

## CLAIM RESERVING WITH GENERALIZED LINEAR MIXED MODELS BY USING R SOFTWARE

### CÁLCULO DE RESERVAS CON MODELOS LINEALES GENERALIZADOS MIXTOS HACIENDO USO DEL SOFTWARE R<sup>†</sup>

Eva Boj del Val<sup>1</sup> y Judit Esquinas Figueras<sup>2</sup>

#### Abstract

It is presented an application of generalized linear mixed models to the claim reserving problem, when data is of individual type, that generally corresponds to RBNS (Reported But Not Settled) claims. Reserves by years of origin and total are calculated and, with parametric bootstrap, predictive distributions of these reserves are estimated. Generalized linear mixed models are estimated using frequentist statistic. The used software is R, especially the *lme4* package, although it is also used SAS. Results are compared with those of the Chain-Ladder method.

**Key words:** Generalized linear mixed model; Claim Reserving; Reported but not settled claims; Chain-Ladder; R.

#### Resumen

Se presenta una aplicación de los modelos lineales generalizados mixtos al cálculo de provisiones cuando los datos son de tipo individual que, en general, se corresponden con datos de siniestros RBNS (*Reported But Not Settled*). Se calculan las reservas por años de ocurrencia y total y, con *bootstrap* paramétrico, se estiman las distribuciones predictivas de dichas reservas. Los modelos lineales generalizados mixtos se estiman utilizando estadística frecuentista. El software utilizado es R, en especial el paquete

---

<sup>†</sup> Trabajo financiado por el Ministerio de Educación y Ciencia, proyecto número MTM2014-56535-R, y por la Generalitat de Catalunya, AGAUR, proyecto número 2014SGR152. El presente artículo está derivado del Trabajo de Fin de Máster de Judit Esquinas, dirigido por Eva Boj.

<sup>1</sup> Autora de correspondencia. Profesora Titular de Universidad. Departamento de Matemática Económica, Financiera y Actuarial. Facultad de Economía y Empresa. Universidad de Barcelona. Avenida Diagonal 690, 08034\_Barcelona. España. E-mail: evaboj@ub.edu

<sup>2</sup> Máster en Ciencias Actuariales y Financieras en la Universidad de Barcelona, promoción 2014-15. E-mail: judit\_esquinas@hotmail.com

lme4, aunque también se utiliza SAS. Se comparan los resultados con los del método Chain-Ladder.

**Palabras clave:** Modelo lineal generalizado mixto; Reservas; siniestros RBNS; Chain-Ladder; R.

## 1. Introducción

El estudio se centra en el cálculo de provisiones pendientes de declaración de Seguros No Vida. En particular, en el caso en que los datos disponibles hacen referencia a siniestros individuales como es el caso de cálculo de reservas de siniestros de tipo RBNS. Las siglas RBNS proceden de las palabras en inglés *Reported But Not Settled* y se refieren a siniestros declarados a la Compañía pero todavía no pagados.

El objetivo principal del trabajo es profundizar en el uso de los modelos lineales generalizados mixtos, en inglés *Generalized Linear Mixed Models*, en adelante, GLMM, en aplicación al cálculo de provisiones de siniestros RBNS. Los GLMM (ver, por ejemplo, Booth y Hobert, 1998, Jiang, 2007, Molenberghs y Verbeke, 2005 y Verbeke y Molenberghs, 2000) incluyen en el predictor lineal una parte de efectos fijos y una parte de efectos aleatorios y pueden ser ajustados mediante métodos basados en estadística frecuentista o bien bayesiana. En este trabajo se utiliza estadística frecuentista para el ajuste de modelos, aunque los programas informáticos utilizados también permiten aplicar ajustes bayesianos (ver Wikidot, 2015).

El trabajo se estructura del siguiente modo: en el Apartado 2 se describe a grandes rasgos el GLMM y se muestran, en el sub-Apartado 2.1, las formulaciones para el caso de distribución Poisson y enlace logarítmico. En el Apartado 3 se realiza la aplicación al cálculo de reservas. Primeramente, en el sub-Apartado 3.1, se generan unos datos de siniestros RBNS que ayudan a ilustrar el modelo teórico propuesto. En el sub-Apartado 3.2 se estima el GLM que da como caso particular la misma estimación de reservas que el método Chain-Ladder cuando se aplica sobre el triángulo de desarrollo. Seguidamente, en el sub-Apartado 3.3 se estima el GLMM a los datos individuales de los pagos generados, realizando en 3.3.1 un supuesto particular ilustrativo que ayuda a comprender los elementos del GLMM propuesto en la práctica. En el sub-Apartado 3.4 se realiza propiamente el cálculo de reservas, incluyendo la estimación *bootstrap* de las distribuciones predictivas de las reservas por años de ocurrencia y total. Finalmente, el Apartado 4, contiene las principales aportaciones del trabajo.

Tal y como se ha comentado al principio del capítulo, este trabajo se centra en la aplicación de GLMM en cálculo de provisiones de siniestros RBNS, pero existen más problemas actuariales en que se aplica GLMM como son por ejemplo: construcción de modelos de riesgo de crédito, estimación de modelos de credibilidad y tarificación de Seguros No Vida, especialmente cuando se predice la frecuencia de siniestralidad. Nos referimos a Antonio *et al.* (2005, 2006), Antonio y Beirlant (2007, 2008), Antonio y Valdez (2012), Boucher y Denuit (2006), Boucher *et al.* (2008), Frees (2010), Frees *et al.* (2014), Godecharle y Antonio (2014), Kroon (2012), Pigeon *et al.* (2013, 2014), Verrall *et al.* (2010) y a McNail y Wendl (2005) para el detalle de todas estas aplicaciones actuariales.

Si nos centramos en la bibliografía relativa al problema que aborda el trabajo, el cálculo de provisiones, ésta es escasa. De las dos referencias consultadas, en Antonio *et al.* (2006) se propone el tratamiento de datos individuales con un modelo que generaliza la curva de Hoerl de la fórmula (3.4) de dicha referencia al caso de aplicar GLMM, incluyendo en la parte de efectos aleatorios un término independiente, los años de desarrollo y su logaritmo. También, en el mismo artículo, se asume la distribución Normal y la función de enlace logarítmica, y se realiza todo el estudio desde un punto de vista y de análisis bayesiano. Por otro lado, en otra línea, es de resaltar la referencia Ayuso y Santolino (2008) que, aunque no aplica GLMM, estudia también el problema del cálculo de provisiones de siniestros RBNS en el ámbito de Solvencia II, todo ello mediante la aplicación de modelos GLM heterocedásticos para el coste de compensación que se distribuye lognormalmente para diferentes niveles de gravedad del daño corporal.

Seguidamente, para tener una visión general del tema que aborda el trabajo, vemos la Figura 1, extraída de Antonio y Plat (2014), dónde se describen las diferentes fases a las que se puede enfrentar una Compañía de Seguros desde la ocurrencia de un siniestro hasta la finalización o cierre de su pago.

Según el esquema de la Figura 1, la ocurrencia de un siniestro tiene lugar en el momento del tiempo  $t_1$  y posteriormente es declarado a la Compañía en  $t_2$ . Seguidamente, la Compañía efectúa uno o varios pagos,  $t_3, t_4, t_5$ , hasta el cierre del siniestro en  $t_6$ . Pero, según el tipo de siniestro y los resultados de los procesos judiciales, puede haber siniestros reabiertos por ejemplo en  $t_7$  con el objetivo de efectuar pagos adicionales en  $t_8$ , justo antes de que finalmente se realice la clausura definitiva del siniestro en  $t_9$ .

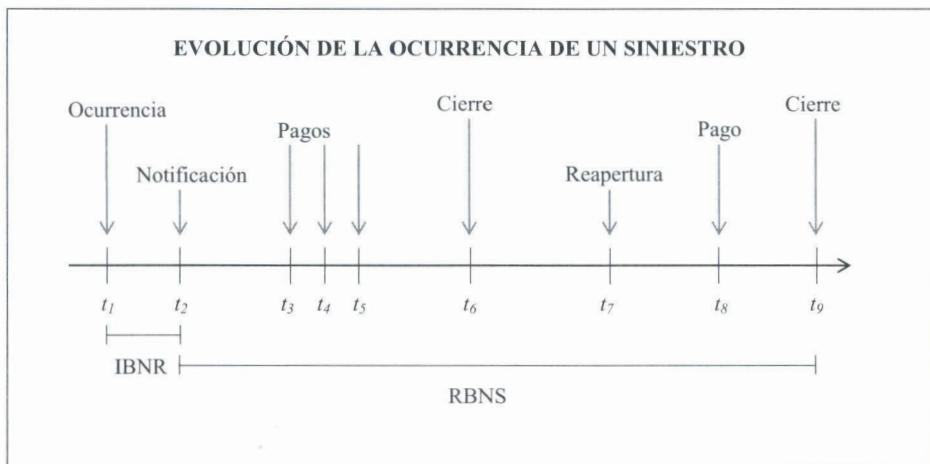


Figura 1. Fases de la evolución de los pagos de un siniestro.

Debido a que un siniestro no se expira con un solo pago y que hay varios que en un futuro son inciertos, la Compañía tiene el objetivo de realizar una buena predicción de reservas.

Se pueden distinguir dos tipos de reservas según el tipo de siniestros:

- Reservas de siniestros RBNS: son las reservas en las que se centra el estudio, y se calculan para siniestros cuya ocurrencia tienen lugar antes del momento de cálculo de reservas, aunque la totalidad de su pago se efectúe en un momento futuro, por ejemplo, en el período  $t_2 \leq \tau < t_6$  o bien  $t_2 \leq \tau < t_9$ , según la Figura 1. En este tipo de reservas se incluye una estimación de los pagos de aquellos siniestros que han ocurrido y se han comunicado a la Compañía pero que aún no se han terminado de pagar. Para esta clase de siniestros se conoce el momento de ocurrencia y se dispone de información relativa a los factores de riesgo del objeto asegurado, del tomador de la póliza y del entorno (ver Frees *et al.*, 2014, capítulo 18).

Por lo tanto, para realizar la predicción de los pagos futuros se puede utilizar la información individual del entorno del siniestro que vendrá explicada por los factores de riesgo, los cuales aportarán la información predictora de la siniestralidad del coste del siniestro.

Año de ocurrencia	Año de desarrollo								
	0	1	...	j	...	k-i	...	k-1	k
0	$c_{00}$	$c_{01}$	...	$c_{0j}$	...	$c_{0k-i}$	...	$c_{0k-1}$	$c_{0k}$
1	$c_{10}$	$c_{11}$	...	$c_{1j}$	...	$c_{1k-i}$	...	$c_{1k-1}$	
...	...	...	...	...	...	...	...	...	
i	$c_{i0}$	$c_{i1}$	...	$c_{ij}$	...	$c_{ik-i}$			
...	...	...	...	...	...				
$k-j$	$c_{k-j0}$	$c_{k-j1}$	...	$c_{k-jj}$					
...	...	...	...						
$k-1$	$c_{k-10}$	$c_{k-11}$							
$k$	$c_{k0}$								

Figura 2. Triángulo de desarrollo con cuantías totales no acumuladas.

