8	9	10
0	0,062	0,198	0,312	0,406	0,511	0,630	0,670	0,710	0,770	0,799	0,850
1	0,042	0,151	0,270	0,425	0,490	0,541	0,622	0,670	0,742		
2	0,036	0,165	0,287	0,427	0,535	0,565	0,645	0,697			
3	0,041	0,193	0,304	0,529	0,578	0,640	0,683				
4	0,064	0,168	0,321	0,444	0,531	0,616					
5	0,050	0,172	0,286	0,398	0,501						
6	0,047	0,140	0,268	0,393							
7	0,039	0,157	0,322								
8	0,053	0,194									
9	0,064										

en donde se supone que la siniestralidad límite para el ejercicio 0 es equivalente a un 85% de las primas -dato de partida-

Paso 2: Cálculo de los $\Delta\Lambda_{ij}$

Este cálculo facilita la siniestralidad afluada en cada año de desarrollo y se define como:

$$\Delta\Lambda_{i,j} = \Lambda_{i,j} - \Lambda_{i,j-1}$$

Obteniéndose los siguientes valores:

	0	1	2	3	4	5	6	7	8	9	10
0	0,062	0,136	0,113	0,094	0,106	0,119	0,040	0,040	0,059	0,029	0,051
1	0,042	0,109	0,119	0,155	0,065	0,051	0,081	0,048	0,072		
2	0,036	0,128	0,123	0,139	0,109	0,030	0,080	0,052			
3	0,041	0,152	0,111	0,225	0,049	0,062	0,043				
4	0,064	0,104	0,152	0,124	0,087	0,085					
5	0,050	0,122	0,114	0,112	0,103						
6	0,047	0,093	0,128	0,125							
7	0,039	0,118	0,165								
8	0,053	0,142									
9	0,064										

Paso 3: estimación de los $\Lambda_{i,j}$ que faltan

En este paso es donde se utilizan las regresiones mínimo cuadráticas. Se impone la condición de que haya un número mínimo de datos para poder hacer la regresión. En este ejemplo, se considerará que se podrán ajustar los datos siempre y cuando existan al menos cinco datos. Por este motivo, esta vía de ajuste se podrá utilizar únicamente en los años de desarrollo 0 a 5. Desde el sexto hasta el octavo se supondrá que los $\Lambda_{i,j}$ serán iguales al promedio de los existentes y para el noveno, será igual al último valor obtenido. En este último caso, a $\Lambda_{i,9}$ se le agrega el correspondiente al paso de 9 en adelante. Con ello, los valores que definen las regresiones, los promedios y el último dato son los recogidos en la siguiente tabla:

año de desarrollo	ordenada en origen	pendiente de la recta	promedio	último dato
1	0,1261	-0,0009		
2	0,1105	0,0050		
3	0,1401	-0,0003		
4	0,0868	-0,0002		
5	0,0807	-0,0057		
6			0,0609	
7			0,0470	
8			0,0658	
9				0,0802

Con todo ello, los valores de $\Lambda_{i,j}$ que se obtienen son los recogidos en la siguiente tabla:

	0	1	2	3	4	5	6	7	8	9
0										
1										0,080
2									0,066	0,080
3								0,047	0,066	0,080
4							0,061	0,047	0,066	0,080
5						0,052	0,061	0,047	0,066	0,080
6					0,086	0,047	0,061	0,047	0,066	0,080
7				0,138	0,086	0,041	0,061	0,047	0,066	0,080
8			0,151	0,138	0,085	0,035	0,061	0,047	0,066	0,080
9	0,118	0,156	0,137	0,085	0,030	0,061	0,047	0,066	0,080	

Paso 4: cálculo de la siniestralidad estimada en cada año

Una vez que se tienen estimados los incrementos anuales en la ratio entre siniestralidad y primas, se procede a calcular la ratio global para cada año de origen, que no es más que la suma de los incrementos a lo largo de los años de desarrollo, correspondientes a un mismo año de origen. Así se consiguen los valores de Λ_i , con lo que la siniestralidad estimada será:

$$\hat{C}_{i,\infty} = \hat{\Lambda}_i \cdot P_i$$

Los resultados, por año de origen, se recogen en la siguiente tabla:

año de origen	Λ	P	$\hat{C}_{i,\infty}$
0	85,0%	3.905	3.319
1	82,3%	5.640	4.639
2	84,3%	5.416	4.567
3	87,6%	5.069	4.441
4	87,0%	4.657	4.052
5	80,8%	5.352	4.324
6	77,9%	6.332	4.934
7	84,0%	6.250	5.251
8	85,8%	4.837	4.148
9	84,4%	3.660	3.089

con lo que la siniestralidad total esperada se obtiene sin más que sumar la columna $\hat{C}_{i,\infty}$, llegándose a un resultado de 42.763

Paso 5: evolución prevista de la siniestralidad pendiente de aflorar

Considerando los porcentajes anteriormente obtenidos, se puede rellenar la parte inferior del triángulo de siniestros. Los resultados en diferencias son los siguientes:

	0	1	2	3	4	5	6	7	8	9
0										
1										452
2									356	434
3								238	333	407
4							284	219	306	374
5						280	326	252	352	429
6					543	296	386	298	417	508
7				862	535	257	381	294	411	501
8			729	666	413	171	295	228	318	388
9	433	570	503	312	109	223	172	241	294	

Paso 6: cálculo de las dotaciones por años

Como en todos los casos que se han realizado hasta ahora, se calculan las diferencias respecto del ejercicio inmediatamente anterior. Los resultados, tanto en valor nominal como descontado se recogen en la siguiente tabla:

año	dotación anual	
	valor nominal	valor actual
1	4.377	4.323
2	3.380	3.289
3	2.523	2.413
4	1.887	1.727
5	1.543	1.365
6	1.370	1.137
7	992	771
8	629	463
9	294	207
a dotar	16.994	15.694
Saldo total	21.181	19.881

5. EL MÉTODO CHAIN LADDER

Trata de estimar la proporción de cambio de un ejercicio a otro. Para ello, se realiza una media ponderada de las link ratios, en donde cada valor se pondera con la siniestralidad que le precede. Es decir, dados unos $R_{i,j}$ definidos como:

$$R_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$$

entonces, para el ejercicio j de desarrollo, la tasa de modificación de la liquidación de siniestros vendrá dada por la siguiente expresión:

$$R_j = \frac{\sum_{t=j}^{l-j-1} C_{t,j} R_{t,j}}{\sum_{t=j}^{l-j} C_{t,j}}$$

en donde, sustituyendo $R_{t,j}$ por su valor, obtenemos una expresión alternativa para calcular las proporciones buscadas:

$$R_j = \frac{\sum_{t=j}^{l-j-1} C_{t,j} R_{t,j}}{\sum_{t=j}^{l-j} C_{t,j}} = \frac{\sum_{t=j}^{l-j-1} C_{t,j} \frac{C_{t,j+1}}{C_{t,j}}}{\sum_{t=j}^{l-j} C_{t,j}} = \frac{\sum_{t=j}^{l-j-1} C_{t,j+1}}{\sum_{t=j}^{l-j} C_{t,j}}$$

A partir de estas proporciones, se calculan los factores de proyección siguiendo la forma habitual, es decir:

$$F_k = \prod_{j=k}^J R_j$$

A modo de ilustración, en el ejemplo que estamos siguiendo, R_0 se calcula como sigue:

$$R_0 = \frac{\sum_{t=0}^8 C_{t,1}}{\sum_{t=0}^8 C_{t,0}} = \frac{8.006}{2.250} = 3,5582$$

Procediendo de igual manera para el resto de los R_j , se obtienen los siguientes valores:

	0	1	2	3	4	5	6	7	8
R_j	3,558	1,778	1,484	1,195	1,124	1,107	1,074	1,097	1,038

con lo que los factores de proyección son:

	0	1	2	3	4	5	6	7	8
F_k	17,092	4,804	2,701	1,821	1,523	1,355	1,224	1,139	1,038

Con estos factores, los valores de los $\hat{C}_{i,J}$ son los que se recogen en la siguiente tabla:

año de origen	$\hat{C}_{i,J}$
1	4.347
2	4.303
3	4.238
4	3.887
5	4.089
6	4.528
7	5.429
8	4.515
9	3.982

siendo la suma de estas cantidades igual a 39.319 y la provisión total la diferencia entre este importe y lo ya dotado, obteniéndose un valor de 16.671. La evolución en el tiempo de los gastos ligados a los siniestros se recoge en el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9
0										
1										4.347
2									4.144	4.303
3								3.719	4.082	4.238
4							3.175	3.411	3.744	3.887
5					3.018	3.340	3.588	3.938	4.089	
6			2.972	3.342	3.699	3.974	4.361	4.528		
7		2.982	3.564	4.007	4.435	4.765	5.229	5.429		
8	1.672	2.480	2.964	3.333	3.689	3.963	4.349	4.515		
9	829	1.474	2.187	2.614	2.939	3.253	3.495	3.836	3.982	

A partir de estas cifras, las dotaciones anuales, tanto en valor nominal como descontado son los siguientes:

año	dotación anual	
	valor nominal	valor actual
1	4.211	4.159
2	3.485	3.391
3	2.734	2.615
4	1.991	1.822
5	1.549	1.369
6	1.219	1.012
7	828	643
8	507	373
9	147	104
a dotar	16.671	15.488
Saldo total	42.440	41.257

6. MÉTODOS QUE INCORPORAN CIERTA INFORMACIÓN PREVIA

6.1. Método *Bornhuetter-Ferguson (B/F)*

Este método fue enunciado por primera vez en el trabajo de los autores de los que toma su nombre en 1972. En líneas generales, la siniestralidad se divide en dos grandes bloques:

- por un lado, la ya aflorada, la cual se recoge en la última diagonal del triángulo de siniestros, es decir, son los valores $C_{i,i}$
- por otro, la pendiente de aflorar, de la cual, se tiene información previa que permite conocer su volumen.

La expresión general para la estimación de las cifras futuras puede enunciarse como:

$$\hat{C}_{i,j+k} = C_{i,j} + (\beta_{j+k} - \beta_j) \hat{C}_{i,\infty}$$

Siendo $\beta_j = \prod_{k=1}^{j-1} f_k^{-1}$ en donde f_k^{-1} son los factores de paso de un ejercicio a otro, en concreto, el correspondiente al ejercicio k . A su vez, este método se puede formular como una mezcla del *loss ratio* ingenuo y de métodos elementales de proyección, como son el *grossing up* o el *link ratio*.

La estimación del porcentaje se realiza aplicando los métodos de proyección arriba indicados -*grossing up* o *link ratio*-, de la siguiente forma:

- si se hace a partir del *grossing up*, entonces, $Z = 1 - g$, siendo Z el porcentaje buscado y g el factor resultante de aplicar el citado método de proyección
- si se hace a partir del *link ratio*, entonces, $Z = 1 - \frac{1}{f}$, teniendo Z el mismo significado que antes, y siendo f el factor resultante de aplicar este método de proyección.

A continuación, se indica cómo se obtendría el nivel de reservas en cada uno de los casos indicados, para lo cual se seguirá un ejemplo numérico.

6.1.1. B/F general

En este caso, el primer paso consiste en obtener los factores de paso de un ejercicio a otro, los cuales se pueden calcular de igual forma que en el *CL*. Una vez obtenidos, se calculan los β_j de acuerdo con la definición dada anteriormente. Los valores resultantes, para cada año de desarrollo, son los siguientes:

	0	1	2	3	4	5	6	7	8
f_k	3,5582	1,7784	1,4835	1,1952	1,1244	1,1068	1,0743	1,0975	1,0383
β_k	0,0585	0,2082	0,3702	0,5492	0,6564	0,7381	0,8169	0,8776	0,9632

A partir de aquí, se puede rellenar la parte inferior derecha del triángulo de siniestros, siendo los resultados los siguientes:

	0	1	2	3	4	5	6	7	8	9
0										
1										4.364
2									4.170	4.340
3								3.723	4.092	4.251
4							3.181	3.421	3.760	3.906
5						3.056	3.414	3.690	4.079	4.247
6				3.064	3.504	3.928	4.254	4.715	4.913	
7			2.961	3.530	3.964	4.383	4.705	5.160	5.356	
8		1.606	2.342	2.783	3.119	3.443	3.692	4.044	4.196	
9	699	1.203	1.760	2.093	2.347	2.592	2.781	3.047	3.162	

de modo que, teniendo en cuenta la siniestralidad aflorada, el nivel de reservas a dotar es la diferencia entre la provisión total estimada (38.733) y la última diagonal (22.648). La dotación por años de accidente, tanto en valor nominal como descontado, se recoge en la siguiente tabla:

año	dotación anual	
	valor nominal	valor actual
1	4.175	4.124
2	3.386	3.295
3	2.629	2.514
4	1.950	1.784
5	1.529	1.352
6	1.148	953
7	736	572
8	418	308
9	115	81
a dotar	16.085	14.982
Saldo total	41.854	40.751

6.1.2. B/F con *grossing up*

Para obtener el nivel de reservas en este caso, es necesario seguir los siguientes pasos:

Paso 1: cálculo de los coeficientes de grossing up

Cada uno de los cuales, se obtiene a partir de la siguiente expresión:

$$g_{i,j} = \frac{C_{i,j}}{C_{0,\infty}}$$

Así, para el ejercicio 0, se obtiene que:

	0	1	2	3	4	5	6	7	8	9
$g_{i,j}$	7,29	23,34	36,66	47,71	60,12	74,07	78,80	83,55	90,54	94,01

con lo que las siniestralidades límite son las siguientes:

año de origen	$\hat{C}_{i,\infty}$
0	3.320
1	4.624
2	4.519
3	4.394
4	3.874
5	4.464
6	5.213
7	5.483
8	4.027
9	3.197

A partir de estas cantidades, se calculan los coeficientes $g_{i,j}$ para el resto de los ejercicios, obteniéndose la siguiente matriz, en la cual se vuelve a incluir los correspondientes al ejercicio 0 (cifras expresadas en porcentaje):

	0	1	2	3	4	5	6	7	8	9
0	7,29	23,34	36,66	47,71	60,12	74,07	78,80	83,55	90,54	94,01
1	5,15	18,42	32,91	51,79	59,73	65,98	75,84	81,74	90,54	
2	4,34	19,72	34,43	51,12	64,13	67,73	77,29	83,55		
3	4,78	22,26	35,05	61,07	66,69	73,86	78,80			
4	7,74	20,21	38,54	53,41	63,87	74,07				
5	6,00	20,59	34,27	47,76	60,12					
6	5,72	17,04	32,57	47,71						
7	4,43	17,87	36,66							
8	6,33	23,34								
9	7,29									

Paso 2: cálculo del coeficiente medio para cada año de desarrollo

La media aritmética de los coeficientes de cada columna son los que se recogen en la siguiente tabla:

	0	1	2	3	4	5	6	7	8	9
$g_{i,j}$	5,91	20,31	35,14	51,51	62,44	71,14	77,68	82,95	90,54	94,01

Estos coeficientes indican la siniestralidad media aflorada hasta el momento presente.

Paso 3: cálculo de la diferencia $1 - g_j$

Restando a uno las medias antes obtenidas, se logra la siguiente tabla:

	0	1	2	3	4	5	6	7	8	9
$g_{i,j}$	94,09	79,69	64,86	48,49	37,56	28,86	22,32	17,05	9,46	5,99

Estos coeficientes indican la siniestralidad que se estima que aún queda por aflorar.

Paso 4: estimación de la siniestralidad total de cada ejercicio

Dado el valor de la ratio entre siniestralidad y primas y el nivel de estas últimas, la siniestralidad estimada se obtiene sin más que multiplicar la ratio citada por las primas. En nuestro caso, se logran los siguientes valores:

0	1	2	3	4	5	6	7	8	9
3.319	4.794	4.604	4.309	3.958	4.549	5.382	5.313	4.111	3.111

Paso 5: evolución prevista de la siniestralidad pendiente de aflorar

Teniendo en cuenta los porcentajes de paso de un ejercicio a otro, se puede rellenar la parte inferior derecha del triángulo de siniestros. Los resultados son los siguientes:

	0	1	2	3	4	5	6	7	8	9	∞
0											3.319
1										4.507	4.794
2									4.168	4.328	4.604
3								3.574	3.901	4.050	4.309
4						3.075	3.284	3.584	3.721	3.958	
5					3.236	3.534	3.774	4.119	4.277	4.549	
6				3.361	3.829	4.181	4.465	4.873	5.060	5.382	
7			2.736	3.317	3.779	4.127	4.407	4.810	4.994	5.313	
8		1.445	2.118	2.567	2.925	3.194	3.410	3.723	3.865	4.111	
9	632	1.093	1.602	1.943	2.213	2.417	2.581	2.817	2.925	3.111	

Paso 6: cálculo de las dotaciones por años

Como en todos los casos que se han realizado hasta ahora, se calculan las diferencias respecto del ejercicio inmediatamente anterior. Los resultados, tanto en valor nominal como descontado se recogen en la siguiente tabla:

año	dotación anual	
	valor nominal	valor actual
1	4.284	4.231
2	3.463	3.370
3	2.738	2.619
4	2.070	1.893
5	1.623	1.435
6	1.283	1.065
7	983	764
8	697	513
9	354	250
10	186	125
a dotar	17.681	16.266
Saldo total	43.450	42.035

6.1.3. B/F con link ratio

Para obtener el nivel de reservas en este caso, es necesario seguir los siguientes pasos:

Paso 1: Cálculo de las link ratios y de los factores de proyección

Tal y como se ha explicado anteriormente, el cálculo de las link ratios arroja los siguientes resultados:

	0	1	2	3	4	5	6	7	8	9
0	3,2025	1,5703	1,3016	1,2601	1,2320	1,0638	1,0604	1,0836	1,0383	1,0638
1	3,5798	1,7864	1,5736	1,1532	1,1046	1,1495	1,0778	1,1077		
2	4,5459	1,7464	1,4846	1,2545	1,0562	1,1411	1,0810			
3	4,6571	1,5746	1,7422	1,0921	1,1075	1,0669				
4	2,6100	1,9068	1,3858	1,1957	1,1597					
5	3,4291	1,6649	1,3935	1,2589						
6	2,9799	1,9122	1,4647							
7	4,0329	2,0510								
8	3,6863									
9										

Las *link ratios* medias y los factores de cambio resultantes para este conjunto de datos son las siguientes:

	0	1	2	3	4	5	6	7	8	9
LR_k	3,636	1,777	1,478	1,202	1,132	1,105	1,073	1,096	1,038	1,064
f_k	18,652	5,130	2,888	1,954	1,625	1,435	1,299	1,210	1,104	1,064

Paso 2: cálculo de $1/f$ y de $1 - 1/f$

El inverso de los factores de proyección representan la siniestralidad media aflorada en el momento de hacer el cálculo. Por tanto, 1 menos el inverso será la siniestralidad que falta por manifestarse. Los resultados en este caso son los siguientes:

	0	1	2	3	4	5	6	7	8	9
$1/f_k$	0,054	0,195	0,346	0,512	0,615	0,697	0,770	0,826	0,905	0,940
$1-1/f_k$	0,946	0,805	0,654	0,488	0,385	0,303	0,230	0,174	0,095	0,060

Paso 3: estimación de la siniestralidad total de cada ejercicio

El cálculo a realizar es el mismo que en el paso 4 del caso anterior, por lo que los resultados son los mismos, es decir:

	0	1	2	3	4	5	6	7	8	9
	3.319	4.794	4.604	4.309	3.958	4.549	5.382	5.313	4.111	3.111

Paso 4: evolución prevista de la siniestralidad pendiente de aflorar

Considerando el porcentaje correspondiente de siniestralidad no manifestada aún, se puede rellenar la parte inferior del triángulo de siniestros. Los resultados son los siguientes:

	0	1	2	3	4	5	6	7	8	9	∞
0											3.319
1										4.507	4.794
2									4.168	4.328	4.604
3								3.561	3.901	4.050	4.309
4							3.048	3.271	3.584	3.721	3.958
5						3.169	3.503	3.759	4.119	4.277	4.549
6					3.313	3.750	4.145	4.448	4.873	5.060	5.382
7				2.719	3.270	3.701	4.091	4.390	4.810	4.994	5.313
8			1.424	2.104	2.530	2.864	3.166	3.398	3.723	3.865	4.111
9		606	1.077	1.592	1.915	2.167	2.396	2.571	2.817	2.925	3.111

Paso 5: cálculo de las dotaciones por años

Como en todos los casos que se han realizado hasta ahora, se calculan las diferencias respecto del ejercicio inmediatamente anterior. Los resultados, tanto en valor nominal como descontado se recogen en la siguiente tabla:

año	dotación anual	
	valor nominal	valor actual
1	4.066	4.015
2	3.483	3.390
3	2.762	2.641
4	2.104	1.925
5	1.674	1.480
6	1.339	1.111
7	1.007	782
8	707	520
9	354	250
10	186	125
a dotar	17.681	16.241
Saldo total	43.450	42.010

6.2. Método *Benktander - Hovinen (BH)*

Este sistema de cálculo de las reservas lleva el nombre de sus dos autores, los cuales publicaron sendos trabajos por separado -Benktander (1976) y Hovinen (1981) -, pero llegando a idénticos resultados.

La principal diferencia entre este método y los anteriores es que aquí se supone un cierto conocimiento previo sobre el valor final de $C_{i,J} - \mu_i$, esto es, se tiene una idea previa de cuál es el nivel de los pagos esperados totales correspondientes al año i -ésimo de accidente. Además, como en otros métodos, se supone conocido el patrón de evolución de los siniestros, el cual viene dado por los β_j , definidos como:

$$\beta_j = \prod_{k=j}^{J-1} f_k^{-1} \quad \forall j = 0, \dots, J-1$$

Este método obtiene resultados a medio camino entre los que se puedan conseguir con el *CL* y con el *B/F*. Dado que el primero de ellos prescinde de la información que pudiera aportar la estimación previa de $C_{i,J}$ y el segundo hace lo propio con la información que se incorpora en la última diagonal de datos disponibles, esto es, en $C_{i,l-i}$, este sistema puede verse como una combinación de los anteriores. El estimador propuesto por Benktander tiene la siguiente expresión:

$$\hat{C}_{i,J} = C_{i,l-i} + (1 - \beta_{l-i}) \left[\beta_{l-i} \hat{C}_{i,J}^{CL} + (1 - \beta_{l-i}) \mu_i \right] \quad \forall i = 1, \dots, l$$

en donde $\beta_{l-i} = \prod_{k=l-i}^{J-1} f_k^{-1}$, por lo que la anterior expresión puede escribirse como:

$$\begin{aligned}\hat{C}_{i,J} &= \beta_{l-i} \hat{C}_{i,J}^{CL} + (1-\beta_{l-i}) \hat{C}_{i,J}^{BF} \\ &= C_{i,l-i} + (1-\beta_{l-i}) \hat{C}_{i,J}^{BF}\end{aligned}$$

es decir, puede obtenerse bien a partir de las estimaciones conseguidas con los métodos *CL* y *B/F*, bien únicamente a partir de los resultados logrados con este último método.

Ejemplo:

Partiendo del conjunto de datos utilizados en los ejemplos anteriores, se obtiene los factores *CL* de desarrollo, a partir de los cuales se calculan los β_j , los cuales se recogen en la siguiente tabla:

	0	1	2	3	4	5	6	7	8
f_j	3,558	1,778	1,484	1,195	1,124	1,107	1,074	1,097	1,038
β_j	0,059	0,208	0,370	0,549	0,656	0,738	0,817	0,878	0,963

Para obtener las reservas siguiendo este método se calcula, para cada año de accidente, la combinación lineal entre los resultados del *CL* y del *B/F* usando los β_{l-i} como factor de ponderación. El siguiente triángulo recoge la senda estimada de siniestralidad:

	0	1	2	3	4	5	6	7	8	9
0										
1										4.348
2									4.146	4.307
3								3.719	4.083	4.240
4							3.176	3.413	3.748	3.892
5						3.022	3.355	3.614	3.983	4.143
6				2.987	3.384	3.774	4.079	4.513	4.702	
7			2.975	3.549	3.986	4.407	4.730	5.187	5.383	
8		1.643	2.394	2.840	3.179	3.506	3.756	4.110	4.262	
9	735	1.246	1.805	2.140	2.394	2.640	2.829	3.095	3.210	

con lo que las dotaciones a realizar, tanto en valor nominal como descontado, son las que se recogen en la siguiente tabla:

año	dotación anual	
	valor nominal	valor actual
1	4.104	4.054
2	3.326	3.237
3	2.584	2.471
4	1.912	1.749
5	1.499	1.326
6	1.141	947
7	739	574
8	419	308
9	115	81
a dotar	15.839	14.747
Saldo total	41.608	40.516

6.3. Método Cape-Cod (CC)

Este sistema trata de paliar una de las deficiencias del CL, que es la referida a la existencia de un posible valor atípico en la última diagonal. Dado que las reservas del CL se calculan a partir de este último dato, la presencia de un valor anormalmente alto o bajo conlleva que la proyección del mismo genere un nivel de reservas igualmente anómalo en idéntica dirección.

En este método se supone lo siguiente:

- independencia entre los importes $C_{i,j}$ de años de accidente diferentes
- existe un conjunto de parámetros $\Pi_0, \Pi_1, \dots, \Pi_l > 0$, $\kappa > 0$ y un patrón de desarrollo de siniestros $\beta_j \forall j = 0, \dots, J$, tales que:

$$E(C_{i,j}) = \kappa \Pi_i \beta_j \quad \forall i = 0, \dots, l$$

Π_i puede interpretarse como el nivel de primas del año i -ésimo, mientras que κ puede verse como una ratio de siniestralidad media. Para llevar a cabo su cálculo es preciso obtener previamente los κ_i correspondientes a cada año, para lo cual el cálculo a realizar es el siguiente:

$$\hat{\kappa}_i = \frac{\hat{C}_{i,J}}{\Pi_i} = \left(\frac{\hat{C}_{i,l-i}}{\prod_{j=l-i}^{J-1} f_j^{-1}} \right) \frac{1}{\Pi_i} = \frac{C_{i,l-i}}{\beta_{l-i} \Pi_i} \quad \forall i = 0, \dots, 9$$

Con lo que un estimador menos dependiente de los atípicos -se podría decir que más robusto- se obtendría como:

$$\hat{C}_{i,l-i} = \hat{\kappa} \beta_{l-i} \Pi_i$$

con lo que el estimador CC queda como sigue:

$$\hat{C}_{i,J}^{CC} = \prod_{j=l-i}^{J-1} f_j \hat{C}_{i,l-i} + (C_{i,l-i} - \hat{C}_{i,l-i}) \quad \forall i = 1, \dots, l$$

que puede interpretarse como la clásica proyección del CL realizada a partir del $C_{i,l-i}$ estimado más la diferencia entre el valor real de la última diagonal y el estimado.

Ejemplo:

Partiendo de los datos habituales, se hace preciso incorporar como información adicional el nivel de primas correspondiente a todos los años. Los datos utilizados son los mismos que en el caso del *Loss ratio* ingenuo y son los siguientes:

año	primas	año	primas
0	3.905	5	5.352
1	5.640	6	6.332
2	5.416	7	6.250
3	5.069	8	4.837
4	4.657	9	3.660

A partir de aquí, se calculan las ratios de siniestralidad anuales como el cociente entre la estimación $\hat{C}_{i,J}^{CL}$ con el CL y las primas de cada año. Los resultados son los recogidos en la siguiente tabla:

año	κ_j	año	κ_j
0	79,92%	5	76,40%
1	77,08%	6	71,51%
2	79,44%	7	86,87%
3	83,60%	8	93,35%
4	83,47%	9	108,81%

Con estos resultados se procede a calcular la ratio media, utilizando como factores de ponderación los productos $\beta_{l-i} \Pi_i$, obteniéndose un resultado de 80,04%. Ya se está en condiciones de estimar el valor de la última diagonal aplicando la expresión $\hat{C}_{i,l-i} = \hat{\kappa} \beta_{l-i} \Pi_i$. Los resultados que se obtienen, junto con las diferencias respecto a los datos originales, son los recogidos en la siguiente tabla:

año	$\hat{C}_{i,l-i}$	$\hat{C}_{i,l-i} - C_{i,l-i}$
0	3.126	-5
1	4.348	-161
2	3.804	-28
3	3.315	147
4	2.751	118
5	2.812	-128
6	2.784	-297
7	1.852	158
8	806	134
9	171	62

y a partir de estas cifras, ya se puede proceder a rellenar la parte inferior derecha del triángulo, sin más que utilizar la expresión:

$$\hat{C}_{i,j}^{CC} = \prod_{j=l-i}^{j-1} f_j \hat{C}_{i,j-i} + (C_{i,l-i} - \hat{C}_{i,l-i}) \quad \forall i = 1, \dots, l$$

Obteniéndose la siguiente senda estimada de siniestralidad:

	0	1	2	3	4	5	6	7	8	9
0										
1										4.353
2									4.147	4.307
3								3.708	4.055	4.205
4							3.163	3.389	3.708	3.845
5						3.034	3.372	3.631	3.998	4.156
6				3.030	3.444	3.844	4.151	4.585	4.772	
7			2.906	3.442	3.850	4.245	4.548	4.976	5.161	
8		1.567	2.260	2.675	2.992	3.297	3.532	3.863	4.006	
9	671	1.146	1.671	1.985	2.224	2.455	2.633	2.883	2.991	

Con lo que las dotaciones a realizar, tanto en valor nominal como descontado, son las que se recogen en la siguiente tabla:

año	dotación anual	
	valor nominal	valor actual
1	3.932	3.883
2	3.189	3.103
3	2.476	2.368
4	1.836	1.680
5	1.439	1.273
6	1.081	897
7	693	539
8	393	290
9	108	76
a dotar	15.147	14.108
Saldo total	40.916	39.877

CAPÍTULO 3 LA DISTRIBUCIÓN LIBRE DE MACK

1. FUNDAMENTOS TEÓRICOS DE LA PROPUESTA

Este método aparece por primera vez en Mack (1993). Su objetivo es la estimación de la desviación típica de la reserva estimada mediante el método del *Chain Ladder*. Se basa en este método porque, si bien el *Chain Ladder* se caracteriza por ser de distribución libre y, por tanto, susceptible de ser usado con cualquier conjunto de datos, también es cierto que las estimaciones realizadas mediante este mecanismo para los años de accidente más recientes resultan ser más sensibles a los cambios en los datos observados. Por tanto, puede resultar muy interesante conocer el valor de la desviación típica de la reserva estimada mediante dicho método y así tener una medida del grado de incertidumbre que tienen los datos. Ello ayudará a saber si las diferencias obtenidas entre el *Chain Ladder* y otros métodos resultan estadísticamente significativas o no.

Para llevar a cabo su objetivo, Mack siguió un proceso en varias fases, en las que, partiendo de las hipótesis básicas del *Chain Ladder*, llega hasta la estimación de la varianza de la reserva calculada, aunque lo que realmente obtiene es el error cuadrático medio.

Sea $C_{i,j}$ el importe acumulado de siniestros del año i de accidente que, o bien se han pagado o bien se ha incurrido en ellos hasta el año j . En ambos casos se cumple que $0 \leq i \leq l$ y que $0 \leq j \leq l$, siendo l el número máximo de años de los que se dispone de información, i es el año de accidente, j es el año de desarrollo y se cumple que $i + j \leq l$. El objetivo que se persigue en el *Chain Ladder* es de la estimación del importe total que se deberá pagar por los siniestros ocurridos en el año i , que se denomina $C_{i,l}$. Por tanto, se trata de estimar el valor de la reserva que hay que dotar para hacer frente a dicho importe, a la que se denominará R_i , la cual se obtiene como:

$$R_i = C_{i,l} - C_{i,l+1-i} \quad / \quad i = 1, \dots, l$$

Según Mack, las dos hipótesis básicas del método *Chain Ladder* son las siguientes:

Hipótesis 1: existen unos factores de desarrollo $f_0, f_2, \dots, f_{l-1} > 0$, los cuales verifican que:

$$E(C_{i,j+1} / C_{i,1}, \dots, C_{i,j}) = C_{i,j} \cdot f_j \quad / \quad 0 \leq i \leq l \quad 0 \leq j \leq l-1$$

El valor estimado de f_k, \hat{f}_k , se obtiene como:

$$\hat{f}_j = \frac{\sum_{i=0}^{l-j-1} C_{i,j+1}}{\sum_{i=1}^{l-j} C_{i,j}} \quad 0 \leq j \leq l-1$$

de modo que el valor total de los siniestros del año i se estima como:

$$\hat{C}_{i,l} = C_{i,l-i} \cdot \hat{f}_{l-i} \cdots \hat{f}_{l-1}$$

con lo que la reserva asociada a los siniestros de ese año se calcula como:

$$\begin{aligned} \hat{R}_i &= \hat{C}_{i,l} - C_{i,l-i} \\ &= C_{i,l-i} \cdot (\hat{f}_{l-i+1} \cdots \hat{f}_{l-1} - 1) \end{aligned}$$

Hipótesis 2: El método Chain Ladder considera que cada año de ocurrencia de accidente es independiente de los demás, es decir:

$$\{C_{i,1}, C_{i,2}, \dots, C_{i,l}\}, \{C_{j,1}, C_{j,2}, \dots, C_{j,l}\} \quad i \neq j \text{ son independientes}$$

A partir de ambas hipótesis, Mack demuestra que:

$$1.- \quad E(C_{i,l} | D) = C_{i,l+1-i} \cdot f_{l+1-i} \cdots f_{l-1}$$

siendo $D = \{C_{i,j} / i + j \leq l\}$ el conjunto de los datos disponibles

2.- La estimación de los factores de desarrollo, \hat{f}_j , es insesgada y sin correlación entre ellos

A partir de aquí se trata de estimar el valor de la dispersión ligada a la estimación del nivel de reservas, para lo cual, en primer lugar se obtiene el error asociado a la reserva de cada año de accidente, R_i . Para realizar esta tarea, utiliza como medida el error cuadrático medio de $\hat{C}_{i,l}$, a partir de ahora,

$MSE(\hat{C}_{i,l})$, definido como:

$$MSE(\hat{C}_{i,l}) = E \left[(\hat{C}_{i,l} - C_{i,l})^2 | D \right] \quad / \quad D = \{C_{i,j} / i + j \leq l\}$$

Partiendo de la definición anterior se llega a que el error cuadrático medio de $\hat{C}_{i,l}$ coincide con el de la reserva estimada, \hat{R}_i . En efecto:

$$\begin{aligned}
MSE(\hat{C}_{i,l}) &= E\left[\left(\hat{C}_{i,l} - C_{i,l}\right)^2 \mid D\right] \\
&= E\left[\left(\hat{R}_i + C_{l-i} - E(\hat{R}_i) - C_{l-i}\right)^2 \mid D\right] = \\
&= E\left[\left[\hat{R}_i - E(\hat{R}_i) \mid D\right]^2\right] = MSE(\hat{R}_i)
\end{aligned}$$

Además, como de forma general, para cualquier variable aleatoria se cumple que:

$$E(x - a)^2 = \text{var}(x) + [E(x) - a]^2$$

es decir, que el *MSE* es igual a la suma de la varianza de la variable más su sesgo, entonces:

$$MSE(\hat{C}_{i,l}) = \text{var}(C_{i,l} \mid D) + [E(C_{i,l} \mid D) - \hat{C}_{i,l}]^2$$

Con el fin de calcular posteriormente el *MSE*, como primer paso se necesita una expresión que permita obtener la varianza de $C_{i,k}$. Teniendo en cuenta que \hat{f}_j se puede considerar como la media de los factores individuales de desarrollo $C_{i,j+1}/C_{i,j}$ / $0 \leq i \leq l - j$, cada uno de ellos ponderado por los importes $C_{i,j}$, se puede deducir que la varianza de dicho factor individual debería ser inversamente proporcional a $C_{i,j}$. Expresado de otra forma, se tiene la Hipótesis 3:

Hipótesis 3: Expresión de la varianza de $C_{i,j+1}$ condicionada a la información disponible:

$$\text{var}(C_{i,j+1} \mid C_{i,0}, C_{i,1}, \dots, C_{i,j}) = C_{i,j} \cdot \sigma_j^2 \quad 0 \leq i \leq l \quad 0 \leq j \leq l - 1$$

en donde las σ_j^2 son parámetros desconocidos, por lo que se necesita un estimador de las mismas. La expresión para el estimador insesgado de σ_j^2 que se recoge en Mack(1993) es la siguiente:

$$\hat{\sigma}_j^2 = \frac{1}{l - j - 1} \sum_{i=0}^{l-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2$$

que es válida para todo j entre 0 e $l - 2$, no siéndolo si $j = l - 1$ ya que entonces el denominador se igualaría a cero, con lo que el valor de la varianza sería infinito. Para esta situación, Mack sugiere las siguientes posibilidades:

- a) si $\hat{f}_{l-1} = 1$ y se espera que el desarrollo de los siniestros finalice tras $l - 1$ años, entonces, sugiere que $\hat{\sigma}_{l-1}^2 = 0$.

b) si lo anterior no es el caso, es necesario extrapolar el siguiente valor de la secuencia: $\{\hat{\sigma}_0, \hat{\sigma}_1, \dots, \hat{\sigma}_{l-3}, \hat{\sigma}_{l-2}\}$, que suele ser una sucesión de valores exponencialmente decrecientes. Para ello, se sugieren, a su vez, dos posibilidades:

– mediante el uso de una regresión loglineal del tipo:

$$\ln \hat{\sigma}_j^2 = \hat{\alpha} + \hat{\beta}j$$

que implica que:

$$\hat{\sigma}_{l-1}^2 = \exp\{\hat{\alpha} + \hat{\beta}(l-1)\}$$

– imponiendo la siguiente condición:

$$\frac{\hat{\sigma}_{l-3}}{\hat{\sigma}_{l-2}} = \frac{\hat{\sigma}_{l-2}}{\hat{\sigma}_{l-1}} \quad \text{si } \hat{\sigma}_{l-3} > \hat{\sigma}_{l-2}$$

Es decir, que el ritmo de caída de la varianza se mantenga en las últimas observaciones. Expresado de otra forma:

$$\hat{\sigma}_{l-1}^2 = \min\left\{\frac{\hat{\sigma}_{l-2}^4}{\hat{\sigma}_{l-3}^2}; \min\{\hat{\sigma}_{l-2}^2; \hat{\sigma}_{l-3}^2\}\right\}$$

Una vez que ya se tiene la secuencia completa de varianzas por año de ocurrencia, se pasa a estimar el $MSE(\hat{R}_i)$ y, a partir de ellos, el correspondiente a toda la reserva, es decir, $MSE(\hat{R})$.

