Anales del Instituto de Actuarios Españoles, 3ª Época, 22, 2016, 1-24

Cuantificación del riesgo para la tarificación en seguros de automóvil

Alemar Padilla[12], Catalina Bolancé[1], Montserrat Guillen[1]

Resumen

El cálculo del precio del seguro del automóvil se centra en identificar los factores que determinan una mayor o menor siniestralidad, sin embargo, una concepción integral del riesgo tendrá en cuenta también la posibilidad de que el asegurado no renueve su póliza. Esta visión holística se basa en la concepción exigida en los requerimientos de capital de solvencia cuyo enfoque contempla riesgo de suscripción y de caída de cartera. En este trabajo mostramos una forma de cuantificar el riesgo que aporta cada asegurado con la finalidad de que se pueda identificar su contribución al riesgo total de la cartera. Los resultados obtenidos se ilustran con una muestra de asegurados del mercado español y muestran la importancia de la elección de los factores de tarificación, la medida de riesgo y el modelo de comportamiento aleatorio para las pérdidas.

Palabras clave: Prima, seguros generales, retención de clientes, caída de cartera, Solvencia II, riesgo de suscripción no vida

Abstract

Calculation of the price of car insurance focuses on identifying the factors that determine a high or low accident rate, however, an integrated concept of risk must consider the possibility to avoid renewing the insurance policy. This holistic view is based on the design imposed by solvency capital requirements, which consider underwriting risk and lapse risk. In this paper we show a form of calculating the risk contributed by each insured with the aim to quantify and identify their contribution to the total risk of the portfolio. The results are illustrated with a sample of insureds in the Spanish market and they show the importance of choosing the ratemaking factors, the risk measure and the stochastic behavior model for losses.

Keywords: Premium, general insurance, customer retention, portfolio lapse, Solvency II, non-life underwriting risk.

---

[1] Dpto. Econometría, Estadística y Economía Aplicada, Riskcenter-IREA; Universitat de Barcelona, Av. Diagonal, 690, 08034 Barcelona.

[2] Zurich España, Vía Augusta, 200, 08021 Barcelona.

1. Introducción

La principal misión de las compañías de seguros es gestionar el riesgo de pérdidas que les transfieren los asegurados, especialmente en lo que concierne a la severidad de los accidentes en los seguros de no-vida. El riesgo de suscripción se define como la posibilidad de que las primas recibidas no sean suficientes para cubrir las pérdidas que se producen por la declaración de siniestros. En general, y debido a la influencia de la normativa regulatoria, se establece una perspectiva de un año, lo que significa que las decisiones sobre la forma como se analizan las primas y las pérdidas consideran un período de doce meses.

Para simplificar el concepto de pérdidas y ganancias en un contrato de seguro, consideremos un cliente que compra un seguro de automóvil por un año. La compañía de seguros tiene que cubrir los gastos de administración, los márgenes de solvencia para atender a las exigencias normativas, la publicidad, los sistemas de comercialización y de gestión, incluyendo la tecnología informática. Una aseguradora que vende seguros para vehículos de motor tendrá miles de contratos de un año de duración y deberá pagar por las compensaciones a aquellos clientes que tienen un accidente cubierto por su correspondiente póliza.

El concepto de *pricing* se refiere a la fijación de precios y por lo tanto es tan sólo una parte de la producción de un contrato de seguro. En este punto, se entiende que las estrategias de cálculo de precios son esenciales para garantizar la estabilidad y la solidez financiera de una compañía de seguros. El precio depende de las características del contrato y del cliente. Por lo tanto, en todo lo que sigue, vamos a suponer que el contrato de seguro ya está diseñado y que no tenemos que contabilizar información adicional, como los gastos generales de la empresa, nos limitaremos a concentrarnos en la parte del precio que tiene que ver con las circunstancias alrededor del objeto asegurado, el tomador y el asegurado que cubre la póliza. Vamos a considerar sólo el seguro de automóviles, como un ejemplo concreto de las pólizas más habituales, dentro del ramo de los seguros generales, a nivel de las familias españolas.

En una evaluación global del riesgo de un contrato de seguro, y aparte del precio de la prima, se debe tener en cuenta la vinculación del cliente con su entidad. Dicha actitud es, de lejos, una de las prioridades más importantes en la mayoría de las compañías de seguros. Cuando un cliente decide cancelar un contrato de seguro para buscar otra compañía, la entidad original pierde la oportunidad de generar beneficios futuros.

Esta parte del riesgo se conoce como riesgo de caída de cartera y se considera explícitamente en la regulación de las exigencias de solvencia de las compañías aseguradoras en Europa.

Hasta ahora sólo unos pocos autores han considerado los riesgos derivados de la suscripción y la caída de cartera conjuntamente, a pesar del hecho de que están muy conectados. Como se demuestra en el trabajo de Guelman (2014) entre otros, cuanto mayor es el precio, mayor es la probabilidad de que el cliente abandone la entidad, y lo contrario ocurre si el precio disminuye. Además se ha analizado que el efecto es heterogéneo en función de algunas variables observadas. Por ejemplo, en el artículo de Guelman (2014) se estudia un caso del mercado canadiense y se encuentra que los asegurados jóvenes tienen una mayor sensibilidad al precio, por lo que reaccionan antes que los asegurados mayores con la cancelación de su póliza si se produce un aumento de la prima. Este análisis supone que a jóvenes y mayores se les aplica el mismo incremento de prima y el resto de circunstancias del contrato son iguales entre ambos tipos de clientes.

Una prima baja es un incentivo para que un cliente permanezca vinculado a su compañía. Sin embargo, si una entidad decidiera aplicar una determinada política de descuentos de forma masiva destinada a un excesivo número de clientes, entonces el flujo de ingresos podría llegar a ser insuficiente para cubrir las pérdidas y la entidad aseguradora podría sufrir tensiones de solvencia, e incluso antes de llegar a este extremo, ver incrementados significativamente sus requerimientos de capital.

Así pues, el precio relativo a la naturaleza del riesgo y la renovación de la póliza son dos fenómenos que no pueden desvincularse. En otras palabras, en el análisis del riesgo para la tarificación es fundamental analizar conjuntamente el comportamiento asociado al riesgo de suscripción, viendo las reclamaciones esperadas e incluyendo su severidad y, a la vez, el comportamiento derivado del riesgo de caída o, lo que es lo mismo, la probabilidad de "renovar la póliza de seguro". Algunos ejemplos de este tipo de análisis se muestran en (Thuring *et al.*, 2012) y (Thuring *et al.*, 2013).

