



FACULTAD DE CIENCIAS ECONÓMICAS Y
EMPRESARIALES

TARIFICACIÓN DE MICROSEGUROS: UNA APLICACIÓN DEL MODELO TWEEDIE

Autora: Inmaculada Peña Sánchez

Directores: Dra. Cristina Lozano Colomer y Dr. José Luis Vilar Zanón

Madrid
Mayo 2017

Inmaculada
Peña
Sánchez

TARIFICACIÓN DE MICROSEGUROS: UNA APLICACIÓN DEL MODELO TWEEDIE



No es lo mismo observar los medios de la vida que observar directamente las vidas de la gente que se las arregla para vivir.

Amartya Sen (2014)

AGRADECIMIENTOS

Esta investigación ha sido posible gracias a la aportación de muchas personas que no han dudado en compartir su conocimiento y experiencia, brindándome todo su apoyo y colaboración para que esta tesis doctoral se convierta en realidad.

En primer lugar me gustaría agradecer a mis directores, Dra. Cristina Lozano y Dr. José Luis Vilar, por su dedicación y valiosos consejos aportados en todo el recorrido de este proyecto, y que gracias a ellos ha sido posible elaborar esta tesis.

A MAPFRE Insular Insurance Corporation, por su inestimable ayuda en facilitarme la información del microseguro analizado, así como su continuo soporte en todas mis peticiones para poner en práctica la propuesta metodológica.

A la Universidad Pontificia de Comillas, a sus profesores, que me han brindado su apoyo e interés en la realización de esta tesis. Y a todo el personal, que tan amablemente siempre me han atendido y ayudado en la resolución de todo el trámite administrativo.

A la Universidad Complutense, a D. Antonio Heras y Dña. María Pérez, por el material didáctico sobre cómo implementar los modelos GLM con la herramienta SAS.

A muchos profesionales del sector asegurador, por todas las referencias facilitadas que me han abierto el camino en este nuevo mundo de los microseguros. A expertos en este campo, tanto profesionales como investigadores, por compartir su conocimiento y experiencia y ayudarme a centrar la línea de mi investigación. Al grupo de trabajo de Microseguros de la *Casualty Actuarial Society* que me ha acompañado durante más de un año en adentrarme en la investigación de este nuevo campo.

A mis compañeros de doctorado, por su acompañamiento y compartir conmigo este camino de la investigación.

A mis amigos y antiguos compañeros de trabajo, por sus ánimos y por su apoyo en perseguir mi objetivo.

Y en especial, a mis padres y hermanos por transmitirme fuerza y ánimos a seguir y no parar hasta lograr este reto de investigación que culmina ya con mi tesis doctoral.

A todos, muchas gracias por compartir conmigo esta ilusión.

ÍNDICE

RESUMEN.....	11
ABSTRACT	12
INTRODUCCIÓN.....	13
1. Contexto y motivación de la investigación	13
2. Delimitación y justificación de la investigación.....	17
3. Objetivos de la investigación.....	20
4. Estructura de la tesis.....	21
PARTE I. MARCO TEÓRICO: LOS MICROSEGUROS.....	24
CAPÍTULO 1. LA PERSPECTIVA SOCIAL	24
1.1 Introducción: contexto social.....	25
1.2 Origen, definición y características del microseguro.....	28
1.3 Contribución al desarrollo económico, social y cultural.....	34
1.4 Panorama del mercado microasegurador.....	37
1.5 Tipología de microseguros	39
1.5.1. Microseguro de vida	40
1.5.2. Microseguro de salud	42
1.5.3. Microseguros generales	44
1.5.4. Seguros agrícolas.....	45
CAPÍTULO 2. LA PERSPECTIVA ASEGURADORA	49
2.1 Estructura de la actividad microaseguradora	50
2.2 Marco regulatorio del microseguro.....	53
2.2.1. Objetivos de la regulación	53
2.2.2. Limitaciones.....	54
2.3 Proceso de diseño del microseguro	58
2.4 La información como fuente de conocimiento para la tarificación.....	61
2.4.1. Fuentes de información	61
2.4.2. Descripción de la información	65
2.4.3. Razonabilidad, trazabilidad y conservación	67
2.5 Problemas asociados con la sostenibilidad de los microseguros	69
2.6 Medidas de eficiencia del microseguro	73

PARTE II. ESTUDIO EMPÍRICO.....	76
CAPÍTULO 3. METODOLOGÍA DE TARIFICACIÓN	77
3.1 La tarificación de microseguros.....	78
3.1.1. El papel del actuario	78
3.1.2. Estimación de la prima	79
3.1.3. Métodos clásicos	80
3.1.4. Desarrollos en la tarificación de microseguros.....	83
3.2 Metodología de la investigación	84
3.2.1. Modelo Tweedie	84
3.2.2. Definición y tratamiento de variables	87
3.2.3. Validación del modelo	92
3.2.4. Medidas estadísticas del modelo	95
3.3 Modelos lineales generalizados	99
3.3.1. Guía de elaboración de GLMs	100
3.3.2. Modelo Tweedie	103
CAPÍTULO 4. TRATAMIENTO DE DATOS	110
4.1 Origen y descripción de la información	111
4.2 Tratamiento y depuración de datos	112
4.3 Factores	115
4.3.1. Variable objetivo	115
4.3.2. Variable respuesta.....	116
4.3.3. Variables explicativas	116
4.4 Análisis descriptivo: datos y variables	118
4.4.1. Datos: visión general	118
4.4.2. Variables explicativas	120
4.4.3. Variables a posteriori.....	125
4.5 Segmentación de las variables explicativas.....	126
4.6 Estudio de influencia de las variables explicativas.....	131
4.6.1. Análisis de dependencia	131
4.6.2. Análisis de interacción.....	136

CAPÍTULO 5. ANÁLISIS PREDICTIVO	137
5.1 Metodología del caso de estudio: el modelo Tweedie	138
5.2 Hipótesis del modelo	140
5.2.1. Variables explicativas	141
5.2.2. Partición de la base de datos	142
5.3 Modelización de la prima pura	143
5.4 Comparación de los modelos seleccionados	147
5.4.1. Análisis de equilibrio financiero	150
5.5 Modelo óptimo	153
5.5.1. Análisis de residuos y de valores atípicos	156
5.6 Análisis comparativo de primas: modelo óptimo vs modelo real.....	158
5.6.1. Análisis de equilibrio financiero	159
5.6.2. Viabilidad del modelo	163
CONCLUSIONES.....	168
1. Respecto al marco teórico bajo un enfoque social.....	168
2. Respecto al marco teórico bajo un enfoque asegurador.....	170
3. Respecto al objetivo general	172
4. Respecto a los objetivos específicos	173
5. Líneas abiertas y repercusión para futuras investigaciones	179
BIBLIOGRAFÍA.....	184
ANEXOS	193
Anexo A. Cuadros resumen de la base de datos del microseguro	193
Anexo B: Información demográfica nacional de Filipinas.....	194
Anexo C: Código SAS del modelo Tweedie	197

ÍNDICE DE FIGURAS Y GRÁFICOS

FIGURAS

Figura 2.1. Estructura de la actividad microaseguradora.....50

GRÁFICOS

CAPÍTULO 4

Gráfico 4.1. Histograma de frecuencia para la variable *edad*. 120

Gráfico 4.2. Distribución de la frecuencia del coste último total..... 126

CAPÍTULO 5

Gráfico 5.1. Gráfico de elevación de las primas puras del modelo Tweedie. 164

Gráfico 5.2. Curva de Lorenz: modelo óptimo vs modelo real. 166

ÍNDICE DE TABLAS

CAPÍTULO 1

Tabla 1.1. Panorama global de asegurados por producto y región.	39
Tabla 1.2. Visión general de la tipología de productos.	40

CAPÍTULO 2

Tabla 2.1. Resumen de categorías e indicadores clave de rendimiento.	74
---	----

CAPÍTULO 3

Tabla 3.1. Principales variables objetivo de los modelos GLM.	88
Tabla 3.2. Resumen de los modelos Tweedie.	106
Tabla 3.3. Funciones de enlace de las distribuciones de la familia exponencial.	109

CAPÍTULO 4

Tabla 4.1. Resumen general de la base de datos.	119
Tabla 4.2. Resumen de la base de datos por cobertura.	119
Tabla 4.3. Resumen descriptivo para la variable <i>sexo</i>	121
Tabla 4.4. Resumen descriptivo para la variable <i>estado civil</i>	122
Tabla 4.5. Resumen descriptivo para la variable <i>antigüedad de la póliza</i>	122
Tabla 4.6. Resumen descriptivo para la variable <i>región</i>	124
Tabla 4.7. Resumen descriptivo para el número de siniestros.	125
Tabla 4.8. Resumen de las variables explicativas disponibles.	127
Tabla 4.9. Segmentación de la variable explicativa <i>edad</i>	128
Tabla 4.10. Segmentación de la variable explicativa <i>densidad</i>	130
Tabla 4.11. Análisis de colinealidad (variables segmentadas <i>age2</i> y <i>den3</i>).	132
Tabla 4.12. Análisis de colinealidad (variables segmentadas <i>age5</i> y <i>den5</i>).	133
Tabla 4.13. Matriz de coeficientes V de Cramer.	135

CAPÍTULO 5

Tabla 5.1. Resumen de las variables explicativas potenciales del modelo.	141
Tabla 5.2. Resumen de muestras para la validación del modelo.	143
Tabla 5.3. Análisis univariante de los predictores.	144
Tabla 5.4. Análisis bivariante de los predictores.	145
Tabla 5.5. Análisis multivariante de los predictores.	146
Tabla 5.6. Análisis comparativo de modelos: parámetros y medidas.	148
Tabla 5.7. Resultados del modelo 1: análisis de los estimadores.	148
Tabla 5.8. Resultados del modelo 2: análisis de los estimadores.	149
Tabla 5.9. Análisis comparativo de modelos: variabilidad de primas.	151

Tabla 5.10. Análisis comparativo de modelos: ECM.....	152
Tabla 5.11. Análisis comparativo de modelos: ratio de siniestralidad.....	152
Tabla 5.12. Modelo óptimo: medidas de bondad del ajuste.....	154
Tabla 5.13. Modelo óptimo: parámetros y medidas.	154
Tabla 5.14. Modelo óptimo: estimadores.	155
Tabla 5.15. Modelo óptimo: análisis descriptivo de residuos estandarizados.	157
Tabla 5.16. Modelo real: plan de tarifas del microseguro.	159
Tabla 5.17. Modelo óptimo: análisis de variabilidad de primas.....	160
Tabla 5.18. Modelo óptimo vs modelo real: ECM.	161
Tabla 5.19. Modelo óptimo vs real: análisis del ratio de siniestralidad.	161
Tabla 5.20. Mercado de microseguros asiático: ratios de siniestralidad.....	162

ANEXOS

Tabla A.1. Descripción de la base de datos de microseguros para el año 2014.	193
Tabla A.2. Resumen descriptivo de la información por año/cobertura.	194
Tabla A.3. Descripción de la población y densidad a nivel provincial.	195

RESUMEN

En el contexto de las microfinanzas está cobrando fuerza la figura del microseguro, como un producto asegurador novedoso de carácter inclusivo, que ofrece protección a las personas de bajos ingresos. Este tipo de productos presenta una tarificación básica debida a la falta de información, lo que obliga a las compañías a incluir recargos adicionales para suplir la incertidumbre asociada en la valoración de los riesgos cubiertos. Siguiendo con esta línea, se plantea la necesidad de mejorar el diseño del producto desde el campo actuarial de la tarificación. Para ello, se propone utilizar las técnicas de modelización predictivas, de uso extendido en el seguro tradicional, como una potencial solución para mejorar la tarificación de los microseguros. En particular, esta investigación se centra en la metodología de tarificación basada en los modelos lineales generalizados que siguen una distribución Tweedie para aplicarla a un caso real, un microseguro combinado de Filipinas. En primer lugar, se analizarán las características de la base de datos para identificar cuál será la metodología adecuada. Después, se aplicará esta técnica a nuestro colectivo con el fin de obtener el modelo que mejor se ajuste a la experiencia acumulada. El modelo óptimo proporcionará una tarifa multivariante ajustada al perfil de riesgo individual, en función de las variables edad, antigüedad de la póliza y densidad de la población asegurada. Para finalizar, se comprueba mediante un doble análisis que el modelo óptimo supera al modelo real que actualmente aplica la compañía.

El análisis predictivo confirma que la distribución Poisson compuesta con Gamma ajusta a nuestra cartera. A partir de aquí se deduce que el modelo óptimo estima unas primas más precisas, aportando un doble beneficio. Por un lado, las compañías aseguradoras disponen de un mayor conocimiento, eliminando los recargos asociados a la incertidumbre en la valoración. Además, se contribuye con un efectivo control de riesgos al considerar la variable antigüedad de la póliza en la tarifa, solventando la selección adversa existente en nuestro microseguro. Por el lado social, se promueve el buen uso del producto con una tarifa acorde al perfil de riesgo, lo que ayuda en la educación financiera del cliente y favorece también la consecución de los objetivos globales de desarrollar productos inclusivos y sostenibles. En definitiva, la aplicación del modelo Tweedie en la tarificación de microseguros debería ser considerada en la práctica actuarial para la estimación de las primas de riesgo.

Palabras clave: microseguro, tarificación, modelo lineal generalizado, modelo Tweedie, prima de riesgo, inclusivo, sostenibilidad.

ABSTRACT

Within the microfinance sector, microinsurance emerges like an innovative and inclusive insurance that provides protection for low-income policyholders against specific risks. This kind of products presents a basic pricing due to the lack of data, which require adding additional charge to premiums in order to compensate unexpected variations in the risks covered. A research need for microinsurance pricing arises as a challenge to boost business growth. For this purpose, we apply predictive modelling techniques taking advantage of conventional methodologies as potential solutions to improve the pricing in the microinsurance market. In particular, we focus on Generalized Linear Models for the Tweedie distribution in a particular case study, a microinsurance in the Philippines. For this, we firstly analyze the features of our available database in order to identify the suitable technique. After that, we run the model for our specific portfolio to obtain the model that best fits our data. The optimal model estimates risk premiums with three risk factors, age of insured, age policy and density of the population. Finally, we prove our optimal model outperforms the current one applied for the company.

The predictive analysis confirms the Tweedie's compound Poisson-Gamma distribution fits our particular portfolio. And hence, fair and accurate premiums are estimated with a twofold benefit: in one hand, any additional charge could be now removed. On the other hand, new estimates enhance risk control through risk factors like age policy that penalizes the bad risks and rewards good risks, avoiding the adverse selection existing in our microinsurance. All in all, the application of the Tweedie model for microinsurance pricing should be considered to carry out in the actuarial practice of microinsurers.

Keywords: microinsurance, pricing, Generalized Linear Models, Tweedie model, risk premium, inclusive, sustainability.

INTRODUCCIÓN

1. Contexto y motivación de la investigación

Esta investigación sobre microseguros pretende continuar una línea innovadora de desarrollo de productos en materia de inclusión social y bajo una perspectiva empresarial. En ella se van a combinar dos aspectos fundamentales. Por una parte, un enfoque social entendido como contexto de desarrollo, enfocado a un vasto y desconocido mundo formado por las personas más vulnerables. Y por otra, un enfoque empresarial desde la práctica actuarial y con el gran valor añadido de poder contar con datos reales de un microseguro. Toda esta labor, entra de lleno en la actividad aseguradora cada vez más motivada por integrarse en la economía inclusiva y que además contribuye al impulso de un nuevo negocio para las compañías.

Esta combinación de ambos aspectos marca con claridad el verdadero sentido de realizar este trabajo. La motivación surge de la confluencia de una serie de ideas y pensamientos que he ido madurando a lo largo de las diferentes etapas de mi vida. En el ejercicio de mi trabajo profesional, siempre he expresado mi creciente inquietud por investigar e innovar en nuevas formas de realizar las tareas, con el fin de contribuir al desarrollo de un trabajo técnico y especializado como es la labor actuarial. A esto se añade, ya desde un contexto social, mi interés general por ayudar a la gente más necesitada, que se materializó colaborando en una experiencia no muy lejana en el tiempo. En ella tuve la oportunidad de participar en un proyecto solidario enfocado en al desarrollo humano e inclusión social en un país donde son muchas las personas especialmente desfavorecidas.

Allí pude observar de cerca cómo esta población vulnerable y con clara escasez de recursos, solventaba la mayoría de sus necesidades más apremiantes gracias a la solidaridad de la comunidad, que se convertía así en su principal fuente de recursos. Fue esta experiencia la que despertó en mí una especial sensibilización a este problema. Esto me hizo plantearme en qué forma podría yo contribuir a mejorar la situación de este sector de la población tan vulnerable. Y fue entonces cuando tomó forma la idea de materializar todas estas cuestiones en la presente investigación, cuya punto clave giraría en torno a la tarificación de microseguros. Y así fue cómo se concretó este trabajo en el que se combinaba mi interés personal por lograr un mayor bienestar y calidad de vida para las personas más vulnerables con mi inquietud

profesional de utilizar el conocimiento y experiencia adquiridos en el campo actuarial para mejorar la tarificación de este tipo de productos.

Esta línea de investigación me abría un nuevo campo de estudio hasta entonces para mí desconocido y que además veía con un gran potencial de crecimiento, como era el microseguro. Este novedoso producto se puede definir, y adelanto un concepto sobre el que después se profundizará, como aquel contrato de seguro que ofrece protección a las personas de bajos ingresos a cambio de unas primas reducidas. La propia descripción del microseguro ofrece las dos vertientes de interés en esta investigación, por un lado, el aspecto social y humano derivado del mercado objetivo al que va dirigido, y por otro lado, el aspecto asegurador en el sentido de mejorar el diseño del producto desde el campo técnico-actuarial. Dos aspectos importantes que, ya de entrada, creo oportuno profundizar un poco más.

La vertiente social y humana

En el transcurso de nuestra vida estamos siempre amenazados por la ocurrencia de hechos inesperados que pueden producir un efecto negativo en nuestra salud y en nuestro bienestar, como también en el de las personas que nos rodean. Es obvio pensar que las personas con menos recursos están expuestas a un mayor número de riesgos, bien por sus limitados recursos económicos, o bien por sus peores condiciones de vida. Esto hace que la población con menos recursos se convierta en un segmento de población especialmente débil y desprotegido, que por si fuera poco, cuenta con menos capacidad para afrontar las consecuencias de su mayor exposición al riesgo.

De ahí su mayor necesidad de la protección social. No solo como mecanismo integral de prevención de riesgos, destinado a reducir la pobreza y la vulnerabilidad de los más desfavorecidos, sino también para promover la equidad y solidaridad entre las personas, fomentar cierta seguridad en los ingresos, y tratar de dar un mayor acceso a los servicios sociales básicos, especialmente los sanitarios. Inevitablemente, este aspecto social debe ir acompañado en su medida de las oportunas facilidades financieras con las que impulsar una mejor distribución de recursos.

En la inacabada tarea de responder a esta creciente demanda de protección social, cabe mencionar también las múltiples iniciativas lideradas por organismos nacionales e internacionales, que desde diversos ámbitos tratan de impulsar acciones a favor de

las personas más vulnerables. Y es precisamente en este punto, donde aparecen los microseguros como producto asegurador concebido especialmente para la protección de las personas de bajos ingresos, que además tienen un acceso mucho más limitado a las medidas formales de prevención de riesgos. El microseguro se configura así en un formidable instrumento de protección de riesgos para este amplio segmento de población, y a la vez contribuye al desarrollo económico sostenible de países emergentes.

Justo en esta línea se sitúa esta investigación, que está enfocada precisamente al mercado de microseguros, que es un segmento novedoso y con un enorme potencial de crecimiento. Con ella queremos proporcionar medidas de gestión de riesgo para estas personas, y al mismo tiempo contribuir a la consecución de los objetivos globales de desarrollo sostenible.

La vertiente del negocio asegurador

Es un hecho constatable que las compañías de seguros presentes en todo el mundo se mueven en un mercado saturado y altamente competitivo. En este sentido, la aparición de los microseguros en países emergentes puede convertirse para ellas en una clara oportunidad de expansión del negocio. Con un doble objetivo: por un lado, para contribuir al desarrollo socio-económico de estos países, y por otro lado, para acceder a un nuevo y latente segmento de mercado, pero con un gran potencial de desarrollo. Ambos retos deben estar equilibrados para lograr un eficaz desarrollo sostenible, y que sea exitoso en el tiempo (Swiss Re, 2010).

Abundando en la misma idea, cabe destacar que el microseguro, como producto inclusivo, brinda numerosas oportunidades en la industria del seguro. El segmento de población de bajos ingresos supone un mercado nuevo, desconocido en gran medida, pero muy extenso y aún por explorar. Se abre entonces una nueva vía que, además de fomentar la responsabilidad social corporativa de las compañías tomando en consideración a la población más vulnerable, se configura como una estrategia de crecimiento del negocio asegurador a largo plazo. Adicionalmente, este nuevo mercado contribuye a crear una buena relación entre el regulador y los responsables de los gobiernos, siempre en aras de una mejora sostenida del desarrollo económico en los países emergentes.

Desde el punto de vista actuarial, la mayoría de microseguros presentan una tarificación básica, un tanto alejada del perfil de riesgo de este colectivo. El principal problema radica en la generalizada limitación de información existente en el mercado microasegurador. Esto se debe, en primer lugar, a la propia novedad del negocio que cuenta con escasa trayectoria. Otra razón es el menor tamaño de la mayoría de las compañías que por lo mismo cuentan con una pobre infraestructura. Todas estas barreras les impiden disponer de suficiente información para realizar los correspondientes análisis predictivos (Biener, 2013).

En todo caso, un buen producto debe adecuarse a las necesidades del asegurado y a la vez ser rentable para la compañía. En la práctica habitual se observa que la falta de datos condiciona en buena medida la aplicación de una tarificación más sofisticada, lo que conduce a los microaseguradores a añadir elevados recargos para cubrir el riesgo de incertidumbre asociado en la previa estimación de las primas. Con la repercusión nada deseable de una sobrevaloración del riesgo, encareciendo las primas del producto y haciendo que la cobertura resulte menos atractiva para el cliente. Por este motivo, se impone la necesidad de diseñar productos de mayor calidad a través del desarrollo de nuevas metodologías de tarificación. De esta forma se obtendrán unas primas más precisas y ajustadas a la distribución real del riesgo, para así poder crear productos más eficientes y atractivos para el cliente (Biener, 2013; Roth et al., 2007).

Ante esta situación, descubrimos que se nos abre aquí un sugestivo campo de investigación, relacionado precisamente con el desarrollo actuarial de los microseguros. Lo que quiere decir que, como paso previo, será fundamental conocer las referencias científicas existentes en este ámbito de estudio. Tenemos, pues, por delante un amplio recorrido cuyo trayecto nos llevará a descubrir el origen, la evolución histórica, los principales problemas y limitaciones del producto hasta alcanzar el momento actual con toda la problemática asociada al terreno actuarial. Este será nuestro punto de partida, tras el cual estaremos en condiciones de proponer una línea de investigación que nos permita avanzar en el desarrollo de los microseguros. Ya de entrada, algunos autores ponen de manifiesto la necesidad de más investigación desde varios ámbitos (Biener, 2013; Clarke, 2011; Churchill, 2006; Purcal et al., 2012), y en particular en el terreno de la tarificación (Biener, 2013; Purcal et al., 2012; Wrede et al., 2016). Es a partir de aquí, donde nuestra investigación toma el relevo para continuar con el objetivo de explorar nuevas metodologías de

tarificación en un mercado nuevo y desconocido aún para una gran parte del sector asegurador como son los microseguros.