En su artículo Mack demuestra que:

$$\widehat{MSE}(\hat{R}_i) = \hat{C}_{i,l}^2 \sum_{j=l+1-i}^{l-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \left(\frac{1}{\hat{C}_{i,j}} + \frac{1}{\sum_{k=0}^{l-j} C_{k,j}} \right)$$

siendo $\hat{C}_{i,j} = C_{i,l+1-i} \cdot \hat{f}_{l+1-i} \cdots \hat{f}_{j-1} \quad \forall j > l+1-i$ y siendo $\hat{C}_{i,l-i} = C_{i,l-i}$, es decir, los valores registrados de la última diagonal. Finalmente, Mack demuestra que el error cuadrático medio de toda la reserva, $\hat{R} = \hat{R}_1 + \cdots + \hat{R}_l$, se estima como:

$$\widehat{MSE}(\hat{R}) = \sum_{i=1}^I \left\{ \left[se(\hat{R}_i) \right]^2 + \hat{C}_{i,I} \left(\sum_{j=i+1}^I \hat{C}_{j,I} \right) \sum_{k=l-i}^{l-1} \frac{2\hat{\sigma}_k^2 / \hat{f}_k^2}{\sum_{n=1}^{l-k} C_{n,k}} \right\}$$

2. EJEMPLO

Para ilustrar el cálculo correspondiente a la reserva total tanto del *MSE* como de su raíz cuadrada, a la que, de ahora en adelante denominaremos error típico, se utilizarán los datos aplicados hasta ahora en todos los ejemplos. El triángulo de liquidación de siniestros, expresado en cifras acumuladas, es el siguiente:

	0	1	2	3	4	5	6	7	8	9
0	242	775	1.217	1.584	1.996	2.459	2.616	2.774	3.006	3.121
1	238	852	1.522	2.395	2.762	3.051	3.507	3.780	4.187	
2	196	891	1.556	2.310	2.898	3.061	3.493	3.776		
3	210	978	1.540	2.683	2.930	3.245	3.462			
4	300	783	1.493	2.069	2.474	2.869				
5	268	919	1.530	2.132	2.684					
6	298	888	1.698	2.487						
7	243	980	2.010							
8	255	940								
9	233									

en donde, como es habitual, las filas representan los años de ocurrencia y las columnas los de desarrollo. A partir de estos datos, los pasos a seguir son los siguientes:

1. cálculo de los f_j estimados
2. cálculo de la reserva mediante Chain Ladder tradicional
3. obtención de la varianza por año de desarrollo
4. cálculo del error cuadrático medio y del error típico por cada año de ocurrencia, es decir, $\widehat{MSE}(\hat{R}_i)$ y $se(\hat{R}_i)$ respectivamente
5. cálculo del error cuadrático medio y del error típico del total de la reserva, es decir, $\widehat{MSE}(\hat{R})$ y $se(\hat{R})$ respectivamente.

Los dos primeros se corresponden con lo que sería el uso del *Chain Ladder* determinista mientras que para llevar a cabo los otros tres, se utiliza el procedimiento propuesto por Mack en su trabajo aquí explicado.

Por lo que se refiere al cálculo de los f_j estimados, se emplea la ya conocida expresión:

$$\hat{f}_j = \frac{\sum_{i=0}^{l-j-1} C_{i,j+1}}{\sum_{i=1}^{l-j} C_{i,j}} \quad 0 \leq j \leq l-1$$

obteniéndose los siguientes resultados:

año de ocurrencia							
0	1	2	3	4	5	6	7
3,5582	1,7784	1,4835	1,1952	1,1244	1,1068	1,0743	1,0975

a partir de aquí, sabiendo cuál es el valor de la provisión actual y aplicando el método Chain Ladder, se obtienen los siguientes resultados:

Año de ocurrencia	Dotación actual	Reserva
1	4.187	160
2	3.776	527
3	3.462	776
4	2.869	1.018
5	2.684	1.405
6	2.487	2.041
7	2.010	3.419
8	940	3.575
9	233	3.749

con lo que el total de la reserva debe ser de 16.671

Una vez calculado el valor de la reserva mediante el uso del *Chain Ladder* tradicional, comienza propiamente el cálculo del error típico de la misma por el procedimiento propuesto por Mack. Primeramente hay que calcular la varianza por año de desarrollo, es decir, por columna -paso 3-. Para ello, se necesita la siguiente información estadística:

- datos originales
- factores estimados

ya que la expresión a utilizar es:

$$\hat{\sigma}_j^2 = \frac{1}{l-j-1} \sum_{i=0}^{l-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2$$

Por ejemplo, para $\hat{\sigma}_5^2$ se tendría:

$$\hat{\sigma}_5^2 = \frac{1}{9-5-1} \sum_{i=0}^{9-5} C_{i5} \left(\frac{C_{i6}}{C_{i5}} - \hat{f}_5 \right)^2 =$$

$$= \frac{1}{3} \left[C_{05} \left(\frac{C_{06}}{C_{05}} - \hat{f}_5 \right)^2 + C_{15} \left(\frac{C_{16}}{C_{15}} - \hat{f}_5 \right)^2 + C_{25} \left(\frac{C_{26}}{C_{25}} - \hat{f}_5 \right)^2 + C_{35} \left(\frac{C_{36}}{C_{35}} - \hat{f}_5 \right)^2 \right]$$

sustituyendo cada variable por su valor se obtiene como resultado:

$$\hat{\sigma}_5^2 = 6,29$$

Haciendo esto mismo para todos los valores de j entre 0 e $l-2$ se obtiene la siguiente tabla de valores:

	0	1	2	3	4	5	6	7
$\hat{\sigma}_k^2$	113,55	26,94	30,49	11,24	10,39	6,29	0,35	0,92

Por tanto, falta por calcular la varianza $\hat{\sigma}_8^2$. Como se ha indicado en la exposición teórica, aquí caben dos posibilidades: la utilización de la expresión propuesta por Mack o bien la estimación mediante el empleo de la regresión logarítmica. A continuación se recogen los resultados tanto de las varianzas de $l-1$ como de los *MSE* de cada fila y del total en cada uno de las dos posibilidades.

Alternativa 1: a partir de la expresión propuesta por Mack

En este caso se obtiene que:

$$\hat{\sigma}_8^2 = \min \left\{ \frac{\hat{\sigma}_7^4}{\hat{\sigma}_6^2}; \min \{ \hat{\sigma}_6^2; \hat{\sigma}_7^2 \} \right\}$$

$$= \min \left\{ \frac{0,92^2}{0,35}; \min \{ 0,35; 0,92 \} \right\} =$$

$$= 0,35$$

Una vez calculadas las varianzas por columna, se puede proceder a abordar el paso 4, es decir, el cálculo de los errores cuadráticos medios y de los errores típicos de cada año de ocurrencia. Para ello, se necesita la siguiente información estadística:

- datos originales,
- datos estimados $\hat{C}_{i,j}$ $j = l+1-i \dots l$
- las varianzas estimadas

ya que se utiliza la siguiente expresión:

$$\widehat{MSE}(\hat{R}_i) = C_{i,i}^2 \sum_{j=i+1-i}^{l-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \left(\frac{1}{\hat{C}_{i,j}} + \frac{1}{\sum_{k=1}^{l-j} C_{k,j}} \right)$$

Por ejemplo, para obtener el error cuadrático medio de la reserva correspondiente al segundo año se hará lo siguiente:

$$\begin{aligned} \widehat{MSE}(\hat{R}_2) &= \hat{C}_{2,9}^2 \sum_{j=9-2}^{9-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \left(\frac{1}{\hat{C}_{2,j}} + \frac{1}{\sum_{k=1}^{9-j} C_{k,j}} \right) = \\ &= \hat{C}_{2,9}^2 \left[\frac{\hat{\sigma}_7^2}{\hat{f}_7^2} \left(\frac{1}{\hat{C}_{2,7}} + \frac{1}{C_{1,7} + C_{2,7}} \right) + \frac{\hat{\sigma}_8^2}{\hat{f}_8^2} \left(\frac{1}{\hat{C}_{2,8}} + \frac{1}{C_{1,8}} \right) \right] \end{aligned}$$

que, en este caso, resulta un valor de 9.417,35. Una vez realizado este mismo proceso para todos los años, se calcula la raíz cuadrada de cada uno de los valores obtenidos con lo que se obtiene la desviación típica de cada reserva. Para hacerse una idea del grado de dispersión asociado al cálculo realizado, se procede a obtener qué porcentaje de la reserva estimada representa la desviación típica. Los resultados se recogen en la siguiente tabla:

Fila	$MSE(R_i)$	error (R_i)	reserva	%
1	3.543	60	160	37%
2	9.417	97	527	18%
3	11.375	107	776	14%
4	43.687	209	1.018	21%
5	108.214	329	1.405	23%
6	199.622	447	2.041	22%
7	490.616	700	3.419	20%
8	604.060	777	3.575	22%
9	1.197.024	1.094	3.749	29%

La última tarea a realizar es el cálculo del error estándar global -paso 5-. En este caso, la información necesaria es la siguiente:

- datos originales
- factores estimado
- datos estimados $\hat{X}_{i,l}$
- varianzas estimadas
- todos los $\left[se(\hat{R}_i) \right]^2$ o los $\widehat{MSE}(\hat{R}_i)$

ya que se utiliza la siguiente expresión:

$$\widehat{MSE}(\hat{R}) = \sum_{i=1}^l \left\{ [se(\hat{R}_i)]^2 + \hat{C}_{i,l} \left(\sum_{j=i+1}^l \hat{C}_{j,l} \right) \sum_{k=l-i}^{l-1} \frac{2\hat{\sigma}_k^2 / \hat{f}_k^2}{\sum_{n=1}^{l-k} C_{n,k}} \right\}$$

Los resultados intermedios se recogen en la siguiente tabla:

Fila	$[se(R_i)]^2$	producto	suma global
1	3.543	33.186	36.730
2	9.417	49.994	59.411
3	11.375	46.016	57.391
4	43.687	65.036	108.723
5	108.214	84.148	192.362
6	199.622	87.674	287.296
7	490.616	83.003	573.619
8	604.060	36.445	640.506
9	1.197.024	0	1.197.024

y el error cuadrático medio de la reserva total es de 3.153.062. El error típico es la raíz cuadrada de la cifra anterior, resultando 1.776. Sabiendo que la reserva total estimada es de 16.671, entonces el error típico en este caso es equivalente a un 10,65% de la misma.

Alternativa 2: uso de la regresión logarítmica

En este caso se trata de estimar el valor de $\hat{\sigma}_8^2$ mediante una regresión de la forma:

$$\ln \hat{\sigma}_j^2 = \hat{\alpha} + \hat{\beta}j$$

En este caso, j oscila entre 0 y 7. Utilizando el método de los mínimos cuadrados ordinarios, la regresión estimada es la siguiente:

$$\ln \hat{\sigma}_j^2 = 4,62 - 0,72j$$

con lo que, el valor estimado de $\hat{\sigma}_8^2$ es:

$$\hat{\sigma}_8^2 = \exp(4,62 - 0,72 \times 8) = 0,68$$

A partir de aquí, y repitiendo los mismos pasos en el caso de la alternativa 1 se procede a estimar el valor de los error cuadrados medios y los errores típicos de cada reserva anual, obteniéndose los siguientes resultados:

Fila	$MSE(R_i)$	error (R_i)	reserva	%
1	6.777	82	160	51%
2	12.599	112	527	21%
3	14.482	120	776	16%
4	46.401	215	1.018	21%
5	111.151	333	1.405	24%
6	203.072	451	2.041	22%
7	495.240	704	3.419	21%
8	607.495	779	3.575	22%
9	1.199.842	1.095	3.749	29%

Finalmente, para el total de la reserva los resultados intermedios son los recogidos en la siguiente tabla:

Fila	$[se(R_i)]^2$	producto	suma global
1	6.777	63.474	70.251
2	12.599	76.283	88.882
3	14.482	68.331	82.813
4	46.401	82.493	128.894
5	111.151	99.182	210.332
6	203.072	100.238	303.310
7	495.240	92.195	587.435
8	607.495	40.028	647.523
9	1.199.842	0	1.199.842

y el error cuadrático medio de la reserva total es de 3.319.283. El error típico es la raíz cuadrada de la cifra anterior, resultando 1.822. Sabiendo que el valor de la reserva total estimada -16.671-, entonces el error típico en este caso es equivalente a un 10,93% de la misma.

Como puede apreciarse, estos resultados son ligeramente superiores a los de la alternativa 1, lo cual es coherente puesto que el valor de $\hat{\sigma}_8^2$ es superior en este segundo caso.

CAPÍTULO 4

INTRODUCCIÓN A LOS MODELOS LINEALES GENERALES

1. CONCEPTO Y FINALIDAD

Los modelos lineales generalizados (a partir de ahora, *GLM*) se utilizan para evaluar y cuantificar la relación entre una variable, denominada respuesta o dependiente, y un conjunto de variables independientes o explicativas. Todo esto también lo hace la regresión lineal habitual. Sin embargo, hay dos grandes diferencias con ella:

1. La variable respuesta no tiene porqué seguir una distribución normal, de hecho, podría ser claramente distinta a la normal. De forma general, su distribución es alguna encuadrada dentro de la familia exponencial, la cual se detallará más adelante.
2. Existe una relación lineal entre la variable explicativa y una transformación de la media de la variable dependiente o respuesta.

Una consecuencia del comportamiento no normal de la variable dependiente es que, generalmente, será heterocedástica. Por tanto, la varianza se modificará conforme lo haga la media, la cual, a su vez, cambiará según lo hagan los valores de las variables independientes. Por tanto, los *GLM* son una alternativa a las transformaciones en la variable dependiente, motivadas por la falta de linealidad y de homogeneidad en la varianza.

Las hipótesis básicas de un modelo de esta clase son las siguientes:

- H1*: independencia entre las respuestas
- H2*: la esperanza de la variable respuesta cambia con las condiciones, pero no la forma funcional de la distribución
- H3*: la esperanza de la variable respuesta, o alguna transformación de ella, cambia de forma lineal cuando cambian las condiciones.

Es decir, se busca tanto la distribución como la relación funcional para la respuesta media de la variable dependiente que mejor se acople a los datos manejados. Dicho de otra manera, es necesario especificar por separado lo siguiente:

- a) la distribución de los datos con los que se trabaja, la cual va a determinar la relación entre la media y la varianza (función de varianza)

- b) la relación lineal entre la respuesta media -esperanza de la variable respuesta- y las variables independientes. Es lo que se denomina función de enlace o *link*, tal y como aparece generalmente en la literatura científica.

Por todo ello, un modelo *GLM* consta de los siguientes elementos:

- 1) una variable respuesta, Y , de la cual se tienen n observaciones (y_i , $i = 1, \dots, n$), todas ellas con la misma distribución -alguna de la familia exponencial-
- 2) un conjunto de variables explicativas X y sus correspondientes parámetros a estimar, β
- 3) una función monótona, llamada *link*, $g(\cdot)$, tal que si $E(y_i) = \mu_i$, entonces,

$$g(\mu_i) = \eta_i = \mathbf{X}_i \boldsymbol{\beta}$$

Habitualmente, la función exponencial se suele expresar como:

$$f(y, \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad (4.1)$$

siendo $a()$, $b()$ y $c()$ funciones específicas de la distribución que recoge el comportamiento de la variable dependiente, y θ es el denominado parámetro natural o canónico. En cuanto a ϕ , si es conocido, entonces se tiene un modelo de la familia exponencial lineal, y en caso contrario, se dice que pertenece a la clase de los modelos de dispersión exponencial.

En el caso de la familia exponencial es interesante analizar la relación entre la media y la varianza. Así, sea $l(\theta, \phi; y)$ el logaritmo neperiano de la función de verosimilitud de la observación. Por tanto:

$$l(\theta, \phi; y) = \ln f(\theta, \phi; y) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$$

Sea U la función de primeras derivadas de l respecto a θ , comúnmente denominada *score*, definida como:

$$U = \frac{\partial \ln f(\theta, \phi; y)}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)}$$

Se tiene que el valor esperado de la función *score* es cero y, por tanto, la respuesta media es igual a la derivada de $b(\theta)$, es decir:

$$E(U) = 0 \rightarrow E(Y) = b'(\theta)$$

Demostración:

$$\begin{aligned}
 E(U) &= \int \underbrace{\frac{\partial \ln f(\theta, \phi; y)}{\partial \theta}}_{\frac{f'}{f} = \frac{1}{f} \frac{\partial f}{\partial \theta}} f(\theta, \phi; y) dy = \int \frac{\partial f(\theta, \phi; y)}{\partial \theta} dy = \\
 &= \frac{\partial}{\partial \theta} \left[\underbrace{\int f(\theta, \phi; y)}_{\substack{1 \text{ por ser suma} \\ \text{de densidades}}} \right] = 0
 \end{aligned}$$

$$\Rightarrow E(U) = 0 = E \left[\frac{y - b'(\theta)}{a(\phi)} \right] \Rightarrow E(y) = E[b'(\theta)] = b'(\theta)$$

Por su parte, la varianza de la función U es igual al valor esperado de U' con el signo cambiado. Por tanto, la varianza de la variable dependiente puede expresarse como el producto de la función de escala $a(\phi)$ por la derivada segunda de $b(\theta)$, es decir:

$$\text{var}(U) = -E(U') \rightarrow \text{var}(Y) = a(\phi) b''(\theta)$$

Demostración:

Como $E(U) = 0$ y $\text{var}(U) = E(U^2) - E(U)^2$, entonces, $\text{var}(U) = E(U^2)$. Por tanto, se trata de demostrar que $E(U^2) = -E(U')$ siendo $U' = \frac{\partial U}{\partial \theta}$

Sabemos que $E(U) = 0 = \int \frac{\partial \ln f(\theta, \phi; y)}{\partial \theta} f(\theta, \phi; y) dy$. Tomando derivadas nuevamente:

$$\begin{aligned}
 0 &= \frac{\partial}{\partial \theta} \left[\int \frac{\partial \ln f(\theta, \phi; y)}{\partial \theta} f(\theta, \phi; y) dy \right] = \int \frac{\partial}{\partial \theta} \left[\frac{\partial \ln f(\theta, \phi; y)}{\partial \theta} f(\theta, \phi; y) \right] dy = \\
 &= \int \left(\frac{\partial^2 \ln f}{\partial \theta^2} f + \frac{\partial \ln f}{\partial \theta} f' \right) dy = \int \left(\frac{\partial^2 \ln f}{\partial \theta^2} f + \frac{\partial \ln f}{\partial \theta} \frac{\partial f}{\partial \theta} \right) dy
 \end{aligned}$$

$$\text{Como } \frac{\partial \ln f}{\partial \theta} = \frac{1}{f} \frac{\partial f}{\partial \theta} \Rightarrow \frac{\partial f}{\partial \theta} = f \frac{\partial \ln f}{\partial \theta} \Rightarrow$$

$$\begin{aligned}
 0 &= \underbrace{\int \left(\frac{\partial^2 \ln f}{\partial \theta^2} \right) f dy}_{E \left(\frac{\partial^2 \ln f}{\partial \theta^2} \right) = E(U')} + \underbrace{\int \left(\frac{\partial \ln f}{\partial \theta} \right)^2 f dy}_{E \left[\left(\frac{\partial \ln f}{\partial \theta} \right)^2 \right] = E(U^2)} \Rightarrow E(U') = -E(U^2)
 \end{aligned}$$

Como $E(U') = -\frac{b''(\theta)}{a(\phi)} = \text{var}(U)$ y también

$$\text{var}(U) = \text{var}\left(\frac{y - b'(\theta)}{a(\phi)}\right) = \text{var}\left(\frac{y}{a(\phi)}\right) = \frac{\text{var}(y)}{a(\phi)^2}$$

Entonces:

$$\frac{\text{var}(y)}{a(\phi)^2} = \frac{b''(\theta)}{a(\phi)} \Rightarrow \text{var}(y) = a(\phi)b''(\theta)$$

A $b''(\theta)$ se le denomina función de varianza - $\text{var}(\mu)$ - por lo que $\text{var}(y)$ también se expresa como: $\text{var}(y) = a(\phi)\text{var}(\mu)$

1.1. Representación en este formato de las distribuciones más habituales

La forma más fácil consiste en tomar neperianos de la función de densidad -o de la de cuantía, si la variable es discreta- e igualar a la expresión entre llaves de la familia exponencial, recogida en la ecuación (4.1). En todos los casos, se trata de encontrar la expresión de θ , $b(\theta)$, $a(\phi)$ y de $c(y, \phi)$, además de la esperanza y la varianza de la variable respuesta. Veamos algunos casos:

Normal general:

tiene como función de densidad:

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

Tomando neperianos e igualando términos queda que:

$$\begin{aligned}\mu &= \theta \\ b(\theta) &= \frac{\mu^2}{2} \\ a(\phi) &= \sigma^2 \Rightarrow \phi = \sigma^2 \\ c(y; \phi) &= -\frac{y^2}{2\phi} - \frac{1}{2} \ln(2\pi\phi)\end{aligned}$$

La esperanza y la varianza de y son:

$$\begin{aligned}E(y) &= b'(\theta) = \mu \\ \text{var}(y) &= a(\phi)b''(\theta) = \sigma^2 \cdot 1 = \sigma^2\end{aligned}$$

Binomial (n, p):

tiene como función de cuantía:

$$P(y, p) = \binom{n}{y} p^y (1-p)^{n-y}$$

Tomando neperianos e igualando términos:

$$\begin{aligned} a(\phi) &= 1 \\ \theta &= \ln\left(\frac{p}{1-p}\right) \\ b(\theta) &= n \ln(1 + e^\theta) \end{aligned}$$

La esperanza y la varianza de y son:

$$\begin{aligned} E(y) &= np \\ \text{var}(y) &= a(\phi) b''(\theta) = np(1-p) \end{aligned}$$

Gamma(α, β):

Habitualmente, esta función de densidad se expresa como:

$$f(y, \alpha, \beta) = \frac{y^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{y}{\beta}} \quad y > 0$$

Igualando términos:

$$\begin{aligned} \mu &= -\frac{1}{\theta} \\ \phi &= \frac{1}{\alpha} \\ b(\theta) &= -\ln(-\theta) \\ c(y, \theta) &= \left(\frac{1}{\phi} - 1\right) \ln y - \ln \Gamma\left(\frac{1}{\phi}\right) + \frac{1}{\phi} \ln\left(\frac{1}{\phi}\right) \end{aligned}$$

La esperanza y la varianza de y son:

$$\begin{aligned} E(y) &= b'(\theta) = \mu = \alpha\beta \\ \text{var}(y) &= a(\phi) b''(\theta) = \alpha\beta^2 \end{aligned}$$

Poisson:

En este caso, la función de cuantía es:

$$f(y; \mu) = e^{-\mu} \frac{\mu^y}{y!}$$

Tomando neperianos e igualando términos:

$$\mu = e^\theta$$

$$a(\phi) = 1$$

$$b(\theta) = \mu = e^\theta$$

$$c(y, \theta) = -\ln y!$$

La esperanza y la varianza de y son:

$$E(y) = b'(\theta) = e^\theta = \mu$$

$$\text{var}(y) = a(\phi)b''(\theta) = e^\theta = \mu$$

Binomial Negativa:

En este caso, la función de cuantía es:

$$P(X = y) = \frac{\Gamma(r+y)}{\Gamma(r)y!} \left(\frac{\beta}{1+\beta} \right)^y \left(\frac{1}{1+\beta} \right)^r$$

que es la correspondiente a la distribución desplazada, siendo:

$$p = \frac{1}{1+\beta}, \quad q = \frac{\beta}{1+\beta}$$

Sea:

$$\mu = \frac{r(1-p)}{p} \quad \text{y} \quad k = \frac{1}{r}$$
$$\Rightarrow \mu p = r - rp \Rightarrow \begin{cases} p = \frac{r}{\mu+r} = \frac{1/k}{\mu+1/k} = \frac{1}{1+\mu k} \\ q = 1-p = 1 - \frac{1}{1+\mu k} = \frac{\mu k}{1+\mu k} \end{cases}$$

Llevando a la función de cuantía y tomando neperianos, agrupando e igualando a la expresión general (4.1) queda:

$$\begin{aligned} \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \theta) &= \\ &= \ln \Gamma(1/k + y) - \ln \Gamma(1/k) - \ln y! + \\ &+ y \ln k + y \ln \left(\frac{\mu}{1 + \mu k} \right) - \frac{1}{k} \ln(1 + \mu k) \\ a(\phi) &= 1 \\ \mu &= \frac{e^\theta}{1 - ke^\theta} \\ b(\theta) &= -\frac{1}{k} \ln(1 - ke^\theta) \\ c(y, \theta) &= \ln \Gamma(1/k + y) - \ln \Gamma(1/k) - \ln y! + y \ln k \end{aligned}$$

La esperanza y la varianza de y son:

$$\begin{aligned} E(y) = b'(\theta) &= -\frac{1}{k} \cdot \frac{-ke^\theta}{1 - ke^\theta} = \frac{e^\theta}{1 - ke^\theta} = \mu \\ \text{var}(y) = a(\phi) b''(\theta) &= \frac{e^\theta}{(1 - ke^\theta)^2} = \frac{\mu}{1 - ke^\theta} \end{aligned}$$

$$\text{como } e^\theta = \frac{\mu}{1 + \mu k} \quad y \quad \mu = \frac{e^\theta}{1 - ke^\theta} \Rightarrow \text{var}(y) = \mu(1 + \mu k)$$

Inversa gaussiana:

En este caso, la función de densidad se expresa como:

$$f(y) = \frac{1}{\sqrt{2\pi y^3 \sigma^2}} \exp \left\{ -\frac{1}{2y} \left(\frac{y - \mu}{\mu \sigma} \right)^2 \right\} \quad y > 0$$

Tomando neperianos, igualando a la expresión general e igualando términos se obtiene que:

$$\begin{aligned} a(\phi) &= \sigma^2 \\ \mu &= \left(-\frac{1}{2\theta} \right)^{1/2} \\ b(\theta) &= -\sqrt{-2\theta} \end{aligned}$$

$$c(y, \theta) = -\frac{1}{2} \ln(2\pi y^3) - \ln \sigma - \frac{1}{2y\sigma^2}$$

La esperanza y la varianza de y son:

$$E(y) = b'(\theta) = \mu$$

$$\text{var}(y) = a(\phi) b''(\theta) = \sigma^2 \mu^3$$

1.2. Utilización del modelo anterior

Con lo visto hasta ahora, un *GLM* puede expresarse como:

$$f(y, \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \theta) \right\} \quad (4.2)$$

$$g(\mu) = \mathbf{X}'\boldsymbol{\beta} \quad (4.3)$$

es decir, (4.2) indica la distribución que sigue la variable dependiente, dentro de la familia exponencial, mientras que (4.3) indica la relación lineal que existe entre las variables independientes y la transformación de la media, $g(\mu)$.

Hay que tener en cuenta que:

- a) La elección de $b(\theta)$ determina la distribución de la variable respuesta.
- b) La elección de $g(\mu)$ determina la relación entre la media y las variables independientes. En un modelo lineal habitual se tiene que $\mu = \mathbf{X}'\boldsymbol{\beta}$, mientras que en un *GLM*, $g(\mu)$ es una función monótona y diferenciable, como puede ser el logaritmo neperiano o la raíz cuadrada.
- c) Las dos ecuaciones anteriores implican que:
 - dada X , μ se fija a través de $g(\mu)$
 - dada μ , θ se determina como $b'(\theta) = \mu$
 - dado θ , la variable dependiente y se determina como una consecuencia a partir de la densidad exponencial especificada en $b(\theta)$.
- d) Se supone independencia entre todos los valores de la variable dependiente, y_i .

Los pasos que habitualmente se siguen para la construcción de un *GLM* son los siguientes:

- 1) Se selecciona una función de distribución para la variable y , por tanto, se escoge $b(\theta)$. La función depende de los datos de cada situación.
- 2) Se selecciona un link $g(\mu)$. La elección, a veces, se simplifica sin más que seleccionar el denominado *link canónico* o *estándar*, el cual se verá posteriormente qué es.

- 3) Selección de las variables explicativas, a partir de las que se modelará $g(\mu)$
- 4) Obtención de los datos y_1, y_2, \dots, y_n y las correspondientes x_{ij} ($i = 1 \dots n, j = 1 \dots k$)
- 5) Ajuste del modelo para lo cual se estiman las β y ϕ si fuera desconocida.
- 6) A partir de los valores estimados de β se obtienen los valores ajustados de la variable dependiente y se evalúa el grado de ajuste del modelo obtenido.

Comparando con una regresión habitual, los pasos adicionales son los numerados como 1) y 2). La elección de $b(\theta)$ depende de la naturaleza de la variable dependiente mientras que la de $g(\mu)$ depende de la relación funcional entre la variable dependiente y las independientes.

La secuencia anterior de fases rara vez se sigue estrictamente. Por ejemplo, en el caso de seguros, la obtención de los datos es anterior a la formulación del modelo. Además, la realización de un análisis previo de los datos puede sugerir diferentes modelos o distintas distribuciones para la variable dependiente. Los ajustes se siguen de distintos procedimientos que permiten decidir la inclusión, exclusión o transformación de una o varias de las variables explicativas. Todo ello a su vez acompañado de una juiciosa selección o exclusión de datos que permitan llevar a cabo un ajuste aceptable.

1.3. Enlaces y complementos añadidos (*links* y *offsets*)

Ya se ha indicado que el link es una función que une la media de la variable dependiente con la combinación lineal de variables explicativas seleccionadas. Las funciones más habituales son las que se recogen en la tabla 1:

Tabla 1: Links más frecuentes

Función	$g(\mu)$	Canónico para
Identidad	μ	Normal
logaritmo neperiano	$\ln \mu$	Poisson
Potencia	μ^p	Gamma ($p = -1$) Inversa gaussiana ($p = -2$)
Raíz cuadrada	$\mu^{1/2}$	---
Logit	$\ln(p/\mu - p)$	Binomial

Cuando $g(\mu) = \theta$, entonces se dice que es el enlace canónico correspondiente a $b(\theta)$. En este caso, $\theta = X'\beta$. Las constantes se suelen omitir en el *link* canónico. Por ejemplo, en el caso de la Inversa gaussiana, sabemos que $\theta = -1/2\mu^2$ mientras que el canónico utilizado es μ^{-2} .

Complementos añadidos

La modelación de eventos tales como el número de siniestros o de muertos en un grupo de riesgo exige correcciones en el número de exposiciones al riesgo. Si μ es la media de y , entonces la tasa de ocurrencia es μ/n y se tiene que $g\left(\frac{\mu}{n}\right) = \mathbf{X}'\boldsymbol{\beta}$. Si, por ejemplo, g fuese la función logarítmica, entonces:

$$\ln\left(\frac{\mu}{n}\right) = \mathbf{X}'\boldsymbol{\beta} \Rightarrow \ln \mu = \ln n + \mathbf{X}'\boldsymbol{\beta}$$

A la variable n se le denomina *exposición* y a $\ln n$ se le conoce como *offset* (complemento), el cual se puede definir como otra variable en la regresión con β unitario. Con un *offset*, el valor esperado de y es proporcional a la exposición:

$$\text{si } \ln \mu = \ln n + \mathbf{X}'\boldsymbol{\beta} \Rightarrow \mu = ne^{\mathbf{X}'\boldsymbol{\beta}}$$

Este tipo de operadores se utiliza para corregir los datos por factores tales como el tamaño de los grupos o por diferimientos en el tiempo.

2. ESTIMACIÓN DEL MODELO

El proceso de estimación de un *GLM* se puede realizar tanto de manera analítica como numérica, pues no siempre se obtendrán ecuaciones lineales en los parámetros a estimar. Para ilustrar esta cuestión, veamos dos casos, en el que la estimación máximo verosímil obtiene expresiones que, en un caso, son lineales en β_j y en el otro no es así.

Ajuste del modelo normal

Sea el modelo $E(Y) = \mathbf{X}\boldsymbol{\beta}$. Entonces:

$$l(\theta; y) = \sum_{i=1}^n \frac{y_i \theta - b(\theta)}{a_i(\phi)} + \sum_{i=1}^n c(y_i, \theta)$$

Sabemos que, en el caso de la distribución normal, se tiene:

$$\theta = \mu = \mathbf{X}\boldsymbol{\beta}$$

$$b(\theta) = \frac{\mu^2}{2}$$

$$a_i(\phi) = \sigma^2$$

$$c(y_i; \phi) = -\frac{y_i^2}{2\phi} - \frac{1}{2} \ln(2\pi\phi)$$

con lo que:

$$l(\theta; y) = \sum_{i=1}^n \left[\frac{y_i x_i \beta - (x_i \beta)^2 / 2}{\sigma^2} - \frac{y_i^2}{2\phi} - \frac{1}{2} \ln(2\pi\sigma^2) \right]$$

derivando y despejando se obtiene que:

$$\begin{aligned} \frac{\partial l(\theta; y)}{\partial \beta_j} &= \sum_{i=1}^n \frac{y_i x_{ij} - x_i' \beta x_{ij}}{\sigma^2} = 0 \Rightarrow \sum_{i=1}^n y_i x_{ij} - x_i' \beta x_{ij} = 0 \\ &\Rightarrow \mathbf{X}'(\mathbf{Y} - \mathbf{X}\beta) = \mathbf{0} \rightarrow \mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\beta \\ &\Rightarrow \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \end{aligned}$$

que es una expresión lineal en β .

Ajuste de un modelo Poisson

Sea el modelo $\ln[E(Y)] = \mathbf{X}\beta$. Como $E(y_i) = \mu_i = e^{\theta_i} \Rightarrow \ln \mu_i = \theta_i$ y, por tanto:

$$l(\beta; y) = \sum_{i=1}^n y_i (\beta_0 + \beta_1 x_i) - \exp(\beta_0 + \beta_1 x_i)$$

Derivando y despejando:

$$\frac{\partial l(\beta; y)}{\partial \beta_0} = \sum_{i=1}^n y_i - \exp(\beta_0 + \beta_1 x_i) = 0 \rightarrow \sum_{i=1}^n y_i = n\bar{y} = \sum_{i=1}^n \exp(\beta_0 + \beta_1 x_i)$$

$$\Rightarrow n\bar{y} = e^{\beta_0} \sum_{i=1}^n \exp(\beta_1 x_i)$$

$$\frac{\partial l(\beta; y)}{\partial \beta_1} = \sum_{i=1}^n x_i y_i - x_i \exp(\beta_0 + \beta_1 x_i) = 0$$

$$\Rightarrow \sum_{i=1}^n x_i y_i = e^{\beta_0} \sum_{i=1}^n x_i \exp(\beta_1 x_i)$$

Ninguna de las dos ecuaciones son lineales en β , por lo que se hace necesaria su estimación mediante métodos numéricos.

2.1. Estimación por máxima verosimilitud

Sea un conjunto de datos $\{y_1, y_2, \dots, y_n\}$ que siguen una de las distribuciones de la familia exponencial, es decir: $y_i \sim f(\theta_i; \phi)$, cuyo neperiano de la función de verosimilitud se puede expresar como:

$$l(\theta; y) = \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + \sum_{i=1}^n c(y_i, \theta) \quad (4.4)$$

donde θ_i es el parámetro canónico, relacionado con $\mu_i = E(y_i)$ y siendo ϕ un parámetro de escala. Se pretende realizar el ajuste a un modelo lineal generalizado del tipo:

$$g(\mu_i) = \eta_i = \mathbf{x}_i' \beta \quad / \quad \beta \in \mathbb{R}^k$$

Para obtener la estimación máximo verosímil, se optimiza (4.4). Teniendo en cuenta que las y_i son independientes, se obtiene el siguiente conjunto de condiciones necesarias:

$$\frac{\partial l(\theta; y_i)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i(\theta; y_i)}{\partial \beta_j} = 0 \quad \forall j = 1, 2, \dots, k$$

Para obtener cada derivada parcial, es necesario tener en cuenta que:

1. existe una relación entre β_j y η_i de la forma: $\eta_i = \mathbf{x}_i' \beta$
2. existe una relación entre η_i y $g(\mu_i)$
3. a su vez, μ_i está ligada con el parámetro canónico θ_i de la forma: $b'(\theta_i) = \mu_i$
4. finalmente, θ_i está unido a $l(\theta_i, y_i)$ mediante la expresión:

$$l(\theta_i; y_i) = \frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)$$

Por tanto, la parcial se obtendrá aplicando la regla de la cadena del siguiente modo:

$$\frac{\partial l(\theta; y_i)}{\partial \beta_j} = \frac{\partial l(\theta; y_i)}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \quad (4.5)$$

Veamos, a continuación, cuánto vale cada miembro del lado derecho:

- $\frac{\partial l(\theta; y_i)}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a_i(\phi)}$, pero $b'(\theta_i) = E(y_i) = \mu_i$, con lo que queda:

$$\frac{\partial l(\theta; y_i)}{\partial \theta_i} = \frac{y_i - \mu_i}{a_i(\phi)}$$

- se sabe que $\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i)$. Asumiendo que existe la función inversa, se tiene

que:

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b''(\theta_i)}$$

- se sabe que $g(\mu_i) = \eta_i \rightarrow \frac{\partial g(\mu_i)}{\partial \mu_i} = g'(\mu_i) = \frac{\partial \eta_i}{\partial \mu_i}$. Al igual que en el caso anterior, asumiendo que existe función inversa, entonces:

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{g'(\mu_i)}$$

- finalmente, como $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} \rightarrow \frac{\partial \eta_i}{\partial \beta_j} = x_{ij}$

Uniendo todas las expresiones, resulta:

$$\frac{\partial l(\theta; y_i)}{\partial \beta_j} = \frac{y_i - \mu_i}{a_i(\phi)} \frac{1}{b''(\theta_i)} \frac{1}{g'(\mu_i)} x_{ij} = \frac{(y_i - \mu_i) x_{ij}}{\underbrace{a_i(\phi) b''(\theta_i)}_{\text{var}(y_i)} g'(\mu_i)}$$

Por lo que la expresión final de la condición necesaria es:

$$\sum_{i=1}^n \frac{\partial l_i(\theta; y_i)}{\partial \beta_j} = \sum_{i=1}^n \frac{(y_i - \mu_i) x_{ij}}{\underbrace{a_i(\phi) b''(\theta_i)}_{\text{var}(y_i)} g'(\mu_i)} = 0$$

Como caso especial tenemos aquél en el que $g'(\mu_i)$ es el link canónico, es decir, $g(\mu_i) = \theta_i$. Entonces se cumple que $\frac{\partial g(\mu_i)}{\partial \mu_i} = g'(\mu_i) = \frac{\partial \theta_i}{\partial \mu_i}$ y, por tanto,

$\frac{1}{g'(\mu_i)} = \frac{\partial \mu_i}{\partial \theta_i}$ que es la expresión de $\text{var}(y_i)$, con lo que queda que:

$$\begin{aligned} \sum_{i=1}^n \frac{(y_i - \mu_i) x_{ij}}{\text{var}(y_i) g'(\mu_i)} = 0 &\rightarrow \sum_{i=1}^n \frac{y_i x_{ij}}{\text{var}(y_i) g'(\mu_i)} = \sum_{i=1}^n \frac{\mu_i x_{ij}}{\text{var}(y_i) g'(\mu_i)} \\ &\Rightarrow \sum_{i=1}^n \frac{y_i x_{ij} \text{var}(\mu_i)}{\text{var}(y_i)} = \sum_{i=1}^n \frac{\mu_i x_{ij} \text{var}(\mu_i)}{\text{var}(y_i)} \end{aligned}$$

A su vez, como $\text{var}(y_i) = a_i(\phi) \text{var}(\mu_i)$, queda finalmente que:

$$\sum_{i=1}^n \frac{y_i x_{ij}}{a_i(\phi)} = \sum_{i=1}^n \frac{\mu_i x_{ij}}{a_i(\phi)}$$

2.2. Estimación mediante el uso de métodos numéricos

Se utilizan en aquellos casos en los que no es posible despejar β a partir de $\frac{\partial l}{\partial \beta}$, lo cual obliga a utilizar algoritmos, como por ejemplo, el de Newton-Raphson. Este método no es más que una aproximación lineal mediante el desarrollo en serie de Taylor de la derivada del logaritmo de la función de verosimilitud. Se trata de resolver $L'(\beta) = 0$:

$$0 = L'(\beta) = L'[\beta(0)] + [\beta - \beta(0)]L''[\beta(0)] + \frac{[\beta - \beta(0)]^2}{2!}L'''[\beta(0)] + \dots$$

$$0 \approx L'[\beta(0)] + [\beta - \beta(0)]L''[\beta(0)]$$

Por ser $L'[\beta(0)]$ un vector fila -vector gradiente-, entonces:

$$\beta = \beta(0) - \{L''[\beta(0)]\}^{-1} L'[\beta(0)]$$

luego, para la iteración r -ésima, será:

$$\beta^{(r)} = \beta^{(r-1)} - \{L''[\beta^{(r-1)}]\}^{-1} L'[\beta^{(r-1)}]$$

Esta aproximación es exacta si la función de verosimilitud es analítica, tal y como ocurre en la regresión lineal clásica (varianza σ^2 y link identidad), en cuyo caso, sólo se necesita una iteración.