Partiendo de la idea de evaluar la fijación de precios y retención del cliente simultáneamente, en este artículo se pretende proponer una medida del riesgo para cuantificar la aportación global de cada asegurado a las pérdidas que pueden generar en una cartera, sabiendo que el cliente puede haber contratado una o más pólizas con la empresa, posiblemente de distintos ramos. La puntuación depende de las características personales y también del comportamiento observado a lo largo del tiempo para cada cliente y sus correspondientes pólizas.

La idea de la modelización conjunta está inspirada en el hecho de que los clientes podrían querer abarcar más de un riesgo a la vez, ya sea por sí mismos o para otros miembros de su familia.

2. Antecedentes

En la literatura sobre la tarificación en seguros de no-vida, la mayoría de contribuciones focalizan su atención sobre cómo se pueden calcular las primas de seguros. También hay algunos artículos que enfocan el análisis en la satisfacción del cliente y la rotación de la cartera en el contexto de los seguros.

Entre las numerosas publicaciones sobre la modelización estadística para el riesgo individual centrado en las reclamaciones de siniestros podemos citar a Almer (1963), para la tarificación ver Hsiao *et al.* (1990), Brown (2007) y Werner (2010). La utilización de los modelos lineales generalizados se presenta en McCullagh (1989) y sus aplicaciones en las ciencias actuariales se estudian en Haberman (1996), De Jong (2008) y Fahrmeir (2013).

Tradicionalmente, desde la perspectiva de la investigación actuarial, el precio del seguro se ha basado en la modelización del número de siniestros ocurridos en el período contractual. En algunos trabajos más recientes se analizan algunas distribuciones del número de siniestros en el contexto de los modelos lineales generalizados, por ejemplo Bolancé *et al.* (2003), Pinquet *et al.* (2001), Bolancé *et al.* (2008), Denuit *et al.* (2007) y Boucher *et al.* (2007 y 2009).

Como hemos mencionado antes, la fidelización y retención de los clientes representa un factor importante para la estabilidad del balance de las compañías de seguros y, por ello, su estudio aborda el impacto que tienen las cancelaciones de pólizas en el riesgo asumido por la empresa (Guillen *et al.*, 2008). Varios estudios se han desarrollado en torno a este tema, por ejemplo Hallowell (1996) aporta evidencias sobre la doble relación entre la satisfacción del cliente y vinculación a la entidad aseguradora, así como la fidelización del cliente y rentabilidad de dicha compañía. Se ha demostrado que los aumentos en la satisfacción pueden mejorar la rentabilidad. Además, según Beerli *et al.* (2004), se puede utilizar un modelo de ecuaciones estructurales para características personales y también del comportamiento observado a lo largo del tiempo para cada cliente y sus correspondientes pólizas.

La idea de la modelización conjunta está inspirada en el hecho de que los clientes podrían querer abarcar más de un riesgo a la vez, ya sea por sí mismos o para otros miembros de su familia.

2. Antecedentes

En la literatura sobre la tarificación en seguros de no-vida, la mayoría de contribuciones focalizan su atención sobre cómo se pueden calcular las primas de seguros. También hay algunos artículos que enfocan el análisis en la satisfacción del cliente y la rotación de la cartera en el contexto de los seguros.

Entre las numerosas publicaciones sobre la modelización estadística para el riesgo individual centrado en las reclamaciones de siniestros podemos citar a Almer (1963), para la tarificación ver Hsiao *et al.* (1990), Brown (2007) y Werner (2010). La utilización de los modelos lineales generalizados se presenta en McCullagh (1989) y sus aplicaciones en las ciencias actuariales se estudian en Haberman (1996), De Jong (2008) y Fahrmeir (2013).

Tradicionalmente, desde la perspectiva de la investigación actuarial, el precio del seguro se ha basado en la modelización del número de siniestros ocurridos en el período contractual. En algunos trabajos más recientes se analizan algunas distribuciones del número de siniestros en el contexto de los modelos lineales generalizados, por ejemplo Bolancé *et al.* (2003), Pinquet *et al.* (2001), Bolancé *et al.* (2008), Denuit *et al.* (2007) y Boucher *et al.* (2007 y 2009).

Como hemos mencionado antes, la fidelización y retención de los clientes representa un factor importante para la estabilidad del balance de las compañías de seguros y, por ello, su estudio aborda el impacto que tienen las cancelaciones de pólizas en el riesgo asumido por la empresa (Guillen *et al.*, 2008). Varios estudios se han desarrollado en torno a este tema, por ejemplo Hallowell (1996) aporta evidencias sobre la doble relación entre la satisfacción del cliente y vinculación a la entidad aseguradora, así como la fidelización del cliente y rentabilidad de dicha compañía. Se ha demostrado que los aumentos en la satisfacción pueden mejorar la rentabilidad. Además, según Beerli *et al.* (2004), se puede utilizar un modelo de ecuaciones estructurales para mostrar cómo influyen en la satisfacción y el precio de la prima las preferencias del asegurado. También hay una gran cantidad de bibliografía en torno al estudio de los comportamientos de fidelización y de abandono de clientes en diferentes áreas (por ejemplo, Bolton (2000) y Verhoef (2003) y ejemplos similares en los que se utiliza un proceso de modelización y predicción (ver también Smith *et al.*, 2000, Yeo *et al.*, 2001, Kim *et al.*, 2005, Mozer *et al.*, 2000, Au *et al.*, 2003, Vafeiadis *et al.*, 2015, Kumar, 2013, Ekinci, 2015, Günther *et al.*, 2014, Jahromi *et al.*, 2016 y Neslin *et al.*, 2006).

Brockett *et al.* (2008) estudiaron el comportamiento de los clientes asegurados que tenían más de una póliza, posiblemente de tipo diferente, contratadas con la misma compañía y destacaron la importancia de identificar la cancelación de una póliza para poder reaccionar rápidamente y evitar la cancelación del resto de contratos.

La revisión de la literatura muestra que los factores que afectan a la siniestralidad no siempre coinciden con los que afectan a la probabilidad de renovación de las pólizas de un mismo cliente, incluso, idénticos factores puede influir de forma distinta en ambos comportamientos (siniestralidad y fidelidad). Por ejemplo, en los seguros del automóvil, los asegurados de mayor edad y por lo tanto con mayor experiencia en la conducción, suelen tener una menor siniestralidad y, a su vez, presentan mayores índices de fidelidad o mayor probabilidad de renovar la póliza en la misma entidad en la que están asegurados. Por el contrario, los jóvenes suelen tener una siniestralidad alta y menor probabilidad de renovar la póliza en la misma entidad. El número de pólizas contratadas no parece afectar a la siniestralidad, en cambio sí es un indicador de una mayor vinculación y satisfacción con la entidad aseguradora y, por lo tanto, suele asociarse a una menor probabilidad de cancelación de alguna póliza.