2. Delimitación y justificación de la investigación

Con la visión que nos ofrece el contexto hasta aquí expuesto podemos entrar ya en el núcleo de este trabajo, cuyo ámbito nos proponemos delimitar con el fin de fijar la línea de investigación que se pretende abordar.

Una amplia revisión de la literatura existente sobre el tema nos ha llevado a observar que los principales estudios científicos sobre microseguros se centran, por una parte, en aspectos cualitativos (Albarrán et al., 2007; Biener & Eling, 2012; Cohen & Sebstad, 2005; Churchill, 2006; Churchill & Matul, 2014; Garayoa, 2011; Zimmerman & Magnoni, 2014). Y por otra, vemos que se muestra una atención especial en la importancia de la labor actuarial en el diseño de estos productos aseguradores, destacando la necesidad de una mayor investigación enfocada en torno a tres principales retos. El primero se refiere al **análisis y control de riesgos**, sobre el cual autores como Biener y Eling (2012) lo achacan a la falta de recursos disponibles a la hora de evaluar los riesgos. El segundo se centra en el **diseño de una buena base de datos**, objetivo imprescindible para garantizar su calidad y la posibilidad de suplir la escasez de datos a través de fuentes externas (Wipf & Garand, 2006). Y el tercero, y no menos importante, hace especial hincapié en la **mejora en la tarificación**, que lleva a no pocos autores a resaltar el papel determinante del actuario en el desempeño de esta labor, sin que por ello se reste importancia a la necesidad de ir incorporando al cálculo la experiencia acumulada (Biener, 2013; Clarke, 2011; Purcal et al., 2012; Wipf & Garand, 2010; Wrede et al., 2016). Sin embargo, y a pesar de la importancia que se le otorga a la tarificación, se echan en falta más estudios relacionados con el desarrollo de nuevas metodologías de tarificación aplicables a los microseguros.

Nuestro enfoque en este punto resulta especialmente novedoso por cuanto trata de cubrir esa necesidad de investigación, proponiendo mejorar el diseño del producto desde el campo actuarial y con el gran valor añadido de utilizar datos reales de un microseguro combinado con experiencia en el mercado.

Tampoco se puede olvidar, por otra parte, que el negocio de microseguros no está exento de problemas, algunos de los cuales pueden poner en peligro la sostenibilidad

financiera del producto, condicionada mayormente por el segmento de población al que va dirigido (Biener & Eling, 2012; Wrede et al., 2016). Baste aquí citar someramente las principales dificultades en este mercado. Una de ellas, siempre presente, está en la asimetría de la información, con frecuencia asociada a un falso comportamiento del asegurado e incluso a otras partes implicadas en el negocio (Biener & Eling, 2012). Otra dificultad viene por el bajo nivel de exposición del negocio, no solo por la novedad del producto sino por el mismo perfil del cliente. Y por último, está la falta generalizada de información que caracteriza este mercado y que condiciona en gran medida su desarrollo desde diferentes ámbitos del negocio.

Desde el punto de vista de la tarificación, los microseguros utilizan una metodología muy básica, sin identificar el perfil de riesgo individual de los asegurados, lo que lleva a promover los seguros de grupo como los más utilizados (Wrede et al., 2016). La ventaja de estos supone unos costes de administración, distribución y siniestros inferiores al considerar al grupo como tomador de la póliza. Sin embargo, la insuficiente valoración del riesgo termina por repercutir en variaciones que no son debidamente captadas en la estimación de las primas. Con el resultado de que la prima pura termina por ser complementada con un elevado margen de seguridad para cubrir la incertidumbre asociada a las variaciones de la tarifa, sin prever el efecto colateral de ahuyentar a no pocos clientes.

En la búsqueda de posibles soluciones, lo más sencillo siempre es recurrir al conocimiento y la experiencia en el mercado tradicional de los seguros, valorando las alternativas metodológicas existentes para escoger la técnica de tarificación que mejor pueda predecir el comportamiento futuro de riesgo de nuestro microseguro. En todo caso, hay que reconocer que sigue siendo importante aprovechar las sinergias del negocio asegurador para proyectarlo en beneficio de este nuevo mercado microasegurador. La experiencia y la habilidad adquirida en la tarificación de seguros proporciona una fuente de conocimiento que puede ser fácilmente trasladable a este nuevo negocio.

Con todo lo dicho, la presente investigación quiere dar un paso más. Se propone cubrir esa necesidad de investigación, enfocando el estudio hacia una mejora significativa del diseño del producto desde el campo actuarial de la tarificación del microseguro. La aplicación de una nueva metodología de tarificación como son los modelos lineales generalizados (GLMs), comúnmente utilizada en el seguro

tradicional, permite obtener unas primas de riesgo del microseguro más precisas al tarificar los riesgos de forma individual. Esta técnica de tarificación a priori parte de unos factores de riesgo, que se representan por las variables explicativas, sabiendo que con un modelo de regresión lineal general es posible obtener la prima de riesgo. De esta forma se obtiene una tarifa más efectiva, clasificada en función del perfil de riesgo del colectivo asegurado, con el consiguiente alivio de poder eliminar los recargos por incertidumbre asociados a estos productos.

La principal contribución de este nuevo enfoque metodológico repercute en un doble beneficio. Beneficio para el cliente porque se cuenta con un producto más atractivo y asequible al reducir la carga de incertidumbre; y beneficio para la compañía porque adquiere un mayor conocimiento y control sobre el negocio microasegurador (Biener, 2013; Purcal et al., 2012; Roth et al., 2007). En definitiva, se trata de promover una tarificación más dinámica y capaz de adaptarse a las necesidades del mercado, con la que puede no solo contribuir a una mejor gestión de riesgos asociados con la asimetría de información, sino también responder al objetivo global de desarrollar productos sostenibles (Werner & Modlin, 2016).

Finalmente, y según todo lo comentado hasta el momento, esta investigación pretende ser de interés para diversos grupos que aquí se señalan. A las propias compañías aseguradoras, como guía de aplicación de los modelos GLM a los microseguros, así como para despertar ideas sobre otras sinergias del seguro que puedan aplicarse a otros negocios o mercados. A diversas asociaciones y grupos de trabajo interesados en microseguros, para incorporarlo en sus proyectos en el desarrollo de este mercado. También, a otros investigadores y docentes que se ocupan de transmitir y continuar con el estudio de otras líneas de investigación relacionadas con los microseguros desde las dos perspectivas planteadas, la social y la aseguradora. Y como eslabón final de la cadena, para los mismos clientes, que serán siempre los potenciales beneficiarios de estudios como este, incluyendo el beneficio de su educación financiera, que les hará más conscientes de que su comportamiento frente al riesgo repercute directamente en el precio del microseguro.

3. Objetivos de la investigación

Una vez definida la línea de investigación, consistente en la mejora del diseño del microseguro desde el campo actuarial, pasamos a concretar la finalidad que pretendemos conseguir con este estudio mediante la definición de los siguientes objetivos:

El **objetivo general** consiste en:

Conocer la viabilidad de la aplicación de los modelos lineales generalizados (GLMs) a un programa de microseguros, con la idea de obtener unas primas de riesgo más ajustadas a la verdadera distribución de riesgo, de manera que sean más efectivas en este mercado emergente.

Para alcanzar este objetivo, seguiremos una serie de etapas que pasan por:

- Describir las metodologías de tarificación, seleccionando el modelo Tweedie como la mejor alternativa.
- Realizar el análisis de los modelos predictivos de acuerdo a la información disponible.
- Tratar de obtener un modelo óptimo capaz de proporcionar las primas de riesgo para nuestro caso de estudio.
- Y finalmente, justificar que el enfoque metodológico planteado es perfectamente aplicable a nuestro microseguro.

Los **objetivos específicos** que desgranar nuestro objetivo general se concretan en:

1. Comprobar que el modelo predictivo Tweedie es la mejor alternativa en la aplicación de modelos GLM para nuestro caso de estudio.

Para demostrar este objetivo, buscaremos el enfoque metodológico más adecuado mediante los siguientes pasos:

- Valorar las diferentes metodologías de tarificación predictivas capaces de estimar las correspondientes primas de riesgo del seguro.
- Analizar los datos e interpretar las características particulares de nuestro microseguro combinado, de acuerdo con la información disponible.
- Elegir cuál es la alternativa de los modelos lineales generalizados más adecuada, según la información disponible y el contexto de mercado emergente donde nos movemos.

2. Conseguir la certeza suficiente de que el modelo Tweedie proporciona la mejor estimación de las primas de riesgo de nuestro producto.

En relación a la comprobación del modelo Tweedie se debe:

- Realizar todo el proceso de análisis predictivo. Para ello es fundamental establecer las siguientes hipótesis del modelo: definir la variable objetivo y por ende la variable respuesta, valorar y transformar las variables explicativas y fijar el criterio de validación del modelo.
- Evaluar y comparar los resultados de los posibles modelos que surjan del análisis, a través de su nivel de significación y las correspondientes medidas de ajuste.
- Elegir el modelo óptimo utilizando nuestra herramienta de decisión, el estudio sobre el equilibrio financiero.
- Obtener las primas de riesgo del modelo óptimo finalmente seleccionado, verificando completamente sus resultados.

3. Llegar a valorar nuestra propuesta, promoviendo la aplicación de nuestro modelo óptimo a la práctica microaseguradora.

Respecto a esta idea de implementar el modelo propuesto en la práctica actuarial, se debe utilizar como referencia el modelo real de tarificación de la compañía, para:

- Estudiar y comparar las ventajas e inconvenientes de nuestra propuesta metodológica de tarificación frente a la que utiliza actualmente la compañía.
- Realizar el estudio comparativo del análisis de equilibrio financiero en los respectivos modelos, con el fin de comprobar que nuestra propuesta es la que mejor contribuye a una mayor viabilidad y sostenibilidad financiera del producto analizado.

4. Estructura de la tesis

Una vez definidos los objetivos de la investigación, el paso siguiente es describir el procedimiento que se va a seguir en su desarrollo. Para ello, comenzaremos por ofrecer una panorámica general de los microseguros. Seguidamente, realizaremos el estudio empírico, desarrollando una metodología de tarificación para un caso real. Para llevar a cabo este trabajo, la estructura se ha dividido en dos partes fundamentales:

Parte I: Marco teórico. Corresponde a una amplia revisión de literatura de los microseguros desde dos vertientes diferenciadas. Por un lado, el enfoque social y humano y por otro lado, el enfoque de la actividad aseguradora.

- *Capítulo 1: Perspectiva social.*

Se pretende ofrecer una visión global desde el punto de vista social para comprender el objetivo funcional del microseguro. Comenzamos el recorrido partiendo de su origen, las diferentes definiciones y sus características diferenciadoras respecto al seguro. Seguidamente, se resalta su contribución social para el desarrollo sostenible, así como su presencia actual en los diferentes continentes. Finalmente, se hace un repaso de la tipología de productos que se comercializan en la actualidad.

- *Capítulo 2: Perspectiva de la actividad aseguradora.*

Este capítulo trata de complementar el marco teórico desde la posición de las compañías aseguradoras, como parte integrante de la actividad del negocio, con la finalidad de explorar las limitaciones existentes y de detectar las posibles mejoras que se puedan realizar. Partimos con una introducción de la estructura de la actividad microaseguradora, prestando especial atención al entorno regulatorio. Continuamos con una descripción de todo el proceso del diseño del producto, haciendo una parada especial en la limitación de información y su importancia como fuente de conocimiento para desarrollar el negocio. Para terminar se describen los principales problemas, así como algunas medidas para evaluar la eficacia del producto.

Parte II: Estudio empírico. Esta parte consiste en todo el proceso práctico de aplicar la metodología predictiva de tarificación basada en los modelos lineales generalizados, en concreto el modelo Tweedie, a un caso real de estudio, un microseguro combinado de Filipinas.

- *Capítulo 3: Metodología de la investigación.*

Este capítulo supone la fundamentación metodológica para el desarrollo de esta investigación. Se empieza resaltando la importancia de mejorar la tarificación de los microseguros. Seguimos con un repaso por las principales metodologías utilizadas en este negocio. Y finalmente, se justifica la selección del modelo Tweedie como la técnica predictiva de tarificación más adecuada para nuestro caso de estudio.

- *Capítulo 4: Tratamiento de datos.*

Se describe con detalle el tedioso proceso que conlleva todo el tratamiento de los datos. Comenzaremos con la ardua tarea de obtener la información y su posterior tratamiento hasta llegar a disponer de una base de datos fiable. Posteriormente, se realiza un análisis exhaustivo de los datos que permitirá ir seleccionando las variables esenciales que se van a integrar en el siguiente análisis predictivo.

- *Capítulo 5: Análisis predictivo.*

Es en esta sección donde se pone en práctica la metodología de tarificación, previamente justificada, para nuestro caso particular del microseguro combinado de Filipinas. Se inicia con la descripción sobre la implementación del modelo Tweedie en la herramienta informática SAS, adaptado a nuestro estudio. Después, se lleva a cabo, paso a paso, todo el procedimiento de análisis predictivo hasta elegir el modelo óptimo que sea capaz de proporcionar una tarifa lo más ajustada al perfil de riesgo del colectivo analizado. Finalmente, se pretende dar un paso más allá y demostrar las ventajas de nuestro modelo, realizando una comparación con el modelo real que utiliza actualmente la compañía.

Para finalizar el estudio, se resumen las principales *conclusiones* derivadas en la realización de este trabajo de investigación en las que se valora el cumplimiento de los objetivos planteados. Y como broche final, se sugieren algunas recomendaciones sobre futuras líneas de investigación a las que puede dar pie la investigación realizada en la presente tesis.

PARTE I. MARCO TEÓRICO: LOS MICROSEGUROS

CAPÍTULO 1. LA PERSPECTIVA SOCIAL

Este primer capítulo trata de adentrarse en este nuevo mundo de los microseguros, ofreciendo la perspectiva social y humana que brinda protección al sector más vulnerable, la población de bajos ingresos.

Para iniciar este primer acercamiento, vemos esencial comenzar el capítulo describiendo el contexto social en el que nos encontramos con el único objetivo de tomar conciencia de la palpable necesidad que existe por favorecer a las personas más desfavorecidas. Con esta visión se está reclamando la implicación de todos nosotros, y desde los diferentes niveles de la sociedad, para tomar acciones a favor del desarrollo de esta población. En torno a estas medidas, entraremos de lleno ya a contar el origen y las características del microseguro como producto asegurador inclusivo. En ellos, se destaca su indudable contribución al desarrollo económico, social y cultural de la población como herramienta de protección social. Posteriormente, profundizaremos en el crecimiento experimentado por este nuevo mercado, examinando las diferencias en su evolución en los tres principales continentes, Asia, África y América Latina. Y finalmente, explicaremos los principales productos y modalidades que se comercializan en la actualidad, y aunque comparten muchas similitudes con el seguro tradicional no dejan de ofrecer una cobertura más sencilla por la población objetivo a la que va dirigida.

1.1 Introducción: contexto social

Sin lugar a duda, el contexto juega un papel fundamental en la actividad de cualquier negocio. Y este es aún más significativo dentro del segmento de población al que se refiere este estudio. Se trata de personas que están expuestas a la ocurrencia de múltiples riesgos. Riesgos que producen un efecto negativo en su salud y en su bienestar general. Es obvio que aquellos que disponen de menos ingresos estén expuestos a mayor número de riesgos, debido tanto a sus limitados recursos económicos como a sus peores condiciones de vida. Dicho de otro modo, estamos ante un segmento de población especialmente débil y desprotegido, agravado por su limitado acceso a los recursos de gestión y prevención de riesgos.

De acuerdo con las cifras del Banco Mundial¹ en el año 2014 tan solo el 62% de adultos de la población mundial era titular de alguna cuenta bancaria formal, bien a través de un banco, institución financiera o de algún proveedor que opera vía móvil. Y más grave aún, según el Banco Mundial (2013) 766 millones² de personas no dispone de recursos suficientes y permanece en una lucha constante por sobrevivir, debido también, y no en escasa medida, a las presiones financieras que vienen experimentando desde siempre. Su trayectoria en zigzag está marcada por períodos de crecimiento, cuando reciben ingresos de forma irregular, seguido por períodos donde deben hacer frente a cargas económicas que muchas veces superan a los ingresos. Esta situación pone de manifiesto la vulnerabilidad asociada a estas personas de bajos ingresos en la medida que no disponen de suficientes recursos para gestionar los riesgos (Churchill, 2006).

En el debate global sobre medidas destinadas a reducir la pobreza y la vulnerabilidad de este sector poblacional, para así poder contribuir a un desarrollo económico sostenible, se reconoce el acceso universal a la seguridad social y a la protección social. En esta línea, se necesita alguna mayor facilidad financiera que mejore la distribución de recursos y promueva nuevas oportunidades en un proceso de globalización más justo. Se hace cada vez más imperativa una protección social como mecanismo integral y colectivo destinado a reducir la vulnerabilidad, la pobreza y la desigualdad, además de proporcionar un fácil acceso a la asistencia sanitaria, una

¹ La base de datos *Global Findex*, creada por el Banco Mundial en el año 2011, recoge información financiera en materia inclusiva (Demirgüç-Kunt et al., 2015).

² Número de personas pobres por debajo de la línea de pobreza de 1,9\$ al día.

cierta seguridad en los ingresos y una mayor accesibilidad a los servicios sociales básicos (Churchill, 2006).

Adicionalmente, la protección social se concibe también como un factor productivo³. Y esto por tres motivos principales. En primer lugar, porque ayuda a la gente a asumir los riesgos importantes de la vida para poder hacer frente a la insuficiencia de ingresos y eventualmente a la pérdida parcial o total de los mismos. En segundo lugar, porque es un instrumento crítico en la gestión del cambio en la economía y en el mercado laboral. Y finalmente, porque contribuye a estabilizar la economía, convirtiendo así la política social en una estrategia para el desarrollo global. (ILO, 2005 y 2006).

Este enfoque social y humano en el que se mueve la presente investigación, encaja perfectamente con las múltiples iniciativas que, cada vez con más urgencia, están promoviendo organismos nacionales e internacionales. Un ejemplo claro es el innovador programa “*Saving for Change*”, diseñado por Oxfam América, Freedom for Hunger y Fundación Stromme, que anima a fomentar la cultura del ahorro entre la población de bajos recursos con la ventaja añadida de una siempre útil educación financiera (Oxfam América, 2013). Este programa, que en el año 2012 contaba ya con casi 680 mil miembros distribuidos en 13 países, consiste en la formación de comunidades de mujeres con el propósito de crear un fondo de ahorro comunitario mediante la aportación regular de todos los miembros, de manera que, llegado el momento, puedan tener la opción de pedir al fondo un préstamo con un bajo interés. Un hecho que ya de por sí es un buen indicador de la creciente educación financiera que experimenta este colectivo.

En este campo de la educación, cabe reseñar la iniciativa de un grupo de universidades españolas destinada a promover el proyecto internacional *Aristos Campus Mundus*⁴ cuyo objetivo es potenciar la excelencia académica a través de proyectos de investigación dentro de su área de estudio de la innovación social, responsable y sostenible.

³ Véase 8º elemento fundamental de la Agenda de Empleo Global (ILO, 2006).

⁴ Proyecto *Aristos Campus Mundus* 2015 impulsado por las universidades de Deusto, Comillas y Ramón Llull.

La iniciativa más relevante en este ámbito de la protección social de la población menos favorecida viene liderada por Naciones Unidas en alianza mundial con sus Estados miembros mediante su ambiciosa Agenda 2030⁵ para el Desarrollo Sostenible, que establece un plan de acción para afrontar el gran desafío de lograr un desarrollo sostenible centrado en las necesidades de los más pobres y vulnerables. Entre los diecisiete Objetivos de Desarrollo Sostenible establecidos por Naciones Unidas (2015: 22), el octavo habla expresamente de “[p]romover el crecimiento económico sostenido, inclusivo y sostenible, el empleo pleno y productivo y el trabajo decente para todos”, y establece la meta de “[F]ortalecer la capacidad de las instituciones financieras nacionales para fomentar y ampliar el acceso a los servicios bancarios, financieros y de seguros para todos”.

En esos “seguros para todos” entran también y cada vez con más fuerza los microseguros, que fueron concebidos para el colectivo de personas con bajos ingresos. Esta denominación es la que se viene utilizando y se empleará durante este trabajo para designar a la población más vulnerable que dispone de escasos recursos económicos y que tiene un acceso mucho más limitado a las medidas formales de prevención de riesgos. Para determinar dónde se encuentra este segmento de población, se tiene en cuenta el umbral internacional de la pobreza establecido por el Banco Mundial⁶ cuyo dato más reciente lo sitúa en 1,9\$ al día. Por su parte SwissRe (2010), estima que las personas situadas en el rango comprendido entre 1,25\$ y 4\$ al día representan el mercado potencial objetivo para productos de microseguros, que con datos actualizados del Banco Mundial (2013) alcanzan la impresionante cifra de 2.340 millones de personas en el mundo.

El evidente crecimiento de las microfinanzas, y en particular los microseguros que juegan ya un papel crucial dentro de los objetivos de desarrollo, propicia un cambio de perspectiva hacia una mayor inclusión financiera y abre un amplio campo en este numeroso y potencial segmento de población de bajos ingresos (Churchill, 2006; Swiss Re, 2010).

⁵ Véase Naciones Unidas (2015). Resolución aprobada por la Asamblea General de la ONU sobre “Transformar nuestro mundo: la Agenda 2030 para el desarrollo sostenible”.

⁶ Estimación de la pobreza mundial del intervalo de tiempo de 1981 al 2013 basado en el índice de paridad del poder de compra (PPP) del año 2011 (Banco Mundial, 2013).

1.2 Origen, definición y características del microseguro

El origen del microseguro está estrechamente ligado, como es obvio, al nacimiento del seguro. Sabemos que este nació como mecanismo de transferencia de riesgo por medio de un sistema mancomunado de protección de riesgos. Pero fue a partir del siglo XIX cuando empezaron a comercializarse pequeñas pólizas de seguros industriales a modo de esquemas de protección mutua que se comercializaban a la entrada de las fábricas. La proliferación del seguro industrial contribuyó en gran medida al desarrollo de la cultura del seguro y a fomentar el desarrollo sostenible de la economía (Roth et al., 2007). Estos avances logrados con el seguro industrial se identifican perfectamente con los retos y objetivos que persiguen los microseguros.

Un claro antecedente del actual microseguro lo tenemos en las conocidas cooperativas y uniones de crédito, que en algunos casos desembocaron en la creación de sus propias compañías de seguros. Pero el momento más decisivo del origen del microseguro hay que fijarlo en la fecha aún reciente de los años 90, cuando las instituciones microfinancieras (IMF de ahora en adelante) desarrollaron el microseguro como parte de su negocio microfinanciero más allá del conocido microcrédito, con el fin de cubrir las necesidades asegurables de sus clientes de bajos recursos. Es en este momento cuando se empiezan a comercializar los primeros microseguros de vida crédito y de protección de préstamos, ligados a productos microfinancieros y a las instituciones de microfinanzas debido a la red de distribución que disponían y que hacía menos costoso el producto (Siegel et al. 2001). Surge entonces el microseguro como una extensión de las microfinanzas, enfocada a la gestión de bajos riesgos con un coste más reducido.

Entre los diversos servicios financieros diseñados para la población de bajos ingresos, el más conocido y el de mayor aceptación fue el microcrédito, especialmente desde que su creador Mohammed Yunus y el banco Grameen recibieron el premio Nobel de la Paz en el año 2006. El microcrédito se distingue de manera señalada por su objetivo humanitario de ayudar a salir de la pobreza a los más desfavorecidos, facilitando recursos necesarios para crecer económicamente por sí mismos. En cambio, el microseguro contribuye más bien a garantizar que este crecimiento sea sostenible en el tiempo, además de prevenir la recaída en situación de pobreza (Lloyd's, 2009). De esta comparación entre el microcrédito y el microseguro es sencillo

deducir que ambos productos comparten el mismo objetivo, el de ayudar a corregir las propias imperfecciones del mercado y dar servicio a la población de bajos recursos existente en los países en desarrollo.