El elemento crucial de la expresión anterior es el cálculo de la matriz de segundas derivadas. Para ello, existen dos posibilidades:

a) *Uso de la matriz hessiana observada*

De forma general:

$$\frac{\partial^2 l}{\partial \beta_i \partial \beta_j} = \sum_i \frac{1}{a(\phi)} \frac{\partial}{\partial \beta_k} \left\{ \frac{y_i - \mu_i}{\text{var}(\mu_i)} \left(\frac{\partial \mu}{\partial \eta} \right)_i x_{ij} \right\}$$

Para obtener la expresión de la derivada de lo situado entre llaves se opera como si fuera un producto en donde el primer elemento es $\frac{y_i - \mu_i}{\text{var}(\mu_i)}$ y el segundo el resto.

Operando se llega a:

$$\frac{\partial^2 L}{\partial \beta_i \partial \beta_j} = - \sum_i \frac{1}{a(\phi)} \left\{ \frac{1}{\text{var}(\mu_i)} \left(\frac{\partial \mu}{\partial \eta} \right)_i^2 - (\mu_i - y_i) \left[- \frac{1}{\text{var}(\mu_i)} \frac{\partial^2 \mu_i}{\partial \eta_i^2} + \frac{\partial \text{var}(\mu_i)}{\partial \mu_i} \left(\frac{\partial \mu}{\partial \eta} \right)_i \frac{1}{[\text{var}(\mu_i)]^2} \right] \right\} x_{ij} x_{ik}$$

b) *Uso de la matriz hessiana esperada y su relación con los mínimos cuadrados ponderados*

En este caso, la expresión de $\hat{\beta}^{(r)}$ es:

$$\hat{\beta}^{(r)} = \hat{\beta}^{(r-1)} + \mathbf{I}_{\hat{\beta}^{(r-1)}}^{-1} \times \mathbf{L}' \left[\hat{\beta}^{(r-1)} \right]$$

Siendo: $\mathbf{I}_{\hat{\beta}^{(r-1)}}^{-1} = -E(\mathbf{H}) = \sum_{i=1}^n E \left(- \frac{\partial^2 l}{\partial \beta_j \partial \beta_k} \right)_{jk}$

A partir de esta última expresión se puede obtener una expresión del estimador de β basado en mínimos cuadrados ponderados. Para ello, como:

$$E \left(\frac{\partial^2 l}{\partial \beta_i \partial \beta_j} \right) + E \left(\frac{\partial l}{\partial \beta_i} \cdot \frac{\partial l}{\partial \beta_j} \right) = 0 \Rightarrow -E \left(\frac{\partial^2 l}{\partial \beta_i \partial \beta_j} \right) = E \left(\frac{\partial l}{\partial \beta_i} \cdot \frac{\partial l}{\partial \beta_j} \right)$$

y como:

$$\begin{aligned} \frac{\partial l_i}{\partial \beta_i} &= \frac{(y_i - \mu_i)}{a_i(\phi) \text{var}(\mu_i)} \left(\frac{\partial \mu}{\partial \eta} \right)_i x_{ij} \\ \Rightarrow E \left(\frac{\partial l}{\partial \beta_i} \cdot \frac{\partial l}{\partial \beta_j} \right) &= \sum_{i=1}^n \frac{x_{ij} x_{ik}}{g'(\mu_i)^2 \text{var}(y_i)^2} \end{aligned}$$

Por tanto, si:

$$\Delta \beta^{(r-1)} = - \left\{ \frac{\partial^2 l}{\partial \beta^{(r-1)'} \partial \beta^{(r-1)}} \right\} \frac{\partial l}{\partial \beta^{(r-1)}}$$

y se sustituye la matriz hessiana por su esperanza, es decir, $\frac{\partial l}{\partial \beta^{(r-1)}}$ por

$$\sum_{i=1}^n \frac{(y_i - \mu_i)}{\text{var}(\mu_i)} \left(\frac{\partial \mu}{\partial \eta} \right)_i x_{ij}, \text{ queda:}$$

$$\Delta\beta^{(r-1)} = \left\{ \sum_{i=1}^n \frac{x_{ij}x_{ik}}{g'(\mu_i)^2 \text{var}(y_i)} \right\}^{-1} \sum_{i=1}^n \frac{(y_i - \mu_i) \left(\frac{\partial \mu}{\partial \eta} \right)_i}{\text{var}(\mu_i) \left(\frac{\partial \mu}{\partial \eta} \right)_i} x_{ij}$$

$$\Rightarrow \left\{ \sum_{i=1}^n \frac{x_{ij}x_{ik}}{g'(\mu_i)^2 \text{var}(y_i)} \right\} \Delta\beta^{(r-1)} = \sum_{i=1}^n \frac{(y_i - \mu_i) \left(\frac{\partial \mu}{\partial \eta} \right)_i}{\text{var}(\mu_i) \left(\frac{\partial \mu}{\partial \eta} \right)_i} x'_i \quad (4.6)$$

Como de forma general, $\eta_i = \sum_{k=1}^p x_{ik}\beta_k + \text{offset}$, se tiene que, para en la

iteración $r-1$ se obtiene que $\eta_i^{r-1} - \text{offset} = \sum_{k=1}^p x_{ik}\beta_k^{r-1}$, por lo que:

$$\left\{ \sum_{i=1}^n \frac{x_{ij}x_{ik}}{a(\phi) \text{var}(\mu_i) \left(\frac{\partial \mu}{\partial \eta} \right)_i^2} \right\} \beta^{(r-1)} = \sum_{i=1}^n \frac{1}{a(\phi) \text{var}(\mu_i) \left(\frac{\partial \mu}{\partial \eta} \right)_i^2} \left(\eta_i^{(r-1)} - \text{offset}_i \right) x'_i \quad (4.7)$$

Sumando (4.6) y (4.7) queda que:

$$\left\{ \sum_{i=1}^n \frac{x_{ij}x_{ik}}{a(\phi) \text{var}(\mu_i) \left(\frac{\partial \mu}{\partial \eta} \right)_i^2} \right\} \frac{\Delta\beta^{(r-1)} + \beta^{(r-1)}}{\beta^{(r)}} =$$

$$\sum_{i=1}^n \frac{1}{a(\phi) \text{var}(\mu_i) \left(\frac{\partial \mu}{\partial \eta} \right)_i^2} \left\{ (y_i - \mu_i) \left(\frac{\partial \mu}{\partial \eta} \right)_i + \left(\eta_i^{(r-1)} - \text{offset}_i \right) \right\} x'_i$$

Por otra parte, sean:

$$\mathbf{W}^{(r-1)} = \text{diag} \left\{ \frac{1}{a(\phi) \text{var}(\mu_i) \left(\frac{\partial \mu}{\partial \eta} \right)_i^2} \right\} \quad \text{matriz de orden } (n \times r)$$

$$\mathbf{Z}^{(r-1)} = \left\{ (y - \mu) \left(\frac{\partial \eta}{\partial \mu} \right) + \left(\eta^{(r-1)} - \text{offset} \right) \right\} \quad \text{matriz de orden } (n \times 1)$$

Sustituyendo en lo anterior se obtiene que:

$$\left(\mathbf{X}' \mathbf{W}^{(r-1)} \mathbf{X} \right) \beta^{(r)} = \mathbf{X}' \mathbf{W}^{(r-1)} \mathbf{Z}^{(r-1)}$$

$$\Rightarrow \beta^{(r)} = \left(\mathbf{X}' \mathbf{W}^{(r-1)} \mathbf{X} \right)^{-1} \mathbf{X}' \mathbf{W}^{(r-1)} \mathbf{Z}^{(r-1)}$$

que indica que un nuevo estimador del vector β se puede obtener mediante mínimos cuadrados ponderados. A diferencia de los ponderados habituales, en este caso es necesario iterar, de ahí que se les denomine mínimos cuadrados ponderados iterados (MCP I).

Ejemplo de MCPI

Para entender mejor este sistema de estimación, vamos a ajustar un conjunto de datos mediante una regresión de Poisson. En este caso, se sabe que $E(y_i) = \mu_i$ y que $var(y_i) = \mu_i$, siendo $E(y_i) = \beta_1 + \beta_2 x_{2i}$, por lo que en este caso se toma como enlace la función identidad, es decir, $g(\mu_i) = \mu_i = \eta_i = \beta_1 + \beta_2 x_{2i}$, con lo que:

$$w_{ii} = \text{diag} \left\{ \frac{1}{var(y_i)} \left(\frac{\partial \mu}{\partial \eta} \right)_i^2 \right\} \quad / \quad \left(\frac{\partial \mu}{\partial \eta} \right)_i = 1$$

$$= \frac{1}{\beta_1 + \beta_2 x_{2i}}$$

Por otra parte:

$$z = \left\{ (y - \mu) \left(\frac{\partial \mu}{\partial \eta} \right) + (\eta - \text{offset}) \right\}$$

En este caso, $\left(\frac{\partial \eta}{\partial \mu} \right) = 1$, con lo que:

$$y_i - \mu_i = y_i - \beta_1 + \beta_2 x_{2i}$$

$$\eta_i - \text{offset} = \beta_1 + \beta_2 x_{2i}$$

y, por tanto:

$$z_i = y_i - \beta_1 - \beta_2 x_{2i} + \beta_1 + \beta_2 x_{2i} = y_i$$

Como $\hat{\beta}^{(r)} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}\mathbf{Z}$, queda que:

$$\mathbf{X}'\mathbf{W}\mathbf{X} = \begin{bmatrix} \sum \frac{1}{\beta_1 + \beta_2 x_{2i}} & \sum \frac{x_{2i}}{\beta_1 + \beta_2 x_{2i}} \\ \sum \frac{x_{2i}}{\beta_1 + \beta_2 x_{2i}} & \sum \frac{x_{2i}^2}{\beta_1 + \beta_2 x_{2i}} \end{bmatrix}$$

$$\mathbf{X}'\mathbf{W}\mathbf{Z} = \begin{bmatrix} \sum z_i w_{ii} \\ \sum z_i x_{2i} w_{ii} \end{bmatrix} = \begin{bmatrix} \sum \frac{y_i}{\beta_1 + \beta_2 x_{2i}} \\ \sum \frac{y_i x_{2i}}{\beta_1 + \beta_2 x_{2i}} \end{bmatrix}$$

Los datos de este ejemplo son los siguientes:

X	1	2	3	4	5	6	7	8	9
Y	70	88	97	91	104	110	113	145	159

Como cualquier valor del vector β depende del obtenido en la iteración anterior, es preciso introducir alguna condición inicial para iniciar el proceso. Se toma como $\hat{\beta}^{(0)}$ los obtenidos mediante el uso de MCO, es decir:

$$\hat{\beta}^{(0)} = \begin{bmatrix} \hat{\beta}_1^{(0)} \\ \hat{\beta}_2^{(0)} \end{bmatrix}$$

En la iteración 1 tenemos que:

$$\mathbf{X}'\mathbf{W}\mathbf{X} = \begin{bmatrix} 0,0875 & 0,3826 \\ 0,3826 & 2,2481 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{W}\mathbf{Z} = \begin{bmatrix} 9,0209 \\ 44,8727 \end{bmatrix}$$

Operando se obtiene que: $\hat{\beta}^{(1)} = \begin{bmatrix} 61,7344 \\ 9,4531 \end{bmatrix}$

El siguiente cuadro recoge los resultados hasta su convergencia final:

	iteración número				
	0	1	2	3	4
$\hat{\beta}_1$	59,83	61,73	61,67	61,68	61,68
$\hat{\beta}_2$	9,83	9,45	9,47	9,47	9,47

siendo la matriz de varianzas-covarianzas la inversa de la matriz de información, es decir:

$$\Gamma^{-1} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} = \begin{bmatrix} 45,1876 & -7,6669 \\ 7,6669 & 1,7437 \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{\sigma}_{\hat{\beta}_1} \\ \hat{\sigma}_{\hat{\beta}_2} \end{bmatrix} = \begin{bmatrix} 6,72 \\ 1,32 \end{bmatrix}$$

3. INFERENCIA CON GLM

3.1. Fundamentos teóricos

Sean y_1, y_2, \dots, y_N observaciones, todas ellas independientes entre sí, entre las que se verifica que $E(y_i) = \mu_i$ y que $g(\mu_i) = \mathbf{x}'\beta = \eta_i$. Como:

$$U_j = \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^n \frac{(y_i - \mu_i)}{\text{var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \quad \forall j = 1, \dots, k$$

y como $E(y_i) = \mu_i$, entonces $E(U_j) = 0$ para cualquiera que sea j . La matriz de varianzas-covarianzas de U es la matriz de información \mathfrak{I} , en la que sus elementos constitutivos son $\mathfrak{I}_{i,j} = E(U_i U_j)$.

Si sólo hay que estimar un parámetro, entonces la distribución asintótica del estadístico U es:

$$\frac{U - E(U)}{\sqrt{\text{var}(U)}} = \frac{U}{\sqrt{\mathfrak{I}}} \approx N(0,1)$$

o también: $\frac{U^2}{\mathfrak{I}} \approx \chi^2_{(1)}$

Si fuese un vector de parámetros $\beta = [\beta_1 \ . \ . \ . \ \beta_k]'$, entonces el vector de estadísticos $U = [U_1 \ . \ . \ . \ U_k]'$ se distribuiría asintóticamente según una normal multivariante $N(0, \mathfrak{I})$ y por tanto, para muestras de gran tamaño se cumplirá que: $U' \mathfrak{I}^{-1} U \approx \chi^2_{(k)}$. Veamos a continuación dos ejemplos:

Estadístico U para la distribución normal

En este caso, $y_i \approx N(\mu_i, \sigma)$ siendo σ desconocida. El neperiano de la función de verosimilitud será:

$$l = -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_i)^2 - N \ln(\sigma\sqrt{2\pi})$$

y el estadístico U será:

$$U = \frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu_i)$$

o, de otra forma:

$$U = \frac{n(\bar{y} - \mu)}{\sigma^2}$$

Como $E(y_i) = \mu_i$, entonces $E(U) = 0$. Por lo que se refiere a la varianza:

$$\text{var}(U) = \text{var}\left(\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu_i)\right) = \text{var}\left[\frac{n(\bar{y} - \mu)}{\sigma^2}\right] = \frac{n^2}{\sigma^4} \text{var}(\bar{y}) = \frac{n^2}{\sigma^4} \frac{\sigma^2}{n} = \frac{n}{\sigma^2}$$

Por tanto:

$$\frac{U - E(U)}{\sqrt{\text{var}(U)}} = \frac{n(\bar{y} - \mu)}{\frac{\sigma^2}{\sqrt{\frac{n}{\sigma^2}}}} = (\bar{y} - \mu) \sqrt{\frac{n}{\sigma^2}} = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}}$$

Con una normal estándar como distribución asintótica, o bien:

$$\frac{(\bar{y} - \mu)^2}{\sigma^2/n} \approx \chi_{(1)}^2$$

Estadístico U para la distribución binomial

En este caso, si $y_i \approx B(1, p)$, entonces el conjunto de observaciones y_1, y_2, \dots, y_N sigue un $B(n, p)$ y el neperiano de la función de verosimilitud será:

$$l(y_1, \dots, y_n; p) = \ln \binom{n}{y} + y \ln p + (n - y) \ln(1 - p)$$

con lo que U es:

$$U = \frac{\partial l}{\partial \mu} = \frac{y}{p} - (n - y) \frac{1}{1 - p} = \frac{y - np}{p(1 - p)}$$

$$E(U) = \frac{E(y) - np}{p(1 - p)} \quad \text{como } E(y) = np \Rightarrow E(U) = 0$$

$$\text{var}(U) = \frac{\text{var}(y - np)}{p^2(1 - p)^2} = \frac{\text{var}(y)}{p^2(1 - p)^2} = \frac{np(1 - p)}{p^2(1 - p)^2} = \frac{n}{p(1 - p)}$$

$$\Rightarrow \frac{U - E(U)}{\sqrt{\text{var}(U)}} = \frac{\frac{y - np}{p(1 - p)}}{\sqrt{\frac{n}{p(1 - p)}}} = \frac{(y - np) \sqrt{p(1 - p)}}{p(1 - p) \sqrt{n}} = \frac{y - np}{\sqrt{np(1 - p)}} \approx N(0, 1)$$

A veces es útil usar desarrollos en serie de Taylor para obtener distribuciones asintóticas. Para el caso del neperiano de la función de verosimilitud con un único parámetro β , los primeros términos del desarrollo en serie son:

$$l(\beta) \approx l(\hat{\beta}) + (\beta - \hat{\beta}) \underbrace{\frac{dl}{d\beta}}_{U(\hat{\beta})} + \frac{1}{2} (\beta - \hat{\beta})^2 \underbrace{\frac{d^2l}{d\beta^2}}_{U'(\hat{\beta})}$$

Siendo $\hat{\beta}$ un estimador de β . Si $U' = \frac{d^2 l}{d\beta^2}$ se aproxima mediante su esperanza, $E(U') = -\mathfrak{I}$, entonces:

$$l(\beta) \approx l(\hat{\beta}) + (\beta - \hat{\beta}) \frac{dl}{d\beta} \Big|_{U(\hat{\beta})} - \frac{1}{2} (\beta - \hat{\beta})^2 \mathfrak{I}(\hat{\beta})$$

Si en vez de un único parámetro es un vector, entonces:

$$l(\beta) \approx l(\hat{\beta}) + (\beta - \hat{\beta})' U(\hat{\beta}) - \frac{1}{2} (\beta - \hat{\beta})' \mathfrak{I}(\hat{\beta}) (\beta - \hat{\beta})$$

siendo U el vector de primeras derivadas y \mathfrak{I} la matriz de información. Haciendo este mismo desarrollo para la función score $-U$ se obtiene:

$$U(\beta) = U(\hat{\beta}) + (\beta - \hat{\beta}) U'(\hat{\beta})$$

y si $U'(\hat{\beta}) = -\mathfrak{I}(\hat{\beta})$, entonces: $U(\beta) = U(\hat{\beta}) - (\beta - \hat{\beta}) \mathfrak{I}(\hat{\beta})$ que, en notación vectorial es:

$$U(\beta) = U(\hat{\beta}) - \mathfrak{I}(\hat{\beta}) (\beta - \hat{\beta})$$

3.2. La ratio entre neperianos de verosimilitud

Una forma de evaluar si un modelo es o no adecuado es mediante la comparación con otro en el que se estima el número máximo posible de parámetros -tantos como observaciones haya-. A este último se le conoce como *modelo saturado, máximo o completo*, al cual se le exige que, para poder realizar la comparación, comparta con el modelo que se desea evaluar, tanto la distribución como la función de enlace.

Sea m el número máximo de parámetros a estimar, β_{max} el vector de parámetros del modelo saturado, $\hat{\beta}_{max}$ su vector de estimadores y sea $L(\hat{\beta}_{max}; y)$ la función de verosimilitud evaluada en $\hat{\beta}_{max}$, siendo aquí donde alcanza el mayor valor posible. Este valor se compara con el que logra la citada función en el modelo que se evalúa. Se define la ratio de verosimilitud como:

$$\lambda = \frac{L(\hat{\beta}_{max}; y)}{L(\hat{\beta}; y)}$$

y es una forma de evaluar la calidad del ajuste. Puesto en forma de neperianos, se tiene que $\ln \lambda = \ln l(\hat{\beta}_{max}; y) - \ln l(\hat{\beta}; y)$, de modo que, cuanto mayor es el valor de este estadístico, peor es la calidad del ajuste.

En la práctica, más que el $\ln \lambda$ se utiliza el doble de éste, medida conocida como *deviance* -que traduciremos como divergencia, para no confundir con la tradicional desviación-. No obstante, para determinar con precisión la región crítica ligada a este estadístico, es necesario conocer la función de distribución subyacente.

Sea $D = 2 \left[\ln l(\hat{\beta}_{max}; y) - \ln l(\hat{\beta}; y) \right]$. A partir del desarrollo en serie de

Taylor de $\ln(\beta)$, si $\hat{\beta}$ es EMV de β , entonces $U(\hat{\beta}) = 0$, con lo que:

$$2 \left[\ln(\hat{\beta}) - \ln(\beta) \right] = (\beta - \hat{\beta})' \mathfrak{I}(\hat{\beta})(\beta - \hat{\beta}) \approx \chi_{(k)}^2$$

en donde k es el número de parámetros. A partir de este resultado, se puede obtener la distribución del estadístico D . Sumando y restando $\ln l(\beta_{max}; y)$ y $\ln l(\beta; y)$ queda:

$$D = 2 \left[\underbrace{\ln(\hat{\beta}_{max}; y) - \ln(\beta_{max}; y)}_{\chi_{(m)}^2} \right] - 2 \left[\underbrace{\ln(\hat{\beta}; y) - \ln(\beta; y)}_{\chi_{(k)}^2} \right] + 2 \left[\underbrace{\ln(\beta_{max}; y) - \ln(\beta; y)}_{\substack{v = \text{constante positivo} \rightarrow 0 \\ \text{si el modelo ajusta bien.}}} \right]$$

por lo que $D \approx \chi_{(m-k, v)}^2$ siendo v el parámetro de no centralidad²².

Una forma de generalizar este concepto es mediante la expresión:

$$S = \frac{D}{\phi}$$

²² Sean Z_1, Z_2, \dots, Z_n variables aleatorias independientes e idénticamente distribuidas, tales que cada una de ellas es una normal estándar, y sea $Y_i = Z_i + \mu_i \forall i = 1, \dots, n$, en donde, al menos, un μ_i es distinto de cero. A partir de estos datos, se calcula la suma de todos los Y_i al cuadrado. Se dice entonces que $\sum_i Y_i^2$ sigue una distribución chi cuadrado no centrada con n grados de libertad y parámetro λ de no centralidad, siendo su esperanza y su varianza:

$$E \left(\sum_i Y_i^2 \right) = E \left(\sum_i Z_i^2 + \sum_i \mu_i^2 + 2 \sum_i Z_i \mu_i \right) = n + \lambda \quad / \quad \lambda = \sum_i \mu_i^2$$

$$\text{var} \left(\sum_i Y_i^2 \right) = \text{var} \left(\sum_i Z_i^2 + \sum_i \mu_i^2 + 2 \sum_i Z_i \mu_i \right) = 2n + 4 \sum_i \mu_i^2 = 2n + 4\lambda$$

Siendo D el estadístico de divergencia y ϕ el factor de escala. De forma general, el estadístico D se calcula como:

$$\begin{aligned} D &= 2 \left[\ln l(\hat{\beta}_{max}; y) - \ln l(\hat{\beta}; y) \right] \\ &= 2 \left[\frac{y\hat{\beta}_{max} - b(\hat{\beta}_{max})}{a(\phi)} + c(y; \phi) - \frac{y\hat{\beta} - b(\hat{\beta})}{a(\phi)} - c(y; \phi) \right] \\ &= 2 \sum \frac{y[\hat{\theta}_{max} - \hat{\theta}] - b(\hat{\theta}_{max}) + b(\hat{\theta})}{a(\phi)} \end{aligned}$$

Ejemplos:

Estadístico D para un modelo binomial

Si $Y_i \approx B(n_i; p)$ entonces:

$$\ln(\beta; y) = \sum_{i=1}^N \left[y_i \ln p_i - y_i \ln(1-p_i) + n_i \ln(1-p_i) + \ln \binom{n_i}{y_i} \right] \quad / \quad N = \sum n_i$$

En el caso del modelo saturado, habrá tantas probabilidades distintas como observaciones se tengan, por lo que: $\beta = [p_1, p_2, \dots, p_n]'$, y su EMV será $\hat{p}_i = \frac{y_i}{n_i}$ con lo que el máximo valor de la función de verosimilitud será:

$$l(\hat{\beta}_{max}; y) = \sum_{i=1}^N \left[y_i \ln \left(\frac{y_i}{n_i} \right) - y_i \ln \left(\frac{n_i - y_i}{n_i} \right) + n_i \ln \left(\frac{n_i - y_i}{n_i} \right) + \ln \binom{n_i}{y_i} \right]$$

En cualquier otro caso el que el número de parámetros es $n < N$, entonces \hat{p}_i es el EMV por lo que $\hat{y}_i = n_i \hat{p}_i$ recoge el valor ajustado. En este caso, la función $\ln l$ será:

$$l(\hat{\beta}; y) = \sum_{i=1}^N \left[y_i \ln \left(\frac{\hat{y}_i}{n_i} \right) - y_i \ln \left(\frac{n_i - \hat{y}_i}{n_i} \right) + n_i \ln \left(\frac{n_i - \hat{y}_i}{n_i} \right) + \ln \binom{n_i}{y_i} \right]$$

Por lo tanto, D será:

$$\begin{aligned} D &= 2 \left[\ln l(\hat{\beta}_{max}; y) - \ln l(\hat{\beta}; y) \right] \\ &= 2 \sum_{i=1}^N \left[y_i \ln \left(\frac{y_i}{\hat{y}_i} \right) - (n_i - y_i) \left(\frac{n_i - y_i}{n_i - \hat{y}_i} \right) \right] \end{aligned}$$

Estadístico D para un modelo de Poisson

Si $Y_i \approx P(\lambda_i)$ entonces:

$$l(\beta; y) = \sum_i y_i \ln \lambda_i - \sum_i \lambda_i - \sum_i \ln y_i!$$

En el modelo saturado $\beta = [\lambda_1, \lambda_2, \dots, \lambda_N]'$ siendo el EMV $\hat{\lambda}_i = y_i$ con lo que:

$$l(\hat{\beta}_{max}; y) = \sum_{i=1}^N y_i \ln y_i - \sum_{i=1}^N y_i - \sum_{i=1}^N \ln y_i!$$

Sea un modelo con $k < N$ parámetros. En este caso:

$$l(\hat{\beta}; y) = \sum_{i=1}^N y_i \ln \hat{y}_i - \sum_{i=1}^N \hat{y}_i - \sum_{i=1}^N \ln y_i!$$

con lo que el estadístico D es igual a:

$$\begin{aligned} D &= 2 \left[\ln l(\hat{\beta}_{max}; y) - \ln l(\hat{\beta}; y) \right] \\ &= 2 \left[\sum_{i=1}^N y_i \ln \left(\frac{y_i}{\hat{y}_i} \right) - \sum_{i=1}^N (y_i - \hat{y}_i) \right] \end{aligned}$$

Como $\sum_i y_i = \sum_i \hat{y}_i$ entonces $D = 2 \sum_{i=1}^N y_i \ln \left(\frac{y_i}{\hat{y}_i} \right)$

Estadístico D para un modelo normal

Sea el modelo $E(y_i) = \mu_i = x' \beta$ / $y_i \approx N(\mu_i, \sigma)$ $i = 1, \dots, N$ en el que, además, todas las y_i son independientes. En este caso, $l(\beta; y)$ es:

$$l(\beta; y) = -\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \mu_i)^2 - \frac{1}{2} N \ln(2\pi\sigma^2)$$

Para el modelo saturado se tiene que:

$$l(\hat{\beta}_{max}; y) = -\frac{1}{2} N \ln(2\pi\sigma^2)$$

Para el modelo no saturado, el *EMV* de β es: $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, con lo que:

$$l(\hat{\beta}; y) = -\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - x_i'\hat{\beta})^2 - \frac{1}{2} N \ln(2\pi\sigma^2)$$

y el estadístico D queda como:

$$\begin{aligned} D &= 2 \left[\ln l(\hat{\beta}_{max}; y) - \ln l(\hat{\beta}; y) \right] \\ &= \frac{1}{\sigma^2} \sum_{i=1}^N (y_i - \hat{\mu}_i)^2 \quad / \quad \hat{\mu}_i = x_i'\hat{\beta} \end{aligned}$$

es decir, es dependiente de σ . Al producto $D\sigma^2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2$ se le conoce como divergencia escalada. El estadístico D también se relaciona con la suma de residuos estandarizados al cuadrado:

$$\sum_i r_i^2 = \frac{1}{\sigma^2} \sum_{i=1}^N (y_i - \hat{\mu}_i)^2$$

Esto proporciona una regla fácil para calibrar la calidad del ajuste ya que, como D se distribuye según una chi cuadrado con $N - k$ grados de libertad, en media,

$$\sum_i r_i^2 \text{ debería valer } N - k.$$

4. DIAGNOSIS DEL MODELO

4.1. Estudio de los residuos

Una vez estimado el modelo parece oportuno analizar si es adecuado para recoger el fenómeno estudiado. Para ello, hay que evaluar si son adecuadas las elecciones de la función de varianza, de la función de enlace y de los regresores incluidos. Para ello, una de las técnicas más ampliamente usadas es la del estudio de los residuos generados tras la estimación.

A diferencia del modelo clásico de regresión, la definición de residuo en los *GLM* es más amplia debido a la multiplicidad de distribuciones que pueden utilizarse en este contexto. Así, podemos encontrar los siguientes tipos de residuos:

- *residuos de la variable respuesta:*
que son la diferencia entre el valor real y el estimado, es decir:

$$r_i^R = y_i - \hat{\mu}_i$$

- *residuos sintéticos (working residuals)*:
son la diferencia entre el valor sintético de la variable respuesta y el obtenido mediante el predictor lineal en convergencia:

$$r_i^W = (y_i - \hat{\mu}_i) \left(\frac{\partial \eta}{\partial \mu} \right)_i$$

- *residuos de Pearson*:
definidos como:

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{var}(y_i)}}$$

siendo la suma de sus cuadrados el estadístico chi cuadrado de Pearson. Dado que la variabilidad de los residuos es función de las variables explicativas empleadas, dicho efecto puede corregirse mediante el empleo de los residuos de Pearson estandarizados, los cuales se definen como:

$$r_i^{PS} = \frac{r_i^P}{\sqrt{1 - h_{ii}}} = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{var}(y_i) \cdot (1 - h_{ii})}}$$

siendo h_{ii} el elemento i -ésimo de la diagonal principal de la matriz $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Un problema de los residuos de Pearson es que si la distribución es distinta de la normal, pueden presentar un marcada asimetría, con lo que sus propiedades pueden diferir de las de los residuos del modelo normal.

- *residuos parciales*:
se utilizan para evaluar la forma de un predictor por lo que son calculados para cada uno de ellos. Su expresión es:

$$r_{ik}^T = (y_i - \hat{\mu}_i) \left(\frac{\partial \eta}{\partial \mu} \right)_i + (x_{ik} \hat{\beta}_k) \quad k = 1 \dots K$$

- *residuos de Anscombe*:
tratan de lograr normalidad en los residuos. Se definen como:

$$r_i^A = \frac{A(y_i) - A(\hat{\mu}_i)}{A'(\hat{\mu}_i) \sqrt{\text{var}(\hat{\mu}_i)}}$$

siendo $A(\cdot)$ una transformación de la forma:

$$A(\cdot) = \int \frac{1}{\sqrt[3]{\text{var}(\mu)}} d\mu$$

Ejemplo:

Si los datos siguen una distribución de Poisson, entonces $var(\mu) = \mu$, con lo que:

$$\begin{aligned} A'(\hat{\mu}) &= var^{-1/3} = \hat{\mu}^{-1/3} \\ A'(\hat{\mu})\sqrt{var(\hat{\mu})} &= \hat{\mu}^{-1/3} \cdot \hat{\mu}^{1/2} = \hat{\mu}^{1/6} \\ A(\hat{\mu}) &= \int \hat{\mu}^{-1/3} d\mu = \frac{3}{2} \hat{\mu}^{2/3} \\ \Rightarrow r_i^A &= \frac{3/2(y^{2/3} - \hat{\mu}^{2/3})}{\hat{\mu}^{1/6}} \end{aligned}$$

- *residuos asociados al estadístico D (Deviance residuals):*
recogen el incremento en D debido a cada observación. Su expresión es:

$$r_i^D = sign(y_i - \hat{\mu}_i) \sqrt{\hat{d}_i^2}$$

este tipo de residuos se suele calcular así o estandarizados, es decir, dividiendo el resultado entre $\sqrt{1-h_{ij}}$, o estudentizados.

- *residuos de probabilidad:*
son una combinación de residuos estandarizados de Pearson, r_i^{PS} , y los estandarizados del estadístico D , r_i^{DS} , teniendo como expresión:

$$r_i^L = sign(y_i - \hat{\mu}_i) \left\{ h_i (r_i^{PS})^2 + (1-h_i) (r_i^{DS})^2 \right\}^{1/2}$$

4.2. El fenómeno de la sobredispersión (*overdispersion*)

Es una situación que se puede presentar con datos que siguen una distribución binomial, Poisson o Binomial Negativa. Si la dispersión estimada tras el ajuste es diferente a la esperada, entonces puede ocurrir que exista sobre o infradispersión dependiendo de si el valor de esta característica es, respectivamente, superior o inferior al previsto.

Para la sobredispersión, la solución es más fácil y simple. Consiste en suponer que la varianza sigue un modelo multiplicativo. Sin embargo, es preciso señalar que este fenómeno puede estar indicando una mala especificación del modelo.

En un modelo de Poisson, el siguiente test se utiliza para detectar la posible existencia de este fenómeno:

$$T_1^2 = \frac{\sum_{i=1}^N [(y_i - \hat{\mu}_i)^2 - (1 - \hat{h}_i) \hat{\mu}_i]^2}{2 \sum_{i=1}^n \hat{\mu}_i^2} \approx \chi_1^2$$

En donde se compara la varianza de una distribución de Poisson con la de una Binomial Negativa. La hipótesis nula en este test es que $\phi = 0$ y la alternativa es que $v(\mu, \phi) = \mu + \phi\mu^2$, es decir, que la varianza es la de la binomial negativa. Otro test usado en los modelos de Poisson es el siguiente:

$$T_2^2 = \frac{1}{2N} \left\{ \frac{\sum_{i=1}^N (y_i - \hat{\mu}_i)^2 - (1 - \hat{h}_i) \hat{\mu}_i}{\hat{\mu}_i} \right\}^2 \approx \chi_1^2$$

En el cual, la hipótesis nula a contrastar es que $\phi = 0$, mientras que la alternativa es que $\phi = 1$.

Existe abundante literatura científica sobre este tema. Los trabajos de Ganio y Schafer (1992), Lambert y Roeder (1995), así como Dean y Lawless (1989), abordan el problema de la sobredispersión en los GLM.

4.3. Estadísticos del modelo

4.3.1. Estadísticos de comparación de modelos

Básicamente son:

- *AIC (Akaike information criterion)*

Debido a Akaike (1973). Se define como:

$$AIC = \frac{2 \ln l[M_k] + 2p}{n}$$

siendo $\ln l[M_k]$ el neperiano de la función de verosimilitud del modelo k , p el número de predictores empleados y n el número de observaciones. El criterio es que, cuanto menor sea el estadístico, mejor será el ajuste.

De forma general, cuando se comparan dos modelos y la diferencia entre ellos, medida por $AIC_1 - AIC_2$, es mayor que dos, entonces se recoge una marcada preferencia por el modelo 2.

- *BIC (Bayesian information criterion)*

Debido a Raftery (1996). Se define como:

$$BIC = D(M_k) - gl[\ln(n)]$$

Siendo gl los grados de libertad del modelo.

Como en AIC , cuanto menor sea el valor del estadístico, mayor será la calidad del ajuste. Raftery en su trabajo dio la siguiente tabla de preferencia relativa de un modelo:

diferencia	Grado de preferencia
0 - 2	Débil
2 - 6	Positiva
6 - 10	Fuerte
> 10	Muy fuerte

- Medidas de calidad del ajuste

Son diferentes variantes del R^2 , entre las que se pueden citar las siguientes:

- *Pseudo R^2 de Efron (1978)*

$$R_{Efron}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

- *Índice de la ratio de verosimilitud de McFadden (1974)*

$$R_{Efron}^2 = 1 - \frac{L(M_\beta)}{L(M_\alpha)}$$

en donde M_α es el modelo solamente con término constante y M_β el modelo que incluye tanto el término constante como los regresores seleccionados. Puede usarse en cualquier modelo ajustado mediante máxima verosimilitud.

- *Índice ajustado de la ratio de verosimilitud de Ben-Akira y Lerman (1985)*

$$R_{BA\&L}^2 = 1 - \frac{L(M_\beta) - k}{L(M_\alpha)}$$

en donde k es el número de parámetros. Puede usarse en cualquier modelo ajustado mediante máxima verosimilitud.