Reservas de siniestros IBNR: sus siglas proceden de las palabras en inglés *Incurred But Not Reported*, y se corresponden con siniestros que se producen antes del momento actual de cálculo de reservas, pero su declaración, y en consecuencia, el pago correspondiente, tiene lugar en un momento futuro. Es el caso en que el momento de cálculo de reservas está situado en el intervalo  $[t_1, t_2]$ , en el que los siniestros han ocurrido pero aún no se han comunicado a la Compañía. Para este tipo de siniestros no se dispone de la información en el momento de ocurrencia ni de los factores de riesgo individuales, únicamente se dispone de información histórica agregada de los últimos períodos que se suele recoger en un triángulo de desarrollo (ver Figura 2). De este modo, su predicción se realiza a partir de la cuantía total de los siniestros por año de ocurrencia y por año de desarrollo, tal y como se explica por ejemplo en Boj *et al.* (2014). En esta situación, se utiliza el modelo lineal generalizado, GLM, incluyendo en el predictor lineal una parte de efectos fijos a partir de las variables binarias que definen los años de ocurrencia y de desarrollo. Paralelamente se obtiene la misma estimación de reservas que con el método clásico Chain-Ladder.

En este trabajo se propone utilizar para GLMM una parte de efectos fijos constituida por los años de ocurrencia, los años de desarrollo y un factor de riesgo (hasta aquí igual que el predictor lineal de un GLM), con la adición de una parte de efectos aleatorios en que se define qué pagos son los que pertenecen al mismo siniestro (que hace que el modelo pase a ser GLMM).

## 2. Modelos lineales generalizados mixtos

En el presente trabajo, se propone el uso de un modelo GLMM que "generaliza" el caso Chain-Ladder clásico en aplicación al cálculo de provisiones de siniestros RBNS, por lo que supone una aportación bibliográfica dentro de las aplicaciones actuariales de cálculo de reservas con GLMM.

Se realiza una aplicación con datos de siniestros RBNS que se generan para ilustrar las metodologías del estudio, pues no se dispone de una base de datos reales. En el GLMM de la aplicación se asume distribución de Poisson y función de enlace logarítmica para poder realizar una comparativa adecuada con el método de cálculo de provisiones Chain-Ladder clásico (van Eeghen, 1981).

Los GLMM (Booth y Hobert, 1998, Jiang, 2007, Molenberghs y Verbeke, 2005 y Verbeke y Molenberghs, 2000) son GLM que incluyen en el predictor lineal una parte de efectos fijos, como en GLM, y además una parte de efectos aleatorios que puede recoger variables de grupo o temporales.

Supongamos la variable respuesta, variable aleatoria que se corresponde con las observaciones para cada individuo<sup>3</sup>  $i$  con  $i=1,\dots,N$  recogidas por  $Y_{i1}, \dots, Y_{in_i}$ . La expresión general del modelo es:

$$\mathbf{Y}_i = g^{-1}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i), \quad i = 1, \dots, N,$$

con

$$\boldsymbol{\mu}_i = E[\mathbf{Y}_i|\mathbf{b}_i] = \psi'(\boldsymbol{\theta}_i) \quad \text{y} \quad \text{Var}[\mathbf{Y}_i|\mathbf{b}_i] = \emptyset V(\boldsymbol{\mu}_i) = \emptyset \psi''(\boldsymbol{\theta}_i).$$

Donde:

$\mathbf{Y}_i = (y_{i1}, \dots, y_{in_i})$  es el vector de observaciones para el  $i$ -ésimo individuo,  $(n_i \times 1)$ ,

$n_i$  es el número de observaciones para el individuo  $i$ -ésimo,

$N$  es el número total de individuos diferentes de la muestra ( $1 \leq i \leq N$ ),

---

<sup>3</sup> Un individuo o sujeto puede referirse a un grupo de riesgos "similares".

$\boldsymbol{\beta}$  ( $p \times 1$ ) es el vector de efectos fijos que contiene los  $p$  parámetros fijos, desconocidos y comunes para todos los individuos de la muestra,

$\mathbf{b}_i$  ( $q \times 1$ ) es el vector de efectos aleatorios con  $q$  parámetros aleatorios para el individuo  $i$ -ésimo, y por tanto, diferente para cada individuo  $i$ ,

$\mathbf{X}_i$  ( $n_i \times p$ ) es la matriz de diseño para los efectos fijos para cada individuo  $i$ ,

$\mathbf{Z}_i$  ( $n_i \times q$ ) es la matriz de diseño para los efectos aleatorios para cada individuo  $i$ ,

$\boldsymbol{\varepsilon}_i$  ( $n_i \times 1$ ) es el vector que contiene los residuos para cada individuo  $i$ ,

$g(\cdot)$  es la función de enlace monótona y diferenciable,

$g^{-1}(\cdot)$  es la inversa de la función de enlace,

$\emptyset$  es el parámetro de dispersión,

$V(\mu_i)$  es la función de varianza de la distribución supuesta (en principio, de la familia exponencial).

Cabe destacar que el predictor lineal del GLMM contiene la suma de la componente sistemática  $\mathbf{X}_i\boldsymbol{\beta}$  que incluye los efectos fijos y la componente aleatoria  $\mathbf{Z}_i\mathbf{b}_i$  que incluye los efectos aleatorios. En este sentido, el GLMM se puede considerar un GLM multinivel, que aporta por consecuencia la parte multinivel (o mixta) respecto a los GLM clásicos.

Las hipótesis que se asumen son las siguientes:

- Dado el vector  $\mathbf{b}_i$ , se asume que las diferentes observaciones de  $\mathbf{Y}_i$  son independientes con una densidad de la familia exponencial:

$$f(y_{ij} | \mathbf{b}_i, \boldsymbol{\beta}, \emptyset) = e^{\left( \frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{\emptyset} + c(y_{ij}, \emptyset) \right)}, \quad j = 1, \dots, n_i.$$

La respuesta va ligada con el predictor lineal a través de la función de enlace, como en GLM:

$$g(\boldsymbol{\mu}_i) = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i .$$

- ii) Por otra parte,  $\mathbf{b}_i$  y  $\boldsymbol{\varepsilon}_i$  también son independientes y se distribuyen según una normal con vector de medias 0 y matrices de varianzas y covarianzas  $\mathbf{D}$  ( $q \times q$ ) y  $\Sigma_i (n_i \times n_i)$ , respectivamente. Así  $\mathbf{b}_i \sim N(0, \mathbf{D})$  y  $\boldsymbol{\varepsilon}_i \sim N(0, \Sigma_i)$ .

La estimación de parámetros se realiza a través de la función de cuasi-verosimilitud que se define para los efectos fijos y para los aleatorios.

El estimador  $\widehat{\boldsymbol{\beta}}$ , conocido como BLUE (*Best Linear Unbiased Estimator*), para los efectos fijos, se haya maximizando la función de cuasi-verosimilitud y se expresa como:

$$\widehat{\boldsymbol{\beta}} = (\sum_{i=1}^N \mathbf{X}'_i \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{V}_i^{-1} \mathbf{Y}_i .$$

Y del mismo modo, los efectos aleatorios,  $\mathbf{b}_i$ , se estiman con  $\widehat{\mathbf{b}}$  expresado como:

$$\widehat{\mathbf{b}} = \mathbf{D} \mathbf{Z}'_i \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}) .$$

Para la estimación, en general son necesarias aproximaciones a la verosimilitud o técnicas numéricas de integración. En la aplicación práctica y en la función *glmer* del paquete *lme4* de R, se utiliza la cuadratura de Gauss-Hermite. Se trata de una integración numérica con el objetivo de conseguir la solución aproximada con el máximo grado de exactitud, a través de una función con intervalos entre unos puntos de interpolación determinados.

## 2.1. Caso particular: distribución Poisson y enlace logarítmico

El GLMM que se asume en el trabajo sigue una distribución Poisson (sobre-dispersa) junto con enlace logarítmico, pues la idea es comparar los resultados con los del caso Chain-Ladder. Por ello se dedica este sub-apartado a las expresiones particulares de estos supuestos. Básicamente se reproduce la descripción de la página 64 de Antonio y Bierlant (2007), donde se encuentra perfectamente descrito.

Consideremos la variable independiente  $Y_{ij}$  que se refiere a la cuantía del siniestro para la observación  $j$ -ésima para el individuo  $i$ . Asumimos que, condicionado a  $b_i$ ,  $Y_{ij}$  sigue una distribución Poisson con:

$$\mu_{ij} = E [Y_{ij} | \mathbf{b}_i] = e^{(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{b}_i)},$$

y asumimos que  $\mathbf{b}_i \sim N(0, \sigma_b^2)$ . En tal caso, se comprueba que:

$$\begin{aligned} \text{Var}(Y_{ij}) &= \text{Var}(E(Y_{ij} | \mathbf{b}_i)) + E(\text{Var}(Y_{ij} | \mathbf{b}_i)) = \\ &= E(Y_{ij})(e^{(\mathbf{X}_{ij}\boldsymbol{\beta})} \left[ e^{\left(\frac{3\sigma_b^2}{2}\right)} - e^{\left(\frac{\sigma_b^2}{2}\right)} \right] + 1), \end{aligned}$$

$$\begin{aligned} \text{Cov}(Y_{ij_1}, Y_{ij_2}) &= \text{Cov}(E(Y_{ij_1} | \mathbf{b}_i), E(Y_{ij_2} | \mathbf{b}_i)) + E(\text{Cov}(Y_{ij_1}, Y_{ij_2} | \mathbf{b}_i)) = \\ &= e^{(\mathbf{X}_{ij_1}\boldsymbol{\beta})} e^{(\mathbf{X}_{ij_2}\boldsymbol{\beta})} (e^{(2\sigma_b^2)} e^{(\sigma_b^2)}). \end{aligned}$$

Se observa que las expresiones de dentro del paréntesis de la ecuación de  $\text{Var}(Y_{ij})$  son superiores a la unidad y, aunque  $Y_{ij} | \mathbf{b}_i$  se distribuye como una Poisson, la distribución marginal de  $Y_{ij}$  será sobre-dispersa. Es por ello que el hecho de incluir los efectos aleatorios de la variable de grupo, indica la sobre-dispersión de los datos, y, por tanto, la distribución asumida es Poisson sobre-dispersa.