Sin embargo, hay claras diferencias entre ambos productos. De Bock y Gelade (2012) exponen varios motivos que explican la complejidad del microseguro con respecto al microcrédito. Primero, porque el microseguro implica un pago anticipado y cierto a cambio de recibir una indemnización futura y de algún modo incierta, de la que algunos se beneficiarán y otros no. Segundo, porque debido a la misma heterogeneidad del producto, el microseguro adopta diversas formas en razón de los diferentes riesgos que se propone cubrir. A dicha complejidad se unen otras características diferenciadoras bien significativas que cabe señalar. Mientras el microcrédito busca aportar los medios necesarios para favorecer la creación de riqueza, el microseguro se limita a suministrar medios que eventualmente puedan aminorar los efectos derivados de riesgos adversos. Otra importante diferencia reside en su nivel de maduración, que está más consolidado en el mercado de los microcréditos que en el de los microseguros. De ahí que, dada la experiencia del microcrédito y teniendo en cuenta el objetivo común que comparte con los microseguros, son varios los autores (Cohen & Sebstad, 2005; Swiss Re, 2010) que recomiendan a la industria microaseguradora que preste atención a las lecciones aprendidas del microcrédito.

Antes de entrar en el concepto o la definición del microseguro que ofrece la literatura, no será ocioso recurrir a la propia etimología de la palabra microseguro. Vemos que está compuesta de dos elementos: "micro" y "seguro". El primero, "micro", de origen griego, equivale a pequeño y hasta muy pequeño. El segundo elemento, "seguro", derivado del latín, tiene un significado claro por sí mismo, el de ser algo que da seguridad. Según este simple criterio, aplicado al ámbito asegurador, vemos que, por una parte, la medida pequeña se hace extensible tanto a la cobertura y prestación recibida como a la prima pagada, y por otra el seguro es un contrato que ofrece protección frente a riesgos específicos. Por tanto, una primera definición lógica y clara podría formularse así:

El microseguro es un producto básico con coberturas limitadas que ofrece protección frente a riesgos específicos a cambio de una prima más reducida respecto al seguro tradicional y que es proporcional al coste del riesgo asumido por el asegurador.

Dado que el concepto de microseguro está en continua evolución, no existe por el momento una definición comúnmente aceptada (Swiss Re, 2010). No obstante, en la literatura revisada aparecen diversas formas de explicar el significado de microseguro según sea la perspectiva considerada: bien como instrumento de protección de riesgos mancomunados para las familias de bajos ingresos, o bien como un seguro de pequeñas indemnizaciones y bajos niveles de prima. Otras veces se añade referencia expresa al colectivo de trabajadores al que va dirigido y que se mueve en una economía informal. Algunas instituciones, asociaciones y profesionales del sector han dado su propia definición. Entre ellas cabe mencionar las siguientes:

La Asociación Internacional de supervisores de seguros lo define de la siguiente forma (IAIS & CGAP, 2007:10):

Microseguro es el seguro accesible a la población de bajos ingresos, proporcionado por una serie de entidades de acuerdo con los principios y prácticas generalmente aceptados en seguros.

Con esta definición la IAIS desde su perspectiva reguladora, pone de manifiesto que la actividad microaseguradora entra en el ámbito de regulación y supervisión del seguro y bajo las leyes nacionales o según cada jurisdicción.

Craig Churchill, director del programa social y financiero de la Organización Internacional del Trabajo (ILO) y que dirige varias iniciativas para impulsar el desarrollo económico y social del seguro, editó el primer compendio de microseguros (Churchill, 2006:12) donde define el microseguro como:

La protección de las personas de bajos ingresos frente a riesgos específicos, a cambio del pago de una prima periódica proporcional a la probabilidad de ocurrencia y el coste del riesgo cubierto.

Con la idea de profundizar en el concepto de microseguro, es importante fijar sus diferencias con el seguro tradicional. Dentro de su convergencia en el mismo objetivo general de ofrecer cobertura y dar protección frente a riesgos específicos atendiendo a los principios básicos del seguro, cada producto tiene su propia singularidad. Una buena parte de la literatura existente en microseguros (Cohen & Sebstad, 2005; Churchill, 2006; Garayoa, 2011; Swiss Re, 2010) establece sus características en

base a los rasgos diferenciadores respecto al seguro tradicional. Entre los aspectos relevantes de los microseguros se destacan los siguientes:

- *Atención al mercado de bajos ingresos.* La tipología de mercado objetivo siempre es determinante en cualquier negocio. El rasgo distintivo del microseguro viene marcado por el enfoque de protección social al propio mercado al que va dirigido, que es la población de bajos ingresos con un conocimiento financiero muy limitado. Y es precisamente este mercado el que va a diferenciar y condicionar muchos aspectos del producto. En cambio, el seguro tradicional está enfocado a la población con un nivel adquisitivo mayor y que posee más conocimiento de los productos financieros existentes en el mercado.
- *Integración de la población vulnerable.* El microseguro ofrece cobertura a la población más expuesta a los riesgos imprevistos. Y por ser personas de bajos ingresos, es relevante que la cobertura del producto se identifique con las necesidades y preferencias del cliente. El éxito del producto se logra cuando se cumple esta relación. La característica diferenciadora en cuanto al perfil del cliente, es que el microseguro es inclusivo, es decir, se propone no excluir a las personas con mayor riesgo para que el enfoque de protección social tome sentido.
- *Accesibilidad y sencillez del producto,* tanto en la forma como en el fondo. Es un producto enfocado a un perfil de personas con una limitada educación financiera y con una capacidad económica reducida. Por esto mismo, es comprensible su desconfianza ante la oferta de nuevos productos, por lo que hace necesario potenciar la sencillez del producto, con un trato más cercano entre el cliente y el asegurador, y una mayor rapidez en el cobro de las prestaciones. Estos aspectos se tornan mucho más complejos en el seguro convencional donde los trámites resultan más largos y tediosos debido normalmente a un mayor requerimiento de documentación.
- *Asequibilidad de la prima para garantizar el acceso al producto y flexibilidad en el cobro de primas.* La tarificación del microseguro suele realizarse a nivel colectivo aprovechando grupos ya formados, como es el caso de las asociaciones de mujeres o las cooperativas. De este modo, se garantiza que la selección de miembros y la distribución del riesgo ya ha sido previamente realizada por el grupo, obteniendo así una prima más reducida para el colectivo. Además, se hace más fácil el cobro de prima al estar adaptado a los flujos

irregulares de ingresos que son habituales en este segmento de mercado. Por el contrario, en el seguro convencional la tarificación suele estimarse en base a múltiples factores de riesgo debido a la calidad de la información.

Muchos de estos aspectos representan decisiones difíciles de cara a lograr un equilibrio entre una cobertura inclusiva a un precio razonable y que a la vez suponga un coste mínimo para el asegurador. Con vistas a lograr este objetivo de equilibrio entre oferta y demanda, McCord (2011) desarrolla una sencilla regla nemotécnica con las siglas inglesas *SUAVE*, cuyas letras representan los atributos necesarios para garantizar el éxito de un microseguro:

- *Simple*, simplicidad y claridad del producto y fácil de gestionar por los clientes.
- *Understood*, que sea comprensible tanto el producto como su funcionamiento por parte del público al que se dirige.
- *Accesible*, de fácil acceso a los clientes en la obtención de información y en el propio funcionamiento del producto.
- *Valued*, en el sentido que aporte valor y sea así apreciado por el cliente porque se ajusta a sus verdaderas necesidades.
- *Efficient*, eficiencia en el coste de distribución en los diversos canales y fácilmente entendible en su gestión.

En resumen, el microseguro que comparte muchas de las características del seguro ordinario se plantea como una nueva corriente centrada en la protección social y enfocada a las personas más vulnerables.

Dando un paso más en la configuración conceptual del microseguro, vemos que se convierte en **una formidable herramienta de prevención de riesgos**. Por riesgo se entiende todo acontecimiento imprevisible que afecta a la estabilidad personal, económica o social de cualquier persona. Paradójicamente, de toda la pirámide de población, el segmento de bajos ingresos es el que se enfrenta a un mayor número de riesgos al no tener acceso a los cauces formales de financiación y aseguramiento. Como resultado, el bienestar de las personas pobres está en continuo vaivén ante los numerosos sucesos que desequilibran su economía y que hacen peligrar su capacidad financiera para hacer frente a estos gastos imprevisibles (Churchill, 2006).

Esta situación de patente vulnerabilidad que afecta a la población de bajos recursos, está inversamente relacionada con la capacidad para gestionar los riesgos⁷. De hecho se observa que cuanto más alto es el grado de vulnerabilidad, menor es su capacidad de respuesta al riesgo al que están expuestos. En consecuencia, se hace más imperiosa la necesidad de buscar nuevos instrumentos, en su mayoría financieros, como medida de protección. Roth et al. (2007) clasifican la gestión del riesgo en dos categorías según el momento temporal. Por un lado, están las estrategias ex-ante, que recogen aquellas acciones previas realizadas para reducir la probabilidad de que ocurra un determinado riesgo. Por otro lado, las estrategias ex-post que describen las medidas tomadas una vez ya ocurrido el siniestro.

La gestión del riesgo comienza con la tarea de identificar qué riesgos suponen mayor vulnerabilidad para la población objetivo. Entre ellos, la enfermedad, la muerte de un familiar, la pérdida de bienes, un robo, un incendio y otros similares. El siguiente paso consistiría en priorizar aquellos riesgos que supongan una mayor repercusión en el bienestar de las personas para acotar bien sus necesidades. El orden de prioridad de riesgos, que en los países desarrollados suele ser variable, tiende a ser consistente en la mayoría de países en desarrollo. En primer lugar la salud, por la ocurrencia de una enfermedad, fallecimiento o incapacidad, que son casos que suponen una mayor preocupación por el doble efecto derivado de la pérdida de ingresos y por los gastos incurridos hasta recuperarse de dicha situación o enfermedad. Y después le siguen todos aquellos riesgos que suponen una reducción en los ingresos, como son los desastres naturales, el robo y las pérdidas derivadas del seguro agrícola y ganadero.

Todos ellos son riesgos donde el seguro ofrece cobertura y protección, y es aquí precisamente donde el microseguro se plantea como una herramienta eficaz de prevención de riesgos. Esto significa una clara exigencia de cambio de actitud frente al riesgo, pues el cliente acostumbra actuar cuando ya se ha producido el siniestro, en vez de tomar una actitud proactiva, de anticiparse a contratar mecanismos que minimicen el impacto de los previsibles riesgos futuros.

Por el lado de la demanda de microseguros, parece obvio que, si los individuos muestran un alto nivel de aversión al riesgo, la demanda aumentará, con el fin de reducir su futuro riesgo. Sin embargo, y extrañamente, las cifras no dicen lo mismo, ya

⁷ Churchill (2006) asocia la vulnerabilidad a la habilidad de los individuos a gestionar el riesgo.

que tan solo existe un 30% de suscripción a programas de microseguro subsidiado y los ratios de renovación del producto son muy bajos. Paradójicamente, se llega a la conclusión planteada por De Bock y Gelade (2012) respecto a que la demanda de microseguros está negativamente correlacionada con la aversión al riesgo, cuando debería ser un instrumento claro de gestión del riesgo.

1.3 Contribución al desarrollo económico, social y cultural

El microseguro como instrumento de protección de riesgos para la población de bajos recursos, contribuye de manera especial al desarrollo macroeconómico de países emergentes. Las familias que contratan este tipo de productos se ven entonces protegidas ante la ocurrencia de riesgos que puedan mermar su flujo de ingresos, alejando así el peligro de recaer en situación de pobreza. La protección que ofrece el microseguro se traduce además en una oportunidad de creación de valor social y de una apetecible situación de bienestar, que se hace visible en el entorno familiar, con una mayor calidad de vida ya sea en la salud, en la educación de los hijos o incluso en un ahorro a largo plazo. El acceso al producto por las personas de bajos recursos posibilita la contratación de otros productos financieros que les proporcionan un mayor crecimiento económico.

Un ejemplo de esto son los productos de ahorro a largo plazo, donde los clientes confían sus ahorros a una institución como una inversión futura a la vez que las compañías de seguros consiguen disponer de un nuevo capital a invertir en diferentes valores de la economía global. En consecuencia, nos empezamos a encontrar con un hecho alentador con el que el segmento de población vulnerable comienza a participar en la cadena de generación de valor con opciones de inversión ya dentro del nivel macroeconómico.

Resulta altamente instructivo ver cómo la literatura existente revisada (Churchill, 2006; Garayoa, 2011; Swiss Re, 2010) describe aspectos y directrices de cómo pueden contribuir los microseguros al desarrollo socio-económico y cultural de los países emergentes. Son especialmente sugerentes los aspectos que proponen y que se recogen a continuación.

- *Educación financiera para lograr la confianza del cliente.* El consumo de productos inclusivos, tales como son los microseguros, contribuye a expandir el conocimiento sobre cómo funciona el seguro y de qué forma puede beneficiarles a los potenciales clientes. Ocurre que el mayor nivel de conocimiento financiero del cliente influye en gran medida en la decisión de compra del producto, debido sobre todo a la percepción del posible beneficio. De hecho, uno de los objetivos, y no el menor, para reducir la desconfianza al producto es crear estrategias de educación, enfocadas no solo a proporcionar conocimientos técnicos sino también a poner en valor su significación social, contribuyendo así, lenta pero eficazmente, a un cambio de mentalidad que los haga capaces de enfrentarse con éxito a los aspectos culturales de su entorno.

- *Fomenta el bienestar y la protección medioambiental.* El microseguro se extiende a una variedad de productos y coberturas diversas, ya sean de vida, de salud, de riesgos agrícolas y otros similares. Poder gozar de una vida duradera y saludable no solo depende de prevenir riesgos visiblemente dañinos, sino también de evitar algunos otros que puedan venir camuflados, por ejemplo, en el aire tóxico, en el agua contaminada o en la propia alimentación. En este punto, adquieren especial protagonismo los micro-productos de salud que, aunque su objetivo principal es proporcionar el servicio médico apropiado, también incluyen otros servicios como promoción de salud e higiene entre sus clientes. Medidas como estas suponen una clara ganancia para todos, además de mejorar la salud y aumentar la productividad de los clientes y repercute también en una reducción del coste del servicio del producto.

Por otro lado, el riesgo medioambiental, al que el mundo de hoy es especialmente sensible, está captando también la atención de las compañías de seguros y reaseguros debido a los efectos posiblemente devastadores del cambio climático. Los cambios meteorológicos no solo influyen en la vida y salud de las personas sino también en el mayor número de riesgos catastróficos, como sequías, terremotos o tifones y otros eventos incontrolables de la naturaleza, que inciden con mayor fuerza en países emergentes (Churchill, 2006). El microseguro aparece aquí de nuevo ofreciendo productos que pueden dar alguna cobertura frente a tales eventos imprevisibles, limitando en alguna medida las pérdidas ocasionadas.

- *Reducir la desigualdad de mujeres y niños.* Las mujeres y niños suelen ser el sector más vulnerable de la población en vías de desarrollo. En cuanto a las mujeres, las cifras señalan que ellas representan un 70% de los pobres del mundo (Mosleh & Ramm, 2006). Este colectivo se enfrenta a riesgos de mayor virulencia que los hombres, no solo en los abusos sexuales, sino también en la explotación en el trabajo con un menor salario, en una salud más vulnerable, en el más restringido acceso a educación y hasta en su malnutrición. Es evidente que las soluciones a estos problemas hacen necesaria la intervención estatal, pero esta a veces no llega o es insuficiente. Se hace necesario recurrir a otros mecanismos como los microseguros, que actúan como medida de protección y en parte pueden contribuir a reducir esta vulnerabilidad y mejorar las condiciones de vida tanto de ellas como de los niños.

En este colectivo con mayor vulnerabilidad piensan también las compañías aseguradoras cuando diseñan sus productos para cubrir sus necesidades más imperiosas. Por ejemplo, en seguros de salud se incluyen las coberturas de embarazo y parto, las enfermedades de transmisión sexual y los programas de cuidado para sus niños. Por eso, en productos de vida, se busca evitar desprotección de la familia ante el fallecimiento del cónyuge, se modifica el pago de la indemnización en forma de renta para evitar que los niños tengan que abandonar la escuela. Al igual ocurre en el pago de primas, se flexibiliza en pequeñas cuotas según la situación financiera de la mujer.

- *Proporciona el acceso a grupos de la sociedad excluidos.* En los países emergentes, un buen sector de la población lo conforman los trabajadores de la economía informal y los trabajadores rurales. Un sector que vive ajeno al mundo del seguro. Pues bien, con el producto de los microseguros se facilita el acceso a estos colectivos más marginados para que puedan entrar a formar parte de esta cadena de valor. Muchos de estos productos se basan en esquemas colectivos a través de comunidades, con lo que no solo se reducen los costes de transacción, sino que además se combaten los riesgos asociados al uso abusivo del producto y al fraude.

Una vez explicadas las características propias del microseguro, se hace evidente la positiva contribución del producto como herramienta de protección social. Así y todo, no está demás decir que no todos los productos ofrecidos contribuyen de la misma forma a este objetivo de la protección social, ya que todo dependerá de las diferentes condiciones y necesidades de la población. Este es el caso de los microseguros

vinculados a garantizar el crédito, así como los de hogar y también aquellos que protegen el ganado o el cultivo. Todos ellos benefician en cierta forma a la persona, pero no todos ejercen la función de protección social al modo como lo hacen, y con notoria eficacia, los productos de salud, vida, accidentes y los que ofrecen pensiones para mayores (Jacquier et al., 2006).

1.4 Panorama del mercado microasegurador

Es un hecho constatado que los microseguros han experimentado un incremento sustancial en los últimos años en los países en vías de desarrollo. Están difundidos principalmente en los tres continentes Asia, América Latina y África. El primer estudio elaborado por Roth et al. (2007) sobre los cien países más pobres del mundo, detecta que en 77 de ellos se identificaron evidencias de alguna cobertura de microseguro con un volumen de 78 millones de personas. En el año 2009 esta cifra ascendía a 135 millones de asegurados (Churchill & Matul, 2014) y en el 2014, según los estudios de mercado⁸ más recientes, ya son más de 280 millones de personas cubiertas por productos microaseguradores. Este crecimiento se debe a una confluencia de factores que van desde una mayor penetración de las microfinanzas gracias al papel más activo de los gobiernos, a una mayor oferta de productos y canales innovadores que se corresponden con las necesidades del mercado objetivo (Churchill & Matul, 2014; Swiss Re, 2010). Esta situación, junto con los objetivos a medio plazo, apunta a un futuro de inclusión financiera donde los microseguros, entre otros productos similares, cambiarán la forma de ver al mercado de bajos ingresos.

A pesar de que el crecimiento de microseguros ha sido generalizado en todo el mundo, sí se aprecian diferencias significativas entre las diferentes regiones.

- **Asia y Oceanía** representan el mayor volumen de población microasegurada con 170 millones de personas en el año 2012. Esto se debe a una mayor densidad de población pero con un ratio de cobertura muy baja, en torno al 4,3%, y con un nivel de renta medio que corresponde a la población de bajos ingresos. Los principales países donde está más extendida la cobertura del microseguro son India, China y Filipinas, con un fuerte crecimiento en los últimos años (Fundación Munich Re, 2013). El papel de liderazgo que mantiene

⁸ Véase Microinsurance Centre (2015c) disponible en <http://worldmapofmicroinsurance.org/#> .

Asia se debe a las grandes poblaciones con alta densidad demográfica y al papel activo de los gobiernos que fomentan el consumo de estos productos.

- **África** es la segunda región con mayor penetración de microseguros, la cifra recogida en el año 2014 asciende a 61,7 millones de asegurados y representa un ratio medio de cobertura del 5,4%. Sudáfrica es el país con más de la mitad de población asegurada, seguido de Ghana y Zambia (Microinsurance Network, 2015a). África destaca además en dos aspectos, por un lado dispone del mayor número de programas de microseguros debido a la proliferación de proveedores basados en comunidades. Y por otro lado, lidera la mayor parte de iniciativas digitales en microseguros, en parte motivados por la necesidad de reducir los elevados costes de adquisición derivados de su baja densidad de población (BFA, 2015).

- **América Latina y Caribe** que cuentan con el nivel de penetración de microseguros más bajo, concretamente con 48,5 millones de asegurados en el 2013, poseen el ratio de cobertura medio más alto de los tres continentes con un 7,9%. Los países con mayor cuota de mercado son México, Brasil, Colombia y Perú (Microinsurance Network, 2015b). El crecimiento en esta región se debe en parte a una ampliación del mercado objetivo al que van dirigidos los microseguros incluyendo también la clase media baja (Churchill & Matul, 2014). Esta ampliación supone un trasvase del negocio microasegurador al seguro masivo. De hecho en Brasil entre los años 2011 y 2013, se traspasaron 27 millones de dólares en primas de microseguros a los seguros masivos. La razón principal de este movimiento viene marcada por la disponibilidad y motivación de los canales de distribución que encuentran mayor accesibilidad y facilidad en los seguros masivos.

Estas cifras muestran sin lugar a duda un crecimiento esperanzador de los microseguros en los distintos continentes. La información de estos estudios es primordial para conocer la evolución de un producto todavía novedoso, que irá desarrollándose a medida que el mercado madure, y así podremos obtener un análisis más real de la situación, avalado también por una mayor calidad de los datos recogidos (McCord et al., 2014).

1.5 Tipología de microseguros

Los microseguros comprenden un amplio abanico de productos similar al existente en el seguro tradicional. Pero si nos referimos al mercado objetivo centrado en países emergentes, vemos que es diferente al del mercado tradicional. Una característica que condiciona fuertemente las coberturas de cada producto, siendo el nuestro un producto enfocado de modo específico a las necesidades básicas inherentes de la población de bajos ingresos.

Con esta idea por delante, no está demás echar una mirada al panorama global de la tipología de microseguros en las diferentes regiones tal como se presenta en el siguiente resumen.

Tabla 1.1. Panorama global de asegurados⁹ por producto y región.

Producto	Asia y Oceanía		África		América Latina		Total asegurados (miles)
	Asegurados (miles)	Ratio cobertura	Asegurados (miles)	Ratio cobertura	Asegurados (miles)	Ratio cobertura	
Vida ⁽¹⁾	83.861	2,13%	46.376	4,08%	32.489	5,28%	162.726
Salud	29.188	0,74%	8.360	0,74%	7.636	1,24%	45.184
Accidentes	77.832	1,98%	13.113	1,15%	20.947	3,40%	111.891
Propiedad	7.671	0,19%	4.522	0,40%	2.192	0,36%	14.385
Agrícolas	23.757	0,60%	1.129	0,10%	2.246	0,36%	27.132
Total ⁽²⁾	170.393	4,33%	61.760	5,43%	48.577	7,89%	280.729

⁽¹⁾ Excluye vida crédito

Fuente de elaboración propia

⁽²⁾ El total producto incluye sólo la garantía principal, ya que muchos asegurados están cubiertos por más de un tipo de riesgo

Lo que estos datos muestran es que el producto de vida (excluido el seguro de crédito) es el más extendido y con la mayor cuota de mercado, 162 millones de asegurados, posiblemente por la sencillez del producto. En segundo lugar, el seguro de accidentes con 111 millones de asegurados. Y después el de salud, que cuenta con 45 millones de clientes y emerge como el producto más demandado debido lógicamente al deseo de la población a no sufrir ningún tipo de enfermedad. Y ya en las últimas posiciones están los productos agrícolas y de propiedad, que presentan los niveles más bajos de penetración. Puede sorprender el dato de que las cifras totales son inferiores a la suma desglosada por producto. Esto se debe a la existencia de productos combinados que ofrecen cobertura en diferentes tipos de producto. Existen casos concretos en los que tales combinaciones de coberturas no están amparadas bajo la ley dentro del

⁹ Datos obtenidos de Microinsurance Centre (2015c).

mercado de seguros ya que las compañías no disponen de la correspondiente licencia para todos las líneas de negocio.