- *Ratio de varianzas de McKelvey y Zavoina (1975)*

$$R_{MK\&L}^2 = \frac{\hat{V}(\hat{y}^*)}{\hat{V}(\hat{y}^*) + V(\varepsilon)}$$

Siendo:

$$\hat{V}(\hat{y}^*) = \hat{\beta}' \hat{V} \hat{\beta}$$

$$\hat{V} = \text{var - cov de } \hat{\beta}$$

$$\mathbf{V}(\varepsilon) = \begin{cases} 1 & \text{para modelos probit} \\ \frac{M^2}{3} & \text{para modelos logit} \end{cases}$$

Esta medida se usa en modelos con variables de respuesta ordinales, binarias o censuradas.

- *Transformación de la ratio de verosimilitud*

$$R_{ML}^2 = 1 - \left\{ \frac{L(M_\alpha)}{L(M_\beta)} \right\}^{2/n}$$

$$= 1 - \exp\left(\frac{G^2}{n}\right)$$

siendo $G = -2 \ln \left\{ \frac{L(M_\alpha)}{L(M_\beta)} \right\}$. Puede usarse en cualquier modelo estimado por máxima verosimilitud.

- *Medida normalizada de Cragg y Uhler (1970)*

$$R_{C\&U}^2 = \frac{R_{ML}^2}{\max R_{ML}^2} = \frac{1 - \{L(M_\alpha)/L(M_\beta)\}^{2/n}}{1 - L(M_\alpha)^{2/n}}$$

Puede usarse en cualquier modelo estimado por máxima verosimilitud.

- R^2 de números enteros (count R^2) de Maddala (1992)

$$R_{count}^2 = \frac{1}{n} \sum_j n_{j,j}$$

en donde $n_{j,j}$ es el número de observaciones j correctamente clasificadas como j . Esta medida se utiliza en modelos de variable de respuesta binaria u ordinal.

- R^2 ajustado de números enteros

$$\bar{R}_{count}^2 = \frac{\sum_j n_{j,j} - \max(n_{r+})}{n - \max(n_{r+})}$$

en donde $n_{j,j}$ tiene el mismo significado que en la medida anterior y n_{r+} es la suma de variables observadas de valor j . Esta medida se utiliza en modelos de variable de respuesta binaria u ordinal.

- R^2 de Veall y Zimmermann (1992)

$$R_{VZ}^2 = \left(\frac{\delta - 1}{\delta - R_{MFadden}^2} \right) \cdot R_{MFadden}^2$$

En donde $\delta = \frac{n}{2L(M_\alpha)}$. Esta medida se usa sólo en modelos de variable ordinal.

- R^2 de Cameron - Windmeijer (1997)

$$R_{CW}^2 = 1 - \frac{K(y, \mu)}{K(y, \mu_0)}$$

siendo $K(\)$ la función de divergencia de Kullback-Leibler²³, μ el valor estimado por el modelo y μ_0 es la media estimada con el modelo sólo con término constante. Esta medida asume que el modelo incorpora término constante y se puede utilizar para varias familias de funciones (normal, Bernouilli, Binomial, Poisson, Gamma e Inversa Gaussiana).

²³ la divergencia de Kullback-Leibler es un indicador de la similitud entre dos funciones de distribución y está relacionada con de ajuste de distribuciones por máxima verosimilitud. Se define como:

$$\int f(x) \ln f(x) dx - \int f(x) \ln f_\lambda(x) dx = \int f(x) \ln \frac{f(x)}{f_\lambda(x)} dx$$

siendo f la verdadera función y f_λ la función estimada mediante máxima verosimilitud

CAPÍTULO 5 MÉTODOS BASADOS EN DISTRIBUCIONES DE PROBABILIDAD

1. INTRODUCCIÓN

Los métodos que podríamos llamar clásicos o tradicionales -véase capítulo 2-, se basan en esquemas denominados de distribución libre, es decir, pueden ser usados en cualquier conjunto de datos. A continuación se presenta un conjunto de estimaciones cuyo eje central consiste en suponer que los datos de partida utilizados se rigen según una determinada distribución estadística. Esta característica tiene dos ventajas:

1. por un lado, permite usar técnicas de simulación
2. por otro, hay que tener en cuenta que en la mayoría de los modelos de distribución los cálculos analíticos no van más allá de los segundos momentos, es decir, media y varianza. Por tanto, si se deseara obtener el valor de momentos de orden superior, cuantiles o medidas de riesgo, se necesitaría utilizar algoritmos numéricos que se basaran en alguna distribución concreta.

Las distribuciones que se van a utilizar dependen de cómo sean los datos de partida. Así, si se está trabajando con importes asociados a siniestros, la información puede venir:

- bien en forma acumulada
- bien como incrementos de un periodo de tiempo a otro.

Para el primer caso, la distribución más empleada es la lognormal, mientras que para el segundo existe una mayor diversidad, pudiéndose aplicar modelos basados en diferentes distribuciones tales como la Poisson, la Binomial Negativa o la Gamma, entre otras.

En todos los modelos que se presentan, la calidad del ajuste se mide a través del error cuadrático medio de predicción, que se denotará por *MSE*. A su vez, al tratar este concepto se distinguirá entre el *MSE* condicional y el incondicional. El *MSE* condicional de la estimación \hat{X} vendrá dado por la siguiente expresión:

$$\begin{aligned}MSE_{X|D}(\hat{X}) &= E\left[(\hat{X} - X)^2 | D\right] \\ &= \text{var}(X | D) + \left[\hat{X} - E(X | D)\right]^2\end{aligned}$$

es decir, en este caso, la medición del error está condicionada a cuál sea el conjunto de datos D . El primer sumando del lado derecho de la anterior expresión, es la varianza del proceso y recoge la variación dentro del modelo estocástico escogido para el análisis. Por tanto, en esencia, es el factor aleatorio que no puede ser eliminado. Por su parte, el segundo sumando recoge el error de estimación de los parámetros. Obviamente, para poder llevar a cabo este cálculo se necesita conocer $E(X|D)$. Sin embargo, este término suele ser desconocido, lo cual complica el cálculo de la expresión anterior. Una manera de evaluar la calidad de la estimación de X consiste en estudiar las fluctuaciones de \hat{X} en torno a $E(X|D)$. Así pues, pueden darse dos situaciones:

a) X es independiente de D :

en este caso tendríamos que

$$\begin{aligned}E(X | D) &= E(X) \\ \text{var}(X | D) &= \text{var}(X)\end{aligned}$$

de forma que el llamado MSE incondicional sería:

$$MSE_X(\hat{X}) = E\left[MSE_{X|D}(\hat{X})\right] = \text{var}(X) + E\left[(\hat{X} - E(X))^2\right]$$

y si se cumple que $E(\hat{X}) = E(X)$, entonces:

$$MSE_X(\hat{X}) = \text{var}(X) + \text{var}(\hat{X})$$

Es decir, el error de estimación de los parámetros se estima a través de la varianza de \hat{X} .

b) X no es independiente de las observaciones D :

en este caso el MSE incondicional se expresa como:

$$\begin{aligned}MSE_X(\hat{X}) &= E\left[MSE_{X|D}(\hat{X})\right] = \\ &= E\left[\text{var}(X | D) + E\left[(\hat{X} - E(X | D))^2\right]\right] \\ &= \text{var}(X) - \text{var}[E(X | D)] + E\left[(\hat{X} - E(X | D))^2\right] \\ &= \text{var}(X) + E\left[(\hat{X} - E(X))^2\right] - 2E\left[(\hat{X} - E(X))(E(X | D) - E(X))\right]\end{aligned}$$

Si \hat{X} es estimador insesgado de $E(X)$, entonces:

$$MSE_X(\hat{X}) = \text{var}(X) + \text{var}(\hat{X}) - 2\text{cov}[\hat{X}, E(X|D)]$$

2. MODELO LOGNORMAL PARA DATOS ACUMULADOS

Este modelo fue planteado por primera vez por Hertig (1985). La hipótesis básica es que los factores individuales del *Chain Ladder* ($f_{i,j} = C_{i,j}/C_{i,j-1}$) se distribuyen según una lognormal, con medias μ_j y varianzas σ_j^2 , es decir:

$$\eta_{i,j} = \ln\left(\frac{C_{i,j}}{C_{i,j-1}}\right) \sim N(\mu_j, \sigma_j) \quad \forall i = 0, \dots, I \quad \forall j = 0, \dots, J$$

Por lo que:

$$E(f_{i,j}) = \exp\left\{\mu_j + \frac{1}{2}\sigma_j^2\right\}$$

$$\text{var}(f_{i,j}) = \exp\{2\mu_j + \sigma_j^2\} \left[\exp\{\sigma_j^2\} - 1 \right]$$

Sabemos que en el método *Chain Ladder* se cumple que:

$$E(C_{i,j} | C_{i,j-1}) = f_{j-1} \cdot C_{i,j-1}$$

$$\text{var}(C_{i,j} | C_{i,j-1}) = \sigma_j^2 \cdot C_{i,j-1}$$

Aplicando lo anterior, se tiene que, para este modelo:

$$E(C_{i,j} | C_{i,j-1}) = C_{i,j-1} \cdot E(F_{i,j} | C_{i,j-1}) = C_{i,j-1} \cdot \exp\left\{\mu_j + \frac{1}{2}\sigma_j^2\right\}$$

$$\text{var}(C_{i,j} | C_{i,j-1}) = C_{i,j-1}^2 \cdot \text{var}(F_{i,j} | C_{i,j-1}) =$$

$$= C_{i,j-1}^2 \cdot \exp\{2\mu_j + \sigma_j^2\} \left[\exp\{\sigma_j^2\} - 1 \right]$$

Por tanto, comparando con las hipótesis del *Chain Ladder* tradicional, se observa que este modelo es equivalente a un *Chain Ladder* con factores f_{j-1} iguales a $\exp\{2\mu_j + \sigma_j^2\}$, pero con una función diferente de varianza.

La estimación de los parámetros se realiza a partir de las siguientes expresiones:

$$\hat{\mu}_j = \frac{1}{l-j+1} \sum_{i=1}^{l-j} \ln \left(\frac{C_{i,j}}{C_{i,j-1}} \right)$$

$$\sigma_j^2 = \frac{1}{l-j} \sum_{i=0}^{l-j} \left[\ln \left(\frac{C_{i,j}}{C_{i,j-1}} \right) - \mu_j \right]^2$$

En primer lugar, se definen las variables en escala logarítmica: $Z_{i,j} = \ln(C_{i,j})$, por lo que:

$$Z_{i,j} = Z_{i,l-i} + \sum_{j=l-1+i}^j \ln \left(\frac{C_{i,j}}{C_{i,j-1}} \right) = Z_{i,l-i} + \sum_{j=l-1+i}^j \eta_{i,j}$$

Sabiendo que $\eta_{i,j} \sim N(\mu_j, \sigma_j)$ y que $\hat{\mu}_j \sim N\left(\mu_j, \frac{\sigma_j}{\sqrt{l-j+1}}\right)$ obtenemos que:

$$E(Z_{i,j} | D_l) = Z_{i,l-i} + \sum_{j=l-i+1}^j \mu_j$$

que puede estimarse a partir de la siguiente expresión:

$$\hat{Z}_{i,j} = \hat{E}(Z_{i,j} | D_l) = Z_{i,l-i} + \sum_{j=l-i+1}^j \hat{\mu}_j$$

el cual es un estimador insesgado de $E(Z_{i,j} | D_l) = E(Z_{i,j} | Z_{i,l-i})$, tal y como como se puede ver a continuación.

Demostración:

si el estimador es insesgado, entonces:

$$E(\hat{Z}_{i,j} - Z_{i,j} | Z_{i,l-i}) = 0$$

$$E(\hat{Z}_{i,j} - Z_{i,j} | Z_{i,l-i}) = E \left[Z_{i,l-i} + \sum_{j=l-i+1}^j \hat{\mu}_j - Z_{i,l-i} + \sum_{j=l-i+1}^j \eta_{i,j} \right] =$$

$$= E \left[\sum_{j=l-i+1}^j \hat{\mu}_j - \sum_{j=l-i+1}^j \eta_{i,j} \right]$$

Como $\eta_{i,j} \sim N(\mu_j, \sigma_j)$ para cualquier i, j y además $\hat{\mu}_j \sim N\left(\mu_j, \frac{\sigma_j}{\sqrt{l-j+1}}\right)$, entonces:

$$E\left[\sum_{j=l-i+1}^J \hat{\mu}_j - \sum_{j=l-i+1}^J \eta_{i,j}\right] = E\left[\sum_{j=l-i+1}^J \mu_j - \sum_{j=l-i+1}^J \mu_j\right] = 0$$

Como se quería demostrar.

Por lo que se refiere al *MSE*, en este caso se tiene que, para el caso condicional, se puede expresar como:

$$MSE_{Z_{i,J}|D_l}(\hat{Z}_{i,J}) = \sum_{j=l-i+1}^J \sigma_j^2 + \left[\sum_{j=l-i+1}^J (\hat{\mu}_j - \mu_j)^2 \right]$$

Demostración:

Como $\eta_{i,j} \sim N(\mu_j, \sigma_j)$, entonces:

$$MSE_{Z_{i,J}|D_l}(\hat{Z}_{i,J}) = E\left[(\hat{Z}_{i,J} - Z_{i,J})^2 \mid D_l\right]$$

Desarrollando la expresión se tiene que:

$$\begin{aligned} E\left[(\hat{Z}_{i,J} - Z_{i,J})^2 \mid D_l\right] &= E\left[\left(\sum_{j=l-i+1}^J \hat{\mu}_j - \sum_{j=l-i+1}^J \eta_{i,j}\right)^2 \mid D_l\right] \\ &= E\left[\left(\sum_{j=l-i+1}^J \hat{\mu}_j - \sum_{j=l-i+1}^J \eta_{i,j} + \sum_{j=l-i+1}^J \mu_j - \sum_{j=l-i+1}^J \mu_j\right)^2 \mid D_l\right] \\ &= E\left\{\left[\left(\sum_{j=l-i+1}^J \hat{\mu}_j - \sum_{j=l-i+1}^J \mu_j\right) - \left(\sum_{j=l-i+1}^J \eta_{i,j} - \sum_{j=l-i+1}^J \mu_j\right)\right]^2 \mid D_l\right\} \\ &= \text{var}\left(\sum_{j=l-i+1}^J \eta_{i,j}\right) + \left[\sum_{j=l-i+1}^J (\hat{\mu}_j - \mu_j)^2\right] \\ &= \sum_{j=l-i+1}^J \sigma_j^2 + \left[\sum_{j=l-i+1}^J (\hat{\mu}_j - \mu_j)^2\right] \end{aligned}$$

Tal y como se quería demostrar.

Por lo que se refiere al MSE incondicional, como $\hat{Z}_{i,J} - Z_{i,J}$ es independiente de $Z_{i,l-i}$, se tiene que:

$$\begin{aligned} MSE_{Z_{i,J}|Z_{i,l-i}}(\hat{Z}_{i,J}) &= MSE_{Z_{i,J}}(\hat{Z}_{i,J}) = E\left[MSE_{Z_{i,J}|D_i}(\hat{Z}_{i,J})\right] \\ &= \text{var}\left(\sum_{j=l-i+1}^J \eta_{i,j}\right) + \text{var}\left(\sum_{j=l-i+1}^J \hat{\mu}_j\right) \end{aligned}$$

Como $\hat{\mu}_j \sim N\left(\mu_j, \frac{\sigma_j}{\sqrt{l-j+1}}\right)$, entonces:

$$MSE_{Z_{i,J}|Z_{i,l-i}}(\hat{Z}_{i,J}) = \sum_{j=l-i+1}^J \sigma_j^2 \left(1 + \frac{1}{l-j+1}\right)$$

A partir de las expresiones anteriores, se obtienen:

- estimaciones para el valor de la reserva y
- medidas del error cometido en dicho cálculo.

Por lo que se refiere al cálculo del estimador de la reserva, es necesario tener en cuenta que $E(C_{i,J} | C_{i,l-i}) = E\left[\exp(Z_{i,J}) | C_{i,l-i}\right]$. Como se supone que el proceso estudiado sigue una distribución lognormal, entonces se tiene que

$E(C_{i,j} | C_{i,j-1}) = C_{i,j-1} \cdot \exp\left\{\mu_j + \frac{\sigma_j^2}{2}\right\}$, por lo que:

$$\begin{aligned} E(C_{i,J} | C_{i,l-i}) &= \exp(Z_{i,l-i}) \cdot \exp\left\{\sum_{j=l-i+1}^J \mu_j + \sum_{j=l-i+1}^J \frac{\sigma_j^2}{2}\right\} \\ &= C_{i,l-i} \cdot \exp\left\{\sum_{j=l-i+1}^J \mu_j + \sum_{j=l-i+1}^J \frac{\sigma_j^2}{2}\right\} \end{aligned}$$

Por otra parte, el valor esperado de la estimación del $\ln(C_{i,J})$, es decir, $\hat{Z}_{i,J}$, condicionado a la última información disponible es:

$$\begin{aligned} E\left[\exp(\hat{Z}_{i,J} | C_{i,l-i})\right] &= \exp(Z_{i,l-i}) E\left[\exp\left\{\sum_{j=l-i+1}^J \hat{\mu}_j\right\}\right] \\ &= C_{i,l-i} \cdot \exp\left\{\sum_{j=l-i+1}^J \mu_j + \frac{1}{2} \sum_{j=l-i+1}^J \frac{\sigma_j^2}{l-j+1}\right\} \end{aligned}$$

Combinando ambas expresiones se obtiene el siguiente estimador:

$$\hat{C}_{i,J} = \hat{E}(C_{i,J} | D_l) = \exp \left\{ \hat{Z}_{i,J} + \sum_{j=l-i+1}^J \sigma_j^2 \left(1 - \frac{1}{l-j+1} \right) \right\}$$

Caracterizado por lo siguiente:

a) dado $C_{i,l-i}$, se cumple que $E(C_{i,J} | C_{i,l-i}) = E(C_{i,J} | D_l)$. En efecto:

$$\begin{aligned} E(\hat{C}_{i,J} | C_{i,l-i}) &= E \left[\exp \{ \hat{Z}_{i,J} \} | C_{i,l-i} \right] \exp \left\{ \frac{1}{2} \sum_{j=l-i+1}^J \sigma_j^2 \left(1 - \frac{1}{l-j+1} \right) \right\} \\ &= E(C_{i,J} | C_{i,l-i}) = E(C_{i,J} | D_l) \end{aligned}$$

b) la expresión de su *MSE* es igual a:

$$\begin{aligned} &MSE_{C_{i,J} | C_{i,l-i}}(\hat{C}_{i,J}) = \\ &= E[C_{i,J} | C_{i,l-i}]^2 \cdot \left(\exp \left\{ \sum_{j=l-i+1}^J \sigma_j^2 \right\} + \exp \left\{ \sum_{j=l-i+1}^J \frac{\sigma_j^2}{l-j+1} \right\} - 2 \right) \end{aligned}$$

En efecto: teniendo en cuenta la ausencia de correlación -independencia-, se tiene que:

$$E(C_{i,J} \cdot \hat{C}_{i,J} | C_{i,l-i}) = E(C_{i,J} | C_{i,l-i}) E(\hat{C}_{i,J} | C_{i,l-i})$$

y teniendo en cuenta la insesgadez, obtenemos la expresión del *MSE* condicional como:

$$\begin{aligned} MSE_{C_{i,J} | C_{i,l-i}}(\hat{C}_{i,J}) &= E \left[(C_{i,J} - \hat{C}_{i,J})^2 | C_{i,l-i} \right] \\ &= \text{var}(C_{i,J} - \hat{C}_{i,J} | C_{i,l-i}) = \text{var}(C_{i,J} | C_{i,l-i}) + \text{var}(\hat{C}_{i,J} | C_{i,l-i}) \\ &= C_{i,l-i}^2 \cdot \text{var} \left(\exp \left\{ \sum_{j=l-i+1}^J \eta_{ij} \right\} \right) \\ &\quad + C_{i,l-i}^2 \exp \left\{ \sum_{j=l-i+1}^J \sigma_j^2 \left(1 - \frac{1}{l-j+1} \right) \right\} \text{var} \left(\exp \left\{ \sum_{j=l-i+1}^J \hat{\mu}_j \right\} \right) \end{aligned}$$

Por lo que:

$$E\left[\left(C_{i,J} - \hat{C}_{i,J}\right)^2 \mid C_{i,l-i}\right] = \\ = C_{i,l-i}^2 \cdot \exp\left\{\sum_{j=l-i+1}^J (2\mu_j + \sigma_j^2)\right\} \left[\exp\left\{\sum_{j=l-i+1}^J \sigma_j^2\right\} + \exp\left\{\sum_{j=l-i+1}^J \frac{\sigma_j^2}{l-j+1}\right\} - 2 \right]$$

Como $E(C_{i,J} \mid C_{i,l-i}) = C_{i,l-i} \cdot \exp\left\{\sum_{j=l-i+1}^J \mu_j + \sum_{j=l-i+1}^J \frac{\sigma_j^2}{2}\right\}$, sustituyendo se obtiene la expresión del *MSE*. De esta expresión conviene resaltar que:

- una parte de ella recoge la varianza del proceso, en concreto:

$$\exp\left\{\sum_{j=l-i+1}^J \sigma_j^2\right\} - 1$$

- el resto de la misma recoge el error de estimación de los parámetros, en concreto:

$$\exp\left\{\sum_{j=l-i+1}^J \frac{\sigma_j^2}{l-j+1}\right\} - 1$$

Por lo que se refiere a la agregación del total de años de accidente -filas del triángulo-, operando de igual manera se llega a que su *MSE* es igual a:

$$MSE_{\sum_{i=0}^l C_{i,J}} \left(\sum_{i=0}^l \hat{C}_{i,J} \right) = \sum_{i=1}^l MSE_{C_{i,J} \mid C_{i,l-i}} (\hat{C}_{i,J}) + 2 \sum_{1 \leq i < k \leq l} E(C_{i,J} \mid C_{i,l-i}) \cdot \\ E(C_{k,J} \mid C_{k,l-k}) \left[\exp\left\{\sum_{j=l-i+1}^J \frac{\sigma_j^2}{l-j+1}\right\} - 1 \right]$$

2.1. Ejemplo

Partimos del conjunto de datos que se ha utilizado en ejemplos anteriores:

	0	1	2	3	4	5	6	7	8	9
0	242	775	1.217	1.584	1.996	2.459	2.616	2.774	3.006	3.121
1	238	852	1.522	2.395	2.762	3.051	3.507	3.780	4.187	
2	196	891	1.556	2.310	2.898	3.061	3.493	3.776		
3	210	978	1.540	2.683	2.930	3.245	3.462			
4	300	783	1.493	2.069	2.474	2.869				
5	268	919	1.530	2.132	2.684					
6	298	888	1.698	2.487						
7	243	980	2.010							
8	255	940								
9	233									

cuyos factores de desarrollo individuales son:

	0	1	2	3	4	5	6	7	8	9
0	3,202	1,570	1,302	1,260	1,232	1,064	1,060	1,084	1,038	
1	3,580	1,786	1,574	1,153	1,105	1,149	1,078	1,108		
2	4,546	1,746	1,485	1,255	1,056	1,141	1,081			
3	4,657	1,575	1,742	1,092	1,108	1,067				
4	2,610	1,907	1,386	1,196	1,160					
5	3,429	1,665	1,393	1,259						
6	2,980	1,912	1,465							
7	4,033	2,051								
8	3,686									
9										

de los cuales se calcula su logaritmo neperiano. Para cada columna, se calcula su media y su varianza, de acuerdo a las siguientes expresiones:

$$\hat{\mu}_j = \frac{1}{l-j+1} \sum_{i=1}^{l-j} \ln \left(\frac{C_{i,j}}{C_{i,j-1}} \right)$$

$$\hat{\sigma}_j^2 = \frac{1}{l-j} \sum_{i=0}^{l-j} \left[\ln \left(\frac{C_{i,j}}{C_{i,j-1}} \right) - \hat{\mu}_j \right]^2$$

Los resultados obtenidos son los recogidos en la siguiente tabla:

	0	1	2	3	4	5	6	7	8	9
0		1,164	0,451	0,264	0,231	0,209	0,062	0,059	0,080	0,038
1		1,275	0,580	0,453	0,143	0,100	0,139	0,075	0,102	
2		1,514	0,558	0,395	0,227	0,055	0,132	0,078		
3		1,538	0,454	0,555	0,088	0,102	0,065			
4		0,959	0,645	0,326	0,179	0,148				
5		1,232	0,510	0,332	0,230					
6		1,092	0,648	0,382						
7		1,394	0,718							
8		1,305								
9										
		1	2	3	4	5	6	7	8	9
media		1,275	0,571	0,387	0,183	0,123	0,099	0,071	0,091	0,038
d.t.		0,190	0,097	0,096	0,059	0,058	0,042	0,010	0,016	0,016

Para obtener la cifra estimada del global de reservas se lleva a cabo la siguiente secuencia de pasos:

Paso 1: obtención de la estimación tendencial

Haciendo:

$$\hat{Z}_{iJ} = Z_{i,I-1} + \sum_{j=I-i+1}^J \hat{\mu}_j$$

Resultan los siguientes valores:

	0	1	2	3	4	5	6	7	8	9
0										
1										8,377
2									8,328	8,365
3								8,220	8,311	8,349
4							8,061	8,132	8,223	8,261
5						8,018	8,117	8,188	8,279	8,317
6					8,002	8,124	8,224	8,294	8,386	8,423
7				7,993	8,176	8,298	8,398	8,468	8,559	8,597
8			7,416	7,803	7,986	8,109	8,208	8,279	8,370	8,408
9		6,726	7,297	7,683	7,866	7,989	8,088	8,159	8,250	8,288

Paso 2: cálculo del elemento ligado a la varianza

Para cada dato de la tabla anterior se calcula:

$$\sum_{j=l-i+1}^J \hat{\sigma}_j^2 \left(1 - \frac{1}{l-j+1}\right)$$

Siendo $\hat{\sigma}_j^2$ la varianza de cada columna. Los resultados son los siguientes:

	0	1	2	3	4	5	6	7	8	9
0										
1										0
2									12	12
3								7	19	19
4							132	139	151	151
5						272	404	411	423	423
6					286	558	689	697	709	709
7				782	1.070	1.340	1.470	1.480	1.490	1.490
8			816	1.600	1.880	2.160	2.290	2.290	2.310	2.310
9	3.210	4.030	4.810	5.100	5.370	5.500	5.510	5.520	5.520	

Paso 3: obtención del valor final de $\hat{Z}_{i,j}$

Sumando el elemento tendencial más la parte asociada a la varianza, se obtiene el estimado de $\hat{Z}_{i,j} \forall j > l - i$. Los resultados se recogen en la siguiente tabla:

	0	1	2	3	4	5	6	7	8	9
0										
1										8,377
2									8,328	8,365
3								8,220	8,311	8,349
4							8,062	8,132	8,224	8,261
5						8,019	8,119	8,190	8,281	8,319
6					8,003	8,127	8,227	8,298	8,389	8,427
7				7,997	8,181	8,305	8,405	8,476	8,567	8,604
8			7,421	7,811	7,996	8,120	8,220	8,290	8,382	8,419
9	6,742	7,317	7,707	7,892	8,016	8,116	8,186	8,278	8,315	

Paso 4: obtención de la cifra final en niveles

Para ello, se calcula la exponencial de las cifras obtenidas en el paso anterior, obteniéndose los siguientes resultados:

	0	1	2	3	4	5	6	7	8	9
0										
1										4.347
2									4.137	4.295
3								3.715	4.070	4.226
4							3.171	3.403	3.728	3.871
5						3.038	3.358	3.604	3.948	4.099
6				2.991	3.385	3.742	4.015	4.399	4.568	
7			2.971	3.572	4.043	4.469	4.796	5.255	5.456	
8		1.670	2.468	2.968	3.359	3.713	3.985	4.366	4.533	
9	847	1.505	2.225	2.675	3.028	3.347	3.592	3.935	4.086	

Una vez calculadas las cifras estimadas, se computan las desviaciones típicas de cada estimación utilizando la expresión:

$$d.t.(C_{i,j} | C_{i,j-1}) = \sqrt{C_{i,j-1} \cdot \exp(2\xi_j + \sigma_j^2) [\exp(\sigma_j^2) - 1]}$$

y se obtienen los siguientes resultados:

	0	1	2	3	4	5	6	7	8	9
0										
1										67
2									64	67
3								39	63	66
4							133	35	58	60
5						177	141	37	61	64
6					175	198	157	42	68	71
7				285	209	236	187	50	82	85
8			162	236	174	196	156	41	68	70
9	163	146	213	157	177	140	37	61	63	

Paso 5: obtención de la cifra de reservas por año de accidente y total

Para ello en cada fila se resta la cifra obtenida correspondiente a la última columna de la última diagonal -último dato conocido-. Los resultados para los nueve años por separado y para el total son los siguientes:

Año	Reserva
1	160
2	519
3	764
4	1.002
5	1.415
6	2.081
7	3.446
8	3.593
9	3.853
Total	16.833

Y el desglose por ejercicios, tanto en valor nominal como actual es el siguiente:

año	dotación anual	
	valor nominal	valor actual
1	4.239	4.187
2	3.518	3.423
3	2.774	2.653
4	2.029	1.856
5	1.569	1.387
6	1.217	1.011
7	827	642
8	511	376
9	151	106
a dotar	16.833	15.641
Saldo total	42.602	41.410

Por lo que respecta al cálculo del *MSE* de la reserva de cada año, se puede proceder de dos formas:

Alternativa 1: cálculo a partir de la desviación típica del proceso y del error de estimación.

En este caso, haremos la siguiente secuencia de instrucciones:

Paso 1: cálculo de la desviación típica del proceso

Para ello se utiliza la siguiente expresión:

$$d.t._{C_{i,J}|C_{i,l-i}}(\hat{C}_{i,J}) = \sqrt{E(C_{i,J} | C_{i,l-i})^2 \cdot \left[\exp\left(\sum_{j=l-i+1}^J \sigma_j^2\right) - 1 \right]}$$

con lo que los resultados, año a año, son los siguientes:

Año	Desviación típica
1	67
2	94
3	103
4	188
5	311
6	438
7	741
8	759
9	1.048

Pasó 2: cálculo de los errores de estimación

Para ello se utiliza la siguiente expresión:

$$\sqrt{\text{error}_{C_{i,j}|C_{i,l-i}}(\hat{C}_{i,j})} = \sqrt{E(C_{i,j} | C_{i,l-i})^2 \cdot \left[\exp\left(\sum_{j=l-i+1}^J \frac{\sigma_j^2}{l-j+1}\right) - 1 \right]}$$

Siendo los resultados, año a año, los siguientes:

Año	raíz del error de estimación
1	67
2	82
3	84
4	112
5	160
6	209
7	318
8	306
9	379

Paso 3: Obtención de la cifra final del MSE a partir de los resultados anteriores
Realmente, lo que interesa es la raíz cuadrada de la suma de cuadrados, es decir:

$$\sqrt{MSE} = \sqrt{d.t.^2 + \text{error}^2}$$

siendo los resultados obtenidos los siguientes:

Año	raíz del MSE
1	95
2	125
3	133
4	218
5	350
6	486
7	807
8	818
9	1.115

Alternativa 2: estimar el MSE directamente.

Para ello se utiliza la expresión:

$$MSE_{C_{i,j}|C_{i,l-i}}(\hat{C}_{i,j}) = E(C_{i,j} | C_{i,l-i})^2 \cdot \left[\exp\left(\sum_{j=l-i+1}^J \sigma_j^2\right) + \exp\left(\sum_{j=l-i+1}^J \frac{\sigma_j^2}{l-j+1}\right) - 2 \right]$$

y, con las cifras obtenidas, se obtiene su raíz cuadrada.

Finalmente, si se desea calcular el MSE del total de la reserva, es decir, de la agregación de años, entonces, la expresión a emplear es la siguiente:

$$\widehat{MSE}_{\sum_i C_{i,j} | \sum_{i=1}^l \hat{C}_{i,j}} = \sum_{i=1}^l MSE_{C_{i,j}|C_{i,l-i}}(\hat{C}_{ij}) + 2 \sum_{1 \leq i < k \leq l} E(C_{i,j} | C_{i,l-i}) \cdot E(C_{k,j} | C_{k,l-k}) \left[\exp\left(\sum_{j=l-i+1}^J \frac{\sigma_j^2}{l-j+1}\right) - 1 \right]$$

Lo que podrían denominarse términos de covarianza están en el segundo sumando de la expresión anterior. Su valor se recoge en la siguiente tabla:

Año	covarianzas ligadas al año <i>i</i>
1	73.533
2	95.668
3	89.299
4	147.184
5	231.898
6	268.641
7	319.400
8	169.124
Total	1.394.746

Obsérvese que se calculan para ocho y no para nueve años pues el límite inferior del segundo sumatorio es $1 \leq i < k \leq l$. Por tanto, una vez que se conocen tanto los errores como los MSE individuales y se conoce el valor del total de los términos cruzados, el valor global de ambos será la su suma, es decir:

	error de estimación	MSE
valor	1.833.265	4.405.363
raíz	1.354	2.099

3. DISTRIBUCIONES PARA DATOS EN INCREMENTOS

3.1. Modelo de Poisson

Una de las consecuencias básicas de utilizar la distribución de Poisson es que el nivel de reservas resultantes es el mismo que el que se obtiene usando el método *Chain Ladder* tradicional. Este hecho fue estudiado por Hachemeister y Stanard (1975), Kremer (1985) y Mack (1991a). Otros trabajos relacionados con el tema son los elaborados por Mack y Venter (2000) y por England y Verrall (2002).

Dada esta igualdad de resultados, el modelo de Poisson se presenta como modelo estocástico que permite atribuir la distribución de reservas ligadas al *Chain Ladder*. Las hipótesis del modelo son las siguientes:

1. existe un conjunto de parámetros $\mu_0, \mu_1, \dots, \mu_I$ y $\gamma_0, \gamma_1, \dots, \gamma_J$, todos ellos positivos, tales que los incrementos en los importes ligados a siniestros, denotados por $X_{i,j}$ se distribuyen de forma independiente y de acuerdo con una distribución de Poisson, cumpliéndose que:

$$E(X_{i,j}) = x_{i,j} = \mu_i \gamma_j \quad \forall i = 0, \dots, I \quad \forall j = 0, \dots, J$$
$$\text{var}(X_{i,j}) = x_{i,j} = \mu_i \gamma_j \quad \forall i = 0, \dots, I \quad \forall j = 0, \dots, J$$

y además: $\sum_{j=0}^J \gamma_j = 1$

en donde μ_i representa el total esperado de importes ligados al año i mientras que γ_j recoge el porcentaje esperado de gastos ligados al año j en relación al total. Es decir, si I y J llegan hasta 9, $\hat{\mu}_5$ vale 10.000.000 y $\hat{\gamma}_5$ vale 1%, entonces la cifra estimada para $X_{5,5}$ será el 1% de 10.000.000, es decir, 100.000.

2. todos los incrementos han de ser positivos por lo que, si no es así, la aplicación del modelo de Poisson no es adecuada.
3. la acumulación de importes asociados al año de accidente i , que se denotará por $C_{i,l}$, es también Poisson, en donde se cumple que:

$$E(C_{i,l}) = \mu_i$$

La estimación de parámetros puede hacerse por máxima verosimilitud. Así, dado un conjunto de información D_l definido como:

$$D_l = \{X_{i,j} / i + j \leq l \quad 0 \leq j \leq J\}$$

La función de verosimilitud se define de la siguiente forma:

$$L_{D_l}(\mu_0, \dots, \mu_l, \gamma_0, \dots, \gamma_J) = \prod_{i+j \leq l} \exp(-\mu_i \gamma_j) \frac{(\mu_i \gamma_j)^{X_{i,j}}}{X_{i,j}!}$$

Como es habitual, derivando la anterior función respecto de los $l+J+2$ variables desconocidas, se obtienen las siguientes expresiones:

$$\sum_{j=0}^{l-i} \hat{\mu}_i \hat{\gamma}_j = \sum_{j=0}^{l-i} X_{i,j} = C_{i,l-i}$$

$$\sum_{i=0}^{l-j} \hat{\mu}_i \hat{\gamma}_j = \sum_{i=0}^{l-j} X_{i,j}$$

sujetas a la condición que $\sum_{j=0}^J \hat{\gamma}_j = 1$. Por tanto, los estimadores máximo verosímiles vienen recogidos en las siguientes expresiones:

$$\hat{X}_{i,j} = \hat{E}(X_{i,j}) = \hat{\mu}_i \hat{\gamma}_j$$

$$\hat{C}_{i,J} = \hat{E}(C_{i,J} | D_l) = C_{i,l-i} + \sum_{j=l-i+1}^J \hat{X}_{i,j} \quad \forall i+j > l$$

Es necesario señalar que la resolución del sistema que permite obtener los valores de $\hat{\mu}_i$ y $\hat{\gamma}_j$ no es sencilla y suele llevarse a cabo mediante la utilización de métodos numéricos. Sin embargo, en este caso, se puede aplicar la siguiente propiedad:

$$\sum_{k=0}^j \hat{\gamma}_k = \prod_{k=j}^{J-1} \frac{1}{\hat{f}_k} \quad \forall j = 0, \dots, J$$

3.1.1. Ejemplo

A continuación, se va a aplicar este modelo sobre los datos del apartado anterior. Por tanto, el problema se reduce a estimar los $\hat{\mu}_i$ y los $\hat{\gamma}_j$ para lo cual se utilizará la relación existente entre los $\hat{\gamma}_j$ y los \hat{f}_j . Los factores de desarrollo ligados a los datos son los que se recogen en la siguiente tabla:

	año								
	0	1	2	3	4	5	6	7	8
\hat{f}_j	3,5582	1,7784	1,4835	1,1952	1,1244	1,1068	1,0743	1,0975	1,0383

Con lo que los $\hat{\gamma}_j$ son:

	año									
	0	1	2	3	4	5	6	7	8	9
$\hat{\gamma}_j$	5,85%	14,97%	16,20%	17,90%	10,72%	8,17%	7,88%	6,07%	8,56%	

Y, finalmente, los $\hat{\mu}_i$ resultantes son:

año	$\hat{\mu}_i$
1	3.121
2	4.347
3	4.303
4	4.238
5	3.887
6	4.528
7	5.429
8	4.515
9	3.982

Así, la parte inferior del triángulo, en la que $\hat{X}_{i,j} = \hat{\mu}_i \hat{\gamma}_j \quad \forall i + j > l$, queda como:

	0	1	2	3	4	5	6	7	8	9
0										
1										160
2									368	159
3								257	363	156
4							306	236	333	143
5						334	322	248	350	151
6					485	370	357	275	387	167
7				972	582	443	428	329	465	200
8			732	808	484	369	356	274	386	166
9		596	645	713	427	325	314	242	341	147

Sumando por filas se obtiene la reserva asociada a cada año de accidente y agregándolas todas, la reserva global siendo su resultado el siguiente:

Año	Reserva
1	160
2	527
3	776
4	1.018
5	1.405
6	2.041
7	3.419
8	3.575
9	3.749
Total	16.671

Mientras que el desglose por ejercicios futuros, tanto en valor nominal como actual es el siguiente:

año	dotación anual	
	valor nominal	valor actual
1	4.211	4.159
2	3.485	3.391
3	2.734	2.615
4	1.991	1.822
5	1.549	1.369
6	1.219	1.012
7	828	643
8	507	373
9	147	104
a dotar	16.671	15.488
Saldo total	42.440	41.257

3.2. Modelo de Poisson con sobredispersión

Es una generalización del modelo anterior en el sentido de que la varianza puede expresarse como:

$$\text{var}(X_{i,j}) = \phi x_{i,j} \quad / \quad \phi > 0$$

en donde a ϕ se le conoce como factor de dispersión. Obviamente, si ϕ es mayor que 1 estaríamos hablando de sobredispersión, mientras que si fuese inferior a uno, entonces sería infradispersión. La pregunta que nos podemos hacer es saber si la cifra de reservas con este modelo es distinta a la obtenida con lo que se podría denominar Poisson normal o con el *Chain Ladder*. Si se asume que ϕ es una constante, la estimación del resto de parámetros del modelo podría hacerse utilizando *GLM* y lleva a que el nivel de reservas es el mismo que en los otros métodos. Sin embargo, las diferencias aparecen al calcular alguna medida de error, el *MSE*, tal y como se puede apreciar en el ejemplo que se desarrolla en el siguiente epígrafe, en el que se aplica este modelo sobre los datos que venimos manejando a lo largo del texto.