En GLMM los parámetros representan el efecto de las variables explicativas en las respuestas de un individuo específico a través de la inversa de la función de enlace,  $E[Y_{ij}] = E[E[Y_{ij} | \mathbf{b}_i]] = E[g^{-1}(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{b}_i)]$ . En cambio, si únicamente intervienen efectos fijos como en el caso de GLM, entonces  $E[Y_{ij}] = E[g^{-1}(\mathbf{X}_{ij}\boldsymbol{\beta})]$ .

## 2.2. Software utilizado: R y SAS

El software utilizado es R (R Development Core Team, 2016). Tal y como se indica en la referencia Wikidot (2015), hay varios paquetes de R que permiten estimar GLMM. Algunos de estos son: la función *glmer* del paquete *lme4*, la función *glmmML* del paquete *glmmML*, la función *MCMCglmm* del paquete con el mismo nombre y la función *glmmPQL* del paquete *MASS*. No obstante, se utiliza principalmente el paquete *lme4*, que permite ajustar GLMM mediante la función *glmer* (Bates *et al.*, 2014a y 2014b). Con este paquete se realizan todos los cálculos, incluido el ajuste de distribuciones predictivas mediante *bootstrap* paramétrico. Para realizar los cálculos relativos al procedimiento *bootstrap* son de especial ayuda las funciones *simulate* y *refit* del paquete *lme4* (Bates *et al.*, 2014b). Como valor añadido al carácter práctico del trabajo y únicamente para contrastar el

modelo resultante, también se justan los GLMM del estudio con la función *glmPQL* del paquete *MASS* de R (Ripley *et al.*, 2015).

Por otro lado, nos referimos al manual on-line SAS Institute Inc. (2009) donde se explica cómo diversos paquetes comerciales de SAS permiten estimar GLMM. Algunos de ellos son PROC NLMIXED y PROC GLIMMIX. En el trabajo se realizan los ajustes del GLMM con el método *QUAD* de la función PROC GLIMMIX (SAS Institute Inc., 2009) que aproxima la probabilidad marginal con una adaptación de cuadratura de Gauss-Hermite.

Tanto R como SAS son programas generalmente aplicados para la resolución de cálculos actuariales (Charpentier, 2014 y SAS Institute Inc., 2009).

### **3. Aplicación actuarial**

#### **3.1. Generación de datos**

Uno de los métodos deterministas clásicos de cálculo de reservas es el método Chain-Ladder (van Eeghen, 1981). Es conocido que muchos de los métodos estocásticos y algunos de los métodos deterministas generalizan dicho método. En particular, cuando se utiliza para la estimación GLM asumiendo distribución de error Poisson (sobre-dispersa), función de enlace logarítmica y como predictores las variables binarias correspondientes a los años de ocurrencia y de desarrollo de los siniestros, el modelo de regresión proporciona las mismas estimaciones que las obtenidas con el método Chain-Ladder clásico (Boj *et al.*, 2014).

Con el objetivo de realizar una interpretación y comparativa de resultados relativamente sencilla con aquellos resultados que se obtienen con el método Chain-Ladder clásico, en este apartado se generan datos individuales de siniestros RBNS y se aplica GLMM ampliando el predictor lineal con un factor de riesgo en la parte de efectos fijos  $\beta$  (aparte de los años de ocurrencia y desarrollo como en Chain-Ladder clásico), y con un término independiente,  $b_i$ , que distingue a los individuos, en este caso siniestros, en la parte de efectos aleatorios. Las hipótesis respecto de la distribución del error y la función de enlace del GLMM se mantienen en distribución Poisson y enlace logarítmico como en el caso del GLM.

Puesto que no se dispone de datos reales, se generan datos de siniestros RBNS. Para generar los importes de las cuantías de los siniestros se utilizan generadores de números aleatorios de distribución Gamma con diferentes parámetros. Los parámetros se combinan con el objetivo de tener cuantías

mayores en el primer año de desarrollo y, en general, cuantías mayores (o menores) en función del siniestro generado. En el Anexo 1 se encuentran las instrucciones para la generación de las cuantías, así como las cuantías resultantes. En la Tabla 1 se recoge el resumen de agregar las cuantías individuales no acumuladas por año de ocurrencia y año de desarrollo, y se observa cómo en general las cuantías son mayores en cada año de ocurrencia para el primer año de desarrollo. En total se generan 252 cuantías de pagos individuales, con una media de 1216.70 unidades monetarias, un máximo de 2046.78, un mínimo de 641.24 y una desviación estándar de 328.87.

En la base de datos de siniestros RBNS que se genera y en el predictor lineal del GLMM que se estima, se realizan las siguientes hipótesis:

- Se asumen 6 años de ocurrencia, de 0 a 5.
- Se asumen 6 años de desarrollo, de 0 a 5.
- Se asumen exactamente 12 siniestros en cada año de ocurrencia, los cuales se pagan durante los 6 años de desarrollo en su totalidad. Los siniestros de los diferentes años de ocurrencia se suponen independientes entre sí.
- Se asume un factor de riesgo con dos clases. Con el objetivo de que el factor de riesgo sea estadísticamente significativo en el modelo, las dos clases se generan del siguiente modo: si la media de los 6 pagos del siniestro es mayor a la media general de 1216.70, se asigna la clase 1 a dicho siniestro, y en caso contrario se asigna la clase 2. De este modo se obtiene un factor de riesgo con dos clases con 119 y 133 cuantías de pagos cada una, y con medias 1453.012 y 1005.263 respectivamente. Posteriormente en el GLMM se comprueba que realmente es un factor de riesgo significativo, tal y como más abajo se muestra.

Cada celda del triángulo contiene exactamente 12 datos. Siguiendo la nomenclatura del apartado anterior, en total se dispone de  $N = 72$  siniestros,  $6 \times 12$ , pues se dispone de 12 siniestros para cada uno de los años de ocurrencia que son 6. Con un total de 252 pagos individuales, pues tenemos 21 casillas con 12 pagos cada una,  $21 \times 12$ . Con lo que estimaremos los valores de 15 celdas, en total 180 estimaciones de pagos futuros individuales,  $15 \times 12$ .

Los 12 primeros siniestros que están situados en el año de ocurrencia 0, tienen  $n_i = 6$  observaciones individuales cada uno, que se corresponden con los 6 años de desarrollo y no se debe realizar ninguna predicción. Los

siguientes 12 siniestros, situados en el año de ocurrencia 1, están numerados del 13 al 24 y tienen  $n_i = 5$  cuantías de pagos individuales. Para este segundo año de ocurrencia se debe realizar la predicción del pago del año de desarrollo 5 para cada individuo o siniestro, por tanto 12 predicciones individuales. Y así sucesivamente hasta el año de ocurrencia 5 en que se dispone tan sólo de 12 datos de pagos individuales referentes a los individuos que van del 61 al 72, los cuales únicamente tienen una observación,  $n_i = 1$ , situada en el año de desarrollo 0. En este último año de ocurrencia se deben realizar un total de  $12 \times 5 = 60$  predicciones de cuantías de pagos de siniestros, las correspondientes a los años de desarrollo que van del 1 al 5 para los siniestros del 61 al 72.

Año de ocurrencia	Año de desarrollo					
	0	1	2	3	4	5
0	17832.93	13245.26	13241.77	13264.87	13328.32	13304.35
1	17847.71	13148.79	13356.13	13033.24	13289.59	
2	17807.91	13518.14	13160.03	13074.81		
3	18117.04	13250.59	13367.27			
4	17974.42	13108.10				
5	18337.13					

Tabla 1. Triángulo de desarrollo con el resumen de agregar las cuantías individuales no acumuladas por año de ocurrencia y año de desarrollo generadas para el caso de estudio.

A continuación, en la Figura 3, se muestra el resumen genérico en forma de triángulo de desarrollo de las cuantías individuales y no acumuladas de los pagos de los siniestros generadas para el caso del estudio. Cabe notar que, como también se tiene en cuenta adicionalmente un factor de riesgo con dos clases, si quisieramos indicar el tipo de información generada teniendo en cuenta el factor, la información debería resumirse incluyendo una tercera dimensión y subíndice en la Figura 3. Para indicar a qué siniestro hace referencia cada pago, en este ejemplo se incorpora un nuevo subíndice. Por ejemplo,  $c_{0,7,1}$  se refiere al pago que se realiza del siniestro 7 ocurrido en el año 0 y que se paga en el año de desarrollo 1.

Año de ocurrencia	Año de desarrollo				
	0	1	...	5	
0	$c_{0,1,0}, c_{0,2,0},$ $c_{0,3,0}, c_{0,4,0},$ $c_{0,5,0}, c_{0,6,0},$ $c_{0,7,0}, c_{0,8,0},$ $c_{0,9,0}, c_{0,10,0},$ $c_{0,11,0}, c_{0,12,0}.$	$c_{0,1,1}, c_{0,2,1},$ $c_{0,3,1}, c_{0,4,1},$ $c_{0,5,1}, c_{0,6,1},$ $c_{0,7,1}, c_{0,8,1},$ $c_{0,9,1}, c_{0,10,1},$ $c_{0,11,1}, c_{0,12,1}.$	...	$c_{0,1,5}, c_{0,2,5},$ $c_{0,3,5}, c_{0,4,5},$ $c_{0,5,5}, c_{0,6,5},$ $c_{0,7,5}, c_{0,8,5},$ $c_{0,9,5}, c_{0,10,5},$ $c_{0,11,5}, c_{0,12,5}.$	
1	$c_{1,13,0}, c_{1,14,0},$ $c_{1,15,0}, c_{1,16,0},$ $c_{1,17,0}, c_{1,18,0},$ $c_{1,19,0}, c_{1,20,0},$ $c_{1,21,0}, c_{1,22,0},$ $c_{1,23,0}, c_{1,24,0}.$	$c_{1,13,1}, c_{1,14,1},$ $c_{1,15,1}, c_{1,16,1},$ $c_{1,17,1}, c_{1,18,1},$ $c_{1,19,1}, c_{1,20,1},$ $c_{1,21,1}, c_{1,22,1},$ $c_{1,23,1}, c_{1,24,1}.$	...		
5	$\vdots$		$\vdots$		

Figura 3. Resumen en forma de triángulo de desarrollo con las cuantías individuales y no acumuladas generadas para el caso de estudio.

Para el siniestro 7, los pagos que se realizan son  $(c_{0,7,0}, c_{0,7,1}, \dots, c_{0,7,5})$ . Según la nomenclatura del apartado anterior dónde se describía en general el GLMM, las observaciones del siniestro 7 en nuestra aplicación se

corresponderían con el vector  $\mathbf{y}_{7n_7} = (y_{71}, y_{72}, \dots, y_{76})$  que tiene  $n_7 = 6$  observaciones.