Antes de pasar a describir el alcance de cada tipo de microseguro, puede ser orientativo tener delante un resumen general con las principales líneas de negocio y su correspondiente tipología de productos.

Tabla 1.2. Visión general de la tipología de productos.

Línea de negocio	Tipología productos
Vida	Vida-crédito
	Vida-protección
	Vida-ahorro
Salud	Asistencia sanitaria
	Indemnización
	Enfermedades graves
Generales	Hogar
	Comercio
	Otros bienes
Agrícolas	Basados en indemnizaciones/ingresos
	Basados en índices

Fuente de elaboración propia

1.5.1. Microseguro de vida

Vida es sin lugar a dudas la línea más popular de productos, la de mayor demanda y con un volumen de asegurados igual o superior a todo el resto de negocio microasegurador. El crecimiento de este mercado se debe a la confluencia de una serie de motivos, de los que el primero de todos viene impulsado por un crecimiento añadido de las instituciones microfinancieras, que favorece su distribución por su fácil vínculo con los microcréditos y otros productos microfinancieros. El segundo motivo, la tarificación que es relativamente sencilla en comparación a otros productos. Otra razón es la resistencia frente a riesgos que persisten en este mercado, como son el fraude y el uso abusivo del producto, ya que la cobertura depende básicamente del fallecimiento del asegurado. Todos estos aspectos favorecen las características básicas que debe prevalecer en el microseguro, como son su sencillez, mayor facilidad en conocer el producto y una mayor rentabilidad que ofrece tanto a los proveedores como los distribuidores (Roth et al., 2007; Swiss Re, 2010).

Descendiendo al nivel de coberturas, el seguro de vida ofrece protección al asegurado frente al caso de fallecimiento o supervivencia. La población de bajos ingresos está

más expuesta a sufrir dificultades a la hora de afrontar la muerte de un familiar, con la consiguiente reducción de ingresos en la economía familiar. Por este motivo existe una tipología de productos adaptada a las necesidades de esta población de bajos ingresos. Con las siguientes modalidades principales:

- Vida temporal. Comprende productos de *vida-crédito* y *vida-protección* que ofrecen cobertura durante un período determinado de tiempo que suele ser variable, puede ir desde un año o hasta que el asegurado alcance una determinada edad. La tarificación y gestión de riesgos en este tipo de productos suele ser muy sencilla. En la práctica habitual esta cobertura suele estar ligada a otras dando lugar a los conocidos *microseguros combinados*. Entre estos productos destacan los productos de vida-protección con el seguro de decesos y accidentes y productos vida-crédito ligados a microcréditos. En el caso de vida-crédito, modalidad más extendida, no se puede olvidar que está pensada más en satisfacer las necesidades de las IMF que en las de las personas de bajos recursos, lo que hace que pierda el carácter propio de protección social.

- Vida a largo plazo con productos de *vida-ahorro*. El objeto de cobertura es similar al anterior con la diferencia de que el plazo es más extenso. El volumen de asegurados es mucho más limitado debido en parte a que no son útiles para el mercado objetivo de los microseguros, dadas las siempre inestables condiciones macroeconómicas, con devaluaciones de moneda y la inflación de precios propias de los países emergentes. Esta modalidad es preferible para las compañías puesto que les proporciona ingresos regulares durante un largo período de tiempo. Productos que, por otra parte, son más complejos, tanto en su estructura como en su gestión, y que requieren de una mayor monitorización actuarial. El producto principal en esta modalidad suele ser el producto de ahorro enfocado a la jubilación, que combina un seguro de vida temporal con un sistema de ahorro hasta la edad de jubilación.

A pesar de que los seguros temporales con el vida-crédito y seguro funerario lideran el mercado de microseguros, hay que subrayar una mayor concienciación por parte de las compañías de seguros para diseñar productos más enfocados a la protección integral, a canalizar los ahorros y a una potenciación del bienestar (Swiss Re, 2010).

1.5.2. Microseguro de salud

La salud es sin lugar a dudas el motivo de mayor preocupación para cualquier persona, y que se acentúa de modo especial en el caso de familias con bajos ingresos. Para ellos los problemas de salud causan una doble pérdida de ingresos, por un lado el coste del tratamiento no siempre fácil de asumir y por otro lado, la pérdida de ingresos por su inactividad laboral. Por estos motivos, una de las primeras necesidades en materia de salud para esta población vulnerable es disponer cuanto antes del servicio médico para así evitar un coste superior. En consecuencia, el microseguro de salud ofrece un considerable alivio para las familias, a la vez que supone un alto valor añadido en este nicho de mercado tanto por la frecuencia del riesgo sanitario como por el coste que esta atención implica. Esta situación favorece que este tipo de riesgos sitúen al microseguro de salud entre los más demandados del mercado.

Por otro lado, el producto de salud es bastante complejo, ya que depende principalmente del proveedor sanitario y de la infraestructura necesaria para la prestación del servicio, lo que nos lleva a analizar en detalle los factores que influyen sobre su viabilidad (Churchill, 2006; Swiss Re, 2010).

En primer lugar y más que cualquier otro producto, el de salud está estrechamente ligado a una tercera parte, el proveedor sanitario, que puede ser tanto público como privado. La eficacia y el funcionamiento del sistema sanitario estatal condiciona en buena medida la demanda del microseguro que tratará de cubrir la ausencia o deficiencia del sistema público (Roth et al., 2007). El microseguro de salud es un buen ejemplo de cómo seguros de estas características pueden coordinarse junto a sistemas de seguridad social públicos, proporcionando así una completa cobertura sanitaria eficiente tanto para el sistema público como el privado. En segundo lugar, la disponibilidad de infraestructuras que no solo faciliten una adecuada gestión a través de los proveedores sanitarios, sino que sirvan para aminorar la incidencia de riesgos típicos en los microseguros, como es el fraude y el riesgo moral asociados al uso abusivo del producto. Entre otros factores importantes para la viabilidad del producto, se destaca la necesidad de una adecuada tarificación ajustada en base al perfil de riesgo del colectivo con una primas que controlen uno de los riesgos más preocupantes en salud, la selección adversa, donde los riesgos malos permanecen en la cartera mientras que los riesgos buenos acaban anulando la póliza.

En cuanto a las modalidades de microseguro de salud, siguiendo la clasificación de Wrede et al. (2016), las principales son las siguientes:

- Asistencia sanitaria: es la modalidad más extendida en el mercado de microseguros. El producto proporciona una amplia gama de coberturas en función de la modalidad contratada y que van desde la asistencia primaria, hospitalaria o ambulatoria, que se traduce en asumir el coste sanitario, parcial o completo cuando ocurre un accidente o una enfermedad. Como medida de control, las compañías suelen establecer períodos de carencia en los servicios de mayor coste.
- Indemnización o protección de ingresos: que proporciona cobertura en caso de enfermedad o incapacidad para ejercer su profesión habitual tanto a corto como a largo plazo. En los microseguros, las coberturas de hospitalización suelen ser pólizas a corto plazo donde el producto ofrece cobertura ante la pérdida de ingresos mientras el asegurado está hospitalizado. Según la experiencia recogida y la escasa información disponible, los productos de salud suelen estar bastante limitados a coberturas de indemnización (Roth et al., 2007).
- Enfermedades graves: se trata del producto donde el asegurado recibe cobertura en caso de contraer una de las enfermedades graves previstas en la póliza. La prestación que recibe viene a ser un capital como soporte económico para realizar el correspondiente tratamiento médico-hospitalario.

El microseguro de salud ayuda a mejorar los objetivos¹⁰ de protección social en los países en vías de desarrollo donde la seguridad social escasea o es incluso inexistente. La mayoría de estos productos limitan la cobertura a una serie de tratamientos de asistencia primaria a través de otras instituciones, como ONGs y organizaciones de comunidades, en países donde el Gobierno proporciona la asistencia especializada.

Una de las mayores limitaciones del microseguro de salud atiende al problema de escala con vistas a lograr productos sostenibles. La solución parece estar en una

¹⁰ Estudios realizados demuestran el efecto positivo sobre las personas en situación de pobreza en zonas rurales de Bangladesh (Swiss Re, 2010).

perfecta combinación entre el asegurador, el proveedor y el canal de distribución (Roth et al., 2007).

1.5.3. Microseguros generales

Esta línea de negocio engloba los llamados seguros no-vida y cuenta con productos que ofrecen cobertura frente a daños en el hogar, comercios y otros bienes. A pesar de que la población de bajos recursos está más expuesta a sufrir riesgos que supongan una pérdida parcial o total de sus bienes, la cuestión fundamental es hasta qué punto el público objetivo está dispuesto a invertir en la protección de sus posesiones. Además, desde el lado de la oferta, las compañías aseguradoras tampoco han desarrollado lo suficiente este tipo de productos debido a la dificultad que supone la realización de determinadas tareas, como son la tarificación, la gestión y tramitación de los siniestros, así como gestionar los problemas asociados al uso abusivo del producto y el fraude.

Los datos confirman la evidencia de lo comentado, según un estudio realizado por Roth et al. (2007) se afirma que el 99'3% de la población mundial de bajos ingresos de los cien países más pobres no dispone de ningún tipo de seguro general. Otra razón de esta baja penetración de dicho seguro se debe a que la comercialización suele hacerse más a nivel individual que a nivel de grupo, lo que supone un incremento de los costes de transacción y en consecuencia de unas primas más elevadas, que las personas con bajos recursos no pueden permitirse.

A la hora de identificar los riesgos existentes en los países en desarrollo dentro de esta línea seguros generales, se observa que existen cuantiosas pérdidas ocasionadas por los desastres naturales, que en esos países emergentes se producen con más frecuencia e intensidad de lo habitual que en otras regiones. Ante esta nueva necesidad detectada, es explicable que surja un nuevo producto, que no es otro que el microseguro frente a desastres naturales. Con él se ofrecen coberturas por fallecimiento, gastos médicos, pérdida de bienes o daños causados en la cosecha y el ganado ante la ocurrencia de cualquier fenómeno adverso de la naturaleza, ya sea de inundación, terremoto, tifón o riesgo semejante.

Ante situaciones como estas, se hacen pronto evidentes los beneficios que aporta un microseguro de este tipo a la población de bajos ingresos. Pero de igual forma es

necesario sopesar sus limitaciones. Entre ellas, el elevado coste asegurable correlacionado con el riesgo de desastres naturales, sin descartar el riesgo moral en el sentido de que el asegurado no tome las debidas precauciones por el hecho de estar cubierto con el producto. Es más, en este tipo de productos existe un alto riesgo de insolvencia por las compañías de seguros en caso de múltiples catástrofes. No obstante, este microseguro sigue promocionándose como una medida ex-ante frente a riesgos naturales pese al debate que surge en cuanto a quién debería asumir la responsabilidad del coste ocasionado por los desastres naturales. ¿Debería recaer en la población empobrecida, cuando en parte se debe a las propias debilidades del gobierno que no proporciona los recursos necesarios para prevenir o reducir este tipo de riesgos? (Mechler et al., 2006).

1.5.4. Seguros agrícolas

En los países desarrollados el seguro agrícola es un mercado maduro, con una gran variedad de formas, debido en parte al apoyo planificado de los gobiernos. En cambio, en los países en desarrollo existe un largo camino por recorrer. En estas regiones la inmensa mayoría de población vive en zonas rurales donde la agricultura se convierte en el pilar de subsistencia y desarrollo. La producción agrícola se ve continuamente amenazada por diversos factores, unos de orden climático, como la lluvia y el granizo, otros de incidencia ocasional, como pueden ser los económicos y políticos, derivados de la situación del país. Por este motivo, la protección de la producción agrícola se convierte en un riesgo asegurable crítico, muy presente en estos países emergentes. El gran potencial del seguro agrícola está en proporcionar la sostenibilidad de la explotación agrícola, cuyo primer beneficiario es el agricultor. El seguro agrícola ofrece la cobertura necesaria para hacer frente a las pérdidas incurridas y facilita al agricultor la posibilidad de seguir trabajando para la siguiente temporada cuando por factores externos se da una mala cosecha (Sandmark et al., 2014). Toda esta situación ha favorecido el crecimiento de este seguro en los últimos años, en parte a consecuencia sobre todo de la iniciativa de las cooperativas rurales y de algunas organizaciones internacionales.

Por otra parte, el seguro agrícola cuenta con una serie de limitaciones que repercuten directamente en el elevado coste del producto. En primer lugar, por los gastos relacionados con el diseño y la gestión del producto con sus consecuentes costes administrativos y los gastos relacionados con la gestión de siniestros, siempre de difícil

tramitación. En segundo lugar, ya desde el punto de vista de los asegurados, por el riesgo moral donde los asegurados cubiertos tienden a despreocuparse por el simple hecho de que su riesgo está cubierto. A lo que cabe añadir también la selección adversa, donde se atrae en mayor medida a los asegurados con riesgos más altos, lo que provoca un incremento de primas para los asegurados con riesgos menores.

Sandmark et al. (2014) establecen las tres principales tipologías de productos agrícolas:

- Seguros indemnizatorios: caracterizados por el pago de una indemnización en base a la pérdida incurrida por el asegurado. Como ejemplos se mencionan los seguros frente a un único riesgo y los seguros multirriesgo en función del número de riesgos objeto de cobertura.
- Seguros de ingresos: que ofrecen protección económica al agricultor ante caídas de precios en productos básicos con la consiguiente disminución de sus ingresos procedentes de explotaciones agrícolas. Dentro de esta tipología se encuadran los fondos mutuos, los fondos para catástrofes y las cuentas de estabilización.
- Seguros basados en índices: igualmente ofrecen protección al agricultor mediante el pago de una indemnización referenciada al valor de un índice. Esta modalidad que prevalece principalmente en los seguros agrícolas se va a profundizar a continuación.

Seguros basados en índices

Tal y como se ha señalado, esta reciente modalidad de seguros basados en índices supone una categoría que está vinculada principalmente a los seguros agrícolas. De hecho el Fondo Internacional de Desarrollo Agrícola y el Programa mundial de alimentos (FIDA y PMA¹¹) identifica los seguros agrícolas basados en índices como un mecanismo de protección con un doble objetivo: de protección y desarrollo, con las siguientes especificaciones:

- Protección de la vida y subsistencia de las personas ante cualquier amenaza derivada de los desastres de la naturaleza que haga peligrar la vida, la salud o parte de sus bienes.

¹¹ Véase el estudio sobre los seguros basados en índices de FIDA y PMA (Hazell et al., 2010).

- Protección de la inversión de los agricultores ante cualquier siniestro, con objeto de promover el desarrollo económico, de crear nuevas vías de ingresos como una saludable estrategia de evitar caer en la pobreza.

Los seguros agrícolas basados en índices suponen una apuesta significativa para resolver buena parte de los problemas existentes en estos productos, como son el riesgo moral y la selección adversa. Esta modalidad de seguro se relaciona con eventos de ocurrencia lenta, en los que la indemnización se vincula al valor de un índice. Se establece un umbral que determinará el límite por debajo del cual la compañía deberá de indemnizar. El índice debe estar lo suficientemente correlacionado con las pérdidas aseguradas y se requiere que sea objetivo, fácilmente cuantificable, verificable públicamente y no manipulable por ninguna de las partes. La contratación del seguro se puede hacer de forma individual, institucional, nacional o a nivel regional.

En estos seguros el índice de referencia puede atender a dos categorías diferentes (Carter et al., 2014; Sandmark et al., 2014) :

- Índices directos: basados en el rendimiento por área, calculada por su valor medio o, en su caso, por el importe medio que asciende la mortalidad del ganado asegurado. Ofrecen cobertura a múltiples riesgos: climáticos, enfermedades, plagas, etc. La medición se realiza mediante técnicas de teledetección. Según la experiencia, esta modalidad de índices tienen mayor potencial de mercado puesto que son más comprensibles para los agricultores.
- Índices indirectos: son los que utilizan otro tipo de datos, como son los factores climáticos recogidos principalmente por estaciones meteorológicas, tales como la precipitación, el granizo, la temperatura y el viento, además de estar relacionados con las pérdidas comprobadas en el terreno. En esta categoría están los seguros basados en índices climáticos que se correlacionan con el rendimiento del cultivo. Sobre estos se han desarrollado los seguros basados en índices satelitales cuya fuente de información son las imágenes vía satélite. Esta novedosa modalidad está en desarrollo y es aún muy costosa.

Una cuestión importante que se plantea en estos seguros indizados es la existencia del llamado riesgo base que se conforma en sus tres dimensiones: *distancia espacial* respecto a la ubicación del índice; *lapso de tiempo*, que es el que transcurre entre el momento del siniestro y la información proporcionada por el índice; y por último

respecto a la forma, *las propias limitaciones* del índice que no recogen correctamente los factores que afectan al cultivo. En definitiva, el problema se refiere a la discrepancia existente entre el riesgo medido por el índice y la pérdida realmente incurrida por el agricultor en el terreno asegurado (Carter et al., 2014; Sandmark et al., 2014).

Aunque este tipo de productos no ha tenido mucho éxito entre los pequeños agricultores, no deja ser un seguro que se identifica claramente con las necesidades de la población de bajos recursos. Es necesario que, al igual que en países desarrollados, se promueva la ayuda desde los gobiernos para impulsar la implantación de estos seguros con la intención también de fomentar la implicación de los propios agricultores.

CAPÍTULO 2. LA PERSPECTIVA ASEGURADORA

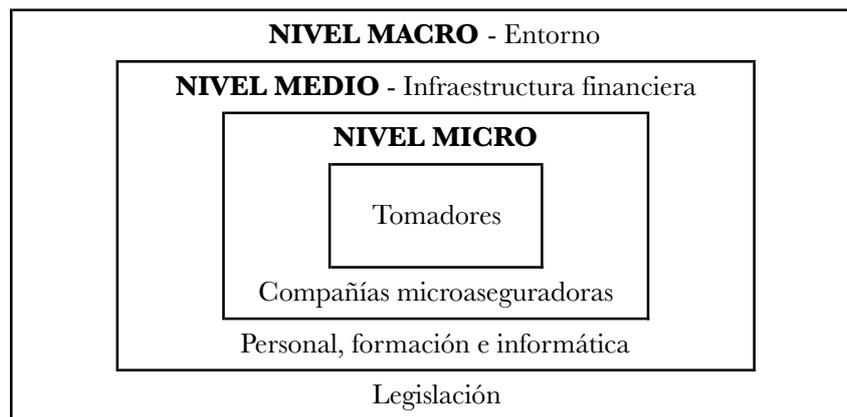
En este capítulo, y siguiendo con el marco teórico de los microseguros, el propósito es aportar una visión general del funcionamiento de este mercado desde la perspectiva de la compañía aseguradora, como principal proveedor y responsable de establecer el precio del producto.

Para realizar este cometido, creemos necesario, en primer lugar, dar a conocer la estructura de la actividad microaseguradora con todos los agentes intervinientes, para que de esta manera sea fácil visualizar toda la cadena de valor del producto. Además, prestaremos una atención especial al marco regulatorio que conforma su entorno de desarrollo. Adentrándonos en la labor de las compañías, se describirá el proceso de diseño del microseguro pasando por las principales fases que requiere su creación. En ellas, se pone de relieve la importancia de los datos como fuente de conocimiento del negocio y también como herramienta básica para la tarificación, sobre todo en un contexto de falta notoria de experiencia, en la que todavía no abunda la información. Este hecho, junto con otras limitaciones, podría hacer peligrar la sostenibilidad financiera del producto, agravada por la aparición de riesgos potenciales asociados a este mercado. Y para prevenirlos o suavizar sus consecuencias, describiremos también algunas medidas que sirven para controlar la eficiencia del microseguro, como son los indicadores clave de rendimiento. Por último, será necesario recalcar que, a pesar de las barreras inherentes a este mercado, existen claras expectativas de beneficios para las compañías aseguradoras al incorporar el microseguro dentro del negocio.

2.1 Estructura de la actividad microaseguradora

La actividad microaseguradora engloba la participación de diferentes instituciones, organizadas en varios niveles, para garantizar el buen funcionamiento de todo el ciclo del negocio. Forma parte del mismo toda la cadena logística por la que circula el producto, desde el cliente, pasando por el proveedor, por el canal de distribución y hasta la más alta institución implicada en el proceso, como es el organismo supervisor. Siguiendo la estructura planteada por Roth et al. (2007) se destacan tres importantes niveles principales, tal y como se puede ver en la siguiente Figura.

Figura 2.1. Estructura de la actividad microaseguradora.



Fuente: elaboración propia a partir de Roth et al. (2007). The Landscape of Microinsurance in the World's 100 Poorest Countries. The Microinsurance Center.

A. Nivel micro

Se sitúa en el centro del sistema, tal y como aparece en la Figura 2.1, y lo protagonizan los tomadores y los proveedores. En estos últimos se incluye también la figura formada por varias organizaciones que se coordinan de manera conjunta para realizar las distintas funciones del negocio. El nivel micro es donde confluye toda la cadena logística del microseguro, cuyo valor repercute a todas las partes intervinientes del negocio. Desde el reasegurador que proporciona a las compañías de seguros cobertura frente a riesgos catastróficos, el asegurador que es quien soporta la mayor carga del riesgo del seguro, hasta el canal de distribución que se encarga de la venta del producto y de todos los servicios básicos necesarios para hacer llegar el producto al consumidor final, que es el tomador de la póliza. Él es el que compra el producto y el asegurado el verdadero receptor de la cobertura del seguro (Roth et al., 2007). Siguiendo con esta cadena del negocio, vamos a detenernos un poco más en cada

uno de los intervinientes en el proceso del microseguro describiendo sus respectivas funciones:

- *Reasegurador.* Su misión es evaluar y asegurar a las compañías aseguradoras ante aquellos riesgos extraordinarios que no tienen capacidad de asumir. Realizan la correspondiente estimación de primas de reaseguro y elaboran los contratos de reaseguro. Entre los riesgos que se reaseguran, suelen estar los siniestros que se producen de forma simultánea o cuando el volumen de pérdidas excede al de los siniestros normales. Por eso mismo, en el negocio microasegurador, donde la cobertura está limitada y los siniestros suelen ser pequeños, la figura del reasegurador no tiene, al menos por el momento, especial protagonismo.
- *Asegurador.* Es el responsable directo de asumir el riesgo que asegura, el que debe asumir todos los compromisos adquiridos, como son el de pagar la indemnización prevista y prestar cualquier otro servicio contratado. También es el que se encarga de planificar estrategias de negocio, de identificar tendencias de mercado, de desarrollar nuevos productos que puedan ser útiles para los asegurados a su vez que rentables para la compañía, y de monitorizar los ya existentes. Centrándonos en la labor actuarial, las funciones principales se centran en estimar las primas de riesgo, en calcular las provisiones necesarias que garanticen la solvencia y suficiencia del negocio, así como realizar una adecuada gestión de riesgos. Es cierto que existen esquemas informales de seguros que operan fuera del ámbito legal. Pero la gran mayoría de productos se ofrece a través de compañías de seguros que cumplen con la legislación vigente en cada país.
- *Canal de distribución.* Es el responsable de la venta y prestación de servicios ofrecidos por el producto. El canal de distribución puede ser un individuo, un grupo o una institución. En el seguro tradicional la figura común es el individuo como agente de seguros que recibe una comisión en función de las ventas que realice. Sin embargo, en microseguros donde la prima suele ser reducida, no existe el mismo margen de beneficio para las comisiones. El canal más utilizado suele ser la misma institución u organización, porque es capaz de atender a un mayor número de personas a la vez de tener la capacidad de hacerlo de una forma rápida y menos costosa.