3.2.1. Ejemplo

Además de utilizar los mismos datos, se tiene en cuenta lo siguiente:

- los parámetros se estiman utilizando *GLM* por lo que para el caso de la Poisson se tiene que:

$$E(X_{i,j}) = x_{i,j} = b'(\theta_{i,j})$$

$$\text{var}(X_{i,j}) = \frac{\phi_{i,j}}{\omega_{i,j}} \text{var}(x_{i,j})$$

Siendo $\phi_{i,j}$ el factor de dispersión -que se asumirá constante-, $\omega_{i,j}$ una medida de volumen y $\text{var}(x_{i,j})$ es la función de varianza correspondiente, que en el caso de la Poisson es $\text{var}(x_{i,j}) = x_{i,j}$.

- El modelo a estimar se puede expresar de la siguiente forma:

$$g(x_{i,j}) = \eta_{i,j} = \Gamma_{i,j} \boldsymbol{\beta} \quad i = 0, \dots, I \quad j = 0, \dots, J$$

en donde $g(x_{i,j})$ es la función de enlace -o link- y $\Gamma_{i,j} \boldsymbol{\beta}$ es la componente sistemática del modelo. $\boldsymbol{\beta}$ es el vector de parámetros a estimar, mientras que $\Gamma_{i,j}$ es la matriz de coeficientes, conocida también como matriz de diseño.

- $x_{i,j}$ media de $X_{i,j}$, adopta la siguiente expresión:

$$x_{i,j} = \mu_i \gamma_j$$

Para este caso, la forma de obtener una expresión lineal es utilizando la función neperiana, con lo que:

$$\eta_{i,j} = \ln(x_{i,j}) = \ln \mu_i + \ln \gamma_j$$

En el modelo simple de Poisson se imponía la condición de que $\sum_{j=0}^J \hat{\gamma}_j = 1$. En

este caso, resulta más fácil imponer como condición que $\mu_0 = 1$, con lo que $\ln \mu_0 = 0$ y, por tanto, $\eta_{0,j} = \ln \gamma_j$, con lo que el vector de parámetros a estimar es:

$$\boldsymbol{\beta} = [\ln(\mu_1), \dots, \ln(\mu_I), \ln(\gamma_0), \ln(\gamma_1), \dots, \ln(\gamma_J)]'$$

y la matriz de coeficientes adopta la siguiente expresión:

$$\Gamma_{i,j} = \begin{cases} \underbrace{0, \dots, 0}_I, \underbrace{0, \dots, 0, 1_{I+j+1}, 0, \dots, 0}_J & \text{si } i = 0 \\ \underbrace{0, \dots, 0, 1_i, 0, \dots, 0}_I, \underbrace{0, \dots, 0, 1_{I+j+1}, 0, \dots, 0}_J & \text{si } i > 1 \end{cases}$$

En nuestro ejemplo, en el modelo $\eta = \Gamma\beta$ cada una de las matrices que lo compone tiene las siguientes dimensiones:

$$\eta = (55 \times 1)$$

$$\Gamma = (55 \times 19)$$

$$\beta = (19 \times 1)$$

Por ejemplo, para el elemento (0,6), es decir, año de accidente igual a cero y año de desarrollo igual a seis, cuyo valor es 157, se tiene que:

$$\eta_{0,6} = \ln(157) = 5,05624$$

y la fila correspondiente de la matriz Γ tiene todas sus valores iguales a cero excepto la posición 16ª que vale 1 ($16 = l + j + 1 = 9 + 6 + 1$). Para el elemento (3,2), que vale 562, se tiene que:

$$\eta_{3,2} = \ln(562) = 6,33150$$

mientras que la fila correspondiente de Γ tiene todos sus elementos iguales a cero excepto las posiciones tercera y décimo segunda que valen uno. Los resultados de la estimación mediante *GLM* se recogen en la siguiente tabla:

Parámetro	$\ln(x_{i,j})$	$x_{i,j}$	Parámetro	$\ln(x_{i,j})$	$x_{i,j}$
			γ_0	5,207293	182,60
μ_1	0,3313742	1,39288	γ_1	6,146606	467,13
μ_2	0,3210877	1,37863	γ_2	6,226008	505,73
μ_3	0,3058929	1,35784	γ_3	6,325584	558,68
μ_4	0,2194868	1,24544	γ_4	5,812795	334,55
μ_5	0,2701039	1,31010	γ_5	5,540913	254,91
μ_6	0,3721629	1,45087	γ_6	5,505475	246,04
μ_7	0,5536358	1,73957	γ_7	5,243408	189,31
μ_8	0,3693259	1,44676	γ_8	5,587406	267,04
μ_9	0,2437455	1,27602	γ_9	4,744933	115,00

Así se puede rellenar el triángulo inferior sin más que aplicar el modelo de partida, es decir, $\hat{x}_{i,j} = \hat{\mu}_i \hat{\gamma}_j$. Los resultados son los siguientes:

	0	1	2	3	4	5	6	7	8	9
0										
1										160
2									368	159
3								257	363	156
4							306	236	333	143
5						334	322	248	350	151
6					485	370	357	275	387	167
7				972	582	443	428	329	465	200
8			732	808	484	369	356	274	386	166
9		596	645	713	427	325	314	242	341	147

y el nivel de reservas para cada año de accidente se calcula sumando por filas los resultados anteriores. Como paso final, el nivel global se logra agregando todas las reservas anuales, con lo que las cifras obtenidas son las siguientes:

Año	Reserva
1	160
2	527
3	776
4	1.018
5	1.405
6	2.041
7	3.419
8	3.575
9	3.749
Total	16.671

que coinciden con las resultantes de la aplicación del modelo simple de Poisson, tanto por años de ocurrencia como por ejercicios futuros. Por lo que se refiere al parámetro de dispersión, como se asume que $\phi_{ij} = \phi$ para cualquiera que sea el valor de i o de j , se obtiene a partir de la siguiente expresión:

$$\hat{\phi} = \frac{\sum_{i+j \leq l} \hat{R}_{i,j}(P)^2}{N - p}$$

en donde N es el total de observaciones empleadas -en este caso 55-, p es el número de parámetros estimados -19 en este caso- y $\hat{R}_{i,j}(P)$ son los residuos de Pearson, cuya expresión es:

$$\hat{R}_{i,j}(P) = \frac{X_{i,j} - \hat{x}_{i,j}}{\sqrt{\text{var}(\hat{x}_{i,j})}}$$

En el caso del modelo de Poisson se tiene que $\text{var}(\hat{x}_{i,j}) = \hat{x}_{i,j}$. Por tanto, la aplicación del modelo sobre las observaciones registradas nos ofrece los valores estimados $\hat{x}_{i,j}$, que son los siguientes:

	0	1	2	3	4	5	6	7	8	9
0	183	467	506	559	335	255	246	189	267	115
1	254	651	704	778	466	355	343	264	372	
2	252	644	697	770	461	351	339	261		
3	248	634	687	759	454	346	334			
4	227	582	630	696	417	317				
5	239	612	663	732	438					
6	265	678	734	811						
7	318	813	880							
8	264	676								
9	233									

Además, en este caso, coinciden con el valor de las varianzas individuales, por lo que los residuos de Pearson que resultan son los siguientes:

	0	1	2	3	4	5	6	7	8	9
0	4,40	3,05	-2,83	-8,11	4,23	13,03	-5,68	-2,28	-2,14	0,00
1	-1,02	-1,44	-1,30	3,40	-4,59	-3,51	6,12	0,57	1,82	
2	-3,51	2,01	-1,22	-0,58	5,90	-10,05	5,04	1,36		
3	-2,41	5,31	-4,76	13,96	-9,72	-1,67	-6,41			
4	4,81	-4,10	3,19	-4,54	-0,57	4,35				
5	1,86	1,58	-2,00	-4,80	5,43					
6	2,03	-3,37	2,81	-0,76						
7	-4,19	-2,65	5,07							
8	-0,56	0,35								
9	0,00									

y el valor del parámetro de dispersión es:

$$\hat{\phi} = \frac{\sum_{i+j \leq l} \hat{R}_{i,j}(P)^2}{N-p} = \frac{1.221,33}{36} = 33,93$$

Por su parte, la desviación típica asociada a la estimación se obtiene como:

$$\sqrt{\widehat{\text{var}}} = \sqrt{\sum_{i+j>l} \hat{\phi} \text{var}(\hat{x}_{i,j})} = \sqrt{\sum_{i+j>l} \hat{x}_{i,j}} = 752,04$$

Equivalente a un 4,51% del total de la reserva.

Finalmente, para estimar el *MSE* se utiliza la siguiente expresión:

$$\widehat{\text{MSE}}_{\sum_i \hat{C}_{i,l}} \left(\sum_{i=1}^l \hat{C}_{i,J} \right) = \sum_{i+j>l} \frac{\phi_{i,j}}{\omega_{i,j}} \text{var}(\hat{x}_{i,j}) + \sum_{\substack{i+j>l \\ n+m>l}} \hat{x}_{i,j} \hat{x}_{n,m} \Gamma_{i,j} \mathbf{H}(\hat{\beta})^{-1} \Gamma'_{n,m}$$

que puede explicarse como la suma de la varianza de la estimación del nivel de reservas -primer sumando- más el cuadrado del error de estimación -segundo sumando-. $\mathbf{H}(\hat{\beta})$ es la matriz de varianzas-covarianzas de los estimadores, cuya expresión es:

$$\widehat{\text{cov}}(\hat{\beta} \hat{\beta}') = \left[\left(\sum_{i+j \leq l} W(\hat{x}_{i,j}) \Gamma_{i,j}^{(k)} \Gamma_{i,j}^{(l)} \right)_{k,l=1,\dots,l+J+1} \right]^{-1}$$

en donde $\Gamma_{i,j}^{(k)}$ y $\Gamma_{i,j}^{(l)}$ son las posiciones k y l -ésima respectivamente de la fila asociada al dato $x_{i,j}$ dentro de la matriz de coeficientes. En este ejemplo, la matriz Γ tiene dimensión (45×19) . $W(\hat{x}_{i,j})$ es igual a:

$$W(\hat{x}_{i,j}) = \left(\frac{\partial \eta_{i,j}}{\partial x_{i,j}} \right)^{-2} \frac{1}{\text{var}(X_{i,j})}$$

Como en este caso, $\eta_{i,j} = \ln(x_{i,j})$ y $\phi_{i,j} = \phi$, nos queda que:

$$W(\hat{x}_{i,j}) = \frac{\hat{x}_{i,j}^2}{\hat{\phi} \text{var}(\hat{x}_{i,j})} = \frac{\hat{x}_{i,j}}{\hat{\phi}}$$

Operando, resulta un valor de 2.292 para la raíz cuadrada del segundo sumatorio, por lo que la raíz del MSE es igual a:

$$\sqrt{MSE} = \sqrt{2.292^2 + 752^2} = 2.412$$

3.3. Modelo basado en la distribución Gamma

En este caso, la variable respuesta sigue una ley estadística de esta clase. Este modelo fue formulado por Mack (1991a). La expresión utilizada de la función gamma es:

$$\Gamma(\gamma; c) = f_{\gamma,c}(x) = \frac{c^\gamma}{\Gamma(\gamma)} x^{\gamma-1} \exp(-cx) \quad \forall x \geq 0$$

siendo γ y c los factores de forma y escala respectivamente.

En este caso, la estimación del nivel de reservas se puede abordar desde dos puntos de vista diferentes:

Alternativa 1: a partir de los siniestros individuales de un cierto año de accidente en un cierto año de desarrollo.

En este caso, existe un número conocido de siniestros, $r_{i,j}$, y supondremos que los importes asociados a cada uno de ellos son independientes entre sí y siguen una distribución gamma, de modo que la siniestralidad total es:

$$X_{i,j} = \sum_{k=0}^{r_{i,j}} x_{i,j}^{(k)} \quad \forall i = 0, 1, \dots, I \quad \forall j = 0, 1, \dots, J$$

De forma que cada $x_{i,j}^{(k)}$ sigue una distribución $\Gamma(a_{i,j}; b_{i,j})$, de modo que su media es igual a:

$$E\left(x_{i,j}^{(k)}\right) = m_{i,j} = \frac{a_{i,j}}{b_{i,j}}$$

y su varianza:

$$\text{var}\left(x_{i,j}^{(k)}\right) = \frac{a_{i,j}}{b_{i,j}^2}$$

con lo que el coeficiente de variación es:

$$cv\left(x_{i,j}^{(k)}\right) = \frac{\sqrt{\text{var}\left(x_{i,j}^{(k)}\right)}}{E\left(x_{i,j}^{(k)}\right)} = \frac{\sqrt{a_{i,j}/b_{i,j}}}{a_{i,j}/b_{i,j}} = \frac{1}{\sqrt{a_{i,j}}} = \frac{1}{\sqrt{v}}$$

Suponiendo que tanto $a_{i,j}$ como $b_{i,j}$ adoptan iguales valores para cualquiera que sea $X_{i,j}^{(k)}$, entonces se tiene que:

$$X_{i,j} \approx \Gamma(r_{i,j}; a, b)$$

o, expresado en función del coeficiente de variación y de la media individual, se tiene que:

$$\begin{aligned} X_{i,j} &\sim \Gamma\left(r_{i,j}v, \frac{v}{m_{i,j}}\right) \Rightarrow \\ E\left(X_{i,j}\right) &= \frac{r_{i,j}v}{\frac{v}{m_{i,j}}} = r_{i,j}m_{i,j} = x_{i,j} \\ \text{var}\left(X_{i,j}\right) &= \frac{r_{i,j}vm_{i,j}^2}{v^2} = \frac{r_{i,j}m_{i,j}^2}{v} = \frac{x_{i,j}^2}{r_{i,j}v} = x_{i,j}^2 r_{i,j}^{-1} v^{-1} \end{aligned}$$

con lo que se obtiene una función cuadrática para la varianza. Si se supuso, como antes, un modelo multiplicativo, es decir:

$$x_{i,j} = E\left(X_{i,j}\right) = r_{i,j}m_{i,j} = \mu_i \gamma_j \quad \leftarrow \text{media de importes}$$

La obtención de los valores de μ_i y γ_j se puede realizar mediante la aplicación de máxima verosimilitud, siendo la función a maximizar la siguiente:

$$L_D(\mu_0, \mu_1, \dots, \mu_l, \gamma_0, \dots, \gamma_J) = \prod_{i+j \leq l} \frac{(vr_{i,j} / \mu_i \gamma_j)^{vr_{i,j}}}{\Gamma(vr_{i,j})} X_{i,j}^{vr_{i,j}-1} \exp\left\{-\frac{vr_{i,j}}{\mu_i \gamma_j} X_{i,j}\right\}$$

Haciendo las primeras derivadas respecto de los parámetros desconocidos e igualando a cero, se consigue el siguiente juego de ecuaciones:

$$\hat{\mu}_i = \frac{\sum_{j=0}^{l-i} r_{i,j} \frac{X_{i,j}}{\hat{\gamma}_j}}{\sum_{j=0}^{l-i} r_{i,j}} = \sum_{j=0}^{l-i} \omega_{i,j} \frac{X_{i,j}}{\hat{\gamma}_j} \quad / \quad \omega_{i,j} = \frac{r_{i,j}}{\sum_{j=0}^{l-i} r_{i,j}}$$

$$\hat{\gamma}_j = \frac{\sum_{i=0}^{l-j} r_{i,j} \frac{X_{i,j}}{\hat{\mu}_i}}{\sum_{i=0}^{l-j} r_{i,j}} = \sum_{i=0}^{l-j} \omega_{i,j} \frac{X_{i,j}}{\hat{\mu}_i} \quad / \quad \omega_{i,j} = \frac{r_{i,j}}{\sum_{i=0}^{l-j} r_{i,j}}$$

en donde se tiene que cumplir que $\sum_{j=0}^J \hat{\gamma}_j = 1$. A partir de aquí se obtiene el estimador de la reserva como:

$$\hat{x}_{i,j} = \hat{E}(X_{i,j}) = \hat{\mu}_i \hat{\gamma}_j$$

$$\hat{C}_{i,J} = \hat{E}(C_{i,J} | D_l) = C_{i,l-i} + \sum_{j=l-i+1}^J \hat{x}_{i,j}$$

Alternativa 2: a partir del importe total ligado a siniestros a un cierto año de accidente en un cierto año de desarrollo.

En este caso, la estimación puede realizarse por GLM. Tal y como se apuntó en el capítulo 3, para el caso de la distribución gamma, la esperanza y la varianza tienen las siguientes expresiones:

$$E(X_{i,j}) = x_{i,j}$$

$$\text{var}(X_{i,j}) = \frac{\phi_{i,j}}{\omega_{i,j}} \text{var}(x_{i,j}) = \hat{\phi} \cdot x_{i,j}^2$$

3.3.1. Ejemplo

Utilizando los mismos datos que en los casos anteriores y aplicando un modelo multiplicativo, los resultados de la estimación *GLM* son los siguientes:

Parámetro	$\ln(x_{i,j})$	$x_{i,j}$	Parámetro	$\ln(x_{i,j})$	$x_{i,j}$
			γ_0	5,2651090	193,47
μ_1	0,2998635	1,34967	γ_1	6,1995360	492,52
μ_2	0,2707465	1,31094	γ_2	6,2576500	521,99
μ_3	0,1857312	1,20410	γ_3	6,3528290	574,11
μ_4	0,2220293	1,24861	γ_4	5,8510710	347,61
μ_5	0,2645327	1,30282	γ_5	5,6199820	275,88
μ_6	0,3478839	1,41607	γ_6	5,5260610	251,15
μ_7	0,4543873	1,57521	γ_7	5,2577490	192,05
μ_8	0,3033793	1,35443	γ_8	5,5864130	266,78
μ_9	0,1859290	1,20434	γ_9	4,744931	115,00

Con lo que los datos estimados, calculados como $\mu_i \gamma_j \forall i + j > 1$ son:

	0	1	2	3	4	5	6	7	8	9
0										
1										155
2									350	151
3								231	321	138
4							314	240	333	144
5						359	327	250	348	150
6					492	391	356	272	378	163
7				904	548	435	396	303	420	181
8			707	778	471	374	340	260	361	156
9		593	629	691	419	332	302	231	321	138

Y el nivel de reservas, tanto por años de accidente como de forma total, son los que se recogen en la siguiente tabla:

Año	Reserva
1	155
2	500
3	691
4	1.030
5	1.434
6	2.051
7	3.186
8	3.446
9	3.658
Total	16.152

Mientras que el desglose por ejercicios futuros, tanto en valor nominal como actual es el siguiente:

año	dotación anual	
	valor nominal	valor actual
1	4.106	4.055
2	3.383	3.293
3	2.674	2.557
4	1.951	1.785
5	1.503	1.329
6	1.146	951
7	774	601
8	477	351
9	138	98
a dotar	16.152	15.020
Saldo total	41.921	40.789

Para estimar el coeficiente de dispersión, primero se deben calcular los residuos de Pearson, los cuales, a su vez, exigen el cálculo previo de las medias y varianzas individuales. Sabiendo que las medias son el valor estimado por el modelo y las varianzas son dicho valor al cuadrado, se obtiene el siguiente triángulo de residuos de Pearson:

	0	1	2	3	4	5	6	7	8	9
0	0,251	0,082	-0,153	-0,361	0,185	0,678	-0,375	-0,177	-0,130	0,000
1	-0,089	-0,076	-0,049	0,127	-0,218	-0,224	0,345	0,053	0,130	
2	-0,227	0,076	-0,028	0,002	0,290	-0,549	0,312	0,124		
3	-0,099	0,295	-0,106	0,653	-0,410	-0,052	-0,282			
4	0,242	-0,215	0,089	-0,196	-0,067	0,147				
5	0,063	0,015	-0,102	-0,195	0,219					
6	0,088	-0,154	0,096	-0,030						
7	-0,203	-0,050	0,253							
8	-0,027	0,027								
9	0,000									

con cuyos valores se estima el coeficiente de dispersión, $\hat{\phi}$ resultando un valor de 0,082. Por lo que se refiere a la desviación típica de la estimación del nivel de reservas, se obtiene aplicando la siguiente expresión:

$$\sqrt{\text{var}} = \sqrt{\sum_{i+j>l} \hat{\phi} \text{var}(x_{i,j})} = \sqrt{\sum_{i+j>l} \hat{\phi} x_{i,j}^2} = 766,76$$

Finalmente, para calcular el *MSE* global, realmente su raíz cuadrada, falta por calcular el error ligado al proceso de estimación. Aplicando la misma expresión que antes, puesto que lo único que se modifica es la expresión de los $W(x_{i,j})$, que en este caso son:

$$W(x_{ij}) = \frac{1}{\hat{\phi}} \frac{\hat{x}_{ij}^2}{\text{var}(\hat{x}_{ij})} = \frac{\hat{x}_{ij}^2}{\hat{\phi} \cdot \hat{x}_{ij}^2} = \frac{1}{\hat{\phi}}$$

Es decir, en este caso, todos los $W(x_{ij})$ tienen un valor igual a una constante que es la inversa del coeficiente de dispersión. Sustituyendo en la expresión del MSE , resulta que la raíz cuadrada del error de estimación es de 1.922, por lo que la raíz del MSE es igual a:

$$\sqrt{MSE} = \sqrt{1.922^2 + 767^2} = 2.069$$

3.4. Modelo basado en la distribución lognormal

Este modelo fue propuesto por primera vez por Kremer (1982) y posteriormente por Renshaw (1989). En estos modelos, la estructura subyacente en el *Chain Ladder* es idéntica a la de un modelo lineal con variable dependiente expresada en forma logarítmica y regresada sobre dos variables sin interacción entre ellas. Es decir:

$$X_{i,j} = \ln(C_{i,j})$$

$$E(X_{i,j}) = m_{i,j} = \mu + \alpha_i + \beta_j \quad i = 0, \dots, I \quad j = 0, \dots, J$$

en donde, para proceder a su estimación, se impone la restricción de suponer que $\alpha_0 = \beta_0 = 0$, por lo que el número de parámetros a estimar es de $I + J - 1$, que son los $I - 1$ α y $J - 1$ β restantes más el término constante del modelo. Una vez que se ha procedido a su estimación, se puede pasar a completar la parte inferior del triángulo. El modelo se ajusta en escala logarítmica. Dado que $\hat{m}_{i,j}$ es lineal en $X_{i,j}$ y que éstas se distribuyen normalmente, se tiene que:

$$E(\hat{m}_{ij}) = E(\hat{\mu}) + E(\hat{\alpha}_i) + E(\hat{\beta}_j)$$

$$\text{var}(m_{ij}) = \text{var}(\hat{\mu}) + \text{var}(\hat{\alpha}_i) + \text{var}(\hat{\beta}_j) + 2\{\text{cov}(\hat{\mu}, \hat{\alpha}_i) + \text{cov}(\hat{\mu}, \hat{\beta}_j) + \text{cov}(\hat{\alpha}_i, \hat{\beta}_j)\}$$

con lo que:

$$\text{var}(m_{ij}) = \text{var}(\hat{m}_{ij}) + \sigma_\varepsilon^2$$

siendo σ_ε^2 la varianza asociada al término de error de la estimación. Como los $\hat{m}_{i,j}$ se distribuyen según una distribución normal general, entonces los $\hat{C}_{i,j}$ son lognormales, con esperanza y varianzas según las siguientes expresiones:

$$E(\hat{C}_{i,j}) = \exp\left\{E(\hat{m}_{i,j}) + \frac{1}{2}\hat{\sigma}_{\hat{m}_{i,j}}^2\right\}$$

$$\widehat{\text{var}}(\hat{C}_{ij}) = E(\hat{C}_{ij})^2 \left[\exp\{\hat{\sigma}_{\hat{m}_{i,j}}^2\} - 1 \right]$$

Una vez obtenidas las estimaciones correspondientes a las cifras de reservas, el siguiente paso es la obtención del *MSE*. El correspondiente a la reserva de un cierto año de accidente se calcula como:

$$\widehat{var}(\hat{R}_i) = \sum_{j=l-i+1}^J \widehat{var}(\hat{C}_{i,j}) + 2 \sum_{j>l-i+1}^J \sum_{k>j}^J \widehat{cov}(\hat{C}_{i,j}, \hat{C}_{i,k})$$

expresión en la que el problema se centra en encontrar el valor de $\widehat{cov}(\hat{C}_{i,j}, \hat{C}_{i,k})$. Usando el teorema aplicado en Aitchison y Brown (1969)²⁴, la expresión a utilizar es:

$$cov(\hat{C}_{i,j}, \hat{C}_{i,k}) = E(\hat{C}_{i,j})E(\hat{C}_{i,k}) \left[\exp\{cov(\hat{m}_{i,j}, \hat{m}_{i,k})\} - 1 \right]$$

en donde el valor de $cov(\hat{m}_{i,j}, \hat{m}_{i,k})$ es:

$$\begin{aligned} cov(\hat{m}_{i,j}, \hat{m}_{i,k}) &= var(\mu) + cov(\mu, \hat{\alpha}_{i_k}) + cov(\mu, \hat{\beta}_{j_k}) + \\ &+ cov(\mu, \hat{\alpha}_{i_j}) + cov(\hat{\alpha}_{i_j}, \hat{\alpha}_{i_k}) + cov(\hat{\alpha}_{i_j}, \hat{\beta}_{j_k}) + \\ &+ cov(\mu, \hat{\beta}_{j_i}) + cov(\hat{\alpha}_{i_k}, \hat{\beta}_{j_i}) + cov(\hat{\beta}_{j_i}, \hat{\beta}_{j_k}) \end{aligned}$$

Finalmente, si se desea obtener el *MSE* del total de la reserva, definida como suma de las reservas de cada año de accidente, no hay más que generalizar el cálculo anterior, es decir:

$$\widehat{var}(\hat{R}) = \sum_{i+j>l}^J \widehat{var}(\hat{C}_{i,j}) + 2 \sum_{j>l-i+1}^J \sum_{k>j}^J \widehat{cov}(\hat{C}_{i,j}, \hat{C}_{i,k})$$

3.4.1. Ejemplo

Aplicando el modelo expuesto a los datos que se vienen utilizando en todos los ejemplos anteriores, se tienen las siguientes estimaciones, obtenidas aplicando mínimo cuadrados ordinarios:

μ	=	5,42726		
α_1	=	0,14560	β_1	= 0,80081
α_2	=	0,25047	β_2	= 0,85374
α_3	=	0,16929	β_3	= 0,77238
α_4	=	0,24713	β_4	= 0,17862
α_5	=	0,29132	β_5	= 0,13612
α_6	=	0,32727	β_6	= 0,08817
α_7	=	0,36555	β_7	= -0,12011
α_8	=	0,20768	β_8	= 0,22772
α_9	=	0,02378	β_9	= -0,68232

²⁴ Aitchinson, J. y Brown, JAC (1969): *The Log-Normal Distribution*. Cambridge University Press

A partir de estos valores se calculan los correspondientes a $\hat{m}_{i,j}$, obteniéndose el siguiente triángulo de valores:

	0	1	2	3	4	5	6	7	8	9
0										
1										4,891
2									5,905	4,995
3								5,476	5,824	4,914
4						5,763	5,554	5,902	4,992	
5					5,855	5,807	5,599	5,946	5,036	
6				5,933	5,891	5,843	5,634	5,982	5,072	
7			6,565	5,971	5,929	5,881	5,673	6,021	5,111	
8		6,489	6,407	5,814	5,771	5,723	5,515	5,863	4,953	
9	6,252	6,305	6,223	5,630	5,587	5,539	5,331	5,679	4,769	

en donde el valor estimado de la varianza del término de error se obtiene como:

$$\hat{\sigma}_\varepsilon^2 = \sum_{i+j \leq l} \frac{(m_{i,j} - \hat{m}_{i,j})^2}{N - p} = 0,1270$$

siendo N el número total de datos disponibles para la estimación y p el número de parámetros estimados -en este caso, 55 y 19 respectivamente-. La varianza de la predicción se obtiene aplicando la expresión:

$$var(\hat{m}_{ij}) = \hat{\sigma}_\varepsilon^2 + var(\hat{\mu}) + var(\hat{\alpha}_i) + var(\hat{\beta}_j) + 2\{cov(\hat{\mu}, \hat{\alpha}_i) + cov(\hat{\mu}, \hat{\beta}_j) + cov(\hat{\alpha}_i, \hat{\beta}_j)\}$$

Con lo que el triángulo con los valores de los incrementos estimados es:

	0	1	2	3	4	5	6	7	8	9
0										
1										133
2									367	148
3								239	338	136
4							318	258	366	147
5						349	333	270	382	154
6					377	362	345	280	396	160
7				710	392	376	358	291	412	166
8			658	606	335	321	306	248	352	142
9	519	547	504	279	267	254	207	293	118	

y el que recoge las esperanzas de los valores anteriores es:

	0	1	2	3	4	5	6	7	8	9
0										
1										153
2									409	171
3								263	377	157
4							349	285	409	171
5						382	366	299	428	179
6				414	398	381	311	446	186	
7			781	432	416	398	325	466	194	
8		732	675	373	359	344	281	403	168	
9	597	630	581	321	309	296	242	347	145	

con lo que los valores de las reservas, tanto por año de ocurrencia como total, son los recogidos en la siguiente tabla:

año	Reserva estimada
1	153
2	579
3	798
4	1.213
5	1.654
6	2.135
7	3.012
8	3.334
9	3.469
Total	16.348

Por su parte, el desglose de las reservas por ejercicios futuros, tanto en valor nominal como actual es el reflejado en la siguiente tabla:

año	dotación anual	
	valor nominal	valor actual
1	4.081	4.031
2	3.334	3.245
3	2.615	2.501
4	1.988	1.818
5	1.603	1.417
6	1.229	1.020
7	839	652
8	515	379
9	145	102
a dotar	16.348	15.165
Saldo total	42.117	40.934

En el siguiente triángulo se recogen las desviaciones típicas de las estimaciones obtenidas:

	0	1	2	3	4	5	6	7	8	9
0										
1										88
2									200	98
3								122	186	91
4							158	133	204	100
5						171	168	142	216	106
6				187	182	178	150	229	112	
7			359	200	197	192	162	246	119	
8		357	330	183	181	176	148	224	107	
9	340	361	332	184	181	176	147	221	103	

Lo último que queda por deducir es la raíz cuadrada del *MSE*, tanto de cada fila como del total de la reserva. Recordemos que, para el caso de la reserva de cada año de accidente, la expresión que se debe aplicar es:

$$\widehat{\text{var}}(\hat{R}_i) = \sum_{j=l-i+1}^J \widehat{\text{var}}(\hat{C}_{i,j}) + 2 \sum_{j>l-i+1}^J \sum_{k>j}^J \widehat{\text{cov}}(\hat{C}_{i,j}, \hat{C}_{i,k}) \quad i = 0, \dots, l$$

siendo *i* el año para el cual se realiza el cálculo. Los resultados obtenidos, son los que se recogen en la siguiente tabla:

año	Reserva estimada	$(MSE)^{1/2}$	% sobre reserva estimada
1	153	88	57,38%
2	579	231	39,81%
3	798	261	32,72%
4	1.213	353	29,06%
5	1.654	453	27,40%
6	2.135	581	27,20%
7	3.012	849	28,20%
8	3.334	1.057	31,70%
9	3.469	1.450	41,81%

Por su parte, para el total de la reserva, la expresión a utilizar es:

$$\widehat{\text{var}}(\hat{R}) = \sum_{i+j>l}^J \widehat{\text{var}}(\hat{C}_{i,j}) + 2 \sum_{j>l-i+1}^J \sum_{k>j}^J \widehat{\text{cov}}(\hat{C}_{i,j}, \hat{C}_{i,k})$$

obteniéndose los siguientes resultados:

Reserva total estimada	16.348
$(MSE)^{1/2}$	2.645
% sobre reserva total estimada	16,18%

CAPÍTULO 6

MÉTODOS DE REMUESTREO

1. INTRODUCCIÓN AL REMUESTREO O BOOTSTRAP

El *bootstrap* o remuestreo es una técnica desarrollada para llevar a cabo ciertas inferencias. En esencia, se trata de sustituir el tradicional sistema de cálculo de ciertas estadísticas mediante la aplicación de complejas expresiones analíticas por un mecanismo basado en la aplicación de un algoritmo numérico de simulación. La simulación consiste en repetir un proceso de generación de muestras un número suficientemente elevado de veces -pongamos 5.000 veces- para poder realizar esas inferencias. Obviamente, este método requiere el uso de un ordenador. En definitiva, mediante la repetición y generación de muestras de datos se trata de estudiar la precisión asociada a la determinados estadísticos que queremos utilizar, por ejemplo, la media o la mediana. El número de posibles muestras diferentes que se pueden extraer se determina de este modo²⁵:

$$\binom{2n-1}{n} = \frac{(2n-1)!}{n!(n-1)!}$$

A la hora de aplicar este mecanismo cabe distinguir entre el *bootstrap* paramétrico del que no lo es. La diferencia fundamental entre ambos es el conocimiento o no de la función de distribución responsable de la generación de los valores que se desea analizar. Si se conoce, entonces estaríamos ante el caso paramétrico, en el caso contrario sería el no paramétrico, situación en la que las probabilidades de ocurrencia vendrían dadas por la función empírica de distribución.

1.1. *Bootstrap* no paramétrico

Sea \hat{F} una distribución empírica tal que, a cada valor observado le otorga una probabilidad igual a $1/n$. Una muestra *bootstrap* se define como aquella que es aleatoria, de tamaño n y extraída de la muestra inicial x . Es decir:

$$\hat{F} \rightarrow (x_1^*, x_2^*, \dots, x_n^*) \quad \text{muestra bootstrap}$$

²⁵ Hall, P. (1992): *The Bootstrap and the Edgeworth Expansion*, Springer-Verlag, Appendix I.

que no es más que una muestra con reemplazamiento obtenida a partir de la inicial $\mathbf{x} = (x_1, x_2, \dots, x_n)$. Es decir, puede ser de la siguiente forma:

$$x_1^* = x_4, x_2^* = x_2, x_3^* = x_{22}, \dots, x_n^* = x_5$$

Por tanto, cualquier dato de la muestra original puede aparecer ninguna, una o varias veces repetido. Con esta muestra \mathbf{x}^* se calcula una réplica de $\hat{\theta}$:

$$\hat{\theta} = e(\mathbf{x}^*)$$

La puesta en marcha de la generación de muestras mediante bootstrap en un ordenador es muy fácil. Un algoritmo selecciona de forma aleatoria números enteros i_1, i_2, \dots, i_n , cada uno de los cuales es un valor entre 1 y n con probabilidad $1/n$. El *bootstrap* consiste en hacer lo siguiente:

$$x_1^* = x_{i_1}, x_2^* = x_{i_2}, x_3^* = x_{i_3}, \dots, x_n^* = x_{i_n},$$

De esta forma, se generan múltiples muestras, con cada una de las cuales se calcula $\hat{\theta} = e(\mathbf{x}^*)$. Por ejemplo, si nos interesara calcular el estimador *bootstrap* del error típico, los pasos a seguir serían los siguientes:

- 1) se seleccionan B muestras independientes $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*B}$, cada una de las cuales tiene n datos obtenidos con reemplazamiento a partir de la muestra inicial \mathbf{x} .
- 2) se calcula $\hat{\theta}^*$ B veces:

$$\hat{\theta}^*(b) = e(\mathbf{x}^{*b}) \quad / \quad b = 1, 2, \dots, B$$

- 3) con estos B valores de $\hat{\theta}^*(b)$ se estima el error típico $\sigma_F(\hat{\theta})$ como la desviación típica muestral de los B elementos:

$$\hat{\sigma}_B = \sqrt{\frac{\sum_{b=1}^B [\hat{\theta}^*(b) - \hat{\theta}^*(\bullet)]^2}{B-1}}$$

siendo:

$$\hat{\theta}^*(\bullet) = \sum_{b=1}^B \frac{\hat{\theta}^*(b)}{B}$$

1.2. Bootstrap paramétrico

Como se ha indicado anteriormente, la diferencia básica con el caso no paramétrico, estriba en que en esta situación se parte de una determinada función de distribución. A efectos prácticos, en vez de obtener las muestras con

reemplazamiento a partir de los datos iniciales, las muestras generadas de tamaño n se obtienen a partir de la estimación paramétrica de F , \hat{F} . Una vez obtenidas las muestras siguiendo este patrón, se procede igual que en el caso no paramétrico, es decir, se calculan las B réplicas del estadístico a estudiar, $\hat{\theta} = e(\mathbf{x}^{*b})$ y a partir de ellas, se calcula aquello que nos interese, por ejemplo, la desviación típica.

2. ENFOQUES MÁS HABITUALES

2.1. England y Verrall (1999): fundamentos de la propuesta

El trabajo que, sin duda, ha marcado la puesta en práctica de este método, ha sido el publicado por ambos autores en 1999. En él, proponen la utilización de esta metodología como mecanismo para solventar el problema que supone el cálculo del error de predicción mediante una compleja expresión analítica, sustituyéndolo por esta técnica numérica, la cual permite llegar a un resultado muy semejante.