### **3.2. Estimación Chain-Ladder clásica a partir de un modelo lineal generalizado con distribución Poisson y función de enlace logarítmica**

A los 21 datos del triángulo de desarrollo agregado que contiene las cuantías totales no acumuladas de cada año de ocurrencia en cada año de desarrollo de la Tabla 1, se aplica GLM con distribución Poisson, enlace logarítmico y se incluyen como predictores las variables binarias referentes a los años de ocurrencia y de desarrollo. Como resultado, obtenemos la estimación de cuantías recogidas en la Tabla 2, las cuales es bien sabido que coinciden con la estimación de cuantías del método Chain-Ladder clásico. A partir de estas estimaciones podemos calcular las reservas por años de ocurrencia y total, que muestran en la segunda columna de la Tabla 6.

Los detalles del ajuste del GLM son los siguientes. El modelo se estima con la función *glm* del paquete *stats* de R (ver Anexo 2). Se denomina "aoglm" a la variable nominal definida como factor en R que contiene los 6 años de ocurrencia numerados del 1 al 6, "adglm" a la variable nominal definida como factor en R que contiene los 6 años de desarrollo numerados del 1 al 6, y "cuantagre" a la variable que contiene las 21 cuantías agregadas de la Tabla 1. La instrucción para el ajuste del modelo es:

```
glm1 <- glm(formula = cuantagre ~ aoglm + adglm, family = quasipoisson).
```

El modelo resultante que en el término independiente,  $\hat{\beta}_0$ , incorpora la primera clase de ambos factores,

$$\hat{\mu} = e^{\left( \hat{\beta}_0 + \hat{\beta}_{ao_2} + \hat{\beta}_{ao_3} + \hat{\beta}_{ao_4} + \hat{\beta}_{ao_5} + \hat{\beta}_{ao_6} + \hat{\beta}_{ad_2} + \hat{\beta}_{ad_3} + \hat{\beta}_{ad_4} + \hat{\beta}_{ad_5} + \hat{\beta}_{ad_6} \right)},$$

se encuentra ejecutado en el Anexo 2, dónde se tiene el valor de los coeficientes y la estimación del parámetro de dispersión, que para este GLM Poisson sobre-disperso es de 1.322037. Por completitud, se utiliza el paquete *ChainLadder* de R (Gesman *et al.*, 2015) para el cálculo de reservas. Los

resultados se muestran en el Anexo 3. Como observación, se muestra el GLM utilizado por la función *glmReserve* a modo interno, *glm2\$model*, y se puede comprobar cómo se corresponde con el modelo *glm1* del Anexo 2, el cuál reproduce Chain-Ladder clásico.

Puesto que los datos de partida son desagregados, calculamos el error cuadrático medio (ECM) de predicción asignando la cuantía media de cada celda a los correspondientes datos originales. Con lo que se obtiene un ECM de 75798.49 (ver Tabla 4). De este modo tendremos cantidades comparables cuando calculemos el ECM cometido con GLMM y datos individuales.

Año de ocurrencia	Año de desarrollo					
	0	1	2	3	4	5
0	17915.65	13253.92	13271.71	13140.59	13331.30	13304.35
1	17855.60	13209.49	13227.22	13096.54	13286.61	13259.75
2	17909.12	13249.09	13266.87	13135.80	13326.44	13299.50
3	18034.02	13341.49	13359.40	13227.41	13419.38	13392.25
4	17865.62	13216.90	13234.65	13103.89	13294.07	13267.20
5	18337.13	13565.73	13583.94	13449.73	13644.93	13617.35

Tabla 2. Estimación del triángulo a partir de las cuantías agregadas por año de ocurrencia y año de desarrollo generadas para el caso de estudio con GLM.

### 3.3. Estimación del modelo lineal generalizado mixto con distribución Poisson y función de enlace logarítmica

En este apartado ajustamos un GLMM que contiene en los efectos fijos los mismos predictores que los que se incluyen en el GLM que reproduce Chain-Ladder del apartado anterior, "ao" y "ad", más el factor de riesgo "fr" definido como factor en R con dos clases numeradas del 1 al 2. En la parte aleatoria se incluye la variable "sin" que es la indicadora de qué pagos pertenecen al mismo siniestro. La variable respuesta en este modelo es la variable "cuantías" que contiene los diferentes pagos individuales. En el Anexo 1 se encuentran las instrucciones de R para la generación de las variables.

El modelo se estima con la función *glmer* del paquete *lme4* de R. La instrucción es:

```
glmm1 <- glmer(cuantías ~ ao + ad + fr + (1|sin), family = poisson, nAGQ=15).
```

El modelo resultante que contiene un término independiente,  $\hat{\beta}_0$ , que incorpora las primeras clases de los tres predictores de la parte de efectos fijos,

$$\hat{\mu} = e^{\left( \hat{\beta}_0 + \hat{\beta}_{ao_2} + \hat{\beta}_{ao_3} + \hat{\beta}_{ao_4} + \hat{\beta}_{ao_5} + \hat{\beta}_{ao_6} + \hat{\beta}_{ad_2} + \hat{\beta}_{ad_3} + \hat{\beta}_{ad_4} + \hat{\beta}_{ad_5} + \hat{\beta}_{ad_6} + \hat{\beta}_{fr_2} + \hat{b}_{sin} \right)},$$

visualizado con el comando *summary(glmm1)* se encuentra ejecutado para los datos del ejemplo en el Anexo 4. En la salida de la ejecución se obtienen las estimaciones de los parámetros de la parte fija del predictor lineal, que también pueden visualizarse con la instrucción *fixef(glmm1)*. Las 252 estimaciones de la parte aleatoria,  $\hat{b}_{sin}$ , debe visualizarse con la instrucción *ranef(glmm1)*.

Para este GLMM podemos estimar el parámetro de dispersión a partir de los residuos de Pearson, obteniendo 2.510621, o bien a partir de los residuos de desvianza, obteniendo 2.594906. Para tener en cuenta la sobre-dispersión en el proceso de *bootstrap* paramétrico que se realiza en el Apartado 3.3, se utiliza el parámetro calculado a partir de los residuos de Pearson por analogía al caso en que se aplica GLM para el cálculo de reservas.

En la Tabla 3, se muestra el resumen agregado por año de ocurrencia y año de desarrollo de las estimaciones a partir de las cuantías individuales generadas para el caso de estudio anterior con GLMM.

Finalmente, se calcula el ECM:

$$ECM = \frac{\sum_{i=1}^N \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_{ij})^2}{\sum_{i=1}^N n_i}.$$

Y se obtiene que es de 2661.638. Si lo comparamos (Tabla 4) con el obtenido con GLM caso Chain-Ladder que trabaja con datos totales o

agregados, observamos una mayor precisión con GLMM como era de esperar.

Año de ocurrencia	Año de desarrollo					
	0	1	2	3	4	5
0	17883.63	13162.19	13195.94	13135.68	13420.70	13399.46
1	17853.32	13139.88	13173.58	13113.42	13397.96	11225.18
2	17758.60	13322.01	13280.89	13199.24	16422.70	16396.71
3	18084.67	13310.15	13344.28	12972.27	13253.75	13232.77
4	17787.80	13298.15	12650.99	12593.22	12866.47	12846.10
5	18342.83	9927.92	9953.38	9907.93	10122.91	10106.89

Tabla 3. Resumen agregado por año de ocurrencia y año de desarrollo de las estimaciones a partir de las cuantías individuales generadas para el caso de estudio con GLMM.

GLM	GLMM
75798.49	2661.638

Tabla 4. Error cuadrático medio de los datos individuales generados para el caso de estudio con GLM y GLMM.

### 3.3.1. Supuesto ilustrativo

Supongamos que estamos interesados en estimar los pagos que ha de realizar la Compañía de los siniestros del año de ocurrencia 5, en el año de desarrollo 2, de los siniestros que pertenecen a la clase de riesgo 1. El modelo que predice dicho caso es:

$$\hat{\mu} = e^{(\hat{\beta}_0 + \hat{\beta}_{ao_6} + \hat{\beta}_{ad_3} + \hat{b}_{sin})},$$

Y sustituyendo los coeficientes de los efectos fijos tenemos:

$$\hat{\mu} = e^{(7.469516 - 0.059577 - 0.303977 + \hat{b}_{sin})}.$$

Finalmente, incluyendo el efecto aleatorio,  $\hat{b}_{sin}$ , que es diferente para cada siniestro, calculamos las estimaciones de los pertenecientes a la clase 1 del factor de riesgo, que son los siniestros 62, 63, 64, 65, 66, 67, 68, 69 y 70, respectivamente:

$$\hat{\mu} = e^{(7.469516 - 0.059577 - 0,303977 + 0.004196)} = e^{(7.110158)} = 1224.34,$$

$$\hat{\mu} = e^{(7.469516 - 0.059577 - 0,303977 - 0.093513)} = e^{(7.012449)} = 1110.37,$$

$$\hat{\mu} = e^{(7.469516 - 0.059577 - 0,303977 - 0.106372)} = e^{(6.99959)} = 1096.18,$$

$$\hat{\mu} = e^{(7.469516 - 0.059577 - 0,303977 + 0.209016)} = e^{(7.314978)} = 1502.64,$$

$$\hat{\mu} = e^{(7.469516 - 0.059577 - 0,303977 + 0,106525)} = e^{(7.212487)} = 1356.26,$$

$$\hat{\mu} = e^{(7.469516 - 0.059577 - 0,303977 + 0.169842)} = e^{(7.275804)} = 1444.91,$$

$$\hat{\mu} = e^{(7.469516 - 0.059577 - 0,303977 - 0.155122)} = e^{(6.95084)} = 1044.03,$$

$$\hat{\mu} = e^{(7.469516 - 0.059577 - 0,303977 - 0.034546)} = e^{(7.071416)} = 1177.81,$$

$$\hat{\mu} = e^{(7.469516 - 0.059577 - 0,303977 + 0.0545336)} = e^{(7.1604956)} = 1287.55.$$

Adicionalmente se estima el mismo GLMM pero con la función *glmmPQL* del paquete *MASS* (Ripley *et al.*, 2015) que estima también por cuasi-verosimilitud. La instrucción es la siguiente:

```
glmm2 <- glmmPQL(cuantías ~ ao + ad + fr, random = ~ 1 | sin, family = poisson).
```

En el Anexo 5 se encuentra la ejecución para los datos del ejemplo visualizada con el comando *summary(glmm2)*.