- *Tomador.* Se trata de la persona física o jurídica que contrata el producto y se compromete a pagar las primas del producto y a reclamar la indemnización o la prestación del servicio cuando ocurra el siniestro. La figura del tomador puede ser un individuo o un grupo. En microseguros, lo habitual es encontrar grupos como tomadores de pólizas porque el precio es más reducido que si se contratase de forma individual.
- *Asegurados.* Son todos los individuos o colectivos que tienen derecho a recibir la cobertura del producto. Normalmente la figura del tomador suele ser también la del asegurado principal, al margen de poder incluir otros asegurados en la póliza, como sus propios familiares. Se debe matizar que la figura del beneficiario no tiene por qué ser el asegurado ni en consecuencia receptor de la cobertura.

B. Nivel medio

En este nivel intermedio de responsabilidad y ejecución, hay que situar a todos aquellos que trabajan para proporcionar la infraestructura financiera, la cual es necesaria para garantizar el buen funcionamiento de la actividad que se desarrolla en el nivel micro. Concretando un poco más, en este nivel medio se incluye, en primer lugar, el personal cualificado que se ocupa de las principales funciones del negocio (actuarios, auditores, tramitadores e informáticos). En segundo lugar y en un sentido amplio, toda la formación del personal requerida para realizar las tareas ejecutivas además de la diversidad de aplicaciones informáticas necesarias para desempeñar las distintas funciones del negocio. Un aspecto crítico y generalmente ausente en este nivel, es la calidad de la información de mercado, que siempre es básica para el desarrollo del producto y para promover una buena comunicación en el negocio a todos los niveles.

C. Nivel macro

Lo integran dos elementos fundamentales con los que siempre hay que contar, que son el propio entorno y la legislación vigente. El entorno lo conforma el lugar, el tipo de clientes y todas las circunstancias que caracterizan el contexto donde se desarrolla el negocio microasegurador. Una circunstancia de primer orden es el país donde se realiza el negocio, cuya figura principal es el Estado, de quien depende la legislación a aplicar. Su función consiste en asegurar el buen desarrollo del negocio mediante la normativa vigente y supervisar que las compañías aseguradoras operen bajo los

límites establecidos. Y siempre con una preocupación primordial, que es la de velar por la protección del consumidor y evitar exponerle a riesgos aún mayores.

2.2 Marco regulatorio del microseguro

El entorno de la actividad microaseguradora está condicionado por la legislación vigente, que se corresponde con el nivel macro de su estructura de negocio que acabamos de comentar. La mayoría de los países emergentes donde se desarrolla el negocio microasegurador cuentan con escasos recursos, y por extensión escasas también son las medidas públicas a favor de esta población. A ello se suma la dificultad de acceder a algunos de sus segmentos, así como a otros que, aún siendo accesibles, son reticentes a contratar esquemas públicos por la falta de información.

En este sentido, se puede observar como la jurisdicción gubernamental diverge entre los distintos países donde se desarrolla este negocio. De hecho, la regulación de microseguros está aún por desarrollar, y donde las instituciones gubernamentales de cada jurisdicción desempeñan un papel fundamental para promover un entorno regulador constructivo de la actividad microaseguradora. La Asociación Internacional de Supervisores de Seguros (IAIS), que representa a supervisores de casi 140 países, es la encargada de promover un sistema global eficiente de supervisión de seguros. Por su parte, el Grupo Consultivo de Asistencia a los pobres (CGAP), cuenta con expertos en promover la creación de sistemas financieros inclusivos (incluido el microseguro). Ambas organizaciones tratan conjuntamente de definir un marco regulador y supervisor del seguro en materia inclusiva, enfocado a personas de bajos ingresos (IAIS & CGAP, 2007).

2.2.1. Objetivos de la regulación

Actualmente no existe una visión uniforme sobre cuáles deberían ser los objetivos en materia de regulación del seguro destinado al segmento de población de bajos ingresos. Sin embargo, la norma generalmente aceptada recoge como objetivos primordiales en materia de regulación los siguientes principios por orden de prioridad (IAIS & CGAP, 2007):

1. Salvaguardar la solvencia de las compañías mediante una adecuada dotación de las provisiones que garantice su estabilidad dentro del sector.
2. Proteger a los asegurados y a los tomadores.

3. Incrementar la competencia y eficiencia del mercado, adoptando nuevas tecnologías y otras formas de innovación. Además de gestionar, operar y evaluar los programas de seguros vigentes. Esta viene siendo la función clásica de los supervisores.
4. Desarrollar el mercado, formalizando servicios financieros inclusivos para los clientes de bajos ingresos.
5. Apoyar otros objetivos estratégicos (no aseguradores) como la inclusión financiera para lograr que el seguro sea accesible y disponible al sector informal, cumpliendo con la ley o con los estándares internacionales.

En la práctica del seguro tradicional, los principios de dar protección al consumidor y velar por la solvencia de las compañías aseguradoras han prevalecido respecto al resto de objetivos. Recientemente, y a partir del impulso generado por la creación de los microseguros, este foco se ha ampliado, propiciando el desarrollo de nuevos mercados y favoreciendo la integración de la inclusión financiera en los seguros para favorecer el acceso a la población de bajos recursos.

2.2.2. Limitaciones

Son muchas e importantes, algo que puede sorprender al tratarse de un nuevo producto asegurador como es el microseguro, que sin embargo se vuelve complicado a la hora de su aplicación. Para empezar, vemos que la creación de un marco regulatorio eficiente para el microseguro está fuertemente condicionada a la situación macroeconómica de los países en desarrollo, que es el mercado principal donde se desarrollan este tipo de productos. Es evidente que la regulación de este producto es una tarea compleja, porque precisa regular toda una serie de aspectos que varían según el riesgo asegurado, el perfil del asegurador y el canal de distribución, así como por la necesidad de controlar todas las partes implicadas en la estructura del negocio y desde sus diferentes niveles (micro, medio y macro).

Dicho esto, las limitaciones existentes en este mercado se plantean bajo diversas perspectivas, desde el plano institucional (las microaseguradoras y los canales de distribución) hasta el plano legal o político (Wiedmaier-Pfister & Chatterjee, 2006). Vamos ahora a profundizar en detalle de dónde y cómo proceden todas estas limitaciones.

A. Respecto a las compañías microaseguradoras

Estas empresas se encuentran enmarcadas en el ámbito institucional y son las responsables de desarrollar el microseguro. Actualmente se observa una notable limitación para crear nuevas instituciones formales para este tipo de seguro, forzada por la vigencia predominante de un modelo único e igual para todos. Es cierto que dicho modelo ahorra trabajo al supervisor, pero con ello no atiende al criterio racional de dar facilidades para el progreso de nuevas y existentes instituciones, que a la postre son fundamentales para el desarrollo de los microseguros (Roth et al., 2007). Entre las desventajas más importantes se distinguen:

- *Una exigencia de altos requerimientos de capital.* Siempre difícil de alcanzar, y mucho más en el campo de los microseguros. Por la sencilla razón de que se necesita un volumen alto de pólizas para generar un retorno positivo. No tiene sentido, por tanto, exigir la misma protección financiera a una póliza normal de seguro que a una póliza de microseguro.
- *La compleja documentación necesaria para desarrollar el negocio.* Lo que eleva considerablemente los costes de gestión y de administración de los operadores. Al margen del impuesto que recae sobre las primas, con el consiguiente aumento del coste del producto, con lo cual se reduce aún más el acceso a la población de bajos ingresos.

Otro punto importante, que también comporta desventaja, es la existencia de una cuota significativa de organizaciones que realizan la actividad microaseguradora fuera del ámbito legal. Tal es el caso de las pequeñas asociaciones y organizaciones no gubernamentales que constituyen los llamados esquemas informales. Es un aspecto que suscita muchas cuestiones y que no debe ser olvidado por los supervisores, a quienes corresponde sopesar la relación coste-beneficio que supondría supervisar este mercado informal de microseguros (Botero et al., 2006). Las limitaciones ya comentadas y que afectan a las compañías microaseguradoras, terminan por dar pie a la proliferación de estos esquemas informales, situando en desventaja a los esquemas formales que sí cumplen con la exigencia de capital. Para afrontar este reto que supone esta anómala situación, sería necesario definir los requisitos mínimos que debe cumplir un esquema informal de microseguro para que se le otorgue una licencia mínima para operar. Claro está que las condiciones siempre van a estar sujetas a las disposiciones que emanen de la jurisdicción propia de cada país.

B. Respecto a los canales de distribución

Como segundo miembro del ámbito institucional están los canales de distribución, como un medio eficaz para acceder a la población de bajos recursos. Pero nos encontramos con que también estos canales se enfrentan a considerables barreras regulatorias. El hecho es que la regulación actual limita de diversas maneras el trabajo de los canales de distribución. Una traba más, que habrá que superar de alguna manera, tal vez concediendo mayor facultades operativas a los canales, pero siempre con la idea de promover el desarrollo del mercado. Entre las restricciones a las que se enfrentan los canales de distribución cabe mencionar las siguientes:

- *Los estrictos requisitos de las licencias de agentes.* Ello supone una fuerte limitación a la venta de pólizas por parte del principal distribuidor de estos productos, y lo mismo a las instituciones microfinancieras.
- *Restricción en el importe de la comisión del agente,* a pesar de que el acceso a la población de bajos ingresos suele generar mayores costes de transacción.
- *Obstáculos normativos para la creación de compañías multirrama.* Estas restricciones se dan en algunas jurisdicciones que no permiten a las compañías de seguros operar a la vez en diferentes líneas de negocio, tales como productos de vida y productos no vida, pese a que de hecho sea la estructura más adecuada para este tipo de mercado.
- *La compleja redacción de las pólizas de seguros,* que suele ser inadecuada para los clientes del microseguro. Sería recomendable utilizar un lenguaje más sencillo y de fácil comprensión, de manera que los clientes puedan entender perfectamente todos los términos y condiciones del contrato.

C. Respecto al ámbito legal y político

Se trata de un espacio especialmente importante porque en él se fija el marco normativo al que hay que atenerse obligatoriamente para el funcionamiento legal del negocio del microseguro. Y sin embargo, de ahí proceden las principales barreras encontradas a nivel global, entre las que se pueden señalar las siguientes:

- *Una sobre-regulación del sector seguros.* Suele ser tan abrumador el exceso de requisitos que en algunos casos impide la inversión en la industria aseguradora y va en detrimento de captar nuevos segmentos de mercado como el correspondiente al negocio microasegurador.
- *Un frecuente solapamiento de legislación.* Con lo que se obstaculiza tanto el diseño como la distribución de productos.
- *Falta de diálogo entre instituciones públicas y privadas.* Por lo mismo, se hace necesario promover acuerdos entre ambos sectores en beneficio de todos, como por ejemplo la creación de microseguros subsidiados que permitan su salida al mercado. Se logra así que la cobertura del seguro alcance a un mayor sector de la población, se garantice un mejor servicio y se reduzcan los costes. Todo lo cual redundará en una mayor eficiencia de los productos.

En resumen, es muy importante poder contar con un marco de regulación estable para todo el desarrollo de la actividad microaseguradora. No solo por el compromiso del seguro con los *Objetivos de Desarrollo Sostenible* de Naciones Unidas, sino por la misma sostenibilidad del crecimiento económico a largo plazo de un sólido mercado microasegurador. A pesar de que algunos supervisores están cada vez más sensibilizados con este sector, no estará demás señalar que predomina una falta de experiencia y también de conocimiento respecto a este segmento de población. Una de las vías para adquirir experiencia es acudir a las lecciones aprendidas en materia de regulación de las microfinanzas, y en particular las relacionadas con los microcréditos. Este escaso conocimiento del mercado pone de manifiesto la necesidad de invertir en educación en todos los niveles. Una actividad que exige un reforzado esfuerzo colaborativo de todas las partes implicadas, aprovechando a su vez las sinergias provenientes de los principios y normas internacionales en materia de regulación y supervisión de seguros. Todo ello para dar un mayor impulso al desarrollo de un marco general de regulación y supervisión de microseguros, que será una alentadora realidad para la población de bajos recursos (IAIS & CGAP, 2007).

2.3 Proceso de diseño del microseguro

La creación o diseño de un nuevo producto es un proceso largo y metódico. Este se va perfilando en un ciclo continuo que recorre de forma iterativa el paso por varias etapas hasta llegar a la correcta definición del producto. Un producto bien diseñado beneficia tanto a la compañía como a los clientes. Más aún, el éxito del mismo va a depender en gran medida de que el producto diseñado se acomode lo más posible a las necesidades y preferencias de la población asegurada. En el campo de microseguros, existen contribuciones diversas, algunas de las cuales se materializan creando productos adaptados al mercado informal o bien incluyendo mejoras funcionales a partir de la experiencia acumulada en los productos ya existentes.

A partir de la literatura revisada sobre el diseño del microseguro (Churchill, 2006; Wrede et al., 2016), se pueden resumir así las principales etapas del ciclo de desarrollo del producto.

1. Definición del producto

El proceso comienza con la recopilación de información sobre el mercado objetivo, cuyos datos nos servirán para poder definir correctamente las características del producto. En microseguros, esta fase suele resultar bastante tediosa por la dificultad que existe en acceder al colectivo de personas de bajos recursos. Más adelante, en este mismo capítulo, abordaremos con detalle el tema de la información como fuente de conocimiento.

Con los datos obtenidos, y una vez tratados y analizados convenientemente, estaremos en condiciones de determinar los rasgos característicos del microseguro, identificar los riesgos asociados al mercado objetivo y aquellos que sean asegurables, así como determinar el perfil del cliente potencial y su capacidad de pago entre otros aspectos. Corresponderá al actuario la tarea de identificar los datos relevantes, así como de priorizar los más determinantes, valiéndose de la información del contexto y del conocimiento adquirido en el mercado para poder realizar así un correcto análisis actuarial.

2. Materialización del producto

Una vez determinadas las principales características del producto, el siguiente paso consistirá en concretar los aspectos relacionados con la idoneidad del producto, en

base a los objetivos definidos por la compañía y las preferencias de los clientes. En el microseguro, esta meta se logra equilibrando una amplia cobertura del producto que proporcione suficientes beneficios para los clientes a cambio de unas primas bajas capaces de garantizar la sostenibilidad del producto.

Dentro todavía de esta fase, habrá que abordar aspectos importantes como son la definición de las coberturas y el beneficio que se espera obtener, procurando articularlo todo de la manera más clara y sencilla posible. También entra en el campo actuarial, la gestión y el control de riesgos, así como - el punto de mayor interés para la presente tesis- la tarificación de primas del microseguro. Asimismo, es necesario contemplar todas las posibles medidas de control de siniestros, bien a través de copagos, bien por el sistema de franquicias o estableciendo algunas exclusiones. En el plano institucional, es igualmente importante también definir cuáles serán los canales de distribución para acceder al mercado objetivo.

3. Planificación del negocio

La compañía deberá elaborar el correspondiente plan financiero de negocio con el fin de garantizar la sostenibilidad del producto. En él se desarrollarán todas las proyecciones de ingresos y gastos bajo las hipótesis del negocio previamente establecidas. Este punto es de gran utilidad por varias razones, por la puesta en marcha del producto, que no se podría llevar a cabo sin la inversión inicial determinada por la compañía. Y también, para garantizar un desarrollo sostenible del producto, dado que se debe cuantificar qué volumen de negocio se prevé, a medio y largo plazo, para obtener el correspondiente retorno de la inversión (Wrede et al., 2016). En este sentido, el actuario también toma aquí un protagonismo especial puesto que es él quien debe tener la habilidad de saber adecuar las primas exigibles de forma que se puedan cumplir las expectativas establecidas en el plan de negocio. De este modo, se hace evidente que una planificación tan minuciosa como la descrita se convierte en una herramienta de validación del producto antes ya de su salida al mercado.

4. Verificación y ensayo

Una vez que el producto ya está perfectamente definido y que puede satisfacer las expectativas del plan de negocio, el paso siguiente consiste en verificar si efectivamente se cumplen todas las hipótesis previamente definidas. Una buena forma de poner a prueba el diseño del producto es realizar un ensayo antes de su salida a

producción. Esta puesta en escena es esencial porque permite localizar fallos que no se habían detectado durante el desarrollo del producto. Y sirve además, para comprobar el buen funcionamiento de todos los procesos implicados. La ventaja añadida es la posibilidad de revisar el producto antes de su salida al mercado corrigiendo los errores detectados. La revisión puede consistir en introducir modificaciones en la prima y/o en las propias características del microseguro, sobre todo cuando la experiencia en el campo revela información importante que no había sido contemplada de antemano (Wrede et al., 2016).

5. Mercado y monitorización

Superada la fase de ensayo, el proceso culminaría con el lanzamiento del microseguro al mercado con la exigencia de un riguroso seguimiento del producto. La salida a producción consiste en poner en práctica todo el proceso de distribución y comercialización del producto, bajo la premisa de que el nuevo microseguro deberá destacar sobre el resto de productos de la competencia.

La fase de seguimiento se ocupará de monitorizar y controlar la buena evolución del producto. Lo que implica ir actualizando la base de datos con el fin de disponer de un completo historial con la experiencia acumulada, que nos será muy útil a la hora de programar los medios más indicados, bien para su conservación, o bien para corregir los puntos débiles observados. Y es que los datos de la experiencia son siempre la fuente más valiosa para controlar y garantizar el correcto funcionamiento del producto y por ende para consolidar de cara al futuro el negocio microasegurador.

No hace falta decir que durante todo el proceso de diseño del producto se requiere un uso intensivo de datos, algo que a menudo recibe poca atención en el proceso estratégico del negocio. Y sin embargo, el éxito y la viabilidad del producto depende, en gran medida, no solo de realizar una buena gestión, sino también y sobre todo de un afinado tratamiento de información. Ambas cosas igualmente necesarias para desempeñar las múltiples tareas en las diferentes áreas de la compañía (World Bank, 2014a).

2.4 La información como fuente de conocimiento para la tarificación

Como se ha venido comentando, la información juega un papel fundamental dentro del negocio asegurador. Sin embargo, es un hecho conocido, que su escasez o dudosa rigurosidad se conviertan en una de las principales limitaciones a la que deben enfrentarse las compañías. Preocupa de manera especial la generalizada carencia de datos en este mercado emergente. A tenor de ello, Biener (2013) describe las principales razones por las que viene justificada esta limitación de información:

- El microseguro es un producto relativamente nuevo, por lo que existe poca experiencia del negocio que pueda ser analizada.
- El tamaño de las microaseguradoras suele ser pequeño, por lo que la información acumulada también es escasa y reducida para realizar un adecuado análisis de tarificación.
- La infraestructura existente en estos países en desarrollo suele ser más bien pobre, e incluso en algunos programas la recogida de datos se hace de forma manual, lo que impide la elaboración de una base de datos robusta y consistente. Este hecho propicia el uso de otras fuentes externas, principalmente macroeconómicas.

2.4.1. Fuentes de información

Como consecuencia de lo dicho, la búsqueda de información se convierte en un auténtico reto para el diseño del microseguro. Y en particular para la labor de tarificación, por lo cual es crucial ampliar el punto de mira y examinar múltiples alternativas como fuente de conocimiento. Wrede et al. (2016) destacan la necesidad de invertir tiempo en aprender de unos y otros, y de investigar detenidamente las condiciones de la población asegurable. Todo con el propósito de poder desarrollar productos que realmente encajen con las necesidades y preferencias de este mercado. Bajando a lo concreto, vamos a explorar cinco categorías de datos que pueden servir a los actuarios para el objetivo de tarificación, que comprenden desde la experiencia acumulada del propio negocio hasta fuentes externas de diversa índole.

1. Datos internos

Al hablar de datos internos, se hace referencia a la experiencia acumulada de una determinada población asegurada. Esta puede venir en primera instancia, procedente

de la propia compañía, o adoptada de otros mercados equivalentes. Lógicamente, la fuente preferida que aporta mayor credibilidad para cualquier análisis es la propia experiencia de la población asegurada de cada compañía, sencillamente porque es la que describe con mayor fiabilidad el comportamiento real del colectivo asegurado.

Sin embargo, y tal como se ha señalado en varias ocasiones, se trata de un negocio nuevo, en el que la gran mayoría de compañías no dispone aún de suficiente experiencia ni tampoco de infraestructura para elaborar una buena base de datos. Es de esperar que en el futuro este aspecto pueda ser reforzado, con el fin de que las compañías de microseguros puedan desarrollar y mantener un buen sistema de información. Lo que no impide que, de forma alternativa, se pueda acudir a otros profesionales que sí cuenten con experiencia adquirida en otros mercados equivalentes.

2. Investigación de mercado

Se trata de un medio eficaz para conocer los riesgos relevantes del mercado asegurable. Nos ayuda a conocer su actitud frente al riesgo, así como reconocer sus preferencias y la capacidad económica del asegurado. Algunos autores (Blacker & Yang, 2015) confirman que una breve muestra recogida a través de este medio puede ser suficiente para realizar una adecuada tarificación. Los métodos cualitativos proporcionan una mayor comprensión del comportamiento de la población y ayudan a explicar mejor los resultados cuantitativos. Biener (2013), por su parte, nos describe las técnicas de mercado más utilizadas en microseguros, tales como:

- La *encuesta*: es la forma más tradicional de recogida de información, mediante cuestionarios a la población general para lograr una mayor representatividad de la muestra. Esta técnica es capaz de proporcionar estimaciones sobre la frecuencia y el coste de los siniestros.
- Los *grupos de discusión*: se trata de técnicas consensuadas, las cuales surgen a través del discurso de un grupo de expertos con igual credibilidad. El resultado se basa en la opinión consensuada del grupo y no de los comentarios realizados de forma individual. Una de las más conocidas es el “método Delphi”, que consiste en realizar entrevistas de forma independiente para ser de nuevo valoradas por la opinión de otros expertos. El resultado aquí será la media o mediana de todas las ideas planteadas por los expertos y así poder determinar mejor los factores de riesgo del producto.

En comparación, aunque ambos métodos proporcionan rasgos característicos de la población objetivo, los grupos de discusión y en concreto el método Delphi, basado en la opinión de expertos, se alza como una mejor alternativa. Entre las ventajas que ofrece frente a la encuesta se distinguen las preguntas de investigación, que se caracterizan por su incertidumbre y especulación con la idea de obtener un discurso más enriquecedor. También ofrecen una mayor riqueza de datos derivada de las sucesivas iteraciones por las que pasan los cuestionarios y por la posterior revisión de las respuestas obtenidas.

3. Asociaciones del sector y organizaciones independientes

Existen diversas instituciones, tanto ligadas al sector como ajenas al mismo, que se dedican, entre otras funciones, a reunir información del negocio asegurador en distintos mercados.

Los **intermediarios**, ya sean compañías consultoras o reaseguradores, disponen también de información relevante sobre microseguros, proveniente de la experiencia de otras compañías y de otros países, aparte de realizar su función principal de consultoría como posibles estudios de las tarifas. Un ejemplo claro es *MicroEnsure*, una entidad especialista en productos y servicios aseguradores creada en el año 2002 y que desarrolla soluciones innovadoras para favorecer la protección de riesgos en países emergentes. De forma similar y más reciente, el consorcio formado por ocho compañías llamado *Blue Marble* se centra en desarrollar microseguros a través de un modelo innovador colaborativo.