Las contribuciones académicas existentes se concentran generalmente en una sola línea de negocio. Ello significa que se analizan, por ejemplo, únicamente pólizas del seguro de automóvil o únicamente del seguro del hogar, pero no ambos. Por lo tanto, lo habitual es encontrar trabajos con modelos estadísticos predictivos separados para cada ramo.

Análogamente, la mayoría de los artículos sólo estudian los modelos de fijación de precios o modelos de retención, pero no los dos a la vez. Sólo existen unas pocas excepciones como es el caso de Guelman (2014) o Thuring *et al.* (2013). Esos artículos mencionan el concepto de elasticidad de los precios, que significa el cambio de la demanda de seguro en función de los cambios de precios.

Para el análisis de los modelos de fijación de precios y de retención de clientes, en forma separada, suponemos que las dos variables dependientes (pérdidas agregadas y renovación de la póliza) son independientes entre sí, pero están influenciadas por variables exógenas observadas que pueden aparecer en ambos modelos.

En los antecedentes analizados no hemos hallado estudios de casos en los que se integren los modelos de fijación de precios y modelos de retención de pólizas para dos líneas de negocio diferentes (por ejemplo, automóviles y seguro del hogar). Algunos artículos sí se centran en el valor de un cliente de seguros, pero no estudian una puntuación de riesgo. No parece que se haya estudiado qué medida de riesgo es la más adecuada y la forma de modelarla. El caso más simple es utilizar la medida basada en el cuantil o valor en riesgo (VaR), pero esta no es la única medida posible.

Al final del proceso de modelización propuesto en este artículo, utilizamos como medida para cuantificar el riesgo individual de cada uno de los asegurados, el VaR, pero es posible emplear otras medidas como el valor en riesgo de la cola (Artzner *et al.*, 1999), u otro tipo de medidas también basadas en los cuantiles de las pérdidas (véase Dowd, 2006) o incluso aquellas que aprovechan, por ejemplo, la existencia de una nueva familia de medidas de riesgo más generales como las GlueVar (Belles-Sampera *et al.*, 2014).

3. Datos

Nuestros datos fueron proporcionados por una compañía de seguros española e incluyen información de pólizas a nivel individual relevantes para este estudio.

Esta base de datos es ilustrativa y corresponde a una muestra no aleatoria de lo que sucedió en la entidad a nivel de cartera de automóviles y hogar durante el período 2010-2014 y se da en dos archivos de datos diferentes: el primero está compuesto por registros de las pólizas de automóvil y el segundo archivo es una relación de dichas pólizas con sus respectivas reclamaciones.

Tabla 1: Algunos factores de riesgo utilizados en el estudio.

En relación con	Variable
Asegurado	Sexo
	Edad
	Número de pólizas vigentes
	Número de pólizas vigentes en otros ramos
Pólizas	Estado de la póliza (C= cancelada, V= vigente)
	Garantías cubiertas
	Prima pagada más reciente
	Prima anterior
	Descuento aplicado o nivel de Bonus-Malus
Vehículo	Tipo de vehículo
	Antigüedad
	Edad del conductor habitual
	Existencia de conductor ocasional (Sí o No)
	Potencia
Siniestros	Número total de siniestros
	Fechas de ocurrencia
	Tipo de siniestros
	Coste agregado de los siniestros

Fuente: Muestra de asegurados en el ramo de autos 2010-2014

Los archivos de pólizas proporcionan características acerca de la póliza, el cliente y el vehículo asegurado. El conjunto de datos de reclamaciones facilita un registro de cada siniestro si este se produjo durante el período de observación. Cada conjunto de datos contiene una variable de identificación de póliza, lo que permite enlazar las dos bases de datos de forma simultánea, por lo que el resultado será un conjunto más grande de datos relacionados entre sí. Algunas variables utilizadas en los modelos posteriores se presentan en la Tabla 1.

El proceso de modelización se inicia vinculando la base de datos de las pólizas y su correspondiente siniestralidad. En cada renovación se analiza si el tomador renovó o no cada uno de los contratos de auto que tenía vigentes, o en su caso, si suscribió algún contrato nuevo.

4. Planteamiento de los modelos

Probabilidades de renovación

Sea Y_{ij} una variable aleatoria de respuesta binaria, cuya observación y_{ij} corresponde a la decisión del tomador i sobre la renovación (valor igual a 1) o no (valor igual a 0) de la póliza j. La variable toma el valor $y_{ij} = 1$ con probabilidad p_{ij} y el valor $y_{ij} = 0$ con probabilidad $(1 - p_{ij})$, donde $p_{ij} \in (0,1)$ y se interpreta como la probabilidad de renovar el contrato j-ésimo para el cliente i-ésimo.

Nuestro objetivo es modelizar la probabilidad de renovación en función de las k variables que conforman el vector columna X_{ij}, donde el primer elemento es igual a 1 y el resto de los elementos contiene información sobre las características observables relacionadas con los asegurados y sus pólizas.

Frecuencia y severidad

En ciencias actuariales la modelización de las pérdidas o el coste del conjunto de las reclamaciones que se asocian a una anualidad de una póliza se suele dividir en dos componentes: la frecuencia, que corresponde a la ocurrencia de un determinado número de accidentes declarados y la cantidad total pagada al asegurado en compensación por cada uno de ellos, que se denomina la gravedad o la severidad del siniestro. Con frecuencia, el análisis del número de siniestros se realiza desde una perspectiva de modelo de frecuencia / gravedad. En este sentido, las distribuciones de ambos factores se pueden estimar de manera aislada, tomando el número de siniestros por un lado y su severidad por otro, o bien se pueden estudiar simultáneamente como veremos a continuación (para más detalles ver Frees *et al.*, 2009, 2013 y 2016).

Modelización del coste agregado de los siniestros

El riesgo de accidentes vinculados a una póliza de un determinado producto se descompone por tipologías de coberturas, por lo que es habitual sumarlas para ofrecer paquetes a los clientes con diferentes tipos de garantías. Habitualmente, cada garantía tiene una o más coberturas asociadas, por lo que, cuando se produce un accidente, los gastos que se generan se clasifican en dichas sub-categorías de riesgo.