Obteniendo, para este supuesto, el modelo:

$$\hat{\mu} = e^{(7.510655 - 0.108062 - 0.30419 + \hat{b}_{sin})}.$$

Siniestro	<i>glmer (lme4)</i>		<i>glmmPQL (MASS)</i>		<i>SAS</i>	
	Predictor lineal	Respuesta	Predictor lineal	Respuesta	Predictor lineal	Respuesta
62	7.11	1224.34	7.11	1222.88	7.11	1223.54
63	7.01	1110.37	7.02	1117.01	7.01	1109.71
64	7.00	1096.18	7.01	1103.88	7.00	1095.54
65	7.31	1502.64	7.30	1483.71	7.31	1501.44
66	7.21	1356.26	7.21	1346.14	7.21	1355.30
67	7.28	14444.91	7.26	1429.37	7.28	1443.87
68	6.95	1044.03	6.96	1055.70	6.95	1043.44
69	7.07	1177.81	7.07	1179.58	7.07	1177.07
70	7.16	1287.55	7.16	1281.85	7.16	1286.67

Tabla 5. Predicción para GLMM con R y SAS en el predictor lineal y en la respuesta para los siniestros que se encuentran situados en el año de ocurrencia 5, año de desarrollo 2 y que pertenecen a la clase 1 del factor de riesgo.

Por otro lado, también se estima el modelo con  $P_{ROC} G_{LIMMIX}$  de SAS (SAS Institute Inc., 2009). En este caso las clases que SAS incorpora en el término independiente por defecto no son 1, 1 y 1 respectivamente para cada predictor definido como factor, sino la 6, 6 y 2, es decir, la clase sexta de "ao", la clase sexta de "ad" y la clase 2 de "fr". Las instrucciones con SAS son:

```
proc glimmix data=data method=quad;
  class ao ad fr sin;
  model cuantías = ao ad fr / link=log s dist=poisson solution;
  random intercept / subject=sin solution ;
run;
```

Obteniendo el modelo:

$$\hat{\mu} = e^{(\hat{\beta}_0 + \hat{\beta}_{ad_3} + \hat{\beta}_{fr_1} + \hat{b}_{sin})},$$

los coeficientes estimados a partir de la cuadratura de Gauss-Hermite son los siguientes:

$$\hat{\mu} = e^{(6.7124 - 0.00292 + 0.391 + \hat{b}_{sin})}.$$

En la Tabla 5 se recogen los valores numéricos de las estimaciones tanto en el predictor lineal como en la respuesta para los casos de este supuesto estimados con las tres funciones: *glmer*, *glmmPQL* de R y P<sub>ROC</sub> G<sub>GLIMMIX</sub> de SAS. Se observa como los valores obtenidos con las tres funciones son muy similares.

### 3.4. Cálculo de reservas

Pasamos a calcular las reservas con el paquete *lme4* de R, por años de ocurrencia y total, tanto con GLM como con GLMM. Los resultados se recogen en la Tabla 6.

Año de ocurrencia	Reserva estimada con GLM	Reserva estimada con GLMM
1	13259.75	11225.18
2	26625.94	32819.41
3	40039.04	39458.79
4	52899.81	50956.78
5	67861.68	50.019.03
Total	200686.22	184479.20

Tabla 6. Reservas por años de ocurrencia y total para el caso de estudio estimadas con GLM y GLMM.

Una vez calculadas las reservas, se estiman mediante *bootstrap* paramétrico las correspondientes distribuciones predictivas para el GLMM. El *bootstrap* se aplica al modelo estimado con la función *glmer* del paquete *lme4*. Son de especial ayuda para el proceso *bootstrap* las funciones *simulate* y *refit* del mismo paquete *lme4*. Dichas funciones simulan y ajustan los pagos bajo los supuestos paramétricos necesarios, nos referimos a Bates *et al.* (2014a, 2104b) para el detalle de ambas funciones. El número de muestras generadas es 1000. Aunque empíricamente se observó que un tamaño de 500 bastaba. Puesto que esta parte es de relevancia, en el Anexo 6 se incluyen las

instrucciones utilizadas en el proceso de estimación de las distribuciones predictivas.

Año de ocurrencia	Media	Desviación típica	Coeficiente de variación (en %)
1	11239.49	536.73	4.78
2	32812.72	1501.59	4.58
3	39512.28	1732.17	4.38
4	51670.39	2238.51	4.33
5	50679.97	2530.50	4.99
Total	185914.90	4066.20	2.19

Tabla 7. Estadísticos descriptivos de las distribuciones predictivas de las reservas por años de ocurrencia y total para el caso de estudio con GLMM.

Año de ocurrencia	Cuantil 50%	Cuantil 95%	Cuantil 99.5%
1	11204.90	12174.50	12738.98
2	32794.98	35365.23	36834.33
3	39425.53	42249.60	44205.12
4	51634.68	55419.94	58076.23
5	50606.58	55111.13	57364.76
Total	185824.80	192620.60	195923.90

Tabla 8. Cuantiles de las distribuciones predictivas de las reservas por años de ocurrencia y total para el caso de estudio con GLMM.

En la Tabla 7 se muestran estadísticos descriptivos de las distribuciones predictivas estimadas. En particular la media, la desviación típica y el coeficiente de variación, entendido como el tanto por ciento de error cometido y estimado con el cociente de la desviación típica de la distribución sobre la media. En la Tabla 8 se muestran varios cuantiles de las distribuciones. En especial es de resaltar el cuantil del 99.5 % que representa el VAR de las distribuciones. Por ejemplo, para un nivel de confianza del 99.5% los pagos futuros totales a los que deberá hacer frente la Compañía no

superaran el valor de 195923.90 unidades monetarias. Finalmente, en el Anexo 7 del trabajo se muestran los correspondientes histogramas que ofrecen información visual de las distribuciones estimadas.

#### **4. Conclusiones**

El objetivo principal del artículo es ilustrar el uso de los GLMM en aplicación al cálculo de provisiones de siniestros RBNS. Estos modelos incluyen en el predictor lineal una parte de efectos fijos y una parte de efectos aleatorios. Para su estimación se pueden utilizar métodos de ajuste basados en estadística frecuentista o bien bayesiana. En este trabajo se utiliza la estadística frecuentista aunque los programas informáticos utilizados también permiten realizar ajustes basados en estadística bayesiana.

Los siniestros RBNS parten de información individual más completa que la que se utiliza cuando se calculan reservas de siniestros IBNR a partir de un triángulo de desarrollo, ya que el siniestro ha sucedido y se tiene información del siniestro, del tomador del seguro y del entorno. Por tanto, a parte del año de ocurrencia y de desarrollo, pueden ser utilizados otros factores de riesgo. En este estudio se pretende generalizar el método Chain-Ladder clásico, por ello se asume distribución Poisson y enlace logarítmico en el modelo. Además, en la parte de efectos fijos del predictor lineal se incluye un factor de riesgo, y en la parte de efectos aleatorios, una variable de grupo que en este caso indica a qué siniestro corresponde cada pago individual.

Puesto que no se dispone de datos reales, se genera una base de datos con la que se ilustra la metodología de cálculo de reservas a partir de siniestros RBNS. Una vez ajustado el GLMM propuesto, se calculan las reservas por años de ocurrencia y total, y sus correspondientes distribuciones predictivas estimadas mediante *bootstrap* paramétrico. Estas distribuciones son de especial interés en el actual contexto de Solvencia II.

Como resultado global en la aplicación presentada, se comparan los ECM de aplicar el GLMM propuesto con los del GLM que reproduce Chain-Ladder a partir de un triángulo de desarrollo, obteniendo, como es de esperar cuando se dispone de información individual, un mejor ajuste con GLMM.

Se utiliza como base el paquete *lme4* de R, en especial las funciones *glmer*, que ajusta GLMM, y las funciones *simulate* y *refit* útiles para la programación de los pasos del *bootstrap*. Como aportación adicional al

carácter práctico del trabajo, se justan los GLMM con la función *glmPQL* del paquete *MASS* de R y con la función *PROC GLIMMIX* de SAS.

La referencia principal en cálculo de reservas de siniestros RBNS es Antonio *et al.* (2006). En ella se propone el tratamiento de datos individuales con un modelo que generaliza la curva de Hoerl al caso de aplicar un GLMM que incluye en la parte de efectos aleatorios un término independiente, los años de desarrollo y su logaritmo. Se asume la distribución Normal, la función de enlace logarítmica y se realiza todo el estudio desde un punto de vista de análisis bayesiano. Por tanto, el presente trabajo donde se propone una generalización del caso Chain-Ladder clásico en aplicación al cálculo de provisiones de siniestros RBNS, representa una aportación a las aplicaciones actuariales de GLMM. Cabe notar que el GLMM propuesto es únicamente uno de los posibles, pues se puede variar tanto el predictor lineal en sus efectos fijos y aleatorios, como la distribución del error o la función de enlace. Una propuesta interesante e interpretable es la inclusión de la variable “año de calendario” en el predictor lineal.

## Bibliografía

- Antonio, K., Beirlant, J. y T. Hoedemakers (2005). Discussion on ‘A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving’. *North American Actuarial Journal* 9:3, 143–145.
- Antonio, K., Beirlant, J., Hoedemakers, T. y R. Verlaak (2006). Lognormal mixed models for reported claims reserves. *North American Actuarial Journal* 10:1, 30–48.
- Antonio, K. y J. Beirlant (2007). Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics* 40:1, 58–76.
- Antonio, K. y J. Beirlant (2008). Issues in claims reserving and credibility: a semiparametric approach with mixed models. *The Journal of Risk and Insurance* 75:3, 643–676.
- Antonio, K. y E. Valdez (2012). Statistical concepts of a priori and a posteriori risk classification in insurance. *Advances in Statistical Analysis* 96, 187–224.
- Antonio, K. y H. J. Plat (2014). Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal* 7, 649–669.

Ayuso, M. y M.A. Santolino (2008). Prediction of individual automobile reported but not settled claim reserves for bodily injuries in the context of Solvency. *Revista de Métodos Cuantitativos para la Economía y la Empresa* 6, 23–41.

Bates, D., Maechler, M., Bolker, B. y S. Walker (2014a). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7. <http://CRAN.R-project.org/package=lme4>

Bates, D., Maechler, M., Bolker, B. and S. Walker (2014b). *Fitting linear mixed-effects models using lme4*. Vignettes of the package lme4. Acceso en septiembre de 2015.  
<http://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf>

Boj, E., Costa, T. y J. Espejo (2014). Provisiones técnicas por años de calendario mediante modelo lineal generalizado. Una aplicación con RExcel. *Anales del Instituto de Actuarios Españoles, Tercera Época*, 20, 83–116.

Booth, J. G. y J. P. Hobert (1998). Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association* 93:441, 262–272.

Boucher, J.P. y M. Denuit (2006). Fixed versus Random Effects in Poisson Regression Models for Claim Counts: Case Study with Motor Insurance. *ASTIN Bulletin* 36, 285-301.

Boucher, J.P., Denuit, M. y M. Guillén (2008). Models of Insurance Claim Counts with Time Dependence Based on Generalisation of Poisson and Negative Binomial Distributions. *Variance* 2:1, 135–162.