Su trabajo arranca a partir del modelo de Kremer (1982) y de sucesivos trabajos basados en éste, tales como los de Renshaw (1989), Verrall (1991), Renshaw y Verrall (1994) o Mack(1991), entre otros. Todos ellos tienen en común que tratan de explicar el valor de los incrementos anuales en el importe de siniestros, $C_{i,j}$, mediante la utilización de modelos lineales generalizados, con una función de enlace logarítmica y una función lineal -el predictor- de la forma $\eta_{i,j} = c + \alpha_i + \beta_j$, en donde i y j van desde 1 hasta I y J respectivamente y en donde α_1 y β_1 son iguales a cero. Las diferencias radican en la función de distribución que se utiliza en cada caso. Así, Kremer supone que los $C_{i,j}$ siguen una distribución normal, con media igual a $\eta_{i,j}$ y varianza igual a σ^2 , Renshaw y Verrall utilizan una función de Poisson sobredispersa, de media $m_{i,j}$ y de varianza $\phi m_{i,j}$, siendo ϕ el factor de escala y $m_{i,j} = \exp\{\eta_{i,j}\}$. Por su parte, Mack aplica una distribución gamma, con esperanza igual a $m_{i,j}$ y varianza igual a $\phi m_{i,j}^2$, teniendo ϕ y $m_{i,j}$ iguales significados que en el modelo anterior.

La utilización de estos modelos para calcular el nivel de reservas implica la estimación de los parámetros de la función lineal, de f y del error cuadrático medio de predicción. Para este último, Renshaw (1994) demuestra que puede expresarse como:

$$E\left[\left(C_{ij} - \hat{C}_{ij}\right)^2\right] \cong \text{Var}\left(C_{ij}\right) + \text{Var}\left(\hat{C}_{ij}\right)$$

Esta expresión puede considerarse como la suma de dos componentes: por un lado, la correspondiente a la variabilidad intrínseca de los datos -varianza del proceso estocástico considerado- y por otro, la debida a la estimación -varianza de la estimación-. Esta expresión es válida para modelos basados en distribuciones lognormales, sobredispersas de Poisson y gamma.

En concreto, para las dos últimas distribuciones se tiene que $var(C_{i,j}) = \phi m_{i,j}^\rho$: si ρ es igual a uno, entonces estaríamos ante una Poisson, mientras que si fuera igual a dos, entonces estaríamos ante una Gamma. Por lo que se refiere a la varianza de la estimación, en ambos casos se tiene que $\hat{C}_{i,j} = m_{i,j} = \exp\{\eta_{i,j}\}$. Utilizando el método de la delta, se obtiene que:

$$var(\hat{C}_{ij}) \cong \left| \frac{\partial m_{ij}}{\partial \eta_{ij}} \right|^2 var(\eta_{ij})$$

Con lo que se tiene que:

$$E\left[(C_{i,j} - \hat{C}_{i,j})^2\right] \cong \phi m_{i,j}^\rho + \left| \frac{\partial m_{i,j}}{\partial \eta_{i,j}} \right|^2 var(\eta_{i,j}) = \phi m_{i,j}^\rho + m_{i,j}^2 var(\eta_{i,j})$$

El componente final de la expresión anterior, la varianza de la función lineal, generalmente se encuentra disponible en cualquier programa estadístico, permitiendo el cálculo del error cuadrático medio sin dificultad. El error estándar de predicción es la raíz cuadrada del error cuadrático medio. También se puede calcular el error estándar de predicción por años de origen y para el estimador de la reserva total. Llamando Δ al triángulo de siniestros predichos que forman la estimación de las reservas, el estimador de las reservas en el año de origen i se obtiene sumando los valores predichos para la fila i en el triángulo Δ , es decir:

$$C_{i+} = \sum_{j \in \Delta_i} C_{ij}$$

Según Renshaw (1994), el error cuadrático medio de predicción de la reserva del año de origen i es:

$$E\left[(C_{i+} - \hat{C}_{i+})^2\right] \cong \sum_{j \in \Delta_i} \phi m_{i,j}^\rho + \sum_{j \in \Delta_i} m_{i,j}^2 var(\eta_{i,j}) + 2 \sum_{\substack{j_1, j_2 \in \Delta_i \\ j_2 > j_1}} m_{i,j_1} m_{i,j_2} cov(\eta_{i,j_1}, \eta_{i,j_2})$$

El estimador de la reserva total es:

$$C_{++} = \sum_{i,j \in \Delta} C_{i,j}$$

y el error cuadrático medio de predicción de la reserva total es:

$$E\left[(C_{++} - \hat{C}_{++})^2\right] \cong \sum_{i,j \in \Delta} \phi m_{i,j}^\rho + \sum_{i,j \in \Delta} m_{i,j}^2 var(\eta_{i,j}) + 2 \sum_{\substack{i_1, i_2 \in \Delta \\ j_1, j_2 \in \Delta \\ i_1 j_1 > i_2 j_2}} m_{i_1, j_1} m_{i_2, j_2} cov(\eta_{i_1, j_1}, \eta_{i_2, j_2})$$

Para calcular las anteriores expresiones se necesita mucho cuidado para hacer las sumas adecuadas. Los términos de las covarianzas no están directamente disponibles a partir de una salida típica de un programa estadístico. Sin embargo, dado que la matriz de coeficientes del modelo (conocida en la literatura como matriz de diseño) y la de varianzas-covarianzas sí pueden obtenerse, entonces sí es posible calcular la matriz completa de varianzas y covarianzas. De hecho, las varianzas de los parámetros de la función lineal son simplemente la diagonal de esa matriz.

Hay que tener en cuenta que el primer sumatorio en las expresiones de los errores de predicción tanto del año de accidente -fila- como de la totalidad de la reserva son las referidas a la varianza del proceso y que el resto se refiere a la varianza de la estimación.

En situaciones como la descrita, en las que los errores de predicción tienen una expresión sumamente compleja, es donde se recomienda el uso del *bootstrap*. De forma habitual, esta técnica regenera muestras a partir de los datos originales. Sin embargo, dado que se estima la reserva a partir de modelos de regresión, el remuestreo se realiza con los residuos de dicho modelo, tal y como sugieren Efron y Tibshirani (1993). El siguiente paso consiste en determinar cuál será la expresión de los residuos para este caso, ya que en los *GLM* existe un amplio abanico de posibilidades (véase Mc Cullagh y Nelder, 1999). Los más habituales son los residuos obtenidos a partir de la medida de dispersión -*deviance*- , los de Pearson y los de Anscombe.

El proceso de *bootstrapping* supone la regeneración de la muestra, con reemplazamiento, a partir de los residuos. Una muestra de este tipo se crea invirtiendo la fórmula de los residuos utilizando los residuos generados junto con los valores ajustados. Por ejemplo, si se usan los residuos de Pearson, entonces el valor de *C* regenerado a partir del remuestreo será:

$$C^* = r_p^* \sqrt{m} + m$$

Tras la obtención de la muestra por regeneración, el modelo se vuelve a ajustar y se calculan los estadísticos de interés. El proceso se repite un número grande de veces, en cada una de ellas se obtiene una nueva muestra y un nuevo valor de los estadísticos. De esta forma, el error estándar por regeneración es la desviación típica de los estadísticos obtenidos por *bootstrapping*. Dicho error típico es un estimador de la raíz cuadrada de la varianza de la estimación. Sin embargo, no se puede comparar directamente con su equivalente analítico puesto que el error por *bootstrapping* obtenido no tiene en cuenta el número de parámetros utilizados en el ajuste del modelo: el proceso de *bootstrap* sólo utiliza los residuos sin tener en cuenta cómo se han obtenido. La estimación analítica sí tiene en cuenta el número de parámetros estimados puesto que en su proceso se incorporan varianzas y covarianzas, las cuales implícitamente incorporan el factor de escala ϕ en su obtención. El parámetro de escala se puede obtener bien como el cociente entre la suma cuadrática de residuos del modelo de desviaciones y el número de grados de libertad o bien como el

cociente entre la suma cuadrática de residuos del modelo de Pearson y el número de grados de libertad. La diferencia entre ambos es mínima. El número de grados de libertad se obtiene como el número de datos (en la muestra original) menos el número de parámetros usados en el modelo ajustado.

Por tanto, el factor de escala en el modelo por desviaciones (*deviance*) se obtiene como:

$$\phi_D = \frac{\sum r_D^2}{n-p}$$

y en el modelo de Pearson como:

$$\phi_P = \frac{\sum r_P^2}{n-p}$$

en donde n es el número de datos de la muestra, p es el número de parámetros estimados y el sumatorio se extiende sobre el número de residuos. Se puede demostrar que un aumento en el número de parámetros utilizados para ajustar el modelo introduce una penalización (*ceteris paribus*).

England y Verrall consideran que es más coherente usar los residuos escalados de Pearson en la expresión de la varianza analítica y los no escalados de Pearson en el proceso de *bootstrapping*. La estimación de la varianza mediante este procedimiento numérico es similar a la obtenida de forma analítica de la varianza pero sin ajustar por el número de parámetros (es decir, lo obtenido es como si se hubiera dividido por n en vez de por $n-p$). Para permitir una adecuada comparación entre las varianzas obtenidas en los dos procedimientos, se hace necesario realizar un ajuste en la estimación de la varianza por regeneración para tener en cuenta el número de parámetros usados en el ajuste del modelo. El ajuste adecuado se realiza multiplicando la varianza estimada obtenida por *bootstrapping* por $n/n-p$.

Finalmente, para obtener el error de predicción del proceso no hay más que sumar a la varianza de la estimación, la correspondiente al proceso (obviamente, al resultado se le calcula su raíz cuadrada). Dado que estos autores utilizan un modelo basado en la distribución sobredispersa de Poisson, la expresión final del error $-EP_{bs}$ - será:

$$EP_{bs} = \sqrt{\phi_P R + \frac{n}{n-p} [SE_{bs}(R)]^2}$$

siendo R la reserva correspondiente al año deseado -fila- o al total y $SE_{bs}(R)$ el error típico *bootstrap* de la estimación de la reserva.

A la hora de llevar a la práctica su modelo, y por motivos de simplicidad en la exposición, England y Verrall lo hacen teniendo en cuenta que la utilización de la distribución sobredispersa de Poisson lleva a que el resultado de la estimación del nivel de reservas es el mismo que en el caso de aplicar el *Chain Ladder* tradicional.

Por ello, lo que hacen es que, en lugar de aplicar estrictamente el bootstrap sobre los residuos de un modelo *GLM*, lo hacen sobre los obtenidos utilizando directamente el *Chain Ladder*, siguiendo la secuencia de pasos que se indican en el epígrafe siguiente.

2.2. England y Verrall (1989): ejemplo

La realización del bootstrap en este esquema supone seguir el siguiente esquema:

Paso 1: Obtención de los factores de desarrollo del Chain Ladder tradicional.

Usando la expresión:

$$f_j = \frac{\sum_{i=0}^{N-j-1} C_{i,j+1}}{\sum_{i=0}^{N-j-1} C_{i,j}}$$

se obtiene la siguiente tabla de valores:

j	Factor	j	Factor
0	3,5582	5	1,1068
1	1,7784	6	1,0743
2	1,4835	7	1,0975
3	1,1952	8	1,0383
4	1,1244	9	1,0000

Paso 2: Obtención de las cifras acumuladas estimadas

Partiendo de la última diagonal, se obtienen los datos de años anteriores de forma recurrente sin más que dividir la cifra del año t entre el factor de desarrollo del año $t - 1$ pues, como se ha indicado en el paso anterior, las cantidades liquidadas en t son f_{t-1} veces las liquidadas el año anterior. Por ejemplo, la cantidad estimada en la posición (7,0) se obtiene como:

- 1) primero se estima la cantidad de la posición (7,1) a partir de (7,2) -dato conocido- de la siguiente forma:

$$\hat{C}_{(7,1)} \cdot f_1 = C_{(7,2)} \Rightarrow \hat{C}_{(7,1)} = \frac{C_{(7,2)}}{f_1} = \frac{2.010}{1,7784} = 1.130$$

- 2) a continuación se estima la posición (7,0) a partir de la estimación de (7,1) como:

$$\hat{C}_{(7,0)} \cdot f_0 = \hat{C}_{(7,1)} \Rightarrow \hat{C}_{(7,0)} = \frac{\hat{C}_{(7,1)}}{f_0} = \frac{1.130}{3,5582} = 318$$

Haciendo lo mismo para todas las posiciones, se obtiene el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9
0	183	650	1.155	1.714	2.049	2.304	2.550	2.739	3.006	3.121
1	254	905	1.609	2.388	2.854	3.209	3.551	3.815	4.187	
2	252	896	1.593	2.363	2.824	3.176	3.515	3.776		
3	248	882	1.569	2.328	2.782	3.128	3.462			
4	227	809	1.439	2.135	2.552	2.869				
5	239	851	1.514	2.246	2.684					
6	265	943	1.676	2.487						
7	318	1.130	2.010							
8	264	940								
9	233									

Paso 3: Cálculo de los incrementos anuales a partir de los valores estimados en el paso 2.

La variación en las cantidades estimadas se calcula por filas, por lo que:

$$\hat{X}_{i,j} = \begin{cases} \hat{C}_{i,j} & \text{si } j = 0 \\ \hat{C}_{i,j} - \hat{C}_{i,j-1} & \text{si } 0 < j \text{ y } j = N-i-1 \end{cases}$$

Por ejemplo, para $i = 7$, se tiene que:

$$\hat{X}_{7,j} = \begin{cases} \hat{C}_{7,0} & \text{si } j = 0 \\ \hat{C}_{7,1} - \hat{C}_{7,0} & \text{si } j = 1 \\ \hat{C}_{7,2} - \hat{C}_{7,1} & \text{si } j = 2 \end{cases}$$

Repetiendo este cálculo para todas la posiciones, se obtiene el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9
0	183	467	506	559	335	255	246	189	267	115
1	254	651	704	778	466	355	343	264	372	
2	252	644	697	770	461	351	339	261		
3	248	634	687	759	454	346	334			
4	227	582	630	696	417	317				
5	239	612	663	732	438					
6	265	678	734	811						
7	318	813	880							
8	264	676								
9	233									

Pasó 4: Cálculo de los residuos adimensionales de Pearson

Para ello se emplea la expresión:

$$\hat{r}_{i,j} = \frac{C_{i,j} - m_{i,j}}{\sqrt{m_{i,j}}}$$

Por ejemplo, para la posición (7,0), $\hat{r}_{i,j}$ será igual a:

$$\hat{r}_{i,j} = \frac{C_{i,j} - m_{i,j}}{\sqrt{m_{i,j}}} = \frac{243 - 318}{\sqrt{318}} = -4,19$$

Repitiendo este proceso para todas las posiciones, se obtiene el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9
0	4,40	3,05	-2,83	-8,11	4,23	13,03	-5,68	-2,28	-2,14	0,00
1	-1,02	-1,44	-1,30	3,40	-4,59	-3,51	6,12	0,57	1,82	
2	-3,51	2,01	-1,22	-0,58	5,90	-10,05	5,04	1,36		
3	-2,41	5,31	-4,76	13,96	-9,72	-1,67	-6,41			
4	4,81	-4,10	3,19	-4,54	-0,57	4,35				
5	1,86	1,58	-2,00	-4,80	5,43					
6	2,03	-3,37	2,81	-0,76						
7	-4,19	-2,65	5,07							
8	-0,56	0,35								
9	0,00									

Con estos residuos se procede al cálculo del factor de escala o varianza de los residuos:

$$\phi = \frac{\sum_{i=1}^N \hat{r}_i^2}{N - p} = \frac{1.221}{55 - 19} = 33,93$$

siendo N el total de observaciones y p el número de parámetros estimados en el modelo²⁶.

Paso 5: Realización del remuestreo de residuos

Ésta es la fase en la que se lleva a cabo el *bootstrap*. La condición para realizarlo es que todos los residuos implicados tengan la misma probabilidad de ser elegidos. Un posible triángulo puede ser el siguiente:

²⁶ Recordemos que el modelo estimado es:

$$\eta_{i,j} = c + \alpha_i + \beta_j \quad / \quad \alpha_0 = \beta_0 = 0 \quad i, j = 0, \dots, 9$$

por lo que se estiman 19 parámetros: nueve alfas, nueve betas y un término constante.

	0	1	2	3	4	5	6	7	8	9
0	-4,80	-2,14	-2,83	1,58	-4,76	-2,65	5,31	5,31	3,05	-3,37
1	-6,41	3,19	-4,54	3,05	4,81	-2,41	-5,68	-4,76	-2,00	
2	1,86	-4,19	-8,11	4,81	5,07	-4,19	0,57	3,05		
3	3,05	1,86	-2,65	-0,57	0,00	2,01	5,07			
4	-4,80	-4,19	-2,65	-0,56	0,00	-4,80				
5	2,03	-2,00	4,81	-1,67	2,81					
6	-0,76	-4,80	-0,56	0,57						
7	2,01	-8,11	-4,59							
8	-9,72	6,12								
9	-8,11									

Paso 6: Regeneración de los incrementos a partir de la muestra obtenida en el paso anterior

Para ello, se deshace la expresión de los residuos de Pearson, es decir:

$$\hat{C}_{i,j} = \hat{r}_{i,j} \sqrt{m_{i,j}} + m_{i,j}$$

Por ejemplo, para la posición (7,0) se hará:

$$\hat{C}_{7,0} = \hat{r}_{7,0} \sqrt{m_{7,0}} + m_{7,0} = 2,01 \sqrt{318} + 318 = 353$$

Haciendo idéntica operación para todas las posiciones se obtiene el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9
0	118	421	442	596	248	213	329	262	317	79
1	152	732	584	863	570	310	238	186	333	
2	281	538	483	904	570	273	350	310		
3	296	681	617	743	454	384	427			
4	155	481	563	681	417	232				
5	271	562	786	687	497					
6	253	553	718	827						
7	353	581	744							
8	106	835								
9	109									

Paso 7: Regeneración de los datos acumulados

Se obtiene sin más que sumar los incrementos obtenidos en el paso anterior, con lo que el triángulo queda como:

	0	1	2	3	4	5	6	7	8	9
0	118	538	980	1.576	1.824	2.037	2.366	2.628	2.945	3.024
1	152	884	1.468	2.331	2.901	3.211	3.449	3.635	3.968	
2	281	819	1.302	2.206	2.776	3.049	3.399	3.709		
3	296	977	1.594	2.337	2.791	3.175	3.602			
4	155	636	1.199	1.880	2.297	2.529				
5	271	833	1.620	2.306	2.803					
6	253	805	1.524	2.351						
7	353	935	1.679							
8	106	941								
9	109									

Paso 8: Obtención de los factores de desarrollo de la muestra regenerada

A partir de los datos acumulados obtenidos en el paso 7 y aplicando la expresión que permitía la obtención de tales factores, resultan los siguientes valores:

<i>j</i>	Factor	<i>j</i>	Factor
0	3,7125	5	1,1171
1	1,7682	6	1,0824
2	1,5471	7	1,1038
3	1,2181	8	1,0268
4	1,1120	9	1,0000

Paso 9: Cálculo de las reservas a partir de la muestra regenerada y sus factores de desarrollo

Por ejemplo, para la reserva correspondiente al año $i = 7$ será:

$$\hat{R}_7 = \hat{C}_{7,2} \left[(\hat{f}_2 \cdots \hat{f}_7 \cdot \hat{f}_8 \cdot \hat{f}_9) - 1 \right] \quad / \quad \hat{f}_9 = 1$$

$$\hat{R}_7 = 1.679 \cdot (2,8719 - 1) = 3.142$$

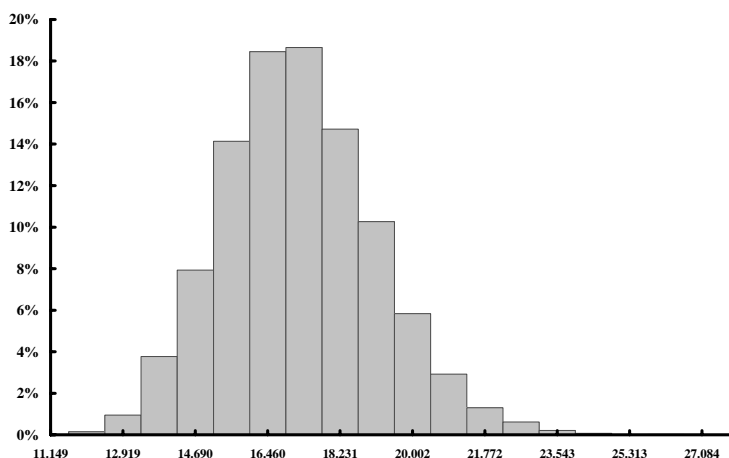
Repitiendo este cálculo para todas las \hat{R}_i se obtiene el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9
0										
1									4.075	
2								4.094	4.203	
3							3.898	4.303	4.418	
4						2.825	3.057	3.375	3.465	
5					3.118	3.483	3.770	4.161	4.272	
6			2.863	3.184	3.557	3.850	4.250	4.363		
7		2.597	3.163	3.518	3.930	4.254	4.695	4.821		
8	1.664	2.574	3.136	3.487	3.896	4.216	4.654	4.779		
9	405	717	1.109	1.351	1.502	1.678	1.817	2.005	2.059	

La reserva total será la suma de las nueve reservas individuales. Su valor, junto con el obtenido por el *Chain Ladder* tradicional se recoge en la siguiente tabla:

Año	Reserva <i>Bootstrap</i>	Reserva <i>Chain Ladder</i>
1	106	160
2	495	527
3	817	776
4	937	1.018
5	1.469	1.405
6	2.013	2.041
7	3.142	3.419
8	3.838	3.575
9	1.950	3.749
Total	14.765	16.671

Este proceso de nueve pasos se repite un número elevado de veces. En este ejemplo se ha llevado a cabo 10.000 veces. Recogiendo el resultado de la reserva total de cada iteración, se puede estimar la distribución de la reserva que, es la reflejada en el siguiente gráfico:



y su desglose por ejercicios futuros, tanto en valor nominal como actual, es el siguiente:

año	dotación anual	
	valor nominal	valor actual
1	4.230	4.177
2	3.503	3.409
3	2.752	2.632
4	2.004	1.834
5	1.560	1.379
6	1.230	1.021
7	836	650
8	514	379
9	150	106
a dotar	16.779	15.587
Saldo total	42.548	41.356

Asimismo, y como se indica en el artículo que sirve de guía, la estimación ha de completarse con la estimación del error de predicción asociado. Para ello se empleará la expresión:

$$EP_{bs} = \sqrt{\phi_P R + \frac{n}{n-p} [SE_{bs}(R)]^2}$$

siendo R el valor de la reserva que se desea calibrar, ϕ el factor de escala y $SE_{bs}(R)$ la desviación típica de las reservas obtenidas por *bootstrap*. El primer elemento de la raíz es la varianza asociada al proceso mientras que el segundo es la asociada al proceso de estimación. Los valores de las raíces cuadradas de cada uno de ellos junto con el error de predicción y su comparación con el valor medio de cada caso se recogen en la siguiente tabla:

Año	Variabilidad (d.t.)		Error de predicción (d.t.)		
	proceso	estimación	Valor	% sobre reservas	
				de C.L.	de <i>Bootstrap</i>
1	74	91	117	73,2%	72,6%
2	134	138	192	36,5%	36,3%
3	162	162	229	29,6%	29,4%
4	186	182	260	25,5%	25,4%
5	218	230	317	22,6%	22,5%
6	263	321	415	20,3%	20,3%
7	341	537	636	18,6%	18,5%
8	348	757	833	23,3%	23,2%
9	357	1.541	1.582	42,2%	41,7%
Total	752	2.306	2.426	14,6%	14,5%

2.3. England (2002): fundamentos de la propuesta

El enfoque anterior se basa únicamente en la consideración de los dos primeros momentos de la distribución. Este trabajo es más ambicioso pues pretende generar la distribución de la reserva estimada utilizando no sólo ambos estadísticos sino la distribución entera. Para ello sugiere un proceso en dos etapas:

- 1) En la primera se utiliza el remuestreo para calcular los errores de estimación. Además, a diferencia del trabajo anterior, es en esta fase en la que se realiza el escalamiento de los residuos de Pearson.
- 2) En la segunda se utilizan técnicas de simulación para conseguir estimar el error del proceso implicado. Es decir, si se supone que los incrementos siguen una distribución del tipo Poisson sobredispersa, entonces se replican tales incrementos generando valores aleatorios que sigan esa distribución.

Todo el proceso es bastante parecido al anterior, de hecho, los cuatro primeros pasos son comunes. A partir de aquí es donde comienzan las diferencias. De hecho, el remuestreo no se realiza con los residuos de Pearson, sino con los ajustados, los cuales tienen como expresión:

$$r'_P = \sqrt{\frac{N}{N-p}} r_P$$

Por ello, al calcular el error de predicción no será necesario ajustar la varianza de la estimación obtenida en el *bootstrap*, por lo que la expresión a utilizar será:

$$EP'_{bs} = \sqrt{\phi'_P R + [SE_{bs}(R)]^2}$$

en donde ϕ'_p hace referencia al factor de escala obtenido con los residuos ajustados.

Para hacer más fácil la comparación con el esquema anterior se realizará el remuestreo de forma que el residuo obtenido en el paso 5 del modelo anterior se multiplicará por $\sqrt{\frac{N}{N-p}}$. Por ello, en el ejemplo que se va a desarrollar se arrancará de ese punto.

2.4. England (2002): ejemplo

Usando los mismos datos que antes, el proceso arranca a partir del cálculo de los residuos ajustados.

Paso 4 bis: obtención de los residuos ajustados de Pearson

Por ejemplo, para la posición (7,0) se tendrá:

$$r'_{7,0} = \sqrt{\frac{55}{36}} r_{7,0} = -4,66$$

Repitiendo para todas las posiciones se tiene el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9
0	4,89	3,39	-3,15	-9,02	4,71	14,49	-6,31	-2,53	-2,38	0,00
1	-1,14	-1,60	-1,44	3,78	-5,10	-3,90	6,80	0,64	2,02	
2	-3,91	2,23	-1,36	-0,65	6,56	-11,17	5,60	1,51		
3	-2,68	5,90	-5,29	15,52	-10,81	-1,86	-7,12			
4	5,35	-4,55	3,55	-5,05	-0,64	4,84				
5	2,07	1,75	-2,23	-5,34	6,04					
6	2,26	-3,75	3,13	-0,84						
7	-4,66	-2,95	5,63							
8	-0,63	0,39								
9	0,00									

Paso 5: Remuestreo a partir de los residuos anteriores:

Un posible triángulo puede ser el siguiente, que es el del paso 5 del modelo anterior multiplicado por $\sqrt{\frac{N}{N-p}}$:

	0	1	2	3	4	5	6	7	8	9
0	-5,34	-2,38	-3,15	1,75	-5,29	-2,95	5,90	5,90	3,39	-3,75
1	-7,12	3,55	-5,05	3,39	5,35	-2,68	-6,31	-5,29	-2,23	
2	2,07	-4,66	-9,02	5,35	5,63	-4,66	0,64	3,39		
3	3,39	2,07	-2,95	-0,64	0,00	2,23	5,63			
4	-5,34	-4,66	-2,95	-0,63	0,00	-5,34				
5	2,26	-2,23	5,35	-1,86	3,13					
6	-0,84	-5,34	-0,63	0,64						
7	2,23	-9,02	-5,10							
8	-10,81	6,80								
9	-9,02									

Paso 6: datos estimados a partir de los residuos obtenidos en el paso 5 y en el paso 3

Es decir, cada posición se obtiene como:

$$\hat{C}_{i,j} = \hat{r}_{i,j} \sqrt{m_{i,j}} + m_{i,j}$$

con lo que se obtiene el siguiente triángulo:

	0	1	2	3	4	5	6	7	8	9
0	110	416	435	600	238	208	339	271	322	75
1	141	741	570	873	582	305	226	178	329	
2	285	526	459	919	582	264	351	316		
3	301	686	609	741	454	388	437			
4	147	469	556	679	417	222				
5	274	557	800	682	504					
6	251	539	717	829						
7	357	556	729							
8	88	853								
9	95									

Paso 7: datos acumulados

Como en el caso anterior, este triángulo se obtiene sin más que acumular incrementos acumulados en cada año de ocurrencia. Los resultados son los recogidos a continuación:

	0	1	2	3	4	5	6	7	8	9
0	110	526	961	1.561	1.799	2.007	2.345	2.616	2.938	3.013
1	141	882	1.452	2.325	2.907	3.211	3.437	3.615	3.944	
2	285	810	1.270	2.188	2.770	3.035	3.385	3.701		
3	301	988	1.597	2.338	2.792	3.180	3.617			
4	147	616	1.172	1.851	2.268	2.490				
5	274	831	1.631	2.313	2.817					
6	251	790	1.507	2.335						
7	357	913	1.642							
8	88	941								
9	95									

Paso 8: obtención de los factores de desarrollo de la muestra regenerada

Son los recogidos en la siguiente tabla:

<i>j</i>	Factor	<i>j</i>	Factor
0	3,7324	5	1,1183
1	1,7670	6	1,0833
2	1,5550	7	1,1046
3	1,2207	8	1,0255
4	1,1106	9	1,0000

Paso 9: proyección de las cifras históricas a partir de los anteriores factores

Se obtienen los siguientes resultados:

	0	1	2	3	4	5	6	7	8	9
0										
1										4.044
2									4.088	4.192
3								3.919	4.328	4.439
4						2.785	3.017	3.333	3.418	
5					3.128	3.498	3.790	4.186	4.293	
6				2.851	3.166	3.541	3.836	4.237	4.345	
7			2.553	3.116	3.461	3.870	4.193	4.631	4.749	
8		1.663	2.586	3.157	3.506	3.921	4.247	4.691	4.811	
9	356	629	978	1.194	1.326	1.483	1.606	1.774	1.820	

Paso 10: obtención de los incrementos estimados a partir de los acumulados anteriores

Sin más que restar respecto del dato del año inmediatamente anterior de desarrollo, con lo que los resultados son:

	0	1	2	3	4	5	6	7	8	9
0										
1										100
2									387	104
3								301	410	110
4							295	232	315	85
5						312	370	292	396	107
6				516	315	375	295	401	108	
7			911	563	345	409	323	438	118	
8		722	923	571	349	415	327	444	119	
9	261	273	349	216	132	157	124	168	45	

Estos resultados se utilizarán como valores medios en la simulación que se realizará en el paso siguiente

Paso 11: Simulación de los pagos correspondientes a cada año

Para ello, se simulará un proceso de Poisson sobredisperso, de la forma:

$$E(C_{i,j}) = m_{i,j}$$

$$var(C_{i,j}) = \phi m_{i,j}$$

En esta fase se generan datos que sigan esta distribución en cada una de las celdas que componen el triángulo inferior derecho. Una forma, que además es la sugerida por England en su trabajo, consiste en generar un proceso de Poisson de parámetro λ igual a $m_{i,j} / \phi$ y multiplicar el resultado por el factor de escala, ϕ . De modo que, siguiendo esta forma de cálculo, una posible colección de resultados sería la siguiente:

	0	1	2	3	4	5	6	7	8	9
0										
1										34
2									305	136
3								339	170	170
4							305	136	339	136
5						441	237	136	204	136
6				475	271	305	509	645	136	
7			543	712	271	373	509	407	102	
8		645	984	509	68	577	271	441	170	
9	305	170	509	237	170	34	136	102	34	

En donde todas y cada una de las celdas son múltiplos de ϕ . Por tanto, para este caso, los posibles valores de la reserva son los recogidos en la siguiente tabla:

Año	Reserva <i>Bootstrap</i>	Reserva <i>Chain Ladder</i>
1	34	160
2	441	527
3	679	776
4	916	1.018
5	1.153	1.405
6	2.341	2.041
7	2.918	3.419
8	3.664	3.575
9	1.696	3.749
Total	13.842	16.671

Este valor se almacenaría y se volvería a repetir el proceso una vez más (realmente, miles de veces) desde el paso 5. Los resultados obtenidos utilizando este mecanismo son muy similares a los que se logran con el primer esquema. Tal y como se ha hecho en todos los ejemplos anteriores, se calcula el desglose por ejercicios, tanto en valor nominal como actual, siendo los resultados los recogidos en la siguiente tabla²⁷:

año	dotación anual	
	valor nominal	valor actual
1	4.234	4.182
2	3.508	3.414
3	2.756	2.636
4	2.008	1.837
5	1.563	1.382
6	1.233	1.024
7	839	652
8	516	380
9	151	106
a dotar	16.808	15.613
Saldo total	42.577	41.382

²⁷ Las diferencias entre los resultados de este método con el anterior y con el *Chain Ladder* tradicional se deben al mecanismo de generación de números aleatorios, a partir de los cuales, se realizan todos los cálculos

Como en el caso anterior, estos resultados han de ir acompañados de una medida del error cometido. Sin embargo, como ya se apuntó, la expresión a utilizar es distinta de la del esquema anterior, pues la corrección por el tamaño de la muestra y los grados de libertad ya está incorporada en la definición de los residuos ajustados. Para este caso, los resultados se recogen en la siguiente tabla:

Año	Variabilidad (d.t.)		Error de predicción (d.t.)		
	proceso	estimación	Valor	% sobre reservas	
				de C.L.	de <i>Bootstrap</i>
1	74	110	133	82,8%	81,6%
2	134	182	226	42,9%	42,5%
3	162	219	272	35,1%	34,8%
4	186	245	307	30,2%	30,0%
5	218	298	369	26,3%	26,1%
6	263	393	473	23,2%	23,1%
7	341	589	681	19,9%	19,8%
8	348	766	841	23,5%	23,4%
9	357	1.436	1.480	39,5%	38,9%
Total	752	2.198	2.323	13,9%	13,8%

que, como puede apreciarse, son muy similares a los obtenidos con el esquema anterior.

CAPÍTULO 7 OTROS MÉTODOS ESTOCÁSTICOS PARA EL CÁLCULO DE LAS PROVISIONES TÉCNICAS

1. MÉTODOS BASADOS EN CREDIBILIDAD

Todos ellos se basan en la utilización de información previa para obtener el resultado final. Los métodos que se van a exponer se basan en la combinación del *CL* y del *B/F*. El principal problema que hay que resolver es el de la determinación del peso que hay que dar a cada uno de los dos métodos anteriores para lograr el resultado final. Es decir, si las reservas según el *CL* pueden definirse como:

$$\hat{R}_i^{CL} = \hat{C}_{i,J}^{CL} - C_{i,I-i} = C_{i,I-i} \left(\prod_{j=I-i}^{J-1} f_j - 1 \right)$$

y según el *B/F* como:

$$\hat{R}_i^{BF} = \hat{C}_{i,J}^{BF} - C_{i,I-i} = (1 - \beta_{I-i}) \mu_i$$

entonces la mezcla de ambos se puede expresar de la siguiente manera:

$$\begin{aligned} \hat{R}_i(c) &= c \hat{R}_i^{CL} + (1-c) \hat{R}_i^{BF} \quad / \quad c \in [0;1] \\ &= \hat{C}_{i,J}^{BF} - C_{i,I-i} + c (\hat{C}_{i,J}^{CL} - \hat{C}_{i,J}^{BF}) \\ &= (1 - \beta_{I-i}) \left[c \hat{C}_{i,J}^{CL} + (1-c) \mu_i \right] \end{aligned}$$

Por tanto, se trata de fijar un criterio que permita deducir el valor óptimo de *c*. El criterio elegido es de la minimización del *MSE* incondicional, definido como:

$$MSE[\hat{R}_i(c)] = E \left\{ \left[R_i - \hat{R}_i(c) \right]^2 \right\}$$

Para lo cual se supondrán las siguientes hipótesis:

- independencia entre los $C_{i,j}$ de diferentes años de accidente
- existencia de una secuencia $\beta_j \forall j = 0, \dots, J$ con $\beta_J = 1$, cumpliéndose que:
$$E(C_{i,j}) = \beta_j E(C_{i,J})$$
- las variables aleatorias U_1, U_2, \dots, U_I son estimadores insesgados de $E(C_{i,j})$. Además tales U_i son independientes tanto de $C_{i,I-i}$ como de $C_{i,J}$

Sabiendo que $\hat{C}_{i,j}^{CL} = \frac{C_{i,l-i}}{\beta_{l-i}}$ y que $\hat{C}_{i,j}^{BF} = C_{i,l-i} + (1-\beta_{l-i})U_i$, se obtiene que:

$$\hat{R}_i(c) = (1-\beta_{l-i})[c\hat{C}_{i,j}^{CL} + (1-c)U_i]$$

En este contexto, se tiene que:

$$\begin{aligned} MSE[\hat{R}_i(c)] &= E\left\{\left[R_i - \hat{R}_i(c)\right]^2\right\} \\ &= var(R_i) + var[\hat{R}_i(c)] - 2cov[R_i, \hat{R}_i(c)] \end{aligned}$$

En Mack (2000) se demuestra que el valor de c que minimiza el MSE incondicional puede expresarse como:

$$c_i^* = \frac{\beta_{l-i} cov(C_{i,l-i}, R_i) + \beta_{l-i}(1-\beta_{l-i})var(U_i)}{1-\beta_{l-i} var(C_{i,l-i}) + \beta_{l-i}^2 var(U_i)}$$

No obstante, para obtener un valor explícito de c^* se necesita un modelo estocástico concreto. Así, en Mack (2000) se supone lo siguiente:

- independencia entre los $C_{i,j}$ de distintos años
- existencia de una secuencia $\beta_j \forall j = 0, \dots, J$ con $\beta_J = 1$ y de una función α^2 que cumple las siguientes condiciones:

1. $E[C_{i,j}/C_{i,J}] = \beta_j C_{i,J}$
2. $var(C_{i,j}/C_{i,J}) = \beta_j(1-\beta_j)\alpha^2(C_{i,J}) \quad \forall i = 0, \dots, l \quad \forall j = 0, \dots, J$

Así, bajo estas hipótesis y aceptando que U_i es un estimador insesgado de $E(C_{i,J})$, a su vez independiente tanto de $C_{i,l-i}$ como de $C_{i,J}$, el valor de c^* que minimiza el MSE incondicional viene dado por la siguiente expresión:

$$c_i^* = \frac{\beta_{l-i}}{\beta_{l-i} + t_i} \quad / \quad t_i = \frac{E[\alpha^2(C_{i,J})]}{var(U_i) + var(C_{i,J}) - E[\alpha^2(C_{i,J})]}$$

por lo que la expresión del MSE es:

$$MSE[\hat{R}_i(c)] = E[\alpha^2(C_{i,J})] \left[\frac{c_i^2}{\beta_{l-i}} + \frac{1}{1-\beta_{l-i}} + \frac{(1-c_i)^2}{t_i} \right] (1-\beta_{l-i})^2$$

De acuerdo con la expresión inicial de $\hat{R}_i(c)$ se tiene que:

- si $c_i = 1$, entonces tendríamos el MSE correspondiente al método B/F, es decir:

$$MSE[\hat{R}_i(1)] = E[\alpha^2(C_{i,J})] \left[\frac{1}{\beta_{l-i}} + \frac{1}{1-\beta_{l-i}} \right] (1-\beta_{l-i})^2$$

- si $c_i = 0$, entonces tendríamos el *MSE* correspondiente al método *CL*, es decir:

$$MSE[\hat{R}_i(0)] = E[\alpha^2(C_{i,J})] \left[\frac{1}{1-\beta_{l-i}} + \frac{1}{t_i} \right] (1-\beta_{l-i})^2$$

- si $c_i = c^*$, entonces tendríamos el *MSE* de la mezcla óptima, es decir:

$$MSE[\hat{R}_i(c^*)] = E[\alpha^2(C_{i,J})] \left[\frac{c_i^2}{\beta_{l-i}} + \frac{1}{1-\beta_{l-i}} + \frac{(1-c_i)^2}{t_i} \right] (1-\beta_{l-i})^2$$

De todo lo anterior se deduce que:

1. en términos de *MSE* y para este modelo, el *B/F* sería más deseable que el *CL* si $t_i > \beta_{l-i}$. Dicho de otro modo, para aquellos años en los que se tenga escasa experiencia en pérdidas -información- es preferible tomar el *B/F* al *CL*
2. el valor de la expresión depende de $E[\alpha^2(C_{i,J})]$ por lo que, para obtener un valor concreto se hace necesario tener una función determinada de dicho operador.