En cuanto a otras **asociaciones**, su presencia y actividad en este campo representa otra fuente de datos y experiencias recopiladas en el sector microasegurador. En Colombia destaca *Fasecolda*, que es la federación de aseguradores y reaseguradores del país, que aparte de recopilar datos del sector, elabora estudios y estadísticas que pueden aportar información relevante en relación con el negocio microasegurador. Por otro lado, *Microinsurance Centre* pone también a disposición del público las lecciones aprendidas de programas de microseguros, que sirven de referencia para la creación y mejora de nuevos productos.

Finalmente, las **organizaciones internacionales para el desarrollo** representan una importante referencia de conocimiento, no solo del mercado objetivo, sino sobre todo

para comprender los principales riesgos a los que se enfrentan estos nuevos productos. Entre ellas destacan Naciones Unidas, Banco Mundial y la Organización Mundial de la Salud (WHO), que facilitan el acceso a una información garantizada tanto económica como demográfica de países en desarrollo y de la población en general. Un buen ejemplo de ello es la WHO, que proporciona las tablas de mortalidad por edad, sexo y causa de fallecimiento para todos los Estados Miembros y que son de gran utilidad para el diseño de la tarifa de microseguros de vida. Adicionalmente, las agencias estatales de estadística tienen también bases de datos accesibles sobre tasas de incidencia y datos demográficos con un alto nivel de detalle. Al igual, los programas públicos de asistencia sanitaria disponen también de gran volumen de información sobre la población aunque en muchos casos resulta inaccesible.

4. Regulador

También, las agencias gubernamentales de cada país, responsables de regular y supervisar el buen funcionamiento del negocio asegurador, pueden suministrar gran cantidad de información. A pesar del ambiente competitivo reinante en la industria aseguradora y de la pesada trayectoria en materia de regulación, el organismo supervisor sigue siendo una fuente alternativa. Hay países que ya cuentan con una regulación específica en microseguros, como es el caso de Perú, con su correspondiente “Autoridad de estabilidad financiera” encargada de establecer un marco regulatorio ya vigente desde el año 2009. Ya antes, en el año 2006, la Comisión de Filipinas comenzó su primera legislación reguladora, que fue completándose posteriormente con una serie de medidas a favor de la protección social y la inclusión financiera. En este caso, el sistema de supervisión utiliza un informe en Excel donde los proveedores de microseguros facilitan sus datos, información que puede ser utilizada para analizar por ejemplo los principales indicadores de rendimiento del producto.

5. Tecnología

Es inevitable destacar aquí el formidable instrumento de la tecnología digital, que está revolucionando todo el proceso de obtención de datos. Su protagonismo ha llegado también al mercado microasegurador, mostrando toda su capacidad y amplitud por llegar a sectores que permanecían hasta entonces inexplorados. La innovación y la tecnología van de la mano y ya se perfilan como fuentes potenciales para la obtención de información.

Entre los servicios digitales más utilizados en microseguros, la infraestructura de telefonía móvil es la que ha mostrado mayor crecimiento tanto en eficiencia como en lograr la distribución a escala (ILO, 2013). El desarrollo de aplicaciones posibilita una multitud de usos en la gestión y análisis de datos. Un ejemplo claro del uso de esta tecnología, es el utilizado por algunas instituciones microfinancieras en África, que utilizan el uso del móvil y la actividad en redes sociales para determinar el nivel de riesgo de los clientes para la concesión de créditos.

Y si nos adentramos en los seguros agrícolas basados en índices, la tecnología vía satélite es capaz de proporcionar información fiable sobre fenómenos meteorológicos a través de imágenes tridimensionales. Esta innovadora tecnología es utilizada por las compañías para mejorar las estimaciones sobre la ocurrencia de eventos extremos causados por fenómenos de la naturaleza. Adicionalmente, las redes sociales junto con las telecomunicaciones contribuyen de forma rápida y eficaz a generar fuentes de información más económicas y de buena calidad. Tal es el caso de las fuentes cruzadas de datos (*crowd mapping*), que permiten agregar datos e imágenes de zonas geográficas para complementar las imágenes vía satélite (Sandmark et al., 2014).

2.4.2. Descripción de la información

Ya conocemos las fuentes donde podemos encontrar la información que necesitamos. Ahora nos falta describir cómo llega esta información a las compañías para su aplicación en el negocio asegurador. Salta a la vista la necesidad de que las mismas dispongan de un sistema de información eficiente para realizar las diversas tareas del negocio microasegurador. Para elaborar una base de datos completa se debe combinar tanto información cuantitativa, con sus variables numéricas cuantificables, como cualitativa, que facilite la correcta interpretación de los valores numéricos. Solo con la conjunción de ambas se puede conseguir un conocimiento suficientemente fundado del riesgo asegurable del colectivo.

Diversos autores (Wipf & Garand, 2006; Wrede et al., 2016) aportan una descripción detallada de los campos principales para la confección de una útil base de datos según los dos niveles ya conocidos: el nivel micro correspondiente al perfil de riesgo de la población objetivo y el nivel macro relacionado con el contexto. A continuación vamos a profundizar más el alcance de cada uno de ellos:

A. Nivel micro: la población objetivo

Esta categoría contempla los aspectos económicos, culturales, profesionales, geográficos y sociológicos del perfil de riesgo de los participantes, así como la experiencia acumulada a partir de los asegurados del microseguro.

Respecto al perfil de los asegurados, la base de datos debe incluir como mínimo los campos siguientes: la clave única de identificación, el nombre, la fecha de nacimiento, el sexo, la fecha de adhesión al producto, el estado civil, el número de hijos, la ocupación y el tipo de residencia si es urbana o rural (Wipf & Garand, 2006). Atributos similares deben ser añadidos para los beneficiarios y dependientes cubiertos, incluyendo la relación de parentesco con el participante asegurado.

También forma parte de este nivel micro, el historial de coberturas aseguradas por cada participante adherido al programa de microseguro. El objetivo, por tanto, es construir una historia completa de cada asegurado por cada riesgo cubierto, para, de esta manera, poder comparar los siniestros esperados con la experiencia real de los siniestros.

A esto hay que añadir el historial de primas del producto por asegurado, con el detalle de la fecha e importe de pago. Este histórico sirve a su vez para realizar diversos estudios, por ejemplo, calcular el patrón de caídas derivado de las anulaciones o rescates de pólizas.

Y finalmente, el historial de siniestros por asegurado y producto. Que debe contener toda la experiencia de siniestralidad del microseguro y con el mayor desglose de datos, porque solo así se podrá realizar un correcto análisis de tarifas así como una adecuada gestión y tramitación de siniestros. Para ello, se contemplan los campos del importe de la reclamación, la valoración del siniestro, los pagos realizados, la causa del siniestro y todas las fechas relativas (ocurrencia, comunicación y liquidación del siniestro).

B. Nivel macro: el contexto

El contexto comprende todas las características que definen el entorno demográfico y económico del mercado del microseguro. Del entorno se deben considerar los

principales indicadores macroeconómicos, tales como el crecimiento económico y la tasa de inflación. Datos generales del país a nivel de desarrollo, indicadores de salud, la proporción de población urbana y rural, así como el porcentaje de población que vive por debajo de la línea de pobreza. Además, es importante tener en cuenta la información financiera, el marco asegurador y la regulación de cada país.

En cuanto al perfil demográfico, hay que considerar factores como el tamaño y el tipo de la población, la distribución por edad, el tamaño de familia media, la pirámide social, la densidad de la población, el nivel de educación, las condiciones de vida y de modo especial el nivel de seguridad. Sin olvidar otros factores que también pueden influir en la contratación, como son la cultura financiera, las creencias religiosas y también las tradiciones populares.

En el contexto más cercano a los asegurados, hay que subrayar la importancia del ambiente en que se mueve su vivir diario así como su particular situación económica. Ampliando el foco contextual a la región, no se pueden desconocer las principales actividades económicas que en ellas se desarrollan y que condicionan la actividad profesional de los distintos estamentos de la población. Todo ello junto nos ayuda a estimar el nivel de ingresos de la población. Por el lado de los gastos, es interesante conocer el coste de vida y paralelamente el nivel medio de ahorro de una familia, con el objetivo de contrastar el rango de prima razonable del microseguro. Sobre este punto, también puede servir de referencia los datos relacionados con el consumo familiar, tales como el coste de la comida o el del alquiler de la vivienda.

Finalmente, el propio contexto del país pone en evidencia el nivel de desarrollo y calidad de la infraestructura que va a condicionar el acceso de la población a los productos y servicios básicos. Este factor de accesibilidad puede influir significativamente en el coste de los siniestros como el nivel de penetración del producto.

2.4.3. Razonabilidad, trazabilidad y conservación

Una vez elaborada la base de datos con una rigurosa descripción de campos, es ya el turno de validar la información frente a otros sistemas independientes. Con el fin de garantizar la razonabilidad, consistencia y veracidad de los datos se utilizan medidas de contraste de eficacia comprobada, tales como las comentadas anteriormente en el

punto 2.4.1. De este modo, se puede utilizar la experiencia de terceros o la procedente de otros contextos, realizando validaciones cruzadas contra el sistema contable u otros medios similares (Wipf & Garand, 2006; Wrede et al. 2016).

Otra alternativa es utilizar información de mercado, en el que la competencia es una referencia clave para el análisis del producto y así conocer en qué rasgos se diferencia respecto al mercado. Otro factor importante a considerar es la capacidad de pago del asegurado para comprobar el nivel de satisfacción de oferta y demanda del producto. Este aspecto es de especial relevancia en la tarificación de microseguros, pues nos obligará a tener en cuenta el segmento de mercado al que va dirigido, para ajustar la cuantía de la prima del seguro a la capacidad de pago del cliente¹². En esta misma línea, Wrede et al. (2016) establecen un índice general de referencia que se estima como un porcentaje de prima mínima sobre el nivel de ingresos de la población objetivo. De acuerdo a los diferentes tipos de microseguro, este porcentaje de prima debe ser inferior al 2% para el seguro de vida, del 4% para productos de salud y del 1% para los seguros no vida en general.

El siguiente paso consiste en garantizar la trazabilidad del producto mediante una continua actualización de la información. Esta propiedad de la trazabilidad se establece a través de un sistema de monitorización que actualice continuamente los datos con cualquier modificación que se vaya produciendo durante la vigencia de las pólizas. Con la ventaja de lograr una mayor eficiencia operacional que permita evaluar la eficacia del producto y mejorar la tramitación de los siniestros (World Bank Group, 2014b). Pese a la importancia de mantener viva la base de datos, las compañías suelen enfrentarse a restricciones que impiden su correcta actualización. Uno de los problemas habituales en microseguros es la inexistencia de un campo de identificación de cada asegurado. Este campo es determinante para monitorizar la exposición y la siniestralidad de cada asegurado, un aspecto de gran relevancia para cualquier análisis actuarial (Wipf & Garand, 2006).

Todo el esfuerzo realizado en elaborar y actualizar la base de datos no debe descuidarse de cara al futuro. Es muy importante su conservación con el paso del tiempo. La infraestructura de las compañías juega un papel crucial para garantizar

¹² En el artículo de Breidert et al. (2006) se proponen diversos métodos para estimar la capacidad de pago.

este objetivo de acumular la experiencia adquirida, ya sea en la misma base de datos o mediante copias de seguridad aparte. El análisis actuarial se basa principalmente en la experiencia histórica para realizar sus oportunas estimaciones a futuro. Por este motivo cobra especial importancia cualquier dato erróneo, ya que nos puede llevar a una mala interpretación y a una consecuente incorrecta estimación. De ahí, que los actuarios estén especialmente sensibilizados con todo lo que concierne al proceso de gestión y tratamiento de información (Wipf & Garand, 2006).

2.5 Problemas asociados con la sostenibilidad de los microseguros

A pesar del aspecto novedoso del microseguro, la realidad es que ya cuenta con un fuerte crecimiento en numerosos países de los diferentes continentes. Un hecho que también ha servido para poner de manifiesto dos problemas importantes, por una parte, la dificultad en lograr la distribución a escala con una mayor implantación del microseguro entre la población de bajos ingresos, y por otra parte, el arduo trabajo que resulta alcanzar la sostenibilidad financiera del negocio (Purcal et al., 2012).

La sostenibilidad financiera depende en gran medida de la eficiencia de las compañías en establecer unas primas del microseguro capaces de cubrir el coste del siniestro, que sean además adecuadas para el mercado objetivo y bajo unos aceptables límites de cobertura. En estos términos es fácil deducir que una buena labor actuarial de tarificación puede contribuir a solucionar una gran parte de los problemas que se enfrenta el mercado microasegurador, hasta eliminar en lo posible la incertidumbre asociada a ellos.

Bajo esta situación de inestabilidad financiera, algunos autores (Biener & Eling, 2012; Purcal et al., 2012) coinciden en señalar como principales problemas los relacionados con la asimetría de información, la capacidad limitada de pago de los asegurados y la baja exposición del microseguro en el mercado. A continuación pasamos a describir en detalle cada uno de ellos.

A. Asimetría de la información

La asimetría de la información persiste en todo el sector asegurador y a todos los niveles económicos, y suele resultar aún más problemático cuando se habla de microseguros. Estos problemas se hacen especialmente visibles en el seguro de salud debido a la subjetividad del propio siniestro, tema donde se concentra un mayor

interés tal y como demuestra una mayor evidencia científica (Biener & Eling, 2012; Radermacher et al., 2006). Gran parte del problema tiene su origen en el falso comportamiento del asegurado del que se derivan tres posibles riesgos: la selección adversa, el riesgo moral y el uso fraudulento del seguro.

- *La selección adversa* ocurre cuando el perfil de riesgo del asegurado es peor al esperado de la población general. Esto se traduce en una mayor propensión de uso del producto por el asegurado, lo que repercute en un incremento del coste de aseguramiento por encima de lo esperado, situándose en ocasiones por debajo de la prima estimada. Un claro ejemplo donde esto se da con mayor frecuencia es en los seguros de salud, en los que se observa una llamativa incidencia de casos en las mujeres en edad fértil, en los niños y en las personas en edad avanzada. Otro tanto ocurre en los grupos con preexistencias, es decir aquellos que ya con anterioridad tengan algún tipo de enfermedad, lo que de entrada advierte de un mayor ratio de utilización del seguro por encima del resto de la población.
- *El riesgo moral*, muy relacionado con la selección adversa, se da en el sentido del uso excesivo del producto por parte de los asegurados simplemente por el mero hecho de disponer del producto y saber que están protegidos. El seguro en algunos casos puede convertirse entonces en el detonante que haga al asegurado incurrir en siniestro. Es frecuente el caso del cliente que, a sabiendas de que viene sufriendo alguna dolencia, se asegure con el objeto de realizarse el correspondiente tratamiento médico.
- *El fraude en seguros* también es un tema muy recurrente en seguros. Ocurre cuando se utiliza el seguro, al margen de las condiciones previamente pactadas, con el fin de obtener un beneficio. El riesgo de fraude puede venir desde cualquiera de las tres partes intervinientes: por el lado del asegurado, que proporciona cobertura a terceros no asegurados; por parte del proveedor, que se excede en la valoración del siniestro para reclamar después a la compañía mejores tratamientos o medicación más cara de lo acordado; y por último, de parte de los propios empleados de la compañía, que emitan falsos siniestros o dupliquen la tramitación de siniestros en acuerdo con el proveedor para obtener después una comisión.

A pesar de que el área de selección de riesgos contribuye en buena parte a controlar este tipo de escollos, es fundamental que las compañías establezcan medidas alternativas enfocadas en la prevención y detección del fraude en el microseguro. Los controles tendrían que ser más efectivos, por ejemplo con una combinación a la vez jerárquica y horizontal, o tal vez gestionados mejor por las propias compañías en vez de recurrir de forma tan inmediata a las consultoras externas (McCord et al., 2006; Radermacher et al., 2006).

B. Limitada capacidad de pago

La capacidad de pago del cliente hace referencia principalmente a los recursos económicos que dispone el cliente para hacer frente al pago de la prima de seguro. Es obvio que la población objetivo del mercado microasegurador dispone de escasos recursos económicos, por lo cual asumir el pago de una prima puede suponer todo un reto para desarrollar este mercado. De ahí la necesidad de contrastar el precio del seguro con los costes de otros productos de consumo para poder ajustar nuestra oferta del microseguro a un precio que resulte asequible al cliente.

Se hacen patentes las limitaciones inherentes a este tipo de mercado que hacen difícil afrontar el pago de la prima del microseguro. Tales dificultades, como señalan Cohen y Sebstad (2005), se deben muchas veces no tanto a la capacidad de pago sino a la forma de afrontarlo. La población de bajos recursos suele recibir sus ingresos no de forma regular sino de forma estacional, período en el cual ya cuenta con los suficientes recursos para pagar ciertos productos o servicios. Por ello, resulta práctico dotar al microseguro de cierta flexibilidad y dejar que se adapte a este habitual flujo irregular de ingresos de estos clientes. Una facilidad que puede contribuir en buena medida a fomentar el consumo de microseguros entre la población de bajos recursos.

Adicionalmente, también influyen en la baja adhesión a este seguro otros factores, como son la percepción subjetiva y social en la adversidad al riesgo, la desigual distribución de recursos y hasta el diseño mismo del producto, sin descontar otros rasgos culturales enraizados en la población objetivo (Biener & Eling, 2012).

C. Baja exposición

Este aspecto hace referencia al reducido número de asegurados que poseen microseguros. En comparación al seguro tradicional, el microseguro sigue presentando un bajo nivel de exposición. La exposición al producto representa para la

compañía una medida temporal del riesgo que asume, no solo por el bajo número de asegurados, sino también por el limitado coste de los siniestros. En todo caso, es siempre necesario establecer las hipótesis sobre la unidad de medida del volumen de asegurados, cuantificar el mercado objetivo y estimar la progresión de asegurados en el tiempo. Con esta finalidad, el actuario debe prestar atención a tres ratios fundamentales que permiten conocer y analizar la evolución de la exposición del producto (Wrede et al., 2016):

- *Ratio de utilización*: es el porcentaje de participación de asegurados en el programa de microseguros. La evolución de este ratio suele crecer de forma progresiva en el tiempo, que al inicio suele ser baja pero una vez que el asegurado comienza a usar el producto, este tiende a incrementarse.
- *Ratio de renovación*: hace referencia al porcentaje de asegurados que continúa asegurado tras el primer vencimiento de la póliza. En especial, los microseguros se caracterizan por tener una baja tasa de renovación debido a diversas razones: bien porque no están satisfechos con el programa, o porque finaliza la cobertura, o por los consabidos factores económicos.
- *Ratio de anulación o de caídas*: este tiene varias acepciones en la literatura de seguros, aunque, siguiendo la definición de Wrede et al. (2016), atiende especialmente al número de asegurados que pierden la cobertura debido a la falta de pago de primas. En este supuesto, también se incluyen aquellos casos donde el pago de la prima se realiza ya transcurrido el período de gracia. Los microseguros se suelen caracterizar por tener una elevada tasa de caídas, lo que obliga a plantear una adecuada estrategia para evitar o ralentizar la elevada tasa de anulaciones, e impulsar a la vez medidas que fomenten los ingresos de los potenciales clientes.

Todos estos ratios en su conjunto nos permiten cuantificar con gran aproximación el riesgo suscrito, así como establecer medidas a favor de la permanencia del asegurado en el producto. Sabiendo que una de las razones principales se debe a motivos económicos, se deben fomentar expresamente acciones que conduzcan a aumentar los ingresos de los clientes, por ejemplo favoreciendo el acceso a servicios financieros. En consecuencia, todas estas medidas deben ir encaminadas a mejorar los ratios de renovación y reducir la elevada tasa de anulaciones (McCord et al., 2006).

2.6 Medidas de eficiencia del microseguro

Por principio, el microseguro, al igual que cualquier producto de cualquier compañía, debe ser rentable, y para ello es indispensable que sea eficiente en su actividad para poder permanecer en el mercado por tiempo prolongado. Algunos programas se mantienen gracias a subsidios estatales y/o donaciones que no dejan de ser ayudas temporales y que por sí solas no garantizan su estabilidad en el tiempo. Por eso, si queremos garantizar una duración permanente en el mercado, es necesario realizar un continuo control del producto, que deberá ser evaluado en términos de eficiencia y rentabilidad.

Durante los años 2006 y 2007, con la idea de disponer de alguna medida estándar que valorase el rendimiento de los microseguros, un grupo representativo de expertos acordó los diez conocidos indicadores clave de rendimiento¹³, aplicables a todo tipo de microseguros. Hoy en día, estos indicadores han sido acogidos con gran aceptación entre las compañías y se han convertido en una magnitud de referencia para el sector. La principal función de estos indicadores, también conocidos por sus siglas KPIs, es medir la eficacia y rentabilidad del producto principalmente desde el punto de vista financiero. Lo que no impide que de alguna forma se tome también en consideración la influencia que ejerce el aspecto social a la hora de interpretar estas medidas. Entre otras interesantes aplicaciones los KPIs pueden servir, bien como una eficaz herramienta para la toma de decisiones internas de la compañía, o bien su evolución en el tiempo permite detectar cambios de tendencias en el negocio de microseguros. El uso generalizado de estos indicadores en el sector puede redundar en provecho para todas las partes interesadas, que de esta manera podrán conocer su situación respecto al mercado, atendiendo a las lecciones aprendidas de otros productos existentes a favor de la mejora continua del microseguro (McCord et al., 2014).

El principal beneficio de la generalizada aplicación de los KPIs se resume en disponer de una visión global y realista del desempeño del programa en las diferentes áreas de la compañía. Wipf y Garand (2010), como los principales precursores de estos indicadores, han elaborado una guía práctica sobre la medición de rentabilidad de los

¹³ En terminología anglosajona se conocen como *Key Performance Indicators* ó KPI recogidos en el manual de Wipf y Garand (2010).

programas de microseguro desde el punto de vista del consumidor. Bajo esta perspectiva, dichos autores proponen una clasificación de los 10 indicadores clave clasificados en cuatro grandes categorías por áreas de acción, tal y como se muestra en la siguiente Tabla.

Tabla 2.1. Resumen de categorías e indicadores clave de rendimiento.

Categoría	Indicador
Valor del producto	Índice de gastos incurridos
	Índice de siniestros incurridos
	Índice de ingresos netos
Conocimiento y satisfacción del producto	Índice de renovaciones
	Índice de cobertura
	Índice de crecimiento
Calidad del servicio	Tiempo de liquidación de siniestros
	Índice de rechazo de siniestros
Prudencia financiera	Índice de solvencia
	Índice de liquidez

Fuente: elaboración propia a partir de Wipf J. y Garand D. (2010). Performance indicators for microinsurance. Segunda edición. ADA con el de Luxembourg Development Cooperation, BRS y Microinsurance Network.

Tal y como se aprecia en la Tabla 2.1, se distinguen cuatro grandes categorías:

- *Valor del producto.* Los tres indicadores bajo esta categoría permiten analizar la viabilidad financiera del microseguro, ya que recogen medidas relacionadas con los ingresos y los gastos derivados del negocio. Desde la perspectiva del cliente, estos índices miden la ganancia percibida a cambio del pago anticipado de la prima.
- *Conocimiento y grado de satisfacción del producto.* Estos indicadores son especialmente necesarios para garantizar la sostenibilidad del programa, por ser condición necesaria el mantener una suficiente cuota de participación de asegurados. Y en ello juega un papel importante el hecho de que el público objetivo conozca y entienda correctamente el buen funcionamiento del producto, lo que le ayudará a mantenerse adherido a la póliza.
- *Calidad del servicio.* Dos son los indicadores que comprende esta categoría y que cuantifican aspectos relacionados con la prestación de servicios por la aseguradora.

- *Prudencia financiera.* Engloba dos índices que miden la solvencia y liquidez del microseguro enfocado a nivel de la organización.