Charpentier, A. (2014). *Computational actuarial science, with R*. Chapman & Hall/CRC. Boca Raton.

Frees, E.W. (2010). *Regression modeling with actuarial and financial applications*. Cambridge University Press. Cambridge.

Frees, E. W., Derrig, R. A. y G. G. Meyers (2014). *Predictive modeling applications in actuarial science. Volume I: Predictive modeling techniques*. Cambridge University Press. Cambridge.

Gesmann, M., Murphy, D. y W. Zhang (2015). *ChainLadder: Statistical methods for the calculation of outstanding claims reserves in general insurance*. Package on CRAN. Version 0.2.2. Publicado 31-08-2015.  
<http://cran.r-project.org/web/packages/ChainLadder/index.html>

Godecharle, E. y K. Antonio (2014). Reserving by conditioning on markers of individual claims: a case study using historical simulation. *AFI Research Report*, AFI1492.

Jiang, J. (2007). *Linear and generalized linear mixed models and their Applications*. Springer Series in Statistics. Springer. New York.

Kroon, R. (2012). Individual reserving by detailed conditioning - A parametric approach. *Aenorm*.

Molenberghs, G. y G. Verbeke (2005). *Models for discrete longitudinal data*. Springer Series in Statistics. Springer. New York.

Pigeon, M., Antonio, K. y M. Denuit (2013). Individual loss reserving with the multivariate skew normal framework. *ASTIN Bulletin* 43:3, 399–428.

Pigeon, M., Antonio, K. y M. Denuit (2014). Individual loss reserving using paid-incurred data. *Insurance: Mathematics and Economics* 58, 121–131.

Verrall, R, Nielsen, J.P. y A. Jessen (2010). Prediction of RBNS and IBNR claims using claim amounts and claim counts. *ASTIN Bulletin* 40:2, 871–887.

Wikidot (2015). *GLMM: DRAFT r-sig-mixed-models FAQ*. Acceso en septiembre de 2015. <http://glmm.wikidot.com/faq#mixed-pkgs>

McNail, A.J., y J. Wendin (2005). Bayesian inference for generalized linear mixed models of portfolio credit risk. *Woking paper*, ETH Zürich.

Ripley, B., Venables, B., Bates, D., Hornik, K., Gbhardt, A. y D. Firth (2015). *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. R package version 7.3-41.  
<http://CRAN.R-project.org/package=MASS>

R Development Core Team (2016). *R: a language and environment for statistical computing*. Vienna. Austria.  
<http://www.R-project.org/>

SAS Institute Inc. (2009). *SAS/STAT® 9.2 User's Guide, Second Edition.* Cary, NC: SAS Institute Inc. Acceso en septiembre de 2015.  
[http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug\\_glimmix\\_sect026.htm](http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_glimmix_sect026.htm)

van Eeghen, J. (1981). Loss reserving methods. *Surveys of Actuarial Studies* 1. National Nederlanden.

Verbeke, G. y G. Molenberghs (2000). *Linear mixed models for longitudinal data.* Springer Series in Statistics. Springer. New York.

## Anexo 1. Construcción de datos a modo de ejemplo ilustrativo con R.

```
R> # Variable de grupo: Etiqueta de siniestro

R> sin<-
c(rep(1,6),rep(2,6),rep(3,6),rep(4,6),rep(5,6),rep(6,6),rep(7,6),rep(8,6),rep(9,6),rep(10,6),rep(
11,6),rep(12,6),
rep(1,5)+12,rep(2,5)+12,rep(3,5)+12,rep(4,5)+12,rep(5,5)+12,rep(6,5)+12,rep(7,5)+12,rep(8,5)
+12,rep(9,5)+12,rep(10,5)+12,rep(11,5)+12,rep(12,5)+12,
rep(1:4)+12*2,rep(2,4)+12*2,rep(3,4)+12*2,rep(4,4)+12*2,rep(5,4)+12*2,rep(6,4)+12*2,rep(7
,4)+12*2,rep(8,4)+12*2,rep(9,4)+12*2,rep(10,4)+12*2,rep(11,4)+12*2,rep(12,4)+12*2,
rep(1,3)+12*3,rep(2,3)+12*3,rep(3,3)+12*3,rep(4,3)+12*3,rep(5,3)+12*3,rep(6,3)+12*3,rep(7
,3)+12*3,rep(8,3)+12*3,rep(9,3)+12*3,rep(10,3)+12*3,rep(11,3)+12*3,rep(12,3)+12*3,
rep(1:2)+12*4,rep(2,2)+12*4,rep(3,2)+12*4,rep(4,2)+12*4,rep(5,2)+12*4,rep(6,2)+12*4,rep(7
,2)+12*4,rep(8,2)+12*4,rep(9,2)+12*4,rep(10,2)+12*4,rep(11,2)+12*4,rep(12,2)+12*4,
rep(1,1)+12*5,rep(2,1)+12*5,rep(3,1)+12*5,rep(4,1)+12*5,rep(5,1)+12*5,rep(6,1)+12*5,rep(7
,1)+12*5,rep(8,1)+12*5,rep(9,1)+12*5,rep(10,1)+12*5,rep(11,1)+12*5,rep(12,1)+12*5)

R> # Generación de los pagos de siniestros RBNS

R> cuantías<-
c(rgamma(1,1200,1),c(rgamma(5,800,1)),rgamma(1,1600,1),c(rgamma(5,1200,1)),rgamma(1,
1500,1),c(rgamma(5,1100,1)),rgamma(1,1530,1),c(rgamma(5,1110,1)),rgamma(1,2000,1),c(rg
amma(5,1500,1)),rgamma(1,1800,1),c(rgamma(5,1300,1)),rgamma(1,1900,1),c(rgamma(5,14
50,1)),rgamma(1,1300,1),c(rgamma(5,1000,1)),rgamma(1,1540,1),c(rgamma(5,1110,1)),rgam
ma(1,1700,1),c(rgamma(5,1250,1)),rgamma(1,900,1),c(rgamma(5,750,1)),rgamma(1,1000,1),c
(rgamma(5,700,1)),
rgamma(1,1200,1),c(rgamma(4,800,1)),rgamma(1,1600,1),c(rgamma(4,1200,1)),rgamma(1,15
00,1),c(rgamma(4,1100,1)),rgamma(1,1530,1),c(rgamma(4,1110,1)),rgamma(1,2000,1),c(rg
amma(4,1500,1)),rgamma(1,1800,1),c(rgamma(4,1300,1)),rgamma(1,1900,1),c(rgamma(4,145
0,1)),rgamma(1,1300,1),c(rgamma(4,1000,1)),rgamma(1,1540,1),c(rgamma(4,1110,1)),rgam
ma(1,1700,1),c(rgamma(4,1250,1)),rgamma(1,900,1),c(rgamma(4,750,1)),rgamma(1,1000,1),c(r
gamma(4,700,1)),
rgamma(1,1200,1),c(rgamma(3,800,1)),rgamma(1,1600,1),c(rgamma(3,1200,1)),rgamma(1,15
00,1),c(rgamma(3,1100,1)),rgamma(1,1530,1),c(rgamma(3,1110,1)),rgamma(1,2000,1),c(rg
amma(3,1500,1)),rgamma(1,1800,1),c(rgamma(3,1300,1)),rgamma(1,1900,1),c(rgamma(3,145
0,1)),rgamma(1,1300,1),c(rgamma(3,1000,1)),rgamma(1,1540,1),c(rgamma(3,1110,1)),rgam
ma(1,1700,1),c(rgamma(3,1250,1)),rgamma(1,900,1),c(rgamma(3,750,1)),rgamma(1,1000,1),c(r
gamma(3,700,1)),
rgamma(1,1200,1),c(rgamma(2,800,1)),rgamma(1,1600,1),c(rgamma(2,1200,1)),rgamma(1,15
00,1),c(rgamma(2,1100,1)),rgamma(1,1530,1),c(rgamma(2,1110,1)),rgamma(1,2000,1),c(rg
amma(2,1500,1)),rgamma(1,1800,1),c(rgamma(2,1300,1)),rgamma(1,1900,1),c(rgamma(2,145
0,1)),rgamma(1,1300,1),c(rgamma(2,1000,1)),rgamma(1,1540,1),c(rgamma(2,1110,1)),rgam
ma(1,1700,1),c(rgamma(2,1250,1)),rgamma(1,900,1),c(rgamma(2,750,1)),rgamma(1,1000,1),c(r
gamma(2,700,1)),
```

```
rgamma(1,1200,1),c(rgamma(1,800,1)),rgamma(1,1600,1),c(rgamma(1,1200,1)),rgamma(1,1500,
1),c(rgamma(1,1100,1)),rgamma(1,1530,1),c(rgamma(1,1110,1)),rgamma(1,2000,1),c(rgamma(1,
1500,1)),rgamma(1,1800,1),c(rgamma(1,1300,1)),rgamma(1,1900,1),c(rgamma(1,1450,1)),rgamm
a(1,1300,1),c(rgamma(1,1000,1)),rgamma(1,1540,1),c(rgamma(1,1110,1)),rgamma(1,1700,1),c(rg
amma(1,1250,1)),rgamma(1,900,1),c(rgamma(1,750,1)),rgamma(1,1000,1),c(rgamma(1,700,1)),
rgamma(1,1200,1),rgamma(1,1600,1),rgamma(1,1500,1),rgamma(1,1530,1),rgamma(1,2000,1),rg
amma(1,1800,1),rgamma(1,1900,1),rgamma(1,1300,1),rgamma(1,1540,1),rgamma(1,1700,1),rga
mma(1,900,1),rgamma(1,1000,1))
```