Al analizar la información de una cartera de pólizas de seguros, los casos se dividen en dos grupos; por un lado, se encuentran aquellas pólizas que no han sufrido ningún siniestro y por lo tanto no se ha efectuado ningún pago de reclamaciones y, por otro lado, se tienen los gastos positivos asociados con el pago de la indemnización correspondiente en las pólizas que sí han generado siniestralidad para la entidad aseguradora. Nótese que en algunas ocasiones y debido a la existencia de convenios entre entidades, la declaración de un siniestro puede acarrear un cobro recibido de otra entidad, por lo que la pérdida asociada resulta ser negativa. Para simplificar la exposición no se consideran estos casos y por lo tanto todas las cuantías son positivas.

En este trabajo únicamente analizamos siniestros que tienen que ver con las coberturas de responsabilidad civil para terceros (lesiones corporales, daños a la propiedad, atención médica y farmacéutica) y consideramos sólo los casos en los que los asegurados tienen la culpa del siniestro declarado. En este sentido, decidimos agregar cada reclamación de información por medio de la suma de las coberturas afectadas con el fin de definir la severidad agregada.

En lo que sigue vamos a emplear la siguiente notación relativa al asegurado *i*-ésimo y su póliza *j*-ésima. Sea N_{ij} el número de siniestros declarados durante una anualidad y L_{ij} la suma de las pérdidas causadas por dichos siniestros asociados a la mencionada póliza *j*.

Modelos para variables con truncamiento inferior en el cero

Como se acaba de exponer, una de las características de las variables relacionadas con el número de siniestros y su severidad es la presencia de un elevado número de ceros debido a la ausencia de accidentes declarados en algunas (generalmente la mayor parte) de las pólizas. Para tratar este comportamiento aleatorio se necesitan modelos que permitan manejar una mezcla de distribución discreta y continua, como por ejemplo los modelos: Tobit (ver Lin, 2007), Tweedie (Jørgensen, 1994) y los modelos en dos partes (para más detalles ver Frees et al., 2016).

Comenzamos nuestro análisis mediante la modelización de la cuantía agregada de los siniestros, es decir, ajustando los modelos para los valores positivos de la variable aleatoria "pérdidas" o también el ajuste de una mezcla utilizando un modelo frecuencia-cuantía conocido como el modelo basado en la distribución Tweedie. Dicho modelo predictivo permite que se pueda efectuar una estimación de cuál será la pérdida esperada que va a generar una póliza dadas las circunstancias conocidas del contrato (tipo, coberturas, etc.), del asegurado (sexo, edad, etc.) y del objeto asegurado si es vehículo, su antigüedad y potencia, por ejemplo y si es una vivienda, su superficie y su tipología, etc.

5. Definición del riesgo del tomador de las pólizas

A partir de la mezcla de las probabilidades de renovación y el coste agregado de las reclamaciones de una póliza, se define la siguiente variable aleatoria:

$$Y_{ij}^{*} = \begin{cases} L_{ij} - \rho_{ij} & si \quad y_{ij} = 1 \\ b \cdot \rho_{ij} & si \quad y_{ij} = 0 \end{cases}$$

donde L_{ij} es la pérdida de la póliza *j*-ésima y el valor y_{ij} indica que se ha renovado la póliza si vale 1, o bien si ha sido cancelada si vale 0. Además, se ha incorporado la cuantía de la prima devengada por el asegurado, que se denota por ρ_{ij}. Además, el factor denotado por b, denota una proporción de coste que puede asociarse a la pérdida generada por la cancelación de una póliza en relación a la cuantía de la prima que hubiera pagado.

Hasta el momento, tenemos una forma de cuantificar el riesgo individual por póliza para cada cliente a partir de la expresión anterior que denota la pérdida que se genera según se renueve o se cancele la póliza. Sin embargo, como hemos mencionado en la introducción, estamos interesados en proponer una nueva medida de riesgo que permita cuantificar el riesgo asegurado por el cliente como una puntuación total, teniendo en cuenta que el mismo individuo podría tener una o varias pólizas en la misma entidad aunque sean de diferentes ramos.

La idea anterior nos lleva a definir el "riesgo general del asegurado (RgA)" a partir de la siguiente variable aleatoria como:

$$RgA_i = \sum_{j=1}^{m_i} Y_{ij}^{*}$$

siendo $\mathrm{m_i}$ el número total de pólizas del asegurado *i*-ésimo.

En este sentido, el concepto RgA permite reflejar la hipótesis de que la experiencia observada en las pérdidas de las pólizas de automóviles pertenecientes al mismo tomador está correlacionada. En general, los miembros de una familia, para los que el tomador del seguro suele ser el cabeza de familia, comparten características comunes y probablemente tienen actitudes de riesgo similares, por lo que no es de extrañar que se comporten de manera similar.

Aunque sea de forma indirecta, la expresión del RgA indica la satisfacción del cliente a través de la posibilidad de renovación o cancelación, que tiene una relación directa con los flujos de ingresos futuros (ver Fornell, 1992) y puede ser afectada por muchas circunstancias exógenas (Fornell, 1988). De hecho, varios estudios combinados con la experiencia de casos reales muestran que un cliente insatisfecho puede cancelar varias pólizas de forma simultánea, por lo tanto, lo que afecta a una póliza puede afectar al resto de las pólizas que tenga contratadas el mismo asegurado, e incluso su familia.

El riesgo general asegurado (RgA) representa una forma de analizar el valor del cliente y permite cuantificarlo, por lo que es un índice que puede resultar clave para mejorar las decisiones de gestión, ya que permite a la compañía entender, clasificar y ser capaz de cuidar a los clientes más valiosos, lo que aumenta inevitablemente la rentabilidad de su negocio.

Bajo este escenario, se aborda el análisis de RgA desde dos perspectivas diferentes. La primera de ellas considera la independencia entre las pólizas contratadas en los ramos de automóviles y hogar, y la segunda consiste en una modelización conjunta de su comportamiento conjunto.

Por otra parte, a priori, consideramos que la probabilidad de renovación de cada póliza, que denotamos por p_{ij} es conocida a través de algún modelo de retención de clientes que ya exista en la entidad. Además, el porcentaje de beneficio b se puede predeterminar. En consecuencia, se puede establecer el comportamiento aleatorio del RgA simplificando las componentes estocásticas de las pérdidas (L_{ij}, variable aleatoria que toma valores no-negativos) y la renovación ($\mathrm{Y_{ij}}$, variable aleatoria binaria que únicamente admite dos posibilidades: renovar o cancelar).

Sin embargo, el reto del modelo RgA es diseñar una optimización de precios para las primas con todos los componentes juntos, lo que plantea un problema de optimización que no es fácil de resolver. En este caso, el uso de la metodología de las cópulas puede ser una alternativa apropiada.