La investigación en torno a la medición del rendimiento de microseguros está todavía en su fase inicial. La guía práctica de los KPIs elaborada por Wipf y Garand, se considera un importante avance, sin perjuicio de que queda todavía mucha trayectoria por recorrer y no pocas lagunas pendientes de cubrir. Una de ellas es la de no diferenciar entre la gran diversidad de proveedores de microseguros; y otra, relacionada principalmente con el aspecto financiero de los indicadores, es que deja en un segundo plano el aspecto social, tan relevante en este negocio. Sobre este aspecto inciden Biener y Eling (2012) en su estudio, en el cual, utilizando una base de datos más reciente se aplican metodologías innovadoras que incorporan la función social realizada por las microaseguradoras para medir toda la eficiencia real del producto.

PARTE II. ESTUDIO EMPÍRICO

Esta sección es la que aporta mayor relevancia a la investigación porque es donde se lleva a la práctica nuestra propuesta metodológica de tarificación, basada en el modelo lineal generalizado que sigue una distribución Tweedie, aplicada a un caso real correspondiente a un microseguro combinado. La finalidad principal será estimar unas primas de riesgo según el perfil de riesgo del colectivo asegurado, de forma que proporcionen una tarifa más ajustada a la verdadera siniestralidad, y así dotar a las compañías de un mayor control y conocimiento del negocio microasegurador.

Para desarrollar este cometido, se ha dividido el estudio empírico en tres capítulos. El primero de ellos, consiste en seleccionar la metodología de tarificación que se va a utilizar en este trabajo. Para ello, se valorarán las diferentes técnicas predictivas así como los datos disponibles del microseguro que permitan elegir convenientemente la técnica predictiva más apropiada para nuestro estudio. El segundo, se centra en ofrecer una visión general de la información disponible de nuestro particular caso de estudio. El cual consistirá en describir todo el proceso que se ha llevado a cabo para la obtención de los datos, su posterior tratamiento y un análisis descriptivo completo de las diferentes variables. El tercer y último capítulo de esta sección corresponde a la aplicación empírica del modelo Tweedie a nuestro producto. En el estudio se irán analizando diferentes alternativas hasta obtener un modelo óptimo lo suficientemente robusto que sea capaz de proporcionar la tarifa más ajustada a la experiencia de siniestralidad de nuestro microseguro.

CAPÍTULO 3. METODOLOGÍA DE TARIFICACIÓN

Este primer capítulo de la parte empírica tiene por objeto sentar las bases metodológicas, en materia de tarificación de seguros, que permitan calcular las primas de riesgo de nuestro microseguro. Se trata de demostrar que es posible utilizar las técnicas predictivas, ampliamente conocidas y empleadas en el seguro tradicional, para la estimación de la tarifa del producto en el nuevo campo de los microseguros.

Con esta finalidad se desarrolla la siguiente estructura. En primer lugar, comenzaremos con una revisión del estado de la cuestión en materia de tarificación de microseguros, resaltando la importancia del papel del actuario como principal protagonista del proceso y describiendo los métodos más utilizados en la práctica, así como los últimos estudios desarrollados sobre esta materia. En segundo lugar, entramos de lleno en el fundamento metodológico de esta investigación, la discusión en torno a la selección de la metodología de tarificación propuesta, el modelo Tweedie. Además, se describen los diferentes procedimientos que van a ser aplicados a lo largo de todo el análisis y que comprenden desde el tratamiento de las variables, las hipótesis de validación del modelo, hasta las medidas estadísticas empleadas tanto en la selección de variables como en evaluar la bondad de ajuste de los modelos. En tercer lugar y para finalizar, se expone el desarrollo matemático de la distribución Tweedie perteneciente a la familia exponencial de distribuciones, lo que permitirá la aplicación de los modelos lineales generalizados.

3.1 La tarificación de microseguros

3.1.1. El papel del actuario

El proceso de tarificación de microseguros consiste en una tarea técnica, que requiere de la participación del actuario como especialista en la estimación de primas. El experto no solo debe tener conocimiento técnico para analizar los riesgos potenciales, sino que debe ir más allá y tener la capacidad y habilidad suficiente para desempeñar otros roles más globales del negocio (Wrede et al., 2016).

- *Tener una visión global del mercado*, que le permita así diseñar el producto combinando su juicio experto en base, no solo a su experiencia acumulada, sino también a otros factores adquiridos en el negocio microasegurador (Wipf & Garand, 2006). El microseguro se enmarca en un entorno sencillo e inclusivo, lo que evita establecer condiciones que limiten su alcance, como son las exclusiones y los copagos.
- *Realizar el trabajo con transparencia y sencillez*. En este mercado existen muchas barreras culturales que dificultan la comprensión del seguro por parte de la población objetivo. El actuario debe saber transmitir de forma adecuada la verdadera filosofía del seguro, en la cual los asegurados pagan por anticipado una prima no reembolsable a cambio de una futura protección ante la ocurrencia de riesgos adversos.
- *Responsabilidad en el diseño y en la monitorización de los datos*. La importancia de disponer de una buena base de datos para realizar los cálculos hace que el actuario sea el responsable de garantizar la calidad de la información. También, es esencial conocer bien el contexto donde se desarrolla este mercado para interpretar correctamente el funcionamiento del negocio y hacer recomendaciones sobre la gestión del producto.
- *Conocimiento sobre la diversidad de productos*. La falta de expertos en microseguros hace que el grado de especialización no sea el mismo que en el seguro tradicional. El actuario adquiere así la capacidad y experiencia de saber tarificar diferentes tipos de productos. Es habitual encontrar en este tipo de negocio productos combinados, que cuentan con diferentes tipos de coberturas, ya sean de ramos de vida, de salud o de seguros generales.

- *Contribuir a la estrategia del plan financiero de negocio.* El actuario debe ofrecer posibles escenarios de prima que puedan adaptarse al plan de negocio definido por la dirección. En relación a esto, existen ciertos indicadores de referencia que pueden servir para realizar una comparación de las primas con otras de otros productos ya existentes en el mercado (McCord et al., 2014).

3.1.2. Estimación de la prima

La prima estimada debe cumplir con los principios básicos de suficiencia y equidad. El principio de suficiencia supone que el valor presente de las primas, y que es equivalente al riesgo que asume la población de bajos ingresos, sea suficiente para cubrir el coste esperado de los futuros siniestros y además sea capaz de generar un justo beneficio para todas las partes implicadas en el negocio. Básicamente se expresa con la siguiente ecuación de seguros:

$$P_{comercial} = P_{pura} + G_i + G_e + B \quad (3.1)$$

Siendo G_i los gastos comerciales, que incluyen las comisiones y los gastos de comunicación. Esta partida suele ser elevada para los microseguros. G_e son los gastos administrativos, que no están relacionados directamente con la distribución del producto. Y B es el margen de beneficio que obtiene la compañía. Lógicamente, el objetivo de toda estimación de primas trata de asegurar que dicha ecuación esté perfectamente equilibrada.

Partiendo de la base de que los siniestros futuros son aleatorios y la prima se establece a priori, puede ocurrir con una cierta probabilidad que la prima pura técnica no sea suficiente para cubrir los costes incurridos derivados de los siniestros. No obstante, la compañía puede controlar la posible insuficiencia de las primas añadiendo un recargo en función de la distribución de los siniestros (Biener, 2013). Dicho esto, la prima pura incluye, no solo la mera valoración del riesgo por la prima estadística P_e , sino que también se añade un margen de seguridad δ que cubre la incertidumbre asociada a posibles variaciones en la valoración de los riesgos objeto de cobertura. Atendiendo a la práctica del mercado, este margen de seguridad δ suele situarse en torno al 5-10% (Blacker & Yang, 2015). Y se expresa a través de la siguiente fórmula:

$$P_{pura} = P_e(1 + \delta) \quad (3.2)$$

Esta consideración hace que sean las propias limitaciones inherentes al mercado microasegurador las que obliguen a las compañías a añadir este margen de seguridad adicional, con el principal propósito de compensar la incertidumbre asociada a la estimación de las primas y mejorar así la sostenibilidad del producto (Biener, 2013; Purcal et al., 2012). Paradójicamente, esta situación conduce a un dilema, al que se enfrentan muchas compañías, que consiste en aplicar recargos superiores a los microseguros que al resto de seguros, cuando la población objetivo per se es más vulnerable y posee una menor capacidad económica inferior respecto al resto de clientes del seguro tradicional. Esto implica una sobrevaloración del riesgo, encareciendo el producto y haciéndolo menos atractivo para este tipo de población. Ante esta disyuntiva, Biener (2013) plantea varias soluciones desde el punto de vista actuarial, por una parte, a través de una estimación de primas más precisa para reducir así la carga por incertidumbre, y por otra parte, estableciendo estrategias de gestión de riesgo para incrementar la capacidad del seguro.

3.1.3. Métodos clásicos

En la práctica, la metodología que se emplea para la tarificación de los microseguros suele ser básica. Se estiman las primas de riesgo en base a una clase o comunidad asegurada y sin identificar el perfil de riesgo individual de los asegurados. Este tipo de tarificación colectiva es utilizada para los seguros de grupo, que vienen siendo los más extendidos en la práctica. Algunas de las razones de su utilización vienen por los bajos costes, tanto de administración, de distribución como de gestión de siniestros, además del requerimiento de alcanzar una cuota mínima de suscripción en el caso de que la cobertura sea obligatoria (Wrede et al., 2016).

Profundizando ahora en la tipología de métodos de tarificación en este campo, algunos autores (Biener, 2013; Wrede et al., 2016) señalan el uso de tres principales metodologías. Por un lado, la tarificación a priori que contempla dos enfoques, uno, respecto a la experiencia en siniestralidad, y el otro, según la exposición del negocio. Y por otro lado, la tarificación a posteriori que utiliza la teoría bayesiana de la credibilidad.

A. Método basado en la experiencia

Este enfoque, también conocido en la literatura anglosajona como *burning cost pricing*, utiliza la experiencia pasada como base para conocer el comportamiento del riesgo de la población asegurada y estimar así la siniestralidad futura (Wrede et al., 2016). La prima final será calculada por la combinación de dos variables, la primera, la siniestralidad obtenida por la experiencia histórica, y la segunda, la exposición basada en la hipótesis de la cartera en vigor.

La aplicación de esta técnica basada en la experiencia, precisa de una buena base de datos con un recorrido suficiente para predecir correctamente el comportamiento de riesgo de la población asegurada. Este método resulta una buena alternativa para la tarificación de microseguros, siempre y cuando no se esperen cambios importantes respecto a la situación inicial. Si bien, es importante realizar el oportuno seguimiento y considerar cualquier posible aspecto susceptible de cambio, ya sea en el producto, en la población asegurada o incluso en factores externos tales como la inflación.

B. Método basado en la exposición

La exposición se considera una de las variables clave a la hora de tarificar, hasta el punto de convertirse en la variable principal de uno de los métodos más utilizados en la práctica cuando no se dispone de suficiente información (Wrede et al., 2016). Este enfoque trata de establecer las hipótesis básicas en el cálculo de la prima, que corresponden a las variables frecuencia y coste medio, utilizando las características propias del colectivo asegurado. También, suele ser habitual emplear otras fuentes para complementar las hipótesis previas asumidas. No obstante, es recomendable ir sustituyendo estas a medida que la compañía vaya adquiriendo mayor experiencia, y que refleja realmente el verdadero comportamiento de los siniestros.

La recurrente falta de información en este mercado hace que esta técnica sea la más utilizada en la tarificación de microseguros (Wrede et al., 2016). Entre sus principales aplicaciones en el seguro directo se utiliza, bien como método de partida para la tarificación de riesgos nuevos y/o de tamaño reducido, o bien como estrategia de validación respecto al método anterior basado en la experiencia. También, es de gran aplicación para el cálculo del reaseguro exceso de pérdidas, donde las primas se estiman realizando un análisis sobre la exposición del negocio.

C. Modelos de credibilidad

En la tarificación a posteriori, se destacan los modelos basados en la teoría de credibilidad. Esta técnica permite combinar distintas fuentes de información, asignando un grado de confianza a la experiencia propia del colectivo a través de una fórmula de credibilidad (Biener, 2013; Déniz & Alegría, 2008; Peña, 2006). El teorema de Bayes, sobre el que gira este enfoque, permite transformar una distribución inicial o a priori (que tiene la información inicial) en una distribución final o a posteriori, que es la que añade experiencia al propio grupo asegurable. En el momento de disponer de la información adicional sobre el colectivo, es cuando conviene de nuevo replantearse las hipótesis iniciales para estimar la distribución final.

Centrando la atención en los microseguros, en donde existe una evidente restricción de información, los modelos de credibilidad resultan de gran utilidad para realizar diversas aplicaciones. Entre ellas, es posible estimar por primera vez la tarifa de un nuevo producto mediante la combinación de distintas fuentes de información, las cuales permiten seleccionar las características del colectivo que mejor representen el comportamiento del riesgo. Estas fuentes pueden venir, bien de otros países u otras regiones cercanas, dependiendo de la cantidad de información disponible y de la distancia respecto a la región de interés (Christopherson & Werland, 1996, citado por Biener, 2013), o bien a través de grupos socio-económicos. Otra posible aplicación consiste en la monitorización de las primas, la credibilidad facilita el hecho de ir ajustando el grado de confianza asignado a la experiencia a medida que se disponga de mayor información del colectivo asegurado. Y es en este punto, donde el actuario debe ser especialmente objetivo para valorar qué información debe ser o no utilizada en el modelo, además de elegir qué grado de confianza va a asignar a esta información adicional (Wrede et al., 2016).

Desde la visión práctica de los seguros agrícolas, y en particular en los innovadores productos basados en índices, algunos autores (Clarke et al., 2012) defienden la eficiencia de la teoría bayesiana de credibilidad frente a otros métodos tradicionales. En concreto, utilizan el modelo de Bühlmann para estimar el umbral por debajo del cual la compañía asume el riesgo y, posteriormente determinan las primas diferenciadas por grupos de riesgo.

En resumen, los tres métodos de tarificación aquí descritos pueden realizar correctas estimaciones de primas basándose en mayor o menor medida de la experiencia acumulada del negocio. No obstante, es necesario hacer especial hincapié en la importancia de ir incorporando la experiencia adquirida del negocio, debido a que cualquier alteración en la situación inicial puede provocar variaciones en el comportamiento del riesgo y en consecuencia incorrectas estimaciones de primas.

3.1.4. Desarrollos en la tarificación de microseguros

La creciente demanda a nivel global de microseguros ha fomentado un mayor interés entre los profesionales del sector en investigar sobre este nuevo negocio. Y en particular, es en el campo actuarial donde impera la necesidad de analizar el comportamiento del riesgo asegurado, motivado principalmente por la escasez de información disponible. Esto hace que el objetivo de desarrollar la tarificación de los microseguros se convierta en todo un reto.

Como solución, para complementar esta falta de información, Wipf y Garand (2006) plantean la aplicación de técnicas de modelización. La metodología bootstrapping (Efron & Tibshirari, 1994; England & Verral, 1999) emplea la técnica del remuestreo con reemplazamiento para obtener diversas muestras alternativas, que se basan en las propiedades de la distribución de datos original. El modelo final resultante se calcula a partir de estos datos remuestreados, concediendo así una mayor estabilidad a la información de base para la tarificación (Biener, 2013; Goldburd et al., 2016).

Siguiendo la línea de investigación sobre la aplicación de nuevos métodos de tarificación en microseguros, Purcal et al. (2012) proponen un sencillo método basado en las cadenas de Markov que permite estimar una tarifa más eficiente para un microseguro de vida, utilizando una matriz de probabilidades de transición respecto a la exposición del colectivo. Los resultados ofrecen primas más reducidas y ajustadas de acuerdo a este modelo, lo que incentiva la cobertura a largo plazo y desincentiva el riesgo de selección adversa que surge de la cobertura a corto plazo.

Desde el punto de vista práctico, la compañía Milliman junto con la Asociación Internacional de Actuarios (IAA) desarrollaron en el año 2015 una **herramienta** básica de tarificación en Excel para un microseguro de salud (Milliman, 2015). El objetivo principal de esta iniciativa era la de ilustrar a los actuarios de microseguros, sobre los

principios básicos que se deben utilizar en la tarificación de un producto de salud. En esta misma línea y con la misma finalidad de formación técnica, el grupo de trabajo de microseguros de la profesión actuarial de Reino Unido ha diseñado también otra herramienta de tarificación similar para el caso de un producto de vida crédito.

3.2 Metodología de la investigación

Este punto constituye el pilar fundamental de la investigación, centrado en desarrollar la metodología de análisis predictivo para la estimación de las primas de riesgo de nuestro microseguro objeto de estudio. Durante su desarrollo se han ido valorando las diversas alternativas según los objetivos de cada fase y en función de la información disponible, hasta llegar a justificar la elección de la metodología seleccionada para esta investigación. La estructura está planteada siguiendo el orden lógico en la realización del análisis predictivo. En primer lugar, se plantea y justifica la elección del modelo Tweedie como la mejor alternativa metodológica en la estimación de primas. En segundo lugar, los criterios fijados en el proceso de definición, tratamiento y segmentación de las variables junto con la validación del modelo. Y finalmente, una descripción de las medidas de bondad de ajuste utilizadas durante todo el análisis.

La herramienta principal para la realización del análisis de los modelos predictivos y demás procesos ha sido SAS Studio, una aplicación web de desarrollo del paquete estadístico SAS. La elección de esta herramienta está justificada por la facilidad que brinda en la construcción de modelos lineales generalizados, gracias a los procedimientos que tiene previamente definidos la propia herramienta. Además, el programa SAS cuenta con la capacidad y la rapidez suficiente para tratar muestras generosas de datos.

3.2.1. Modelo Tweedie

La metodología de tarificación a priori, sobre la que basaremos la presente investigación, depende en primera instancia de la información disponible. En función de la calidad y cantidad de los datos recopilados se podrán aplicar diferentes técnicas predictivas de estimación de tarifas.

Si ahondamos dentro de la información disponible, puede ocurrir que esta venga desagregada por el número y cuantía de siniestros en cuyo caso se podría plantear la aplicación de modelos predictivos bajo dos posibles opciones:

- *Una regresión de Poisson* con los datos del número de siniestros, siempre y cuando se tenga la certeza de que la heterogeneidad se concentra en esta variable respuesta. Acto seguido, se seleccionarán los factores de riesgo que influyen sobre el número de siniestros para determinar finalmente los estimadores que nos aportarán la tarifa del producto. En la práctica habitual, la variable frecuencia de siniestros suele ser bastante estable, lo que favorece en una mayor precisión de cálculo de los estimadores (Denuit et al., 2007; Ohlsson & Johansson, 2010).
- *Una doble regresión* que combina un modelo Poisson con la variable número de siniestros y otro modelo Gamma para los datos de la cuantía del siniestro. Posteriormente, se procederá, de igual manera al caso anterior, seleccionando los factores de riesgo influyentes sobre la variable respuesta y obteniendo así los estimadores para cada uno de los modelos. Finalmente, para calcular la tarifa se suele aplicar un modelo multiplicativo que será el producto de las esperanzas de las estimaciones de ambas regresiones (Denuit et al., 2007).

El cálculo separado proporciona una mayor percepción sobre cómo afectan los factores de riesgo al cálculo de la prima pura (Anderson et al., 2007; Ohlsson & Johansson, 2010). Esta opción suele ser la preferida por las compañías, siempre y cuando se disponga de una buena base de datos con la suficiente calidad y granularidad para poder realizar el análisis separado para las dos variables básicas en el cálculo de primas.

En algunas ocasiones, la información disponible puede no ser suficiente para utilizar las opciones ya contempladas, bien por la escasez de datos o bien porque los campos vienen agregados en una sola variable, coste último total. Esta situación es la que viene planteándose en el mercado de los microseguros y donde se enfocará la actual investigación. En el caso particular de nuestro estudio, es preciso analizar en detalle las limitaciones que presenta la información disponible, ya que van a condicionar la aplicación de una u otra metodología:

- En primer lugar, la dificultad más grave viene por el elevado porcentaje de ceros recogidos en la muestra, debido a que la mayoría de asegurados no presentan siniestro alguno. Esta situación hace que la regresión ordinaria no funcione correctamente debido al cuantioso número de ceros que impide estimar una recta de regresión que resulta en un modelo predictivo realista.
- En segundo lugar, la alta asimetría que presenta la variable respuesta, coste último total ponderado por la exposición. La experiencia de siniestralidad acumulada en el período presenta una clara asimetría a la derecha que denota la falta de normalidad en la distribución.

A pesar de todas estas dificultades, el presente trabajo plantea la posibilidad de aplicar una metodología predictiva que permita modelizar el coste último total por expuesto, seleccionando las variables explicativas según su mayor significación y obteniendo la estimación de las primas de riesgo del producto. Las dos alternativas, que resuelven estas dificultades, parten de dos ópticas diferentes, desde la perspectiva econométrica se contempla el modelo de regresión censurada también llamado de Tobit, y bajo el punto de vista actuarial está el modelo lineal generalizado de la distribución Tweedie.

El primer caso es el **modelo de regresión censurada o Tobit**. La censura es un defecto de la muestra, ocurre cuando existe una limitación o pérdida de datos en la variable dependiente pero no de los predictores. En nuestro caso particular, existe un notable número de ceros en la variable respuesta que podría interpretarse de forma forzada como que los datos nulos están censurados por la izquierda. A partir de aquí, se define una variable latente, no censurada, sobre la que se realiza la regresión lineal. La nueva nube de puntos incorporará los datos no censurados, y bajo la hipótesis de normalidad dará lugar a una recta de regresión con capacidad para explicar la variable respuesta del modelo.

El segundo caso es el **modelo lineal generalizado basado en la distribución Tweedie**, que estima directamente la prima de riesgo del microseguro. Esta técnica predictiva permite que la distribución de la variable respuesta pertenezca a la familia exponencial, lo que significa que la varianza no es constante y por tanto la variable respuesta sea heterocedástica. Esta es una de las características diferenciadoras respecto a la regresión lineal ordinaria, donde la homocedasticidad de la variable debe prevalecer (De Jong & Heller, 2008). La función de distribución Tweedie pertenece a la

familia exponencial de distribuciones, y en concreto cuando la variable respuesta es el coste último total se asume que sigue una distribución Tweedie de potencia p que está comprendido entre $1 < p < 2$ y que permite estimar la prima pura de riesgo (Jørgensen & De Souza, 1994). La distribución Tweedie se define como una función de distribución Poisson compuesta con una Gamma, que, por un lado, asigna una probabilidad al cero igual a la probabilidad de Poisson de que no suceda ningún siniestro. De esta manera, acoge de forma natural la posibilidad de que una muestra tenga un nutrido grupo de ceros, sin necesidad de acudir a interpretaciones forzadas acerca de la censura de una variable latente. Y por otro lado, establece una distribución para el resto de observaciones de la variable.

En vista de las dos soluciones planteadas y analizando nuestra variable respuesta del coste último total por expuesto, se considera que la distribución Tweedie es la mejor solución para la aplicación del modelo predictivo. Aunque el modelo Tobit interpreta el elevado número de ceros, presente en nuestra base de datos, la distribución de nuestra variable respuesta presenta una clara asimetría y una falta de normalidad que impide su correcta aplicación. Además, la solución actuarial ofrece un uso muy extendido en la tarificación a priori del seguro tradicional, lo que hace que pueda ser fácilmente trasladable al campo de los microseguros, a la vez que puede resultar una propuesta innovadora ya que se desconoce su aplicación en este tipo de productos.

Se concluye por tanto que, bajo la limitación de información disponible, el modelo predictivo basado en la distribución Tweedie es el mejor candidato para estimar las primas de riesgo del microseguro, además de aportar un aspecto innovador en su aplicación a la tarificación de microseguros.