En concreto, si se supone que el cociente $C_{i,j}/C_{i,J}$ sigue una ley beta²⁸ de parámetros $a = a_i b_j$ y $b = a_i(1-b_j)$, se tiene que:

$$E(C_{i,j}/C_{i,J}) = C_{i,J} E\left(\frac{C_{i,j}}{C_{i,J}} / C_{i,J}\right) = \beta_j C_{i,J}$$

$$var(C_{i,j}/C_{i,J}) = C_{i,J}^2 E\left(\frac{C_{i,j}}{C_{i,J}} / C_{i,J}\right) = \beta_j (1-\beta_j) \frac{C_{i,J}^2}{1+\alpha_j} \quad \forall i = 0, \dots, l \quad \forall j = 0, \dots, J$$

²⁸ la función de densidad de una beta(a,b) tiene la expresión:

$$f(x) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} \quad \forall x \geq 0 \quad / \quad B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

siendo:

$$E(x) = \frac{a}{a+b}$$

$$var(x) = \frac{ab}{(a+b)^2 (a+b+1)}$$

$$cv(x) = \sqrt{\frac{b}{a(a+b+1)}}$$

Por lo que $\alpha^2(C_{i,j}) = \frac{C_{i,j}^2}{1 + \alpha_i}$, siendo su esperanza:

$$E[\alpha^2(C_{i,j})] = \frac{E(C_{i,j}^2)}{1 + \alpha_i} = \frac{1}{1 + \alpha_i} [E(C_{i,j})^2 + cv^2(C_{i,j})] = \frac{E(C_{i,j})^2}{1 + \alpha_i} [cv^2(C_{i,j}) + 1] \quad / \quad E(C_{i,j}) = \mu_i \text{ conocida}$$

Ejemplo:

Partimos de los datos utilizados en los ejemplos anteriores. Supongamos que se dispone de la siguiente información a priori, la cual no ha sido deducida a partir de los datos disponibles sino a partir de opiniones de expertos:

- $cv(U_i) = 5\%$
- error del proceso = 6% -r-
- $\alpha_i = 525 \quad \forall i$

Además supondremos que $cv(C_{i,j}) = \sqrt{cv^2(U_i) + r^2}$, que con los valores supuestos arroja un resultado del 7,81%. Con todo ello, resulta la siguiente tabla de resultados intermedios:

<i>i</i>	α_i	$cv(U_i)$	<i>r</i>	$cv(C_{i,j})$	$E[\alpha^2(C_{i,j})]$	$var(U_i)$	$var(C_{i,j})$
0	525	5%	6%	7,81%	21.073	27.544	67.206
1	525	5%	6%	7,81%	43.959	57.456	140.193
2	525	5%	6%	7,81%	40.537	52.983	129.278
3	525	5%	6%	7,81%	35.509	46.411	113.243
4	525	5%	6%	7,81%	29.971	39.173	95.583
5	525	5%	6%	7,81%	39.585	51.738	126.241
6	525	5%	6%	7,81%	55.408	72.420	176.705
7	525	5%	6%	7,81%	53.983	70.557	172.158
8	525	5%	6%	7,81%	32.333	42.260	103.115
9	525	5%	6%	7,81%	18.512	24.196	59.038

con los cuales, además de los β_{1-i} , se pueden obtener los valores de t_i y de c_i^* que son los siguientes:

<i>i</i>	t_i	$1-i$	c_i^*
0	28,60%	1,000	77,76%
1	28,60%	0,963	77,10%
2	28,60%	0,878	75,42%
3	28,60%	0,817	74,07%
4	28,60%	0,738	72,07%
5	28,60%	0,656	69,65%
6	28,60%	0,549	65,76%
7	28,60%	0,370	56,41%
8	28,60%	0,208	42,12%
9	28,60%	0,059	16,98%

A partir de los c_i^* estimados, el nivel de reservas asociado se obtiene como combinación lineal de CL y B/F con un factor de ponderación c_i^* . Los resultados son los que se recogen en la siguiente tabla:

	0	1	2	3	4	5	6	7	8	9
0										
1										4.351
2									4.150	4.312
3								3.720	4.084	4.241
4							3.177	3.414	3.748	3.892
5						3.029	3.363	3.619	3.981	4.137
6				3.004	3.397	3.777	4.070	4.482	4.660	
7			2.973	3.549	3.989	4.413	4.739	5.199	5.397	
8		1.634	2.400	2.859	3.209	3.546	3.806	4.172	4.330	
9	721	1.249	1.832	2.182	2.448	2.705	2.902	3.181	3.301	

por lo que la estimación anual de reservas es:

año	Reserva estimada
1	164
2	536
3	779
4	1.023
5	1.453
6	2.173
7	3.387
8	3.390
9	3.068

lográndose un total de 15.974. En cuanto al desglose por ejercicios futuros, distinguiendo entre valor nominal y valor actual, los valores son los recogidos en la siguiente tabla:

año	dotación anual	
	valor nominal	valor actual
1	4.111	4.060
2	3.361	3.270
3	2.609	2.495
4	1.921	1.758
5	1.498	1.325
6	1.154	958
7	762	592
8	437	321
9	120	85
a dotar	15.974	14.865
Saldo total	41.743	40.634

Finalmente, el valor del *MSE* para cada año se obtiene sin más que aplicar la expresión

$$MSE[\hat{R}_i(c^*)] = E[\alpha^2(C_{i,j})] \left[\frac{c_i^2}{\beta_{l-i}} + \frac{1}{1-\beta_{l-i}} + \frac{(1-c_i)^2}{t_i} \right] (1-\beta_{l-i})^2$$

La raíz cuadrada de los resultados de su aplicación para los valores 0, 1 y c_i^* se recogen en la siguiente tabla:

<i>i</i>	<i>c</i> = 0	<i>c</i> = 1	<i>c</i> = c_i^*
1	41	43	41
2	75	84	74
3	89	103	87
4	103	123	99
5	144	173	137
6	213	254	198
7	303	330	269
8	351	311	295
9	546	273	433

y los correspondientes al total de la reserva, obtenido como la raíz de la suma de la elemento al cuadrado, son los siguientes:

<i>c</i> = 0	<i>c</i> = 1	<i>c</i> = c_i^*
640	778	655

2. MODELO DE GOGOL (1993)

Conocido también como modelo lognormal / lognormal. Es decir, que tanto la distribución a priori como a posteriori siguen la citada ley estadística²⁹. Las hipótesis de partida en este modelo son:

- independencia entre los $C_{i,j}$ de diferentes años de accidente.
- $C_{i,j}$ se distribuye según una lognormal de parámetros $\mu^{(i)}$ y $\sigma_i^2 \forall i = 0, \dots, l$, siendo la media de $C_{i,j}$ función de ambos parámetros e igual a:

²⁹ ley con función de densidad:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2 x}} \exp\left\{-\frac{1}{2}\left[\frac{\log(x)-\mu}{\sigma}\right]^2\right\} \quad \forall x > 0$$

con las siguientes características:

$$E(x) = \exp\left\{\mu + \frac{\sigma^2}{2}\right\}$$

$$\text{var}(x) = \exp\{2\mu + \sigma^2\} [\exp\{\sigma^2\} - 1]$$

$$\text{cv}(x) = \sqrt{\exp\{\sigma^2\} - 1}$$

$$\mu_i = E(C_{i,J}) = \exp\left\{\mu^{(i)} + \frac{1}{2}\sigma_i^2\right\}$$

Este modelo también cumple que:

- $E(C_{i,j} | C_{i,J}) = \beta_j C_{i,J}$
- $\text{var}(C_{i,j} | C_{i,J}) = \beta_j (1 - \beta_j) \alpha^2(C_{i,J}) \quad \forall i = 0, \dots, I \quad \forall j = 0, \dots, J$

en donde β_j se define igual que anteriormente, por lo que:

- $E(C_{i,j} | C_{i,J}) = \exp\left\{v_j + \frac{1}{2}\tau_j^2\right\} = \beta_j C_{i,J}$
- $\text{var}(C_{i,j} | C_{i,J}) = \exp\{2v_j + \tau_j^2\} [\exp\{\tau_j^2\} - 1] = \beta_j (1 - \beta_j) \alpha^2(C_{i,J})$

y despejando v_j y τ_j^2 resultan las siguientes expresiones, ambas en función de $C_{i,J}$:

$$\tau_j^2 = \ln\left[1 + \frac{1 - \beta_j}{\beta_j} \frac{\alpha^2(C_{i,J})}{C_{i,J}^2}\right]$$

$$v_j = \ln(\beta_j C_{i,J}) - \frac{1}{2} \ln\left[1 + \frac{1 - \beta_j}{\beta_j} \frac{\alpha^2(C_{i,J})}{C_{i,J}^2}\right]$$

en donde las expresiones finales dependen de cuál sea la forma funcional de $\alpha^2(C_{i,J})$. Así, si la forma es $\alpha^2(C_{i,J}) = a^2 c^2$, las expresiones de ambos parámetros resultan ser:

$$\tau_j^2 = \ln\left(1 + \frac{1 - \beta_j}{\beta_j} a^2\right)$$

$$v_j(c) = \ln(C_{i,J}) + \ln(\beta_j) - \frac{1}{2} \tau_j^2$$

siendo la función de densidad conjunta de $C_{i,j}$ y $C_{i,J}$ la expresión:

$$\begin{aligned} f(x, y) &= f_{C_{i,j}|C_{i,J}}(x|y) f_{C_{i,J}}(y) = \\ &= \frac{1}{\sqrt{2\pi}\tau_j(y)x} \exp\left\{-\frac{1}{2}\left[\frac{\ln(x) - v_j(y)}{\tau_j(y)}\right]^2\right\} \frac{1}{\sqrt{2\pi}\sigma_i y} \exp\left\{-\frac{1}{2}\left[\frac{\ln(y) - \mu^{(i)}}{\sigma_i}\right]^2\right\} = \\ &= \frac{1}{2\pi\sigma_i\tau_j(y)xy} \exp\left\{-\frac{1}{2}\left[\frac{\ln(x) - v_j(y)}{\tau_j(y)}\right]^2 - \frac{1}{2}\left[\frac{\ln(y) - \mu^{(i)}}{\sigma_i}\right]^2\right\} \end{aligned}$$

Partiendo de la anterior función de densidad conjunta, de la forma funcional tomada para $\alpha^2(C_{i,j})$ y de los valores resultantes para τ_j^2 y para v_j , se demuestra que, dada $C_{i,j}$, la distribución a posteriori de $C_{i,j}$ es lognormal con parámetros:

$$\mu_{post(i,j)} = \left(1 - \frac{\tau_j^2}{\sigma_i^2 + \tau_j^2}\right) \left[\frac{1}{2} \tau_j^2 + \ln\left(\frac{C_{i,j}}{\beta_j}\right) \right] + \frac{\tau_j^2}{\sigma_i^2 + \tau_j^2} \mu^{(i)}$$

$$\sigma_{post(i,j)}^2 = \frac{\tau_j^2}{\sigma_i^2 + \tau_j^2} \sigma_i^2$$

en donde $\mu_{post(i,j)}$ es una media ponderada entre el parámetro a priori, $\mu^{(i)}$, y la observación transformada, $\frac{1}{2} \tau_j^2 + \ln\left(\frac{C_{i,j}}{\beta_j}\right)$. Los factores de ponderación son:

$$\alpha_{i,j} = \left(1 - \frac{\tau_j^2}{\sigma_i^2 + \tau_j^2}\right) = \frac{\sigma_i^2}{\sigma_i^2 + \tau_j^2}$$

$$1 - \alpha_{i,j} = \frac{\tau_j^2}{\sigma_i^2 + \tau_j^2}$$

lo cual supone que la media a priori de $C_{i,j}$ se modifica pasando de:

$$E(C_{i,j}) = \exp\left\{\mu^{(i)} + \frac{1}{2} \sigma_i^2\right\}$$

a ser:

$$E(C_{i,j} | C_{i,j}) = \exp\left\{\mu_{post(i,j)} + \frac{1}{2} \sigma_{post(i,j)}^2\right\}$$

Así, si se desean calcular los valores estimados de $C_{i,j} \forall i + j > l$ tomando como dato conocido $C_{i,l-i}$ la anterior esperanza adopta la siguiente forma:

$$\begin{aligned} E(C_{i,j} | C_{i,l-i}) &= \exp\left\{\mu_{post(i,j)} + \frac{1}{2} \sigma_{post(i,j)}^2\right\} \quad \forall i + j > l \\ &= \exp\left\{\left(1 - \alpha_{i,j}\right)\left(\mu^{(i)} + \frac{1}{2} \sigma_i^2\right) + \alpha_{i,j} \left[\ln\left(\frac{C_{i,l-i}}{\beta_{l-i}/\beta_j}\right) + \frac{1}{2} \tau_j^2\right]\right\} \\ &= \exp\left\{\left(1 - \alpha_{i,j}\right)\left(\mu^{(i)} + \frac{1}{2} \sigma_i^2\right) + \alpha_{i,j} \left[-\ln\left(\frac{\beta_{l-i}}{\beta_j}\right) + \frac{1}{2} \tau_j^2\right]\right\} C_{i,l-i}^{\alpha_{i,j}} \end{aligned}$$

que es la expresión a partir de la cual se procede a calcular el nivel de reservas. Otra forma de escribir la anterior expresión puede obtenerse teniendo en cuenta que:

$$\alpha_{i,j} = \frac{\sigma_i^2}{\sigma_i^2 + \tau_j^2}$$

$$1 - \alpha_{i,j} = \frac{\tau_j^2}{\sigma_i^2 + \tau_j^2}$$

$$\mu_j = \exp\left\{\mu^{(i)} + \frac{1}{2}\sigma_i^2\right\}$$

$$\hat{C}_{i,j}^{CL} = \frac{C_{i,l-i}}{(\beta_{l-i}/\beta_j)}$$

con lo que sustituyendo queda:

$$\hat{C}_{i,j}^G = \exp\left\{\left(1 - \alpha_{i,l-i}\right)\mu^{(i)} + \alpha_{i,l-i} \ln\left(\frac{C_{i,l-i}}{\beta_{l-i}/\beta_j}\right) + \alpha_{i,l-i}\tau_{l-i}^2\right\}$$

$$= \mu_j^{1-\alpha_{i,l-i}} \exp\left\{\ln\left(\hat{C}_{i,j}^{CL}\right) + \frac{1}{2}\tau_{l-i}^2\right\}^{\alpha_{i,l-i}}$$

Por lo que se refiere al cálculo del nivel de error asociado a la estimación, *MSE*, puede distinguirse entre:

- *MSE* condicional, cuya expresión es:

$$MSE_{C_{i,j}|C_{i,l-i}}(\hat{C}_{i,j}^G) = \text{var}(C_{i,j} | C_{i,l-i}) =$$

$$= \exp\left\{2\mu_{\text{post}(i,l-i)} + \sigma_{\text{post}(i,l-i)}^2\right\} \left[\exp\left(\sigma_{\text{post}(i,l-i)}^2\right) - 1\right]$$

$$= \left[E(C_{i,j} | C_{i,l-i})\right]^2 \left[\exp\left(\sigma_{\text{post}(i,l-i)}^2\right) - 1\right]$$

$$= \left(\hat{C}_{i,j}^G\right)^2 \left[\exp\left(\sigma_{\text{post}(i,l-i)}^2\right) - 1\right]$$

- *MSE* incondicional, cuya expresión es:

$$MSE_{C_{i,j}}(\hat{C}_{i,j}^G) = E\left[\left(C_{i,j} - \hat{C}_{i,j}^G\right)^2\right]$$

$$= E\left[\text{var}(C_{i,j} | C_{i,l-i})\right] =$$

$$= \hat{E}\left[\left(\hat{C}_{i,j}^G\right)^2\right] \left[\exp\left(\sigma_{\text{post}(i,l-i)}^2\right) - 1\right]$$

que indica que, el valor final depende de cuál sea la esperanza de $\hat{C}_{i,j}^G$, la cual a su vez, depende de cuál sea la distribución de $C_{i,l-i}/\beta_{l-i}$, que es la estimación asociada al uso del CL. Sabiendo que:

$$f_{C_{i,l-i}}(x) = \int_0^{\infty} f_{C_{i,l-i}, C_{i,j}}(x, y) dy$$

$$= k \int_0^{\infty} \frac{1}{\sqrt{2\pi(\sigma_i^2 + \tau_{l-i}^2)} \cdot y} \exp \left\{ -\frac{1}{2} \frac{\left\{ \ln(y) - \left[\frac{\sigma_i^2 c(x) + \tau_{l-i}^2 \mu^{(i)}}{\sigma_i^2 + \tau_{l-i}^2} \right] \right\}^2}{\frac{\sigma_i^2 \tau_{l-i}^2}{\sigma_i^2 + \tau_{l-i}^2}} \right\} dy$$

siendo k :

$$k = \frac{1}{\sqrt{2\pi(\sigma_i^2 + \tau_{l-i}^2)} \cdot x} \exp \left\{ -\frac{1}{2} \frac{\left[\ln(x/\beta_{l-i}) + \frac{1}{2} \tau_{l-i}^2 - \mu^{(i)} \right]^2}{\sigma_i^2 + \tau_{l-i}^2} \right\}$$

Como el valor de la integral es igual a uno, entonces $\hat{C}_{i,j}^{CL}$ se distribuye según una lognormal de parámetros $\frac{1}{2} \tau_{l-i}^2 - \mu^{(i)}$ y $\sigma_i^2 + \tau_{l-i}^2$. Teniendo en cuenta la propiedad reproductiva de la lognormal, se tiene que:

$$\left(\hat{C}_{i,j}^{CL} \right)^\alpha \approx \log N \left(\alpha \mu^{(i)} - \frac{\alpha}{2} \tau_{l-i}^2, \alpha^2 (\sigma_i^2 + \tau_{l-i}^2) \right)$$

A partir de la expresión que permite obtener el valor de $\hat{C}_{i,j}^G$, que $\mu_i = E(C_{i,j}) = \exp \left\{ \mu^{(i)} + \frac{1}{2} \sigma_i^2 \right\}$, y conociendo la distribución de $\hat{C}_{i,j}^{CL}$ se obtiene que:

$$\begin{aligned} MSE_{C_{i,j}}(\hat{C}_{i,j}^G) &= \hat{E} \left[\left(\hat{C}_{i,j}^G \right)^2 \right] \left[\exp(\sigma_{post(i,l-i)}^2) - 1 \right] \\ &= \mu_i^{2(1-\alpha_{i,l-i})} \exp(\alpha_{i,l-i} \tau_{l-i}^2) \left[\exp(\sigma_{post(i,l-i)}^2) - 1 \right] E \left[\left(\hat{C}_{i,j}^{CL} \right)^{2\alpha_{i,l-i}} \right] \\ &= \exp \left[2\mu^{(i)} + (1-\alpha_{i,l-i})\sigma_i^2 + \alpha_{i,l-i}\tau_{l-i}^2 \right] \times \\ &\quad \exp \left[-\alpha_{i,l-i}\tau_{l-i}^2 + 2\alpha_{i,l-i}^2(\sigma_i^2 + \tau_{l-i}^2) \right] \left[\exp(\sigma_{post(i,l-i)}^2) - 1 \right] \end{aligned}$$

Agregando términos y teniendo en cuenta que $\alpha_{i,j}(\sigma_i^2 + \tau_j^2) = \sigma_i^2$, se obtiene la expresión final para el *MSE* incondicional:

$$MSE_{C_{i,j}}(\hat{C}_{i,j}^G) = \exp\left[2\mu^{(i)} + (1 + \alpha_{i,l-i})\sigma_i^2\right] \left[\exp(\sigma_{post(i,l-i)}^2) - 1 \right] \quad \forall i = 1, \dots, l$$

Ejemplo:

A partir de los datos habituales y suponiendo además que:

- se tiene una estimación a priori de $E(C_{i,j})$ - μ_i - dada por los siguientes valores:

año	$\mu_i = E(C_{i,j})$
0	3.319
1	4.794
2	4.604
3	4.309
4	3.958
5	4.549
6	5.382
7	5.313
8	4.111
9	3.111

- $a_i = 525$ para todo i , se supone que $a^2(c) = a^2c^2$ y que $a^2 = \frac{1}{1 + \alpha_i}$
- el coeficiente de variación de $C_{i,j}$ viene dado por: $cv(C_{i,j}) = \sqrt{cv^2(U_i) + r^2}$, teniendo sus dos componentes los mismos valores que en el ejemplo anterior.

A partir de estos datos se obtiene primeramente los valores de σ_i y de $\mu^{(i)}$ de acuerdo a las siguientes expresiones:

$$E(C_{i,j}) = \mu_i = \exp\left\{\mu^{(i)} + \frac{1}{2}\sigma_i^2\right\} \rightarrow \mu^{(i)} = \ln(\mu_i) - \frac{1}{2}\sigma_i^2$$

$$cv(C_{i,j}) = \sqrt{\exp\{\sigma_i^2\} - 1} \rightarrow \sigma_i = \sqrt{\ln[cv^2(C_{i,j}) + 1]}$$

Como los β_j se calculan a partir de los factores de desarrollo del *CL*, entonces es posible obtener los τ_j^2 usando la fórmula $\tau_j^2 = \ln\left(1 + \frac{1 - \beta_j}{\beta_j} a^2\right)$. Con todo ello, resulta la siguiente tabla de resultados intermedios:

i	$cv(C_{ij})$	σ_j	$\mu^{(i)}$	β_{l-i}	a^2	τ_{l-i}
0	7,81%	7,80%	8,1045	1,000	0,19%	0,00%
1	7,81%	7,80%	8,4721	0,963	0,19%	0,85%
2	7,81%	7,80%	8,4316	0,878	0,19%	1,63%
3	7,81%	7,80%	8,3653	0,817	0,19%	2,06%
4	7,81%	7,80%	8,2806	0,738	0,19%	2,60%
5	7,81%	7,80%	8,4197	0,656	0,19%	3,15%
6	7,81%	7,80%	8,5878	0,549	0,19%	3,95%
7	7,81%	7,80%	8,5748	0,370	0,19%	5,68%
8	7,81%	7,80%	8,3185	0,208	0,19%	8,49%
9	7,81%	7,80%	8,0397	0,059	0,19%	17,36%

A partir de aquí y sabiendo que:

$$\begin{aligned}
 E(C_{i,j} | C_{i,l-i}) &= \exp \left\{ \mu_{post(i,j)} + \frac{1}{2} \sigma_{post(i,j)}^2 \right\} \quad \forall i + j > l \\
 &= \exp \left\{ (1 - \alpha_{i,j}) \left(\mu^{(i)} + \frac{1}{2} \sigma_i^2 \right) + \alpha_{i,j} \left[\ln \left(\frac{C_{i,j}}{\beta_{l-i}/\beta_j} \right) + \frac{1}{2} \tau_j^2 \right] \right\} \\
 &= \exp \left\{ (1 - \alpha_{i,j}) \left(\mu^{(i)} + \frac{1}{2} \sigma_i^2 \right) + \alpha_{i,j} \left[-\ln \left(\frac{\beta_{l-i}}{\beta_j} \right) + \frac{1}{2} \tau_j^2 \right] \right\} C_{i,l-i}^{\alpha_{i,j}}
 \end{aligned}$$

se puede proceder a rellenar el triángulo inferior:

	0	1	2	3	4	5	6	7	8	9
0										
1										4.352
2									4.156	4.315
3								3.724	4.087	4.243
4							3.182	3.418	3.752	3.895
5						3.065	3.392	3.644	3.999	4.152
6				3.081	3.464	3.834	4.119	4.521	4.694	
7			2.963	3.541	3.981	4.407	4.734	5.195	5.394	
8		1.591	2.361	2.822	3.173	3.512	3.773	4.140	4.299	
9	677	1.204	1.786	2.134	2.400	2.656	2.853	3.131	3.251	

con lo que el nivel de reservas anuales es el siguiente:

año	Reserva estimada
1	165
2	539
3	781
4	1.026
5	1.468
6	2.207
7	3.384
8	3.359
9	3.018

y el total es de 15.948. El desglose por ejercicios futuros, distinguiendo entre valores nominales y actuales es el recogido en la siguiente tabla:

año	dotación anual	
	valor nominal	valor actual
1	4.143	4.092
2	3.344	3.254
3	2.595	2.481
4	1.908	1.746
5	1.486	1.314
6	1.152	956
7	764	593
8	437	321
9	120	85
a dotar	15.948	14.843
Saldo total	25.769	25.769

Por lo que se refiere a los errores asociados a este cálculo, el *MSE* condicional se obtiene aplicando:

$$MSE_{C_{i,J}|C_{i,l-i}}(\hat{C}_{i,J}) = (C_{i,J})^2 \left[\exp\{\sigma_{post(i,l-i)}^2\} - 1 \right]$$

mientras que, para el incondicional, la expresión es:

$$MSE_{C_{i,J}}(\hat{C}_{i,J}) = \exp\{2\mu^{(i)} + (1 + \alpha_{i,l-i})\sigma_i^2\} \left[\exp\{\sigma_{post(i,l-i)}^2\} - 1 \right] \quad \forall i = 1, \dots, l$$

Los resultados para cada año y para el total son los recogidos en la siguiente tabla:

año	MSE condicional	MSE incondicional
1	37	41
2	69	74
3	85	86
4	96	98
5	121	133
6	165	190
7	248	245
8	247	237
9	232	222
Total	491	493

3. MODELO DE POISSON SOBREDISPERSA CON GAMMA

El modelo de Poisson sobredispersa se ha tratado con un enfoque bayesiano en los trabajos de Verrall (1990, 2000 y 2004) y en Renshaw y Verrall (1998). El modelo que se plantea en ellos introduce un conjunto de variables latentes, Θ_i , las cuales describen las características del año i -ésimo en términos de riesgo, es decir, permite clasificar ese año como bueno o malo. Condicionadas a estas variables, se obtienen los valores estimados de $C_{i,j}$ y de los incrementos anuales $X_{i,j}$.

Se parte de suponer la existencia de Θ_i y de $Z_{i,j}$, variables aleatorias y de $\phi_i > 0$,

$\gamma_0, \dots, \gamma_J > 0$ que cumplen que $\sum_{j=1}^J \gamma_j = 1$, tales que:

- dada Θ_i , los $Z_{i,j}$ son variables aleatorias independientes y distribuidas según una Poisson, en las que los incrementos anuales, definidos como $X_{i,j} = \phi_i Z_{i,j}$ satisfacen que:

$$- E(X_{i,j} | \Theta_i) = \Theta_i \gamma_j$$

$$- var(X_{i,j} | \Theta_i) = \phi_i \Theta_i \gamma_j$$

Por lo que, dado Θ_i , la esperanza y la varianza de $Z_{i,j}$ son:

$$- E(Z_{i,j} | \Theta_i) = E\left(\frac{X_{i,j}}{\phi_i} | \Theta_i\right) = \frac{\Theta_i \gamma_j}{\phi_i}$$

$$- var(Z_{i,j} | \Theta_i) = var\left(\frac{X_{i,j}}{\phi_i} | \Theta_i\right) = \frac{\Theta_i \gamma_j}{\phi_i}$$

- las combinaciones $[\Theta_i, X_{i0}, \dots, X_{iJ}]$ son independientes entre sí y Θ_i se distribuye según una Gamma de parámetros a y b .

Para el total acumulado se tiene que:

$$C_{i,J} = \sum_{j=0}^J X_{i,j} = \phi_i \sum_{j=0}^J Z_{i,j}$$

que, dado Θ_i supone que:

$$\frac{C_{i,J}}{\phi_i} \approx P(\Theta_i / \phi_i)$$

Por lo que la esperanza es igual a Θ_i / ϕ_i y por tanto, $E(C_{i,J} / \Theta_i) = \Theta_i$, lo cual supone que este parámetro puede interpretarse como el importe acumulado esperado para el año i -ésimo. Es necesario reseñar que este modelo presenta como limitación que todos los incrementos $X_{i,j}$ tienen que ser no negativos.

Teniendo en cuenta que, de acuerdo con el modelo presentado la esperanza a priori de los $X_{i,j}$ puede expresarse como:

$$E(X_{i,j}) = E[E(X_{i,j} / \Theta_i)] = E[\Theta_i \gamma_j / \Theta_i] = \gamma_j E(\Theta_i) = \gamma_j \frac{a_i}{b_j}$$

entonces se tiene que la distribución a posteriori de Θ_i , dado $\{X_{i,j}\} \forall i = 0, \dots, I$ $\forall j = 0, \dots, J$, sigue una distribución Gamma de parámetros:

$$a_{i,j}^{post} = a_i + \frac{C_{i,j}}{\phi_i}$$

$$b_{i,j}^{post} = b_i + \sum_{k=0}^j \frac{\gamma_k}{\phi_i} = b_i + \frac{\beta_j}{\phi_i} \quad / \quad \beta_j = \sum_{k=0}^j \gamma_k$$

Para comprobarlo, no hay más que calcular la función de densidad conjunta de $[X_{i0}, \dots, X_{iJ}]$ condicionada a Θ_i para ver que la función resultante sigue una Gamma con los parámetros antes indicados. Por tanto, su esperanza y varianza vienen dadas por las siguientes expresiones:

$$E(\Theta_i | D_i) = \frac{a_{i,l-i}^{post}}{b_{i,l-i}^{post}} = \frac{a_i + \frac{C_{i,l-i}}{\phi_i}}{b_i + \frac{\beta_{l-i}}{\phi_i}}$$

$$= \frac{b_i}{b_i + \frac{\beta_{l-i}}{\phi_i}} \frac{a_i}{b_i} + \left(1 - \frac{b_i}{b_i + \frac{\beta_{l-i}}{\phi_i}}\right) \frac{C_{i,l-i}}{\beta_{l-i}}$$

es decir, es una media ponderada entre la esperanza a priori de Θ_i , a_i/b_i , y la estimación $C_{i,l-i}/\beta_{l-i}$.

Otra consecuencia inmediata es que puede obtenerse la distribución a posteriori de las reservas anuales depuradas de sobredispersión, es decir, de $\frac{C_{i,J} - C_{i,l-i}}{\phi_i}$, condicionada a la información disponible, D_i , es decir:

$$P\left[\frac{C_{i,J} - C_{i,l-i}}{\phi_i} = k | D_i\right] =$$

$$= \int_0^\infty \exp\left[-(1 - \beta_{l-i}) \frac{\theta}{\phi_i}\right] \frac{\left[(1 - \beta_{l-i}) \frac{\theta}{\phi_i}\right]^k}{k!} \frac{(b_{i,l-i}^{post})^{a_{i,l-i}^{post}}}{\Gamma(a_{i,l-i}^{post})} \theta^{a_{i,l-i}^{post} - 1} \exp(-b_{i,l-i}^{post} \theta) d\theta$$

operando se llega a que:

$$P\left[\frac{C_{i,J} - C_{i,l-i}}{\phi_i} = k | D_i\right] =$$

$$= \binom{k + a_{i,l-i}^{post} - 1}{k} \left(\frac{b_{i,l-i}^{post}}{b_{i,l-i}^{post} + \left(\frac{1 - \beta_{l-i}}{\phi_i}\right)}\right)^{a_{i,l-i}^{post}} \left(\frac{\left(\frac{1 - \beta_{l-i}}{\phi_i}\right)}{b_{i,l-i}^{post} + \left(\frac{1 - \beta_{l-i}}{\phi_i}\right)}\right)^k$$

de parámetros $r = a_{i,l-i}^{post}$ y $p = \frac{b_{i,l-i}^{post}}{b_{i,l-i}^{post} + \left(\frac{1 - \beta_{l-i}}{\phi_i}\right)}$.

Teniendo en cuenta la independencia de los $X_{i,j}$, su esperanza condicionada a Θ_i , es:

$$\begin{aligned} E(C_{i,J} | D_l) &= E\{E[C_{i,J} | \Theta_i, D_l] | D_l\} \\ &= C_{i,l-i} + E\left[E\left(\sum_{j=l-i+1}^J X_{i,j} | \Theta_i\right) | D_l\right] \\ &= C_{i,l-i} + \left(1 - \frac{\beta_{l-i}}{\beta_j}\right) E(\Theta_i | D_l) \end{aligned}$$

por lo que el estimador de $C_{i,J}$ en este modelo puede expresarse como:

$$\hat{C}_{i,J}^{PG} = C_{i,l-i} + (1 - \beta_{l-i}) \left[\frac{b_j}{b_j + \left(\frac{\beta_{l-i}}{\phi_i}\right)} \frac{a_j}{b_j} + \left(1 - \frac{b_j}{b_j + \left(\frac{\beta_{l-i}}{\phi_i}\right)}\right) \frac{C_{i,l-i}}{\beta_{l-i}} \right] \quad \forall i = 1, \dots, l$$

Una vez obtenido el nivel estimado de reservas, el paso final es el de la determinación del error asociado a su cálculo. El MSE condicional viene dado por:

$$\begin{aligned} MSE_{C_{i,J}|D_l}(\hat{C}_{i,J}^{PG}) &= E\left[(C_{i,J} - \hat{C}_{i,J}^{PG})^2 | D_l\right] \\ &= E\left\{\left[\sum_{j=l-i+1}^J X_{i,j} - (1 - \beta_{l-i})E(\Theta_i | D_l)\right]^2 | D_l\right\} \\ &= E\left\{\left[\sum_{j=l-i+1}^J X_{i,j} - \gamma_j E(\Theta_i | D_l)\right]^2 | D_l\right\} \end{aligned}$$

Como $\forall j > l - i$ se cumple que:

$$E(X_{i,j} | D_l) = E[E(X_{i,j} | \Theta_i, D_l) | D_l] = E[E(X_{i,j} | \Theta_i) | D_l] = \gamma_j E(\Theta_i | D_l)$$

entonces se tiene que:

$$MSE_{C_{i,J}|D_l}(\hat{C}_{i,J}^{PG}) = var\left(\sum_{j=l-i+1}^J X_{i,j} | D_l\right)$$

que, considerando la independencia entre los $X_{i,j}$, dado Θ_i , se tiene que:

$$\text{var} \left(\sum_{j=l-i+1}^J X_{i,j} \mid D_l \right) = E \left[\text{var} \left(\sum_{j=l-i+1}^J X_{i,j} \mid \Theta_i \right) \mid D_l \right] + \text{var} \left[E \left(\sum_{j=l-i+1}^J X_{i,j} \mid \Theta_i \right) \mid D_l \right]$$

Como $E(X_{i,j} \mid \Theta_i) = \Theta_i \gamma_j$ y $\text{var}(X_{i,j} \mid \Theta_i) = \phi_i \Theta_i \gamma_j$, entonces se tiene que:

$$\begin{aligned} \text{var} \left(\sum_{j=l-i+1}^J X_{i,j} \mid D_l \right) &= E \left[\sum_{j=l-i+1}^J \phi_i \Theta_i \gamma_j \mid D_l \right] + \text{var} \left[\sum_{j=l-i+1}^J \Theta_i \gamma_j \mid D_l \right] = \\ &= \phi_i (1 - \beta_{l-i}) E(\Theta_i \mid D_l) + (1 - \beta_{l-i})^2 \text{var}(\Theta_i \mid D_l) \\ &= \phi_i (1 - \beta_{l-i}) \frac{a_{i,l-i}^{\text{post}}}{b_{i,l-i}^{\text{post}}} + (1 - \beta_{l-i})^2 \frac{a_{i,l-i}^{\text{post}}}{(b_{i,l-i}^{\text{post}})^2} \end{aligned}$$

que es la expresión a aplicar para el cálculo del *MSE* condicional.