Análisis del riesgo global asegurado (RgA)

Deseamos estudiar la distribución de $\mathrm{RgA_i}$ para cada asegurado *i*, suponiendo que a priori son conocidos los valores de $\mathrm{p_{ij}}$ (probabilidad de renovación) y b (porcentaje de la prima que se utiliza para valorar el gasto que supone la cancelación de una póliza). Vamos a suponer también que la prima fijada para el cliente *i*-ésimo y la póliza *j*-ésima es conocida (ρ_{ij}), y que se saben cuántas pólizas ($\mathrm{m_i}$) posee cada asegurado. Por lo tanto, bajo este supuesto y como mencionábamos antes, las únicas componentes que son aleatorias son por un lado $\mathrm{L_{ij}}$, es decir la pérdida generada por la póliza *j*-ésima del cliente i-ésimo, si ésta se renueva, y el hecho de que se renueve ($\mathrm{y_{ij}} = 1$) o no ($\mathrm{y_{ij}} = 0$).

Mediante los modelos predictivos planteados en la sección 4 (ver también Bolancé *et al.*, 2016), se puede tener una aproximación al comportamiento que tienen tanto las pérdidas $\mathrm{L_{ij}}$, como la renovación $\mathrm{Y_{ij}}$, en función de los factores conocidos de cada cliente.

Proposición 1. La esperanza y varianza matemáticas de $\mathrm{RgA_i}$ *son*, respectivamente,

$$\sum_{j=1}^{m_i}\left[\left(E(L_{ij})-\rho_{ij}\right)\mathrm{p_{ij}}+b\cdot\rho_{ij}(1-\mathrm{p_{ij}})\right]$$

y

$$\textstyle\sum_{j=1}^{m_i}\left\{V\left[L_{ij}\right]+\left[E\left(L_{ij}\right)-\rho_{ij}-b\rho_{ij}\right]^2\right\}p_{ij}(1-p_{ij}).$$

Demostración. Podemos calcular la esperanza matemática de la variable $\mathrm{RgA_i}$:

$$E[\mathrm{RgA_i}]=E\left[\textstyle\sum_{j=1}^{m_i}Y_{ij}^*\right]=\textstyle\sum_{j=1}^{m_i}E\left[Y_{ij}^*\right]=$$
$$=\textstyle\sum_{j=1}^{m_i}\left[\left(E(L_{ij})-\rho_{ij}\right)\mathrm{p_{ij}}+b\cdot\rho_{ij}(1-\mathrm{p_{ij}})\right].$$

Además, la varianza de la variable $\mathrm{RgA_i}$ es:

$$V[\mathrm{RgA_i}] = V\left[\sum_{j=1}^{m_i} Y_{ij}^*\right] =$$

$$= \begin{cases} \displaystyle\sum_{j=1}^{m_i} V[Y_{ij}^*] & si\ Y_{ij}^* son\ independientes \\ \displaystyle\sum_{j=1}^{m_i}\left(V[Y_{ij}^*] + 2\sum_{j'>j} Cov[Y_{ij}^*, Y_{ij'}^*]\right) & si\ no\ lo\ son \end{cases}$$

Para obtener la varianza hay que utilizar la expresión de la esperanza de la esperanza condicionada.

$$\mathrm{V}[\mathrm{Y}_{ij}^*] = \mathrm{E}\left[\mathrm{E}\left(\left(\mathrm{Y}_{ij}^* - \mathrm{E}(\mathrm{Y}_{ij}^*)\right)^2 | \mathrm{L}_{ij}\right)\right]$$

Tomando la parte correspondiente a la esperanza condicionada,

$$\mathrm{E}\left(\left(\mathrm{Y}_{ij}^* - \mathrm{E}(\mathrm{Y}_{ij}^*)\right)^2 | \mathrm{L}_{ij}\right) = \mathrm{V}(\mathrm{Y}_{ij}^* | \mathrm{L}_{ij}) = \left(\mathrm{L}_{ij} - \rho_{ij} - \mathrm{b}\rho_{ij}\right)^2 \mathrm{p}_{ij}(1 - \mathrm{p}_{ij}).$$

Por lo tanto,

$$\mathrm{V}[\mathrm{Y}_{ij}^*] = \mathrm{E}\left[\left(\mathrm{L}_{ij} - \rho_{ij} - \mathrm{b}\rho_{ij}\right)^2\right] \mathrm{p}_{ij}(1 - \mathrm{p}_{ij}) =$$

$$= \left\{\mathrm{V}[\mathrm{L}_{ij}] + \left[\mathrm{E}(\mathrm{L}_{ij}) - \rho_{ij} - \mathrm{b}\rho_{ij}\right]^2\right\} \mathrm{p}_{ij}(1 - \mathrm{p}_{ij}).$$

Finalmente,

$$\mathrm{V}[\mathrm{RgA_i}] = \sum_{j=1}^{m_i}\left\{\mathrm{V}[\mathrm{L}_{ij}] + \left[\mathrm{E}(\mathrm{L}_{ij}) - \rho_{ij} - \mathrm{b}\rho_{ij}\right]^2\right\} \mathrm{p}_{ij}(1 - \mathrm{p}_{ij}). \square$$

Cálculo del valor en riesgo en el riesgo global asegurado (RgA)

Como mencionamos anteriormente, estamos interesados en cuantificar el valor en riesgo (VaR) para cada cliente. Simplificando el problema, de modo que se supone que la probabilidad de cancelación es cero, por lo que la póliza siempre se renueva y la prima es fija y conocida, necesitamos centrarnos en el modelo:

$$VaR(RgA_i) = VaR\left(\sum_{j=1}^{m_i} Y_{ij}^*\right) \approx VaR\left(\sum_{j=1}^{m_i} L_{ij}\right)$$

El VaR es una de las medidas de riesgo más conocidas basada en la distribución de la pérdida que es igual al cuantil a un nivel de confianza α, es decir, para un tomador del seguro (*i*) y un contrato (*j*) tenemos, si $F_{L_{ij}}^{-1}(\cdot)$ es la inversa de la función de distribución de la cuantía L_{ij} tenemos:

$$VaR(L_{ij}) = F_{L_{ij}}^{-1}(\alpha).$$

Sin embargo, como no se puede garantizar la subadditividad, entonces no podemos asegurar (Artzner *et al.*, 1999) que siempre se cumpla que:

$$VaR\left(\sum_{j=1}^{m_i} L_{ij}\right) \leq \sum_{j=1}^{m_i} VaR(L_{ij})$$

Las pérdidas agregadas pueden analizarse bajo el supuesto de independencia o dependencia entre las diferentes pólizas para el mismo asegurado.