R> # Pagos resultantes (permiten reproducir cálculos al lector)

```
R> cuantías<- c (1240.26770544529, 791.263512713262, 757.274373942323, 776.93283906011,
801.059472263851, 820.068749294598, 1627.32100902704, 1175.14144817427,
1179.09073908729, 1148.87718088475, 1162.73819082741, 1177.35221419724,
1500.96319989126, 1126.04013933659, 1087.10900333521, 1125.23819084825,
1060.81900997326, 1145.70730538388, 1463.1601243852, 1178.42194013896,
1082.59705514095, 1099.93977983183, 1132.52687303961, 1144.37124370822,
2033.1014870379, 1483.99464135162, 1465.8690994624, 1516.19621370881,
1533.87598764139, 1546.99828919701, 1792.76139278995, 1258.43320160688,
1323.89668143899, 1341.60707504105, 1319.8500435879, 1257.0630653227,
1921.09016552402, 1460.46221129759, 1448.30140199864, 1458.37488103236,
1475.84146836542, 1443.25505659561, 1248.34250778551, 1015.11915334196,
949.44049311398, 964.835878340194, 1021.72880191413, 1044.22456300597,
1495.53859283876, 1069.63654658772, 1108.22083565169, 1125.3476314726,
1100.6027388135, 1063.27471375982, 1644.90735793537, 1220.95629715685,
1346.65727492106, 1254.48006572262, 1292.34052795215, 1253.25104913186,
862.384830522686, 738.395376177227, 766.463347564984, 768.130917145276,
741.765343412171, 733.399446174738, 1003.09498390901, 727.395145244588,
726.849482421583, 684.913260344928, 685.175588276744, 675.384848082673,
1132.33906248826, 805.286146667428, 825.982309898913, 810.672756983991,
801.804931544519, 1622.11086883177, 1193.92703023482, 1221.61049727777,
1200.58453206522, 1149.19566447248, 1522.21823796128, 1093.83674360905,
1110.97178409549, 1084.63342110947, 1139.76590593365, 1544.19501270378,
1073.70181539194, 1080.63391369498, 1100.74426412717, 1078.35737927894,
1987.44276669349, 1479.77842156389, 1501.87999091121, 1424.56219752113,
1482.47407652262, 1798.1701155558, 1287.02986165503, 1321.31953716589,
1250.74458121618, 1301.87240404348, 1942.07913898983, 1474.99411689037,
1525.3023030419, 1379.57194028532, 1482.66732403395, 1304.61044520538,
988.662351111758, 986.305197287558, 1015.69414195311, 1024.84056236939,
1439.66950789331, 1131.81824478759, 1067.49673272294, 1053.35936687893,
1097.40789627014, 1688.75572811056, 1251.50441160425, 1218.31162804245,
1228.59443324313, 1273.72536068485, 873.758465499315, 712.424317016125,
726.334351236943, 769.519912005692, 786.175920083742, 992.358918356816,
655.82305729027, 769.985708357776, 714.557109997774, 671.304405179365,
1191.87950651409, 840.916848663724, 863.745419931888, 763.59441880876,
1649.40102143215, 1197.89663794182, 1156.13581363954, 1223.40407911693)
```

```
R> cuantías<- c (cuantías, 1429.2938207771, 1131.97950871, 1096.60379366189,
1143.03147092032, 1590.89228208763, 1109.51861893971, 1042.32532491853,
1039.77595482766, 1922.05244235421, 1508.27140693863, 1518.4419006449,
1509.61798459329, 1742.35732598567, 1308.25141957961, 1299.47540756031,
1270.6540980771, 1912.91374261223, 1475.97609497905, 1432.56121435306,
1457.80912857645, 1244.28591007574, 999.684963609423, 918.088465746069,
1003.49779147536, 1569.27041322399, 1157.97471413212, 1105.32390646747,
1066.94512319135, 1644.09640906881, 1300.07165009297, 1263.72519710513,
1187.97937959044, 909.095342224074, 757.463540011688, 785.37774055471,
767.255126352393, 1002.36819846154, 730.13325942661, 678.222095933468,
641.24393921779, 1203.25248005575, 794.761365803983, 831.71976141236,
1634.31820237888, 1171.36305407163, 1203.23595990252, 1475.38272728459,
1071.30788782593, 1102.56278504669, 1595.84469956238, 1153.0579224304,
1112.27316231754, 2030.10782795757, 1485.52982487471, 1525.48606586429,
1771.8147325153, 1289.56088013926, 1350.88893833275, 1881.27544223004,
1469.23949778975, 1476.72082035363, 1316.25155460008, 1017.91250341286,
965.192711343622, 1570.23385037918, 1119.08305623769, 1128.86126769259,
1676.61190182022, 1261.24366701285, 1197.96629926087, 957.112172409445,
767.402629802176, 792.159237990612, 1004.83264762756, 650.130631476082,
680.206336351276, 1178.66359197432, 814.115796907195, 1627.97722235234,
1174.75844647372, 1499.73467706434, 1061.73688579565, 1532.91115932518,
1122.6126909805, 1935.45335454555, 1497.1790818509, 1809.69475377629,
1327.77851029431, 1901.47908662499, 1378.74932425675, 1256.9298458733,
994.100112951744, 1572.07300419831, 1096.68422825078, 1720.61105376643,
1194.85003047836, 932.679231079914, 741.990178306869, 1006.21159124262,
703.549636949321, 1189.27452529578, 1659.48218919746, 1500.18562120147,
1480.32226505619, 2046.78357319275, 1843.3312758126, 1966.61151871627,
1407.22180154908, 1594.50822123067, 1747.63514781166, 926.616885225907,
975.158561722517)
```

R> # Generación de los predictores

R> # Año de ocurrencia

```
R> ao<-c(rep(1,12*6),rep(2,12*5),rep(3,12*4),rep(4,12*3),rep(5,12*2),rep(6,12))
```

R> # Año de desarrollo

```
R> ad<-c(rep(1:6,12),rep(1:5,12),rep(1:4,12),rep(1:3,12),rep(1:2,12),rep(1,12))
```

R> # Factor de riesgo

R> fr<-rep(2,252)

```
R> for (k in 1:72) { if (mean(cuantías[sin==k])>=mean(cuantías)) fr[sin==k]=1 }
```

R> sin <- as.factor(sin)

R> ao <- as.factor(ao)

R> ad <- as.factor(ad)

R> fr <- as.factor(fr)

**Anexo 2.** Agregación de los datos del ejemplo y GLM ajustado a los datos agregados con la función *glm* del paquete *stats* de R.

```
R> triangle <- matrix (0, nrow =6, ncol =6)
R> for (i in 1:6) { for (j in 1:6) { triangle[i,j]=sum(cuantías[(ao==i)&(ad==j)]) } }

R> cuantagre <- rep(0, 21)
R> a=0
R> for (i in 1:6) { for ( j in 1:(6-i+1) ) { a=a+1; cuantagre [a]=triangle[i,j] } }

R> aoglm <- rep(1:6, times = 6:1) ; aoglm <- as.factor(aoglm)
R> adglm <- c(1:6,1:5,1:4,1:3,1:2,1); adglm; adglm <- as.factor(adglm)

R> glm1 <- glm(formula = cuantagre ~ aoglm + adglm, family = quasipoisson)

R> summary(glm1)

Call:
glm(formula = cuantagre ~ aoglm + adglm, family = quasipoisson)

Deviance Residuals:
    Min      1Q  Median      3Q     Max 
-0.94767 -0.55357 -0.05903  0.06813  2.32961 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  9.7934298  0.0057018 1717.610 < 2e-16 ***
aoglm2       -0.0033576  0.0061114  -0.549  0.5948  
aoglm3       -0.0003644  0.0066161  -0.055  0.9572  
aoglm4        0.0065854  0.0072542   0.908  0.3853  
aoglm5       -0.0027964  0.0082926  -0.337  0.7429  
aoglm6        0.0232535  0.0102277   2.274  0.0463 *  
adglm2       -0.3013815  0.0058913  -51.157 1.97e-13 ***
adglm3       -0.3000399  0.0064619  -46.432 5.17e-13 *** 
adglm4       -0.3099690  0.0072764  -42.599 1.22e-12 *** 
adglm5       -0.2955598  0.0085365  -34.623 9.57e-12 *** 
adglm6       -0.2975835  0.0114839  -25.913 1.68e-10 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 1.322037)

Null deviance: 6358.199 on 20 degrees of freedom
Residual deviance: 13.183 on 10 degrees of freedom
AIC: NA
```

**Anexo 3.** GLM con datos agregados y el paquete *ChainLadder* de R.

```
R> c0 <- c(triangle[1,1:6]); C0 <- cumsum(c0)
R> c1 <- c(triangle[2,1:5]); C1 <- c(cumsum(c1),NA)
R> c2 <- c(triangle[3,1:4]); C2 <- c(cumsum(c2),NA,NA)
R> c3 <- c(triangle[4,1:3]); C3 <- c(cumsum(c3),NA,NA,NA)
R> c4 <- c(triangle[5,1:2]); C4 <- c(cumsum(c4),NA,NA,NA,NA)
R> c5 <- c(triangle[6,1]); C5 <- c(c5,NA,NA,NA,NA,NA)
R> C <- matrix(c(C0,C1,C2,C3,C4,C5),ncol=6); C <- t(C);C <- as.triangle(C)
R> glm2<-glmReserve(C,mse.method = "formula"); glm2

      Latest Dev.To.Date Ultimate IBNR   S.E     CV
2  70675.46  0.8420215 83935.46 13260 203.8816 0.015375687
3  57560.88  0.6837274 84186.88 26626 296.1045 0.011120876
4  44734.90  0.5276966 84773.90 40039 391.5275 0.009778654
5  31082.52  0.3701070 83982.52 52900 513.1052 0.009699530
6  18337.13  0.2127299 86199.13 67862 739.3601 0.010895053
total 222390.90  0.5256512 423076.90 200686 1358.1806 0.006767690

> summary(glm2$model)

Call:
glm(formula = value ~ factor(origin) + factor(dev), family = fam,
  data = ldaFit, offset = offset)

Deviance Residuals:
    Min      1Q      Median      3Q      Max 
-0.94767 -0.55357 -0.05903  0.06813  2.32961 

Coefficients:
            Estimate Std. Error   t value Pr(>|t|)    
(Intercept) 9.7934298  0.0057018 1717.610 < 2e-16 ***
factor(origin)2 -0.0033576  0.0061114  -0.549  0.5948  
factor(origin)3 -0.0003644  0.0066161  -0.055  0.9572  
factor(origin)4  0.0065854  0.0072542   0.908  0.3853  
factor(origin)5 -0.0027964  0.0082926  -0.337  0.7429  
factor(origin)6  0.0232535  0.0102277   2.274  0.0463 *  
factor(dev)2   -0.3013815  0.0058913  -51.157 1.97e-13 ***
factor(dev)3   -0.3000399  0.0064619  -46.432 5.17e-13 ***
factor(dev)4   -0.3099690  0.0072764  -42.599 1.22e-12 ***
factor(dev)5   -0.2955598  0.0085365  -34.623 9.57e-12 ***
factor(dev)6   -0.2975835  0.0114839  -25.913 1.68e-10 ***

---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Tweedie family taken to be 1.322037)

Null deviance: 6358.199 on 20 degrees of freedom
Residual deviance: 13.183 on 10 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 3
```

**Anexo 4.** GLMM ajustado con la función *glmer* del paquete *lme4* de R.

Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite Quadrature, nAGQ = 15) ['glmerMod']

Family: poisson ( log )

Formula: cuantías ~ ao + ad + fr + (1 | sin)

AIC	BIC	logLik	deviance	df.resid
1026.3	1072.2	-500.1	1000.3	239

Scaled residuals:

Min	1Q	Median	3Q	Max
-8.6036	-0.6300	0.0881	0.7707	4.9611

Random effects:

Groups	Name	Variance	Std.Dev.
sin	(Intercept)	0.0202	0.1421

Number of obs: 252, groups: sin, 72

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	7.469516	0.046500	160.64	<2e-16 ***
ao2	0.037170	0.058646	0.63	0.526
ao3	0.028612	0.058627	0.49	0.626
ao4	-0.010770	0.059104	-0.18	0.855
ao5	-0.026824	0.059202	-0.45	0.650
ao6	-0.059577	0.060380	-0.99	0.324
ad2	-0.306537	0.005167	-59.33	<2e-16 ***
ad3	-0.303977	0.005653	-53.78	<2e-16 ***
ad4	-0.308554	0.006360	-48.51	<2e-16 ***
ad5	-0.287087	0.007417	-38.71	<2e-16 ***
ad6	-0.288672	0.009970	-28.95	<2e-16 ***
fr2	-0.383478	0.034938	-10.98	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '' 1

Correlation of Fixed Effects:

(Intr)	ao2	ao3	ao4	ao5	ao6	ad2	ad3	ad4	ad5	ad6
ao2	-0.638									
ao3	-0.637	0.505								
ao4	-0.679	0.503	0.502							
ao5	-0.678	0.502	0.501	0.510						
ao6	-0.711	0.494	0.493	0.511	0.511					
ad2	-0.048	0.000	0.002	0.000	0.002	0.037				
ad3	-0.050	0.000	0.002	0.000	0.025	0.038	0.391			
ad4	-0.051	0.000	0.002	0.020	0.027	0.039	0.349	0.363		
ad5	-0.052	0.000	0.021	0.024	0.030	0.040	0.297	0.309	0.316	
ad6	-0.054	0.024	0.028	0.030	0.034	0.041	0.221	0.230	0.235	0.240
fr2	-0.441	0.002	0.003	0.102	0.102	0.197	0.002	0.002	0.001	0.001

**Anexo 5.** GLMM ajustado con la función *glmmPQL* del paquete *MASS* de R.

Linear mixed-effects model fit by maximum likelihood  
 Data: NULL  
 AIC BIC logLik  
 NA NA NA

Random effects:  
 Formula: ~1 | sin  
     (Intercept) Residual  
 StdDev: 0.1376462 1.818071

Variance function:  
 Structure: fixed weights  
 Formula: ~invwt

Fixed effects: cuantías ~ ao + ad + fr

	Value	Std.Error	DF	t-value	p-value
(Intercept)	7.510655	0.04657263	175	161.26759	0.0000
ao2	-0.002525	0.05846881	65	-0.04319	0.9657
ao3	-0.007898	0.05870392	65	-0.13453	0.8934
ao4	-0.056525	0.05912566	65	-0.95602	0.3426
ao5	-0.070856	0.05955485	65	-1.18975	0.2385
ao6	-0.108062	0.06122578	65	-1.76497	0.0823
ad2	-0.308137	0.00961371	175	-32.05184	0.0000
ad3	-0.304190	0.01051370	175	-28.93269	0.0000
ad4	-0.313926	0.01184543	175	-26.50182	0.0000
ad5	-0.298971	0.01384339	175	-21.59666	0.0000
ad6	-0.300995	0.01861542	175	-16.16910	0.0000
fr2	-0.386836	0.03536517	65	-10.93832	0.0000

Correlation:

	(Intr)	ao2	ao3	ao4	ao5	ao6	ad2	ad3	ad4	ad5	ad6
ao2	-0.628										
ao3	-0.629	0.498									
ao4	-0.670	0.494	0.494								
ao5	-0.669	0.491	0.491	0.499							
ao6	-0.700	0.477	0.478	0.495	0.495						
ad2	-0.090	0.000	0.004	0.001	0.004	0.068					
ad3	-0.093	0.000	0.004	0.000	0.046	0.070	0.391				
ad4	-0.095	0.000	0.004	0.038	0.050	0.072	0.348	0.362			
ad5	-0.096	0.000	0.039	0.044	0.055	0.073	0.296	0.308	0.314		
ad6	-0.099	0.045	0.052	0.055	0.063	0.075	0.220	0.229	0.233	0.238	
fr2	-0.442	0.000	0.002	0.101	0.102	0.198	0.004	0.003	0.003	0.002	0.001

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-4.66020536	-0.34905044	0.04658667	0.42187435	2.74206569

Number of Observations: 252  
 Number of Groups: 72

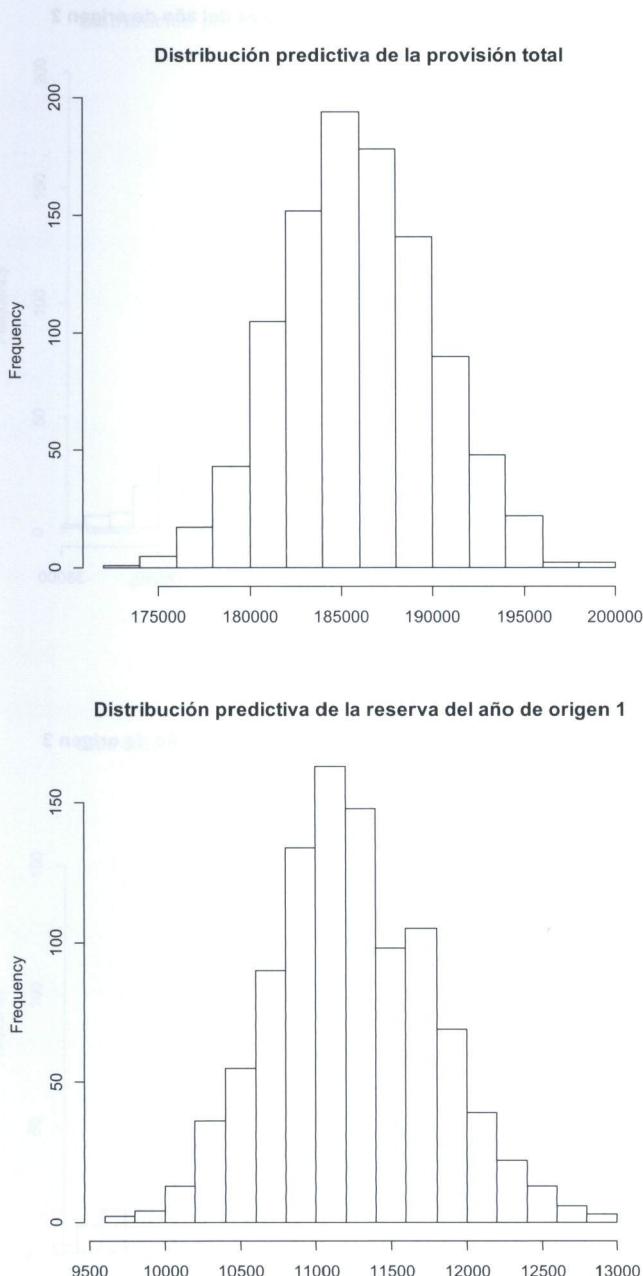
**Anexo 6.** Proceso *bootstrap* para la estimación de las distribuciones predictivas de las reservas total y por año de ocurrencia u origen utilizando los datos individuales del ejemplo para GLMM con R.

```
R> df.resid<-239
R> Prs.resid<-residuals(glmm1,type="pearson") # Pearson
R> phi.P<-sum(Prs.resid^2)/df.resid; phi.P
[1] 2.510621

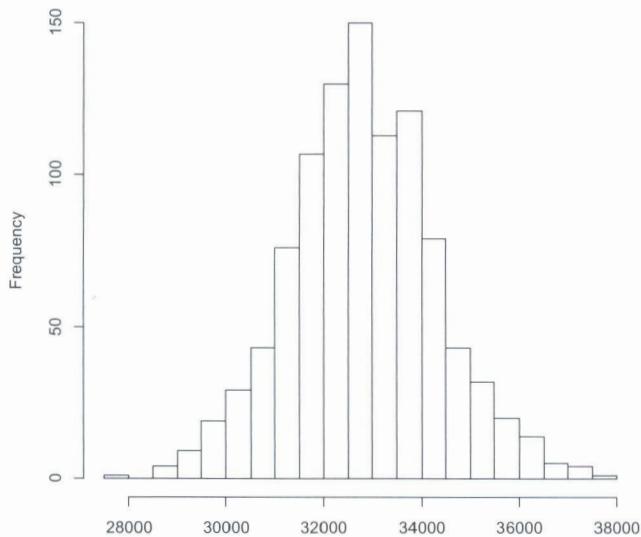
R> Dev.resid<-residuals(glmm1,type="deviance") # Deviance
R> phi.D<-sum(Dev.resid^2)/df.resid; phi.D
[1] 2.594906

R> nBoot<-1000
R> Provtotalb<-rep(0,nBoot)
R> Provorig1b<-rep(0,nBoot)
R> Provorig2b<-rep(0,nBoot)
R> Provorig3b<-rep(0,nBoot)
R> Provorig4b<-rep(0,nBoot)
R> Provorig5b<-rep(0,nBoot)
R> set.seed(1111) # Si se desean reproducir los cálculos del ejemplo del artículo
R> for (boot in 1:nBoot) {
  cuantíasb <- simulate(glmm1, nsim = 1,family=poisson); # función "simulate"
  glmm1b <- refit(glmm1,cuantíasb); # función "refit"
  cuantíaspred <- rep(0,180);
  a=0;
  for (i in 2:6) {
    for (j in (6-i+2):6) {
      aof<-i;
      aof<-as.factor(aof);
      adf<-j;
      adf<-as.factor(adf);
      sinf<-1:12+(i-1)*12;
      sinf<-as.factor(sinf);
      frf<-fr[1:12+(i-1)*12];
      frf<-as.factor(frf);
      newdata<-data.frame(ao=aof,ad=adf,sin=sinf,fr=frf);
      cuantíaspred[(a+1):(a+12)] <- predict(glmm1b,newdata,type="response");
      a<-a+12 }
    }
  for (o in 1:180) { cuantíaspred[o]<-phi.P*rpois(1, cuantíaspred[o]/phi.P) }
  Provtotalb[boot]<-sum(cuantíaspred);
  Provorig1b[boot]<-sum(cuantíaspred[1:12]);
  Provorig2b[boot]<-sum(cuantíaspred[13:36]);
  Provorig3b[boot]<-sum(cuantíaspred[37:72]);
  Provorig4b[boot]<-sum(cuantíaspred[73:120]);
  Provorig5b[boot]<-sum(cuantíaspred[121:180]) }
```

**Anexo 7.** Histogramas de las distribuciones predictivas de las reservas por años de ocurrencia y total con GLMM.



Distribución predictiva de la reserva del año de origen 2



Distribución predictiva de la reserva del año de origen 3