Por lo que respecta al cálculo del *MSE* incondicional, se tiene que:

$$\begin{aligned} \text{MSE}_{C_{i,j}}(\hat{C}_{i,j}^{\text{PG}}) &= E \left[\text{MSE}_{C_{i,j} \mid D_l}(\hat{C}_{i,j}^{\text{PG}}) \right] \\ &= \phi_i (1 - \beta_{l-i}) E \left(\frac{a_{i,l-i}^{\text{post}}}{b_{i,l-i}^{\text{post}}} \right) + (1 - \beta_{l-i})^2 E \left(\frac{a_{i,l-i}^{\text{post}}}{(b_{i,l-i}^{\text{post}})^2} \right) \end{aligned}$$

Como $a_{i,l-i}^{\text{post}}$ es función de $C_{i,j}$, pero $b_{i,l-i}^{\text{post}}$ es una suma de constantes y además $E(C_{i,l-i}) = \beta_{l-i} \frac{a_i}{b_i}$, se obtiene finalmente que:

$$\text{MSE}_{C_{i,j}}(\hat{C}_{i,j}^{\text{PG}}) = \phi_i (1 - \beta_{l-i}) \frac{a_i}{b_i} + (1 - \beta_{l-i})^2 \frac{1 + \phi_i b_i}{\phi_i b_i + \beta_{l-i}}$$

que es la expresión a utilizar para el cálculo del *MSE* incondicional

Ejemplo:

Con los datos habituales, teniendo en cuenta que:

$$\begin{aligned} E(\Theta_i) &= \frac{a_i}{b_i} \\ \text{cv}(\Theta_i) &= \sqrt{a_i} \end{aligned}$$

y suponiendo que $a_i = 325 \forall i$, el primer paso consiste en calcular los parámetros que harán falta en fases subsiguientes, en concreto b_i y ϕ_i . Para encontrar b_i basta con despejarlo de $E(\Theta_i)$, siendo esta última igual a $E(C_{i,j})$ que es un dato conocido. Por su parte, para hallar ϕ_i se parte de:

$$\text{var}(C_{i,j}) = \frac{a_i}{b_i} \left(\phi_i + \frac{1}{b_i} \right)$$

El cálculo es inmediato, teniendo en cuenta que $\text{var}(C_{i,j}) = [\text{cv}(C_{i,j})E(C_{i,j})]^2$ y asumiendo que $\text{cv}(C_{i,j})$ se calcula igual que en el ejemplo anterior. Con todo ello resulta la siguiente tabla de valores:

i	$E(\Theta_i)$	$\text{cv}(\Theta_i)$ (%)	$\text{cv}(C_{i,j})$	a_i	b_i (%)	ϕ_i
0	3.319	5,55%	7,81%	325	9,79%	10,03
1	4.794	5,55%	7,81%	325	6,78%	14,49
2	4.604	5,55%	7,81%	325	7,06%	13,92
3	4.309	5,55%	7,81%	325	7,54%	13,03
4	3.958	5,55%	7,81%	325	8,21%	11,97
5	4.549	5,55%	7,81%	325	7,14%	13,75
6	5.382	5,55%	7,81%	325	6,04%	16,27
7	5.313	5,55%	7,81%	325	6,12%	16,06
8	4.111	5,55%	7,81%	325	7,90%	12,43
9	3.111	5,55%	7,81%	325	10,45%	9,40

La siguiente fase del cálculo consiste en obtener $a_{i,l-i}^{post}$ y $b_{i,l-i}^{post}$, para lo cual se emplean las siguientes expresiones:

$$a_{i,j}^{post} = a_i + \frac{C_{i,j}}{\phi_i}$$

$$b_{i,j}^{post} = b_i + \sum_{k=0}^j \frac{\gamma_k}{\phi_i} = b_i + \frac{\beta_j}{\phi_i} \quad / \quad \beta_j = \sum_{k=0}^j \gamma_k$$

ambas dependientes de β_{l-i} , parámetro que se obtiene a partir de los factores de desarrollo del CL. Los valores obtenidos son:

i	$C_{i,l-i}$	β_{l-i}	$a_{i,l-i}^{post}$	$b_{i,l-i}^{post}$
0	3.121	1,000	636,03	0,1976
1	4.187	0,963	613,91	0,1343
2	3.776	0,878	596,32	0,1337
3	3.462	0,817	590,79	0,1381
4	2.869	0,738	564,75	0,1438
5	2.684	0,656	520,16	0,1192
6	2.487	0,549	477,85	0,0941
7	2.010	0,370	450,15	0,0842
8	940	0,208	400,63	0,0958
9	233	0,059	349,77	0,1107

Finalmente, el triángulo inferior se puede rellenar utilizando la expresión:

$$\hat{C}_{i,j} = C_{i,l-i} + \left(1 - \frac{\beta_{l-i}}{\beta_j}\right) \left\{ \frac{b_i}{b_i + \left(\frac{\beta_{l-i}}{\beta_j} / \phi_i\right)} \frac{a_i}{b_i} + \left[1 - \frac{b_i}{b_i + \left(\frac{\beta_{l-i}}{\beta_j} / \phi_i\right)}\right] \frac{C_{i,l-i}}{\frac{\beta_{l-i}}{\beta_j}} \right\}$$

obteniéndose los siguientes valores:

	0	1	2	3	4	5	6	7	8	9
0										
1										4.355
2									4.158	4.322
3								3.721	4.087	4.245
4							3.178	3.417	3.753	3.898
5						3.038	3.381	3.646	4.022	4.184
6				3.022	3.434	3.835	4.145	4.585	4.775	
7			2.969	3.543	3.979	4.400	4.723	5.179	5.376	
8		1.630	2.381	2.827	3.166	3.493	3.745	4.099	4.251	
9	728	1.240	1.801	2.136	2.391	2.637	2.827	3.093	3.208	

con lo que las reservas anuales son:

año	Reserva estimada
1	168
2	546
3	783
4	1.029
5	1.500
6	2.288
7	3.366
8	3.311
9	2.975

y el total su suma, es decir, 15.966. El desglose por ejercicios futuros, distinguiendo entre valor nominal y actual es el recogido en la siguiente tabla:

año	dotación anual	
	valor nominal	valor actual
1	4.151	4.100
2	3.361	3.270
3	2.604	2.490
4	1.926	1.762
5	1.507	1.333
6	1.144	949
7	740	575
8	419	309
9	115	81
a dotar	15.966	14.869
Saldo total	41.735	40.638

Hay que decir que si se desease calcular únicamente el valor de las reservas en J sin considerar importes por años de desarrollo, se podría haber hecho sin más que calcular:

$$\begin{aligned}\hat{R}_i &= \hat{C}_{i,J} - C_{i,l-i} = C_{i,l-i} + (1 - \beta_{l-i})E(\Theta_i | D_l) - C_{i,l-i} \\ &= (1 - \beta_{l-i}) \frac{a_{i,l-i}^{post}}{b_{i,l-i}^{post}}\end{aligned}$$

Por último, aplicando las expresiones correspondientes se obtiene los valores del *MSE* condicional e incondicional, tanto por años como para el total, obteniéndose los siguientes resultados:

<i>i</i>	$(MSE \text{ condicional})^{1/2}$	$(MSE \text{ incondicional})^{1/2}$
1	50	51
2	90	91
3	106	106
4	119	120
5	158	161
6	220	226
7	281	281
8	262	260
9	231	229
TOTAL	557	559

BIBLIOGRAFÍA

- Aitchinson, J. and Brown, J.A.C. (1969): *The Log-Normal Distribution*. Cambridge University Press.
- Akaike, H. (1973): *Information theory and an extension of the maximum likelihood principle*. *Second International Symposium on Information Theory*. Ed. B.N. Petrov and F. Csaki, 267-281, Budapest: Akademiai Kiado.
- Alonso, P. y Albarrán, I. (2007): *Técnicas estadísticas avanzadas aplicadas al análisis del riesgo en seguros dentro del marco de Solvencia II: Monte Carlo y bootstrapping*, Fundación Mapfre, Madrid.
- APRA (2002): *Future Policy directions for the regulation and prudential supervision of the General Insurance Industry*, September 2002.
- Arjas, E. (1989): “The claims reserving problem in non-life insurance: Some structural ideas”, *Astin Bulletin*, 19, 2, pp: 139-152.
- Artzner, P., Delbaen, J., Eber, M. and Heath, D. (1999): “Coherent measures of Risk”, *Mathematical Finance*, 9, pp: 203-228.
- Babu, G.J. and Singh, K. (1983): “Inference on means using the bootstrap”, *Ann. Statist.*, 11, pp: 999-1003.
- Beard, R.E., Pentikäinen, T. and Pesonen, E. (1984): *Risk Theory*, Chapman and Hall, London.
- Ben-Akiva, M. and Lerman, S.R. (1985): *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.
- Benktander, G. (1976): “An Approach to Credibility in Calculating IBNR for Casualty Excess Insurance”, *The Actuarial Review*, April 1976, pp: 7.
- Beran, R.J. (1995): “Stein confidence sets and the bootstrap”, *Statist. Sinica*, 5, pp: 109–127.
- Beran, R.J. (2003): “The impact of the bootstrap on statistical algorithms and theory”, *Statist. Sci.*, 18, pp: 175–184.
- Birkes, D. and Dodge, Y. (1993): *Alternative Methods of Regression*, John Wiley & Sons.
- Booth, J.G. and Hobert, J.P. (1998): “Standard errors of prediction in generalized linear mixed models”, *J. Amer. Statist. Assoc.*, 93, pp: 262–272.

- Bornhuetter, R.L. and Ferguson R.E. (1972): “The actuary and IBNR”, *Proceedings of the Casualty Actuarial Society*, LIX, pp: 181-195.
- Boss, D. and Monahan, J. (1986): “Bootstrap methods using prior information”, *Biometrika*, 73, pp: 77-83.
- Bretagnolle, J. (1983): “Les limites du bootstrap de certaines fonctionelles”, *Ann. Inst. H. Poincaré Probab. Statist.*, 19, pp: 281–296.
- Buckland, S.T. (1983): “Monte Carlo methods for confidence interval estimation using the bootstrap technique”, *Bull. Appl. Statist.*, 10, pp: 194-212.
- Buckland, S.T. (1984): “Monte Carlo confidence intervals”, *Biometrics*, 40, pp: 811-817.
- Bühlmann, H. (1996): *Mathematical methods in risk theory*. Springer Verlag Berlin Heidelberg, New York.
- Bühlmann, H., Schnieper, R. and Straub, E. (1980): “Claims reserves in casualty insurance based on probabilistic model”, *Mitteilungen*, 1, pp: 21-45.
- Bühlmann, P. (2002): “Bootstraps for time series”, *Statist. Sci.*, 17, pp: 52–72.
- Cameron, A.C. and Windmeijer, F.A.G. (1997): “An R-squared measure of goodness of fit for some common nonlinear regression models”, *Journal of Econometrics*, 77, pp: 329-342.
- Canty, A.J., Davison, A.C., Hinkley, D.V. and Ventura, V. (2002): *Bootstrap diagnostics*, Preprint, Institute of Mathematics, Swiss Federal Institute of Technology, Lausanne.
- Carpenter, J. (1999): “Test inversion bootstrap confidence intervals”, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 61, pp: 159–172.
- CEIOPS (2005): *Answers to the European Commission on the second wave of Calls for Advice in the framework of the Solvency II project*, CEIOPS-DOC 07/05.
- CEIOPS (2005): *Draft Answers to the European Commission on the third wave of Calls for Advice in the framework of the Solvency II project*, Consultation Paper nº 9, CEIOPS-CP-06/05.
- CEIOPS (2005): *EU Solvency II project -the first Quantitative Impact Study*, CEIOPS-FS-12/05.

- CEIOPS (2005): *Report on Possible Need for Amendments to the Insurance Groups Directive*, Consultation paper nº 6, CEIOPS-CP-03/05.
- CEIOPS (2006): *Draft Advice to the European Commission in the framework of the Solvency II project on Pillar I issues – further advice*, Consultation Paper nº 20, CEIOPS-CP-09/06.
- CEIOPS (2006): *Quantitative Impact Study 2. Technical Specifications*, CEIOPS-PI-08/06.
- CEIOPS (2006): *Summary of comments on CEIOPS-CP-06/05. Draft Answers to the European Commission on the third wave of Calls for Advice in the framework of the Solvency II project*, Call for Advice nº 19, CEIOPS-SEC-20/06.
- CEIOPS (2007): *QIS3: Calibration of the underwriting risk, market risk and MCR*, CEIOPS-FS-14/07.
- CEIOPS (2007): *QIS3: Technical Specifications. Part I: Instructions*, CEIOPS-FS-11/07.
- CEIOPS (2007): *QIS3: Technical Specifications. Part II: Background Informations*, CEIOPS-FS-12/07.
- CEIOPS (2008): *Consultation Paper 23*.
- CEIOPS (2008): *QIS 4 Technical Specifications* (MARKT/2505/08).
- Centeno, L. (1989): “The Buhlmann-Straub Model with the premium calculated according to the variance principle”, *Insurance: Mathematics and Economics*, 8, pp: 3-10.
- Chernick, M.R. (1999): *Bootstrap Methods: A Practitioner’s Guide*, John Wiley & Sons.
- Christofides, S. (1990): *Regression models based on log-incremental payments*, *Claims Reserving Manual*, Vol. 2, Institute of Actuaries, London.
- Comisión Europea, D.G. Mercado Interior (2003): *Concepción de un futuro sistema de control cautelar en la UE. Recomendaciones de los servicios de la Comisión*, MARKT/2509/03. Bruselas, 3 de marzo de 2003.
- Comité Européen des Assurances (2006): *CEA Document on Cost of Capital*, Brussels, 21 April.
- Comité Européen des Assurances and Chief Risk Officers Forum (2005): *Solutions to major issues for Solvency II*.

- Cragg, J.G. and Uhler, R. (1970): “The demand for automobiles”, *Canadian Journal of Economics*, 3, pp: 386-406.
- Cummins, J.D. and Derrig, A. (1991): *Managing the insolvency risk of insurance companies*, Proceedings of the Second International Conference on Insurance Solvency, Kluwer Academic Publishers, Boston.
- Davison, A.C. and Hinkley, D.V. (1997): *Bootstrap Methods and Their Application*, Cambridge Univ. Press.
- Dean, C. and Lawless, J.F. (1989): “Tests for detecting overdispersion in Poisson regression models”, *Journal of the American Statistical Association*, 84, pp: 467-472.
- Doray, L. (1994): “IBNR reserve under a loglinear location-scale regression model”, *Casualty Actuarial Society Forum*, Spring, pp: 607-651.
- Efron B. and Tibshiran, R.J. (1993): *An introduction to the Bootstrap*, Chapman and Hall.
- Efron, B. (1978): “Regression and ANOVA with zero-one data: Measures of residual variation”, *Journal of the American Statistical Association*, 73, pp: 113-121.
- Efron, B. (1979): “Bootstrap methods: Another look at the Jackknife”. *Ann. Statist.*, 7, pp: 1-26.
- Efron, B. and LePage, R. (1992): *Introduction to bootstrap*, In: Exploring the Limits of Bootstrap (R. LePage and L. Billard, eds.): 3-10. Wiley.
- Efron, B. and Tibshirani, R. (1993): *An Introduction to the Bootstrap*, Chapman and Hall.
- England, P. and Verrall, R. (1999): “Analytic and bootstrap estimates of prediction errors in claim reserving”, *Insurance: Mathematics and Economics*, 25, pp: 281-293.
- England, P.D. (2002): *Addendum to ‘Analytic and bootstrap estimates of prediction errors in claims reserving’*. Actuarial Research Paper # 138, Department of Actuarial Science and Statistics, City University, London EC1V 0HB.
- England, P.D. and Verrall, R.J. (2002): “Stochastic claim reserving in general insurance”, *British Actuarial Journal*, 8/3, pp: 443-518.
- European Commission (2005): *Amended Framework for Consultation on Solvency II*, MARKT 2506-04.

- European Commission (2006): *Amended Framework for Consultation on Solvency II*, MARKT 2515-06.
- FSA (2003): *Enhanced Capital Requirements and individual capital assessments for non-life insurers*, Consultative Paper 190.
- FSA (2003): *Enhanced Capital Requirements and individual capital assessments for life insurers*, Consultative Paper 195.
- FSA (2006): *Integrated Prudential Sourcebook*.
- FSA (2006): *Solvency II: a new framework for prudential regulation of insurance in the EU. A discussion paper*.
- Ganio, L.M. and Schafer, D.W. (1992): “Diagnostics for overdispersion”, *Journal of the American Statistical Association*, 87, pp: 795-804.
- Gogol, D. (1993): “Using Expected Loss Ratios in Reserving”, *Insurance: Mathematics and Economics*, 12, pp: 297-299.
- Hachemeister, C.A. (1980): “A stochastic model for loss reserving”, *Transactions of the XXI International Congress of Actuaries*, I, pp: 185-194.
- Hachemeister, C.A. and Stanard, J.N. (1975): *IBNR claims count estimation with static lag functions*. Astin Colloquium 1975, Portimao, Portugal.
- Hall, P. (1992): *The Bootstrap and the Edgeworth Expansion*, Springer-Verlag, Appendix I.
- Hertig, J. (1985): “A statistical approach to the IBNR-reserves in marine insurance”, *Astin Bulletin*, 15/2, pp: 171-183.
- Hovinen, E. (1981): *Additive and Continuous IBNR*. Proceedings of the XV ASTIN Coll., Loen, Norway 1981.
- IAA (2004): *A Global Framework for Insurer Solvency Assessment*, International Actuarial Association.
- Jorion, P. (2000): *Valeu at Risk: The New Benchmark for Managing Financial Risk*, MCGraw-Hill.
- KPMG / European Commission (2002): *Study into the methodologies to assess the overall financial position of an insurance undertaking from the perspective of prudential supervision*, May 2002.
- Kremer, E. (1982): “IBNR claims and the two way model of ANOVA”, *Scandinavian Actuarial Journal*, 1, pp: 47-55.

- Kremer, E. (1985): *Einführung in die Versicherungsmathematik*. Vandenhoeck & Ruprecht, Göttingen.
- Kullback, S. (1987): “Letter to the Editor: The Kullback-Leibler distance”, *The American Statistician*, 41 (4), pp: 340–341.
- Kullback, S. and Leibler, R.A. (1951): “On Information and Sufficiency”, *Annals of Mathematical Statistics*, 22 (1), pp: 79–86.
- Lambert, D. and Roeder, K. (1995): “Overdispersion diagnostics for generalized linear models”, *Journal of the American Statistical Association*, 90, pp: 1225-1236.
- Mack, T. (1990): “Improved estimation of IBNR claims by credibility theory”, *Insurance: Mathematics and Economics*, 9 (1), pp: 51-57.
- Mack, T. (1991a): “A simple parametric model for rating automobile insurance or estimating IBNR claims reserves”, *Astin Bulletin*, 21 (1), pp: 93-109.
- Mack, T. (1991b): “Which stochastic model is underlying the chain ladder method?”, *Insurance: Mathematics and Economics*, 15, pp: 133-138.
- Mack, T. (1993): “Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates”, *ASTIN Bulletin*, 23, pp: 213-225.
- Mack, T. (2000): “Credible claims reserves: the Benktander method”. *Astin Bulletin* 30/2, 333-347
- Mack, T. and Venter, G. (2000): “A comparison of stochastic models that reproduce chain ladder reserve estimates”, *Insurance: Mathematics and Economics*, 26(1), pp: 101-107.
- Maddala, G.S. (1992): *Introduction to Econometrics*. 2nd ed. New York: MacMillan.
- McCullagh, P. and Nelder, J.A. (1999): *Generalized Linear Models*. New York: Chapman and Hall/CRC Press.
- McFadden, D. (1974): *Conditional logit analysis of qualitative choice behavior*. *Frontiers of Econometrics*, ed. P. Zarembka, 105-142. New York: Academic Press.
- McKelvey, R.D. and Zavoina, W. (1975): “A statistical model for the analysis of ordinal level dependent variables”, *Journal of Mathematical Sociology*, 4, pp: 103-120.

- Mercer Oliver Wyman (2004): *Life at the end of the tunnel*, June 2004.
- Panjer, H.H. and Willmot, G.E. (1992): *Insurance risk models*, Society of Actuaries, Schaumburg.
- Quenouille, M. (1949): “Aproximate tests of correlation in time series”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 11, pp: 18-44.
- Raftery, A. (1996): *Bayesian model selection in social research*. Sociological Methodology, ed. P.V. Marsden, vol. 25, 111-163, Oxford: Basil Blackwell.
- Renshaw, A.E. (1989): “Chain Ladder and Interactive Modelling (Claims Reserving and GLIM)”, *Journal of the Institute of Actuaries*, 116 (III), pp: 559-587.
- Renshaw, A.E. (1994): “Modelling the claims process in the presence of covariates”, *ASTIN Bulletin*, 24, pp: 265-286.
- Renshaw, A.E. and Verrall, R.J. (1998): “A stochastic model underlying the chain-ladder technique”, *British Actuarial Journal*, 4, pp: 903-923.
- Renshaw, A.E. and Verrall, R. J. (1994): *A stochastic model underlying the chain ladder technique*, Proceedings XXV ASTIN Colloquium, Cannes.
- Sharma, P. (2002): *Prudential Supervision of Insurance Undertakings*, Conference of Insurance Supervisory of the Member States of the European Union, December 2002.
- Straub, E. (1988): *Non-life insurance mathematics*, Springer-Verlag Berlin Heidelberg and Association of Swiss Actuaries, Zurich.
- Swiss Federal Office of Private Insurance (2004): *White Paper of the Swiss Solvency Test*.
- Swiss Re (2006): *Solvency II: an integrated risk approach for European insurers*, Sigma nº 4.
- Taylor, G.C. (1986): *Claim reserving in non-life insurance*, North-Holland, Amsterdam.
- Veall, M. and Zimmermann, K. (1992): “Pseudo-R² in the ordinal probit model”, *Journal of Mathematical Sociology*, 16, pp: 333-342.
- Verrall, R. (1990): “Bayes and the empirical Bayes estimation for the chain ladder model” *Astin Bulletin*, 20/2,217-243

- Verrall, R. (1991): “On the Estimation of Reserves from Loglinear Models”, *Insurance: Mathematics and Economics*, 10, pp: 75-80.
- Verrall, R. (1996): “Claims reserving and generalised additive models”, *Insurance: Mathematics and Economics*, 19 (1), pp: 31-43.
- Verrall, R. (2000): “An investigation into stochastic claims reserving models and the chain-ladder technique”, *Insurance: Mathematics and Economics*, 26, pp: 91-99.
- Verrall, R.J. (1991): “On the estimation of reserves from loglinear models”, *Insurance: Mathematics & Economics*, 10, pp: 75-80.
- Verrall, R.J. (2004): “A Bayesian Generalized Linear Model for the Bornhuetter-Ferguson Method of Claims Reserving”, *North American Actuarial Journal*, 8(3), pp: 67–89.
- Zehnwirth, B. (1989): *The Chain Ladder technique - A stochastic model, Claims Reserving Manual, Vol. 2*, Institute of Actuaries, London.

NORMATIVA CONSULTADA

- Primera Directiva del Consejo 73/239/CEE, de 24 de julio de 1973, sobre coordinación de las disposiciones legales, reglamentarias y administrativas relativas al acceso a la actividad del seguro directo distinto del seguro de vida.
- Primera Directiva del Consejo 79/267/CEE, de 5 de marzo de 1979, sobre coordinación de las disposiciones legales, reglamentarias y administrativas, referentes al acceso a la actividad del seguro directo sobre la vida, y a su ejercicio.
- Directiva del Consejo 91/674/CEE, de 19 de diciembre, relativa a las cuentas anuales y a las cuentas consolidadas de las empresas de seguros.
- Directiva 98/78/CE, del Parlamento Europeo y del Consejo, de 27 de octubre de 1998, relativa a la supervisión adicional de las empresas de seguros y de reaseguros que formen parte de grupos de seguros o de reaseguros.
- Directiva 2000/12/CE del Parlamento europeo y del Consejo, de 20 de marzo, relativa al acceso a la actividad de las entidades de crédito y a su ejercicio.
- Directiva 2001/24/CE del Parlamento Europeo y del Consejo, de 4 de abril de 2001, relativa al saneamiento y a la liquidación de las entidades de crédito.
- Directiva 2002/13/CE del Parlamento Europeo y del Consejo de 5 de marzo de 2002 por la que se modifica la Directiva 73/239/CEE del Consejo en lo que respecta a los requisitos del margen de solvencia de las empresas de seguros distintos del seguro de vida.
- Directiva 2002/83/CE del Parlamento Europeo y del Consejo, de 5 de noviembre de 2002, sobre el seguro de vida.
- Directiva 2002/87/CE, del Parlamento Europeo y del Consejo, de 16 de diciembre de 2002, relativa a la supervisión adicional de las entidades de crédito, empresas de seguros y empresas de inversión de un conglomerado financiero, y por la que se modifican las Directivas 73/239/CEE, 79/267/CEE, 92/49/CEE, 92/96/CEE, 93/6/CEE y 93/22/CEE del Consejo y las Directivas 98/78/CE y 2000/12/CE del Parlamento Europeo y del Consejo.
- Directiva 2005/1/CE, del Parlamento Europeo y del Consejo, de 9 de marzo de 2005, por la que se modifican las Directivas 73/239/CE, 85/611/CEE, 91/675/CEE, 92/49/CEE y 93/6/CEE del Consejo y las Directivas 94/19/CE,

98/78/CE, 2000/12/CE, 2001/34/CE, 2002/83/CE y 2002/87/CE, a fin de establecer una nueva estructura organizativa de los comités de servicios financieros.

- Directiva 2005/68/CE del Parlamento Europeo y del Consejo, de 16 de noviembre de 2005 sobre el reaseguro.
- Directiva 2006/48/CE del Parlamento Europeo y del Consejo de 14 de junio de 2006, relativa al acceso a la actividad de las entidades de crédito y a su ejercicio (refundición).
- Directiva 2006/49/CE del Parlamento Europeo y del Consejo de 14 de junio de 2006, sobre la adecuación del capital de las empresas de inversión y las entidades de crédito (refundición).
- Directiva 2009/138/CE del Parlamento Europeo y del Consejo, de 25 de noviembre de 2009 sobre el seguro de vida, el acceso a la actividad del seguro y de reaseguro y su ejercicio (Directiva de Solvencia II).
- R.D. 239/2007, de 16 de febrero, por el que se modifica el Reglamento de ordenación y supervisión de los seguros privados, aprobado por el Real Decreto 2486/1998, de 20 de noviembre, y el Reglamento de mutualidades de previsión social, aprobado por el Real Decreto 1430/2002, de 27 de diciembre.

COLECCIÓN “CUADERNOS DE LA FUNDACIÓN”
Instituto de Ciencias del Seguro

Para cualquier información o para adquirir nuestras publicaciones puede encontrarnos en:

FUNDACIÓN MAPFRE
Publicaciones
Paseo de Recoletos 23 – 28004 Madrid – (España)
Telf.: + 34 915 818 768 Fax: +34 915 818 409
www.fundacionmapfre.com/cienciasdelseguro

- 158. Métodos estocásticos de estimación de las provisiones técnicas en el marco de Solvencia II. 2010
- 157. Introducción al Reaseguro. 2010
- 156. Encuentro Internacional sobre la Historia del Seguro. 2010
- 155. Los sistemas de salud en Latinoamérica y el papel del seguro privado. 2010
- 154. El Seguro de Crédito en Chile. 2010
- 153. El análisis financiero dinámico como herramienta para el desarrollo de modelos internos en el marco de Solvencia II. 2010
- 152. Características sociodemográficas de las personas con doble cobertura sanitaria. Un estudio empírico. 2010
- 151. Solidaridad impropia y seguro de Responsabilidad Civil. 2010
- 150. La prevención del blanqueo de capitales en las entidades aseguradoras, las gestoras y los corredores de seguros 2010
- 149. Fondos de aseguramiento agropecuario y rural: la experiencia mexicana en el mutualismo agropecuario y sus organizaciones superiores. 2010
- 148. Avaliação das Provisões de Sinistro sob o Enfoque das Novas Regras de Solvência do Brasil. 2010

147. El principio de igualdad sexual en el Seguro de Salud: análisis actuarial de su impacto y alcance. 2010
146. Investigaciones históricas sobre el Seguro español. 2010
145. Perspectivas y análisis económico de la futura reforma del sistema español de valoración del daño corporal. 2009
144. Contabilidad y Análisis de Cuentas Anuales de Entidades Aseguradoras (Plan contable 24 de julio de 2008). 2009
143. Mudanças Climáticas e Análise de Risco da Indústria de Petróleo no Litoral Brasileiro. 2009
142. Bases técnicas dinámicas del Seguro de Dependencia en España. Una aproximación en campo discreto. 2009
141. Transferencia Alternativa de Riesgos en el Seguro de Vida: Titulización de Riesgos Aseguradores. 2009
140. Riesgo de negocio ante asegurados con múltiples contratos. 2009
139. Optimización económica del Reaseguro cedido: modelos de decisión. 2009
138. Inversiones en el Seguro de Vida en la actualidad y perspectivas de futuro. 2009
137. El Seguro de Vida en España. Factores que influyen en su progreso. 2009
136. Investigaciones en Seguros y Gestión de Riesgos. RIESGO 2009.
135. Análisis e interpretación de la gestión del fondo de maniobra en entidades aseguradoras de incendio y lucro cesante en grandes riesgos industriales. 2009
134. Gestión integral de Riesgos Corporativos como fuente de ventaja competitiva: cultura positiva del riesgo y reorganización estructural. 2009
133. La designación de la pareja de hecho como beneficiaria en los seguros de vida. 2009
132. Aproximación a la Responsabilidad Social de la empresa: reflexiones y propuesta de un modelo. 2009
131. La cobertura pública en el seguro de crédito a la exportación en España: cuestiones prácticas-jurídicas. 2009

130. La mediación en seguros privados: análisis de un complejo proceso de cambio legislativo. 2009
129. Temas relevantes del Derecho de Seguros contemporáneo. 2009
128. Cuestiones sobre la cláusula cut through. Transferencia y reconstrucción. 2008
127. La responsabilidad derivada de la utilización de organismos genéticamente modificados y la redistribución del riesgo a través del seguro. 2008
126. Ponencias de las Jornadas Internacionales sobre Catástrofes Naturales. 2008
125. La seguridad jurídica de las tecnologías de la información en el sector asegurador. 2008
124. Predicción de tablas de mortalidad dinámicas mediante un procedimiento bootstrap. 2008
123. Las compañías aseguradoras en los procesos penal y contencioso-administrativo. 2008
122. Factores de riesgo y cálculo de primas mediante técnicas de aprendizaje. 2008
121. La solicitud de seguro en la Ley 50/1980, de 8 de octubre, de Contrato de Seguro. 2008
120. Propuestas para un sistema de cobertura de enfermedades catastróficas en Argentina. 2008
119. Análisis del riesgo en seguros en el marco de Solvencia II: Técnicas estadísticas avanzadas Monte Carlo y Bootstrapping. 2008
118. Los planes de pensiones y los planes de previsión asegurados: su inclusión en el caudal hereditario. 2007
117. Evolução de resultados técnicos e financeiros no mercado segurador iberoamericano. 2007
116. Análisis de la Ley 26/2006 de Mediación de Seguros y Reaseguros Privados. 2007
115. Sistemas de cofinanciación de la dependencia: seguro privado frente a hipoteca inversa. 2007

114. El sector asegurador ante el cambio climático: riesgos y oportunidades. 2007
113. Responsabilidade social empresarial no mercado de seguros brasileiro influências culturais e implicações relacionais. 2007
112. Contabilidad y análisis de cuentas anuales de entidades aseguradoras. 2007
111. Fundamentos actuariales de primas y reservas de fianzas. 2007
110. El Fair Value de las provisiones técnicas de los seguros de Vida. 2007
109. El Seguro como instrumento de gestión de los M.E.R. (Materiales Especificados de Riesgo). 2006
108. Mercados de absorción de riesgos. 2006
107. La exteriorización de los compromisos por pensiones en la negociación colectiva. 2006
106. La utilización de datos médicos y genéticos en el ámbito de las compañías aseguradoras. 2006
105. Los seguros contra incendios forestales y su aplicación en Galicia. 2006
104. Fiscalidad del seguro en América Latina. 2006
103. Las NIC y su relación con el Plan Contable de Entidades Aseguradoras. 2006
102. Naturaleza jurídica del Seguro de Asistencia en Viaje. 2006
101. El Seguro de Automóviles en Iberoamérica. 2006
100. El nuevo perfil productivo y los seguros agropecuarios en Argentina. 2006
99. Modelos alternativos de transferencia y financiación de riesgos "ART": situación actual y perspectivas futuras. 2005
98. Disciplina de mercado en la industria de seguros en América Latina. 2005
97. Aplicación de métodos de inteligencia artificial para el análisis de la solvencia en entidades aseguradoras. 2005

96. El Sistema ABC-ABM: su aplicación en las entidades aseguradoras. 2005
95. Papel del docente universitario: ¿enseñar o ayudar a aprender?. 2005
94. La renovación del Pacto de Toledo y la reforma del sistema de pensiones: ¿es suficiente el pacto político?. 2005
92. Medición de la esperanza de vida residual según niveles de dependencia en España y costes de cuidados de larga duración. 2005
91. Problemática de la reforma de la Ley de Contrato de Seguro. 2005
90. Centros de atención telefónica del sector asegurador. 2005
89. Mercados aseguradores en el área mediterránea y cooperación para su desarrollo. 2005
88. Análisis multivariante aplicado a la selección de factores de riesgo en la tarificación. 2004
87. Dependencia en el modelo individual, aplicación al riesgo de crédito. 2004
86. El margen de solvencia de las entidades aseguradoras en Iberoamérica. 2004
85. La matriz valor-fidelidad en el análisis de los asegurados en el ramo del automóvil. 2004
84. Estudio de la estructura de una cartera de pólizas y de la eficacia de un Bonus-Malus. 2004
83. La teoría del valor extremo: fundamentos y aplicación al seguro, ramo de responsabilidad civil autos. 2004
81. El Seguro de Dependencia: una visión general. 2004
80. Los planes y fondos de pensiones en el contexto europeo: la necesidad de una armonización. 2004
79. La actividad de las compañías aseguradoras de vida en el marco de la gestión integral de activos y pasivos. 2003
78. Nuevas perspectivas de la educación universitaria a distancia. 2003
77. El coste de los riesgos en la empresa española: 2001.

76. La incorporación de los sistemas privados de pensiones en las pequeñas y medianas empresas. 2003
75. Incidencia de la nueva Ley de Enjuiciamiento Civil en los procesos de responsabilidad civil derivada del uso de vehículos a motor. 2002
74. Estructuras de propiedad, organización y canales de distribución de las empresas aseguradoras en el mercado español. 2002
73. Financiación del capital-riesgo mediante el seguro. 2002
72. Análisis del proceso de exteriorización de los compromisos por pensiones. 2002
71. Gestión de activos y pasivos en la cartera de un fondo de pensiones. 2002
70. El cuadro de mando integral para las entidades aseguradoras. 2002
69. Provisiones para prestaciones a la luz del Reglamento de Ordenación y Supervisión de los Seguros Privados; métodos estadísticos de cálculo. 2002
68. Los seguros de crédito y de caución en Iberoamérica. 2001
67. Gestión directiva en la internacionalización de la empresa. 2001
65. Ética empresarial y globalización. 2001
64. Fundamentos técnicos de la regulación del margen de solvencia. 2001
63. Análisis de la repercusión fiscal del seguro de vida y los planes de pensiones. Instrumentos de previsión social individual y empresarial. 2001
62. Seguridad Social: temas generales y régimen de clases pasivas del Estado. 2001
61. Sistemas Bonus-Malus generalizados con inclusión de los costes de los siniestros. 2001
60. Análisis técnico y económico del conjunto de las empresas aseguradoras de la Unión Europea. 2001
59. Estudio sobre el euro y el seguro. 2000
58. Problemática contable de las operaciones de reaseguro. 2000

56. Análisis económico y estadístico de los factores determinantes de la demanda de los seguros privados en España. 2000
54. El corredor de reaseguros y su legislación específica en América y Europa. 2000
53. Habilidades directivas: estudio de sesgo de género en instrumentos de evaluación. 2000
52. La estructura financiera de las entidades de seguros, S.A. 2000
51. Seguridades y riesgos del joven en los grupos de edad. 2000
50. Mixturas de distribuciones: aplicación a las variables más relevantes que modelan la siniestralidad en la empresa aseguradora. 1999
49. Solvencia y estabilidad financiera en la empresa de seguros: metodología y evaluación empírica mediante análisis multivariante. 1999
48. Matemática Actuarial no vida con MapleV. 1999
47. El fraude en el Seguro de Automóvil: cómo detectarlo. 1999
46. Evolución y predicción de las tablas de mortalidad dinámicas para la población española. 1999
45. Los Impuestos en una economía global. 1999
42. La Responsabilidad Civil por contaminación del entorno y su aseguramiento. 1998
41. De Maastricht a Amsterdam: un paso más en la integración europea. 1998

Nº Especial Informe sobre el Mercado Español de Seguros 1997
Fundación MAPFRE Estudios

39. Perspectiva histórica de los documentos estadístico-contables del órgano de control: aspectos jurídicos, formalización y explotación. 1997
38. Legislación y estadísticas del mercado de seguros en la comunidad iberoamericana. 1997
37. La responsabilidad civil por accidente de circulación. Puntual comparación de los derechos francés y español. 1997

36. Cláusulas limitativas de los derechos de los asegurados y cláusulas delimitadoras del riesgo cubierto: las cláusulas de limitación temporal de la cobertura en el Seguro de Responsabilidad Civil. 1997
35. El control de riesgos en fraudes informáticos. 1997
34. El coste de los riesgos en la empresa española: 1995
33. La función del derecho en la economía. 1997

Nº Especial Informe sobre el Mercado Español de Seguros 1996
Fundación MAPFRE Estudios

32. Decisiones racionales en reaseguro. 1996
31. Tipos estratégicos, orientación al mercado y resultados económicos: análisis empírico del sector asegurador español. 1996
30. El tiempo del directivo. 1996
29. Ruina y Seguro de Responsabilidad Civil Decenal. 1996

Nº Especial Informe sobre el Mercado Español de Seguros 1995
Fundación MAPFRE Estudios

28. La naturaleza jurídica del Seguro de Responsabilidad Civil. 1995
27. La calidad total como factor para elevar la cuota de mercado en empresas de seguros. 1995
26. El coste de los riesgos en la empresa española: 1993
25. El reaseguro financiero. 1995
24. El seguro: expresión de solidaridad desde la perspectiva del derecho. 1995
23. Análisis de la demanda del seguro sanitario privado. 1993

Nº Especial Informe sobre el Mercado Español de Seguros 1994
Fundación MAPFRE Estudios

22. Rentabilidad y productividad de entidades aseguradoras. 1994
21. La nueva regulación de las provisiones técnicas en la Directiva de Cuentas de la C.E.E. 1994

20. El Reaseguro en los procesos de integración económica. 1994
 19. Una teoría de la educación. 1994
 18. El Seguro de Crédito a la exportación en los países de la OCDE (evaluación de los resultados de los aseguradores públicos). 1994
- Nº Especial Informe sobre el mercado español de seguros 1993
FUNDACION MAPFRE ESTUDIOS
16. La legislación española de seguros y su adaptación a la normativa comunitaria. 1993
 15. El coste de los riesgos en la empresa española: 1991
 14. El Reaseguro de exceso de pérdidas 1993
 12. Los seguros de salud y la sanidad privada. 1993
 10. Desarrollo directivo: una inversión estratégica. 1992
 9. Técnicas de trabajo intelectual. 1992
 8. La implantación de un sistema de controlling estratégico en la empresa. 1992
 7. Los seguros de responsabilidad civil y su obligatoriedad de aseguramiento. 1992
 6. Elementos de dirección estratégica de la empresa. 1992
 5. La distribución comercial del seguro: sus estrategias y riesgos. 1991
 4. Los seguros en una Europa cambiante: 1990-95. 1991
 2. Resultados de la encuesta sobre la formación superior para los profesionales de entidades aseguradoras (A.P.S.). 1991
 1. Filosofía empresarial: selección de artículos y ejemplos prácticos. 1991