Sea $S_i = L_{i1} + \cdots + L_{im_i}$ la suma de los m_i costes que ha generado el cliente i, a causa de los accidentes en los que haya incurrido. Empezamos a calcular el valor en riesgo suponiendo que los costes L_{ij} son variables aleatorias independientes e idénticamente distribuidas (i.i.d.). En este caso tenemos que deducir la distribución de la suma de i.i.d. y después calcular el cuantil de dicha distribución.

En caso contrario, si suponemos que L_{i1} y L_{i2} son dos pérdidas aleatorias pero idénticamente distribuidas, con respectivas funciones de densidad igual a f_{i1} y f_{i2}, entonces, la densidad de la suma de las dos pérdidas $(L_{i1} + L_{i2})$ es:

$$f_{S_i}(x) = \int_0^{+\infty} f_{i1}(x - v)\, f_{i2}(v) dv$$

Cuando el número de pérdidas es superior a 2 $(m_i > 2)$ entonces se aplica la expresión anterior recursivamente.

Si suponemos que las pérdidas derivadas de las diferentes pólizas que pertenecen a un mismo cliente son dependientes, la densidad de probabilidad de la suma de variables aleatorias se puede calcular a partir de la función de densidad multivariante. Seguidamente, a partir de la densidad podemos deducir la función de distribución acumulativa y el valor en riesgo. Por lo tanto, podemos suponer una distribución multivariante de las pérdidas (Klugman *et al.*, 2012) o una cópula con marginales univariantes dadas y estimar el riesgo asociado a la suma de las pérdidas conociendo los predictores lineales, en los que el vector columna de variables para el individuo *i* y el contrato *j* son conocidos y se ha estimado el vector columna con los parámetros asociados a dichas variables. Hay varias distribuciones que pueden ser buenas candidatas para la modelización de la cuantía de los siniestros, sin embargo, dadas las características de las pérdidas se prefieren algunas distribuciones concretas como por ejemplo: Inversa Gaussiana, Log-Normal, Beta generalizada de segunda clase, etc. Es de señalar que la estimación del valor en riesgo es más complicada cuando se usan distribuciones distintas de la Log-normal (Guillen *et al.*, 2013).

6. Análisis predictivo

La Tabla 2 muestra las variables y su descripción en términos de si la póliza presenta alguna reclamación o no y de la misma manera, en la Tabla 3 se exponen los parámetros estimados para los tres modelos estimados en este trabajo: Inversa Gaussiana, Tweedie y Log-normal.

Tabla 2: Descripción de las variables dependiendo de la existencia de siniestros declarados

Variable	Total	Sin siniestros	Con siniestros
Sexo			
Mujer	31.23%	13.37%	17.86%
Hombre	68.77%	31.84%	36.93%
Edad conductor	53	54	53
Antigüedad del carnet	30	30	29
Potencia	110.10	108.90	111.07
Peso-Potencia	477.00	475.83	478.05
Antigüedad del vehículo	11	11	10

Fuente: Elaboración propia, a partir de una base de datos de 11.700 pólizas y una base de datos de siniestros ocurridos durante el periodo 2010-2014.

Tabla 3: Estimación de los parámetros para tres modelos de cuantía de los siniestros.

	Inv. Gaussian		Tweedie		Log-Normal		
Variable	Estimate	t-value	Estimate	t-value	Estimate	t-value	
Constante	7.384	24.091***	6.925	22.802***	5.961	75.447	***
Sexo	0.060	0.54	0.012	0.112	0.028	0.99	
Edad conductor	-0.009	-1.226	-0.008	-1.034	0.000	-0.077	
Antigüedad permiso	0.010	1.212	0.007	0.817	0.002	0.956	
Potencia	0.002	1.608	0.003	2.532*	0.004	11.598	***
Antigüedad Vehículo	-0.036	-3.591***	-0.043	-4.171***	-0.041	-15.185	***
Peso- Potencia	0.000	-1.211	0.000	-1.455	0.000	-2.598	**
Indice de ajustes	145468.2		-		42274.79		

Fuente: Elaboración propia. Máxima verosimilitud considerando el tiempo de exposición según la vigencia de cada póliza.

Valor en riesgo individual

A partir del os resultados anteriores presentamos un ejemplo del cálculo del VaR asociado a la pérdida global del asegurado y, en su caso, el VaR asociado a las pérdidas ocasionadas por cada una de las pólizas contratadas. Supondremos una distribución Log-Normal para las pérdidas ocasionadas por la siniestralidad en cada una de las pólizas (L_{ij}).

Un modo sencillo de ajustar una distribución Log-Normal a las variables aleatorias L_{ij} es transformar dichas variables en logaritmos, de modo que podemos definir la variable $S_i = \ln(L_{i1}) + \cdots + \ln(L_{im_i})$, que equivale a una suma de normales que a su vez es Normal. Es decir, se sabe que la distribución del logaritmo de una variable Log-Normal es Normal y además, como la distribución de la suma de variables normales también es Normal es sencillo obtener los valores de VaR de la variable S_i suponiendo correlación entre las variables $\ln(L_{ij})$, a dicho valor en riesgo lo denominamos "VaR0.99(multi)" y asume que existe dependencia entre las siniestralidades de las distintas pólizas que tiene contratadas el asegurado i. Además, calculamos los valores del VaR asociados a cada variable $\ln(L_{ij})$, los cuales denominamos VaR0.99 y que representan el riesgo asociado a las pérdidas por siniestralidad del asegurado i en la póliza j.

La Tabla 4 muestra un ejemplo de los resultados para el VaR0.99 y el VaR0.99(multi). Como puede observarse en la Tabla 4, los asegurados que tienen más de una póliza tienen un valor en riesgo agregado (multi) que es inferior al que resulta de sumar el valor en riesgo individual de cada póliza. En este caso, y aunque no siempre sea así en el caso de VaR, se cumple la propiedad de subaditividad de la medida del riesgo.

Si tenemos la certeza de que el asegurado renovará su póliza, los valores en la columna VaR0.99(multi) son una aproximación del riesgo integral del asegurado, definido anteriormente como $VaR(RgA_i)$.

Una aplicación futura del enfoque presentado en éste estudio es la cuantificación del riesgo global de cada asegurado, teniendo en cuenta que éste puede tener una o más pólizas contratadas, ya no sólo en un mismo producto sino en productos diferentes, y que existe una probabilidad p_{ij} inferior a la unidad de que el individuo renueve cada una de sus pólizas contratadas.

Tabla 4: Ejemplo de estimación del valor en riesgo del cliente

Cliente	Póliza	Log(Costes)	Valores estimados	Exposición	VaR0.99	VaR0.99 (multi)
947	63625	7,21	6,09	0,6	5,65	5,65
699	61501	4,62	5,98	0,8	7,45	7,45
653	49646	4,77	5,91	1	9,24	9,24
4035	57714	8,00	6,73	0,8	8,05	8,05
5795	70368	5,56	5,80	0,4	3,65	3,65
6329	58317	7,81	6,25	0,8	7,67	7,67
8433	32219	5,55	6,33	1	9,66	9,66
9796	53602	6,78	5,81	1	9,14	9,14
1985	47658	7,72	6,14	1	9,47	9,47
2236	45631	5,09	6,08	1	9,42	9,42
2411	49760	7,13	5,84	1	9,17	9,17
2521	18866	8,70	6,24	1	9,58	9,58
3407	41855	4,31	6,14	1	9,47	9,47
3600	57180	6,78	5,86	0,8	7,35	7,35
4705	44366	7,35	6,32	1	9,65	9,65
4892	56925	4,10	5,86	0,8	7,35	7,35
6879	12235	8,37	6,62	1	9,95	16,62
6879	59853	8,37	7,17	0,8	8,40	16,62
8976	72142	6,78	5,75	0,4	3,63	3,63
9320	54224	8,41	6,11	1	9,44	9,44
9343	93583	5,15	5,91	1	9,24	9,24
1191	88106	5,10	5,83	1	9,16	16,79
1191	50941	4,05	6,25	1	9,58	16,79

Fuente: Elaboración propia. Sombreados, los clientes con dos pólizas diferentes.

7. Conclusiones

Una compañía aseguradora puede estar interesada en un índice de riesgo individual con el fin de ser capaz de clasificar a los clientes que conforman su cartera de asegurados. Ello permite clasificarlos en función del riesgo que generan para la entidad. Por otra parte, dicha medida puede ser utilizada para mejorar las estrategias y los modelos de precios y, en consecuencia, la política de retención. Idealmente, este concepto puede generar una forma de integrar los modelos de fijación de precios y de fidelización en un solo enfoque. Dicha integración facilitará la cuantificación del riesgo contemplando simultáneamente suscripción y caída de cartera, lo cual es un factor clave para las compañías de seguros que centran sus esfuerzos en la gestión completa de los riesgos.

Sin lugar a dudas, es razonable pensar que precio y renovación son dos conceptos que están relacionados entre sí, debido a la presencia de factores no observables que afectan a ambos. Si se desea modelizar este comportamiento estocástico completo, se puede utilizar un modelo conjunto. Podemos considerar, por ejemplo, un modelo lineal generalizado para la severidad de los siniestros, un modelo lineal generalizado para el número de reclamaciones y un modelo lineal generalizado para el resultado binario sobre la renovación o no de la póliza, sabiendo que los parámetros de asociación entre dichas tres magnitudes pueden ser incluidos en el modelo simultáneo. Alternativamente, podemos utilizar una cópula con diferentes distribuciones marginales en función del tipo de variable, lo cual permitiría incorporar de un modo sencillo la dependencia entre las pérdidas de los diferentes contratos y las probabilidades de renovación de los mismos. Esta será la vía futura de la investigación presentada.

En resumen, la aproximación presentada permite analizar una estructura de datos de alta complejidad con el fin de explicar y predecir el comportamiento de pérdida de clientes y, en especial, su riesgo asociado, teniendo en cuenta la relación entre la siniestralidad de las pólizas contratadas.

Agradecimientos

Las autoras agradecen la ayuda recibida en el Proyecto ECO2013-48326-C2-1-P/FEDER, ICREA Academia y AGAUR Programa de doctorados industriales.

Referencias

Almer, B. (1963). Individual risk theory and risk statistics as applied to fire insurance. *ASTIN Bulletin*, *2*(03), 365-379.

Artzner, P., Delbaen, F., Eber, J. M. y Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*(3), 203-228.

Au, W. H., Chan, K. C. y Yao, X. (2003). A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transactions on Evolutionary Computation*, *7*(6), 532-545.

Bailey, R. A. y Simon, L. J. (1960). Two studies in automobile insurance ratemaking. *Astin Bulletin*, *1*(04), 192-217.

Beerli, A., Martin, J. D. y Quintana, A. (2004). A model of customer loyalty in the retail banking market. *European Journal of Marketing*, 38(1/2), 253-275.

Belles-Sampera, J., Guillen, M. y Santolino, M. (2014). Beyond value-at-risk: Gluevar distortion risk measures. *Risk Analysis*, 34(1), 121–134.

Bermúdez, L., Ferri, A. y Guillen, M. (2013). A correlation sensitivity analysis of non-life underwriting risk in solvency capital requirement estimation. *Astin Bulletin*, *43*(01), 21-37.

Blanco-Morales, P. y Guillen Estany, M. (Dir.). (2010). Estudio sobre el sector asegurador en España (2010): los aspectos cualitativos de solvencia II, Fundación de Estudios Financieros, Madrid.

Bolancé, C., Guillen, M. y Pinquet, J. (2003). Time-varying credibility for frequency risk models: estimation and tests for autoregressive specifications on the random effects. *Insurance: Mathematics and Economics*, *33*(2), 273-282.

Bolancé, C., Guillen, M. y Pinquet, J. (2008). On the link between credibility and frequency premium. *Insurance: Mathematics and Economics*, 43(2), 209-213.

Bolancé, C., Guillen, M. y Padilla-Barreto, A. E. (2016). Predicting probability of customer churn in insurance. In *Modeling and Simulation in Engineering, Economics and Management.* 82-91. Springer International Publishing.

Bolton, R. N., Kannan, P. K. y Bramlett, M. D. (2000). Implications of loyalty program membership and service experiences for customer retention and value. *Journal of the Academy of Marketing Science*, *28*(1), 95-108.

Boucher, J.P., Denuit, M. y Guillen, M. (2007). Risk classification for claim counts: A comparative analysis of various zero-inflated mixed Poisson and hurdle models. *North American Actuarial Journal*, 11(4), 110–131.

Boucher, J.P., Denuit, M. y Guillen, M. (2009). Number of accidents or number of claims? an approach with zero-inflated Poisson models for panel data. *The Journal of Risk and Insurance, 76*(4), 821-846.

Brockett, P. L., Golden, L. L., Guillen, M., Nielsen, J. P., Parner, J. y Perez-Marin, A. M. (2008). Survival analysis of a household portfolio of insurance policies: how much time do you have to stop total customer defection? *The Journal of Risk and Insurance*, 75 (3), 713–737.

Brown, R. L. y Gottlieb, L. R. (2007). *Introduction to ratemaking and loss reserving for property and casualty insurance*. Actex Publications.

Cummins, D. y Rubio Misas, M. (2001). Deregulation, consolidation, and efficiency: evidence from the Spanish insurance industry. Wharton Financial Institutions Center Working Paper No. 02-01.

De Jong, P. y Heller, G. Z. (2008). *Generalized linear models for insurance data* (Vol. 136). Cambridge: Cambridge University Press.

Denuit, M., Maréchal, X., Pitrebois, S. y Walhin, J. F. (2007). *Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems*. John Wiley & Sons.

Dowd, K. y Blake, D. (2006). After VaR: the theory, estimation, and insurance applications of quantile-based risk measures. *Journal of Risk and Insurance*,*73*(2), 193-229.

Ekinci, Y. y Duman, E. (2015). Intelligent classification-based methods in customer profitability modeling. In *Intelligent Techniques in Engineering Management.* 503-527.

Europe, I. (2015). European insurance - key facts. http://www.insuranceeurope.eu/sites/default/files/ attachments/European%20Insurance%20-%20Key%20Facts%20-% 20August%202015.pdf (Agosto del 2015)

Fahrmeir, L. y Tutz, G. (2013). *Multivariate statistical modelling based on generalized linear models*. Springer Science & Business Media. New York.

Fornell, C. y Wernerfelt, B. (1988). A model for customer complaint management. *Marketing Science*, 7(3), 287–298.

Fornell, C. (1992). A national customer satisfaction barometer: The Swedish experience. *the Journal of Marketing*, 6-21.

Frees, E. W. (2009). *Regression modeling with actuarial and financial applications*. Cambridge University Press.

Frees, E. W., Jin, X. y Lin, X. (2013). Actuarial applications of multivariate two-part regression models. *Annals of Actuarial Science*, *7*(02), 258-287.

Frees, E. W., Lee, G. y Yang, L. (2016). Multivariate frequency-severity regression models in insurance. *Risks*, 4(1), 4.

Gourieroux, C. y Jasiak, J. (2007). *The econometrics of individual risk: credit, insurance, and marketing*. Princeton university press.

Guelman, L. y Guillen, M. (2014). A causal inference approach to measureprice elasticity in automobile insurance. *Expert Systems with Applications,* 41(2), 387-396.

Guillen, M., Nielsen, J. P. y Pérez-Marín, A. M. (2008). The need to monitor customer loyalty and business risk in the European insurance industry. *The Geneva Papers on Risk and Insurance Issues and Practice*, 33(2), 207–218.

Guillen, M., Sarabia, J. M. y Prieto, F. (2013). Simple risk measure calculations for sums of positive random variables. *Insurance: Mathematics and Economics*, *53*(1), 273-280.

Günther, C. C., Tvete, I. F., Aas, K., Sandnes, G. I. y Borgan, Ø. (2014). Modelling and predicting customer churn from an insurance company. *Scandinavian Actuarial Journal*, *2014*(1), 58-71.

Haberman, S. y Renshaw, A. (1996). Generalized linear models and actuarial science. *Journal of the Royal Statistical Society. Series D (The Statistician), 45*(4), 407-436.

Hallowell, R. (1996). The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study. *International Journal of Service Industry Management*, *7*(4), 27-42.

Hsiao, C., Kim, C. y Taylor, G. (1990). A statistical perspective on insurance rate-making. *Journal of Econometrics*, *44*(1-2), 5-24.

Jahromi, A. T., Stakhovych, S. y Ewing, M. (2016). Customer churn models: a comparison of probability and data mining approaches. In *Looking forward, looking back: Drawing on the past to shape the future of marketing* (pp. 144-148). Springer International Publishing.

Jørgensen, B. y Paes De Souza, M. C. (1994). Fitting Tweedie's compound Poisson model to insurance claims data. *Scandinavian Actuarial Journa*l, 1994(1), 69-93.

Kim, J. K., Song, H. S., Kim, T. S. y Kim, H. K. (2005). Detecting the change of customer behavior based on decision tree analysis. *Expert Systems*, *22*(4), 193-205.

Klugman, S. A., Panjer, H. H. y Willmot, G. E. (2012). *Loss models: from data to decisions* (Vol. 715). John Wiley & Sons.

Kumar, D. y Garg, A. (2013). A study of data mining techniques for churn prediction. *International Journal of Science, Engineering and Computer Technology*, *3*(1), 1.

Lin, Y. y Grace, M. F. (2007). Household life cycle protection: Life insurance holdings, financial vulnerability, and portfolio implications. *Journal of Risk and Insurance*, 74(1), 141–173.

McCullagh, P. y Nelder, J. A. (1989). *Generalized linear models* (Vol. 37). CRC press.

Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E. y Kaushansky, H. (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on Neural Networks*,*11*(3), 690-696.

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J. y Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, *43*(2), 204-211.

Pinquet, J., Guillen, M. y Bolancé, C. (2001). Allowance for the age of claims in bonus-malus systems. *ASTIN Bulletin*, 31(2), 337–348.

Rees, R., Gravelle, H. y Wambach, A. (1999). Regulation of insurance markets. *The Geneva Papers on Risk and Insurance Theory*, *24*(1), 55-68.

Smith, K. A., Willis, R. J. y Brooks, M. (2000). An analysis of customer retention and insurance claim patterns using data mining: A case study. *Journal of the Operational Research Society*, *51*(5), 532-541.

Thuring, F., Nielsen, J. P., Guillen, M. y Bolancé, C. (2012). Selecting prospects for cross-selling financial products using multivariate credibility. *Expert systems with Applications, 39*(10), 8809-8816.

Thuring, F., Nielsen, J.P., Guillen, M. y Bolancé, C. (2013). Segmenting and selecting cross-sale prospects using dynamic pricing. In *Proceedings of the 2nd International Conference on Operations Research and Enterprise Systems*. 103–108.

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G. y Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, *55*, 1-9.

Vaughan, E. J. y Vaughan, T. (2007). *Fundamentals of risk and insurance*. John Wiley & Sons.

Verhoef, P. C. (2003). Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing*, *67*(4), 30-45.

Werner, G. y Modlin, C. (2010). *Basic ratemaking, 4th Edition*. Casualty Actuarial Society.

Yeo, A. C., Smith, K. A., Willis, R. J. y Brooks, M. (2001, May). Modeling the effect of premium changes on motor insurance customer retention rates using neural networks. In *International Conference on Computational Science*. 390-399.