3.2.2. Definición y tratamiento de variables

La metodología de tarificación que se propone en este trabajo, los modelos lineales generalizados, trata de explicar la relación entre la variable respuesta y (lo que se quiere predecir) y las variables explicativas x (los predictores o covariables). Este análisis predictivo comienza por definir la variable objetivo que representa la finalidad del cálculo que queremos realizar y que se traslada al modelo a través de la variable respuesta. Centrando la atención en el campo de tarificación de seguros, las principales variables objetivo se resumen claramente a continuación.

Tabla 3.1. Principales variables objetivo de los modelos GLM.

Variable objetivo $Y=X/w$	Variable respuesta X	Peso w
Frecuencia	Número de siniestros	Exposición
Coste medio	Coste último total	Número de siniestros
Prima pura	Coste último total	Exposición
Ratio de siniestralidad	Coste último total	Prima imputada

Fuente: elaboración propia a partir de Ohlsson E. y Johansson, B. (2010). *Non-life insurance pricing with generalized linear models*, vol. 21. Berlin: Springer.

Como se observa en la Tabla anterior toda variable objetivo cuenta con una misma estructura, la proporción entre el resultado de una variable aleatoria, la variable respuesta X, y una unidad de medida, el peso w . El término peso puede representar diversas variables según nuestro objetivo de cálculo, y a medida que se disponga de más información sobre esta variable mejores resultados se obtendrán, ya que repercute directamente en una menor variación de la variable objetivo (Ohlsson & Johansson, 2010).

Las variables explicativas del modelo, que van a representar los factores de riesgo de la tarifa del producto, juegan un papel sustancial en el análisis predictivo. Cuanto mejor se definan estos predictores mayor capacidad tendrá nuestro modelo para explicar la variable respuesta, que en nuestro caso particular es el coste último total por expuesto. Los predictores que se pueden utilizar en cualquier modelo lineal generalizado se pueden clasificar según el tipo de elemento¹⁴ que contengan, como variables de carácter cualitativo o cuantitativo:

- *Las variables cualitativas* se refieren a características que no pueden medirse con números. En esta tipología se distingue entre la variable cualitativa nominal, que presenta atributos que no admiten un criterio de orden, y la variable ordinal en las que sí existe un orden. Estas variables como tal no pueden ser incluidas dentro del modelo, salvo que sean previamente tratadas a través de algún procedimiento para que sean consideradas como variables numéricas y poder así incorporarlas al modelo.

¹⁴ Cachero (1993) establece una denominación diferenciada según los valores de una variable estadística, como *datos* a los elementos de carácter cuantitativo y como *atributos* a los elementos de carácter cualitativo.

- *Las variables cuantitativas* se expresan mediante números y se distinguen dos tipos, discretas o continuas. Una variable cuantitativa puede ser discreta, cuando puede tomar un número finito de valores entre dos posibles, o continua, cuando puede ser un número infinito de valores en un determinado intervalo.

En función de esta tipología se definirá su intervención en el modelo. De forma que si la variable es continua, su estimador vendrá representado por un único coeficiente para todos los valores de la variable. De otro modo, si es discreta, su rango de valores determinará los diferentes niveles de la variable que tendrán asociado su correspondiente estimador. Entre todos los niveles de la variable se designará un nivel de referencia, que se conoce como nivel base, sobre el que se calcularán el resto de coeficientes de los diferentes niveles de la variable.

Respecto a la elección del nivel base de las variables numéricas discretas, Goldburd et al. (2016) señalan la importancia de analizar correctamente las categorías de la variable con el fin de seleccionar un nivel base adecuado. De esta forma, se evita escoger un nivel con pocos datos ya que los coeficientes estimados estarían afectados por esa baja frecuencia, provocando una mayor variabilidad en las estimaciones. Por este motivo, y sabiendo que los resultados del modelo varían en función del nivel base elegido, uno de los criterios establecidos es seleccionar aquel nivel que presente mayor frecuencia.

La herramienta de análisis SAS establece por defecto como nivel base el último de los niveles de la variable discreta, pero dispone de cierta flexibilidad ofreciendo otros criterios para su selección. En concreto, para nuestro análisis vamos a considerar como nivel base de las variables discretas aquel valor que presente mayor exposición de la cartera, que casualmente coincide con el criterio de mayor frecuencia. Bajo esta equiparación, SAS contempla, dentro de las opciones de clasificación de variables, el criterio de mayor frecuencia de observaciones a través de la sentencia *DESC ORDER=FREQ*.

Adicionalmente, el modelo también proporciona un término independiente denominado *intercept*, que es un parámetro que se aplica a todas las observaciones. Este valor recoge por defecto los niveles base de todas los predictores del modelo, y lógicamente su valor cambiará en función de los niveles base elegidos para cada variable discreta (Anderson et al., 2007).

Una vez definidas las variables explicativas, también es necesario concretar la estrategia seguida en el tratamiento de estas variables. Para la presente investigación se va a utilizar como modelo de cálculo la distribución Poisson con el número de siniestros. Su elección viene condicionada por varios aspectos: por un lado, el número de siniestros es una variable que forma parte del modelo Tweedie que se corresponde con la distribución Poisson compuesta con Gamma; y por otro lado, por un criterio de sencillez en su aplicación en SAS respecto al modelo Tweedie implementado para la estimación de la prima pura.

A continuación se pasa a explicar todo el proceso seguido en el tratamiento de las variables explicativas con cada una de sus diferentes etapas, que comienza en primer lugar, con la técnica de segmentación empleada para determinar cuáles son los niveles homogéneos de la variable, y en segundo lugar, el criterio elegido para valorar las posibles interacciones en el modelo.

A. Segmentación

Los predictores pueden ser variables numéricas y contener pocos valores, en cuyo caso no es necesario tratamiento alguno, en cambio hay otros casos en los que inevitablemente sí se requiere una agrupación previa para incorporarlos al modelo, bien porque presentan un rango amplio de valores, o bien porque no se disponen de suficientes datos. A este proceso se le denomina segmentación, y consiste en reagrupar los diferentes valores de cada predictor en categorías de riesgo homogéneas en función de la exposición y experiencia de los siniestros (Goldburd et al., 2016). Esta solución tiene un doble efecto, por una parte, dará mayor robustez en las estimaciones y en la validez del modelo, sin embargo, por otra parte, perderá granularidad y finura de las estimaciones (Ohlsson & Johansson, 2010). Por esta razón, hay que buscar el equilibrio en obtener un número de niveles que sea homogéneo y representativo para la variable analizada.

La estrategia de segmentación que se va a utilizar para la presente investigación emplea un proceso iterativo y secuencial de un modelo lineal generalizado de Poisson con la variable número de siniestros para decidir cómo agrupar los diferentes niveles de la variable (Brisard, 2014). Además, este modelo considera la exposición como peso, lo que hace que se definan niveles homogéneos para la variable analizada, no solo por el número de siniestros, sino por la composición de asegurados en la cartera.

El proceso comienza calibrando el modelo para un único predictor y sin previa clasificación. Con los resultados obtenidos, se comparan los valores de sus parámetros estimados junto con su nivel de significación, y en función de esto se realizan las agrupaciones según la mayor proximidad entre sus valores para crear así los niveles homogéneos. Es recomendable ir agrupando poco a poco los diferentes valores para lograr una mayor estabilidad en las nuevas categorías. Después de esto, se crea una nueva variable con los niveles elegidos y se elabora de nuevo el modelo con dicha variable reducida. Este planteamiento se repetirá una serie de veces hasta reducir progresivamente los niveles de la variable y alcanzar un número adecuado para este modelo Poisson, con la ayuda de las medidas de ajuste y el nivel de significación que se vayan obteniendo en los diferentes niveles.

B. Interacción

La interacción se define como el efecto combinado que se puede producir entre dos o más variables explicativas sobre la variable respuesta, y sobre sus efectos individuales. Es decir, existe interacción cuando una variable varía según los niveles del otro predictor (Goldburd et al., 2016; Werner & Modlin, 2016).

Ahora bien, al igual que en el planteamiento seguido para la segmentación, el procedimiento de análisis de la interacción entre variables sigue el mismo criterio, que consiste en utilizar el modelo Poisson con el número de siniestros para analizar su influencia en el modelo. Este proceso comienza en primer lugar, analizando los predictores disponibles en nuestro modelo y planteando una combinación lógica de sus posibles interacciones. Después, se identifica el tipo de interacción que se puede dar según el tipo de variable. Y dado que todas las variables son discretas, la estrategia resulta más sencilla siendo el término interacción la simple combinación de estas variables. En segundo lugar, se realizan dos modelos por cada interacción, uno incluyendo tan solo los predictores individuales y otro incluyendo además esta variable interacción, que tendrá después su correspondiente estimador asociado. Por último, se comparan los resultados de ambos modelos, con y sin el término interacción, mediante las medidas de bondad de ajuste para confirmar si la incorporación de la interacción produce mejores resultados y debe por tanto ser incluida dentro del modelo.

3.2.3. Validación del modelo

La realización del modelo predictivo debe estar siempre respaldada por la bondad de sus resultados, así como de la confianza en su aplicación para la práctica aseguradora. Con este objetivo se contemplan dos enfoques diferentes, el primero de ellos hace referencia a la selección de muestras con las que se va a realizar el análisis, y el segundo de ellos, describe el planteamiento seguido para analizar la estabilidad financiera del modelo analizado.

A. Partición de la base de datos

La información es la herramienta básica para la realización de cualquier análisis predictivo, ya que las estimaciones dependen de los datos disponibles. Por este motivo, es conveniente seleccionar y utilizar bien estos datos para demostrar la eficacia de nuestro modelo.

Para este menester, se propone dividir la base de datos en varias muestras que permitan no solo realizar el modelo, sino también poder comprobar la validez de sus resultados. Para ello, se plantean dos posibles planteamientos y su elección vendrá condicionada, en gran medida, por la disponibilidad de muestras lo suficientemente representativas.

- *La validación cruzada* es un método estadístico que consiste en la partición de la base de datos en k^{15} múltiples muestras, separando en subconjuntos de datos de entrenamiento y de test, con el propósito de calibrar las diferentes muestras y evaluar la bondad del modelo. Se realizarán diferentes modelos por cada una de las $k-1$ muestras y estos se verificarán con los datos test restantes. El resultado de este proceso dará lugar a k estimaciones del modelo, en las que cada una de ellas se habrá realizado con datos que otra no contenga (Goldburd et al. 2016). En nuestro caso particular, esta opción se ha descartado debido a la limitación de datos disponibles a la hora de establecer las k particiones.
- *La partición simple, muestras entrenamiento y test.* Esta alternativa, que es la que se utilizará en la presente investigación, plantea una validación más sencilla y muy utilizada en la práctica, que consiste en dividir la base de datos en dos

¹⁵ El número de k muestras puede ser cualquier valor, aunque lo habitual suele ser de 10. Y la división puede realizarse bien de forma aleatoria, o bien utilizando alguna variable temporal como el año de ocurrencia (Goldburd et al., 2016).

muestras, una de entrenamiento y otra de test (Goldburd et al., 2016; Towers Watson, 2010). La división, aunque se puede realizar de varias formas, en nuestro caso se ha hecho de forma aleatoria. La muestra de entrenamiento, de mayor volumen, contendrá el 70% de los datos y se utilizará para la construcción del modelo con su correspondiente estimación de parámetros. La muestra test lo formará el 30% de los datos restantes y se utilizará para comprobar la efectividad del modelo obtenido, así como para comparar los resultados de los modelos alternativos. Finalmente, el modelo óptimo se construirá con el conjunto total de la muestra, el 100% de los datos, para proporcionar un modelo que plantee resultados más robustos y eficientes.

B. Análisis de equilibrio financiero

Los modelos lineales generalizados proporcionan la mejor estimación de las primas en base a la experiencia de la población asegurada. Sin embargo, es indispensable contrastar que estos resultados son asumibles en la práctica real por las compañías, y para ello se propone realizar un análisis del equilibrio financiero como herramienta de decisión para valorar la eficacia y estabilidad de los modelos, seleccionando así la mejor alternativa. Este estudio contendrá tres partes esenciales, un análisis de la variabilidad de las primas, la estimación del error cuadrático medio y un análisis del ratio de siniestralidad.

1. Análisis de variabilidad de primas

Este análisis utiliza las primas estimadas, que asignaremos como P_i , para calcular un nuevo concepto objeto de estudio, la proporción relativa de prima estimada por suma asegurada A , es decir P_i/A . Este término nos proporciona una medida relativa del coste unitario del microseguro por cada unidad asegurada. Además, tiene la ventaja de ser una medida que puede ser comparable con otros microseguros similares existentes en el mercado. A partir de esta nueva variable se procederá entonces a realizar una estadística descriptiva con las principales medidas de dispersión y de forma, para analizar su distribución y variabilidad, así como una medida comparativa para contrastar varios modelos planteados.

2. Estimación del error cuadrático medio

El proceso predictivo calcula, en base a los parámetros, las correspondientes estimaciones del modelo. Y toda estimación lleva asociada su correspondiente error,

conocido como error cuadrático medio (ECM^{16} de ahora en adelante), que compara el valor estimado con el valor observado (Hastie et al., 2008). La preferencia está en utilizar aquellos estimadores que presenten un menor error cuadrático. Al hilo del estudio que aquí se plantea y siguiendo con el análisis anterior, la fórmula de cálculo del ECM se obtiene en base a las variables ponderadas por la suma asegurada A que vendrá dada por la siguiente relación:

$$ECM = \frac{\sum_{i=1}^n \left[\left(\frac{S_i}{A} - \frac{P_i}{A} \right)^2 \times e_i \right]}{\sum_{i=1}^n e_i} \quad (3.3)$$

Donde S_i/A es el coste último total por cada unidad de suma asegurada A y P_i/A es la prima de riesgo estimada por la suma asegurada. La exposición se define como e_i para cada asegurado $i = 1, \dots, n$. En esta fórmula, el numerador expresa la diferencia entre los valores observados y los estimados, al elevarlo a su cuadrado evitamos que las posibles diferencias negativas compensen a las positivas. Del mismo modo que las primas se estiman en base a la exposición, la medida del ECM también va a ir ponderado por la exposición del asegurado.

Por tanto, el ECM se trata de una buena medida que nos va a permitir comparar los diferentes modelos, analizar su eficacia y como herramienta de decisión para seleccionar el modelo óptimo.

3. Análisis del ratio de siniestralidad

Como última herramienta de decisión para estudiar la estabilidad financiera de los modelos analizados se emplea el ratio de siniestralidad. Se trata de una medida de la parte de cada prima que se utiliza para pagar los siniestros del producto en un período determinado, su fórmula de cálculo es muy sencilla:

$$\text{Ratio siniestralidad} = \frac{\text{Siniestralidad}}{\text{Primas}} \quad (3.4)$$

Esta medida es utilizada tradicionalmente por las compañías para analizar y monitorizar la siniestralidad del negocio, además de ser un excelente indicador para medir la suficiencia de las primas del producto (Werner & Modlin, 2016).

¹⁶ En la literatura anglosajona se conoce como *Mean Squared Error (MSE)* (Hastie et al., 2008).

Para este estudio en particular, se plantea un análisis descriptivo de las primas junto con la siniestralidad del período y el correspondiente ratio de siniestralidad. Esta medida nos permitirá comparar los resultados de los diferentes modelos para analizar la sostenibilidad del producto de acuerdo al plan de tarifas de cada modelo.

3.2.4. Medidas estadísticas del modelo

Dentro de la metodología planteada es preciso establecer también las herramientas que se van a utilizar para calibrar la bondad y eficacia del modelo. En la práctica, puede ocurrir que del análisis predictivo se obtenga no solo un único modelo adecuado, sino que sean varios los que sean capaces de explicar correctamente la variabilidad de la variable respuesta. Bajo esta observación, la idea que se pretende es seleccionar el modelo óptimo que mejor cumpla las hipótesis de validación anteriormente definidas, junto con los mejores resultados de las medidas que aquí se plantean.

Para este trabajo de investigación se plantea utilizar diferentes medidas estadísticas en función del objeto de estudio. En primer lugar, para analizar la significación de los predictores del modelo, y en segundo lugar para calibrar la bondad de ajuste del modelo en su conjunto.

A. Significación de los predictores

Este punto trata de medir el nivel de significación estadístico que presentan las variables explicativas dentro de nuestro modelo. Con frecuencia, al calibrar dicho modelo suele ser habitual ir incorporando variables explicativas para mejorar su ajuste. Sin embargo, llegados a un punto, una variable más puede no tener ningún efecto o incluso provocar un empeoramiento del mismo. Por tanto, el objetivo será lograr el equilibrio entre un buen ajuste sin perder precisión del modelo, lo que se conoce como equilibrio entre el sesgo y varianza (Hastie et al., 2008).

Adaptado a nuestra herramienta principal de análisis SAS, es fácil obtener diferentes magnitudes que permitan analizar la significación estadística de las variables. Entre las medidas estadísticas que se van a utilizar en nuestro modelo se distinguen las dos siguientes: el error estándar asociado al proceso aleatorio, y el contraste de hipótesis, a través del p-valor asociado y el intervalo de confianza.

- El error estándar -

Todo método predictivo, consistente en la estimación de los correspondientes coeficientes de las variables, es un proceso aleatorio que lleva asociado un error estándar conocido como la desviación estándar. Una pequeña desviación estándar indica que el coeficiente estimado se aproxima a su verdadero valor. Por el contrario, una amplia desviación estándar confirma que el coeficiente se aleja del verdadero valor.

El error estándar está directamente relacionado con el parámetro de dispersión ϕ , cuanto más grande sea su valor mayor será el error. El ϕ mide la variabilidad del modelo, un valor alto implica más varianza en la aleatoriedad de los resultados y mayor error, lo que va en perjuicio del modelo.

- Contraste de hipótesis -

El contraste de hipótesis permite realizar inferencias acerca del verdadero valor de un parámetro. Para realizar este contraste o test de hipótesis es necesario establecer la hipótesis nula y la hipótesis alternativa. En nuestro caso, la hipótesis nula sería que el estimador de regresión sea cero, $H_0: \beta_j = 0$ lo que significa que la variable estimada no tiene efecto para nuestra variable respuesta y la hipótesis alternativa cuando el $H_1: \beta_j \neq 0$ (Cachero, 1993).

Mediante nuestra herramienta SAS, el contraste de hipótesis utiliza el estadístico de contraste de la Chi-cuadrado con un nivel de significación del 95%. Con vistas a analizar la significación estadística de los predictores se utilizan dos magnitudes:

- **P-valor:** el estadístico Chi-cuadrado lleva asociado una probabilidad que se denomina p-valor. Esta probabilidad se define como el mínimo nivel de significación al que puede ser rechazada la hipótesis nula, cuanto mayor sea su valor, mayor confianza se depositará en la hipótesis nula. El nivel de significación especificado por defecto, y que se adoptará en el estudio, es de 0,05. Si el p-valor resultante es lo suficientemente pequeño, por debajo de este nivel se rechaza la hipótesis nula, y en consecuencia se acepta la hipótesis alternativa de que el estimador sea distinto de cero.
- **El intervalo de confianza:** está asociado al valor del parámetro estimado. Se define como todo el rango de valores para nuestro estimador, donde la hipótesis

no sería rechazada para un nivel de significación establecido del 95%. Nuestro parámetro estimado estaría situado en el centro del intervalo de confianza. Es conveniente comprobar para garantizar una mayor significación, que el valor cero no está incluido dentro del intervalo de confianza.

B. Bondad de ajuste del modelo

Las medidas de bondad de ajuste permiten evaluar la calidad y eficacia del modelo en su conjunto, además de poder comparar la capacidad explicativa entre los diferentes modelos alternativos. El objetivo es conseguir un equilibrio entre la mayor explicación y la precisión del modelo de acuerdo a la información disponible.

En la práctica, la herramienta SAS nos ofrece una serie de medidas estadísticas destinadas a tal fin. Entre las más importantes, se enumeran los tres criterios principales que serán de aplicación para esta investigación, la desviación (conocida en inglés como *Deviance*), el criterio de información de Akaike (denominado por las siglas *AIC*), y el criterio de información bayesiano (con sus siglas *BIC*). Todas estas magnitudes se valoran de igual forma, de modo que el modelo con mejor ajuste será el que presente el menor valor de estas medidas.

- Desviación y reducción de la desviación -

La desviación se define como la diferencia entre el modelo analizado y el modelo saturado. El modelo saturado representa el mejor ajuste posible con los datos disponibles y se obtiene cuando existen tantos parámetros como observaciones. A pesar de que en la práctica este modelo no es de mayor interés, sí es muy utilizado como medida de referencia para medir la bondad de ajuste.

Para nuestro estudio particular, también se va a utilizar la llamada desviación escalada D que se define como el doble de la diferencia entre las funciones de verosimilitud del modelo saturado $l(y)$ y la del modelo analizado $l(\hat{\mu})$ (De Jong & Heller, 2008; Ohlsson & Johansson, 2010).

$$D = D(y, \hat{\mu}) = 2[l(y) - l(\hat{\mu})] \quad (3.5)$$

El modelo proporciona un buen ajuste cuando la D se hace mínima y $l(\hat{\mu})$ se acerca sin sobrepasar a $l(y)$. El primer término $l(y)$ es constante, puesto que está en función de los datos y de la distribución seleccionada. La desviación para el modelo saturado

es cero, mientras que la D para el modelo nulo se corresponde con la desviación total derivada de los datos. Por tanto, la D del modelo analizado estará comprendida dentro del intervalo entre estos dos extremos (Goldburd et al., 2016).

Por otra parte, la desviación no escalada puede fácilmente calcularse a partir de la desviación escalada. Bastaría con multiplicar esta D por el factor de escala ϕ , tal que $D^* = \phi D$. Además, D^* tiene la ventaja de ser independiente de dicho parámetro ϕ .

La siguiente medida de bondad de ajuste del modelo es **la reducción de la desviación** D^2 , que consiste en la cantidad que va disminuyendo la desviación de nuestro modelo respecto a la del modelo nulo. Se calcula como la variación de la desviación residual del modelo analizado respecto a la desviación nula, según la siguiente fórmula:

$$D^2 = \frac{D_{nulo} - D_{residual}}{D_{nulo}} \times 100 \quad (3.6)$$

- Criterio de información de Akaike y el criterio de información bayesiano -

Las siguientes magnitudes que se van a estudiar incorporan también información sobre la complejidad del modelo, ya que consideran el número de parámetros p . A medida que incrementamos el número de predictores se incrementa también el número de parámetros p , esto repercute en un mejor ajuste aunque también le concede mayor variabilidad a las estimaciones del modelo. Entre los indicadores que ayudan a equilibrar la bondad del ajuste con la menor pérdida de precisión, se presentan el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC) (De Jong & Heller, 2008). Ambos sirven como herramienta de decisión para comparar entre diferentes modelos y seleccionar el que presente mejores resultados.

- **El criterio de información de Akaike:** es una medida que establece el equilibrio entre la bondad del ajuste y la simplicidad derivada del número de parámetros del modelo.

$$AIC = -2l + 2p \quad (3.7)$$

Siendo l el logaritmo de la función de verosimilitud y p el número de parámetros. El primer término se refiere a la bondad del ajuste de los datos, el máximo valor de la función de verosimilitud del modelo estimado que supone el mínimo valor de $-2l$, mientras que el segundo término representa la varianza, que penaliza cuanto mayor número de parámetros tenga el modelo. Así, cuanto menor sea el valor de AIC mayor calidad tendrá el modelo estimado.

- **El criterio de información bayesiano:** es muy similar al criterio anterior, con la diferencia de que este penaliza aún más cuanto mayor sea el número de parámetros que tenga el modelo estimado. Y se define tal que:

$$BIC = -2l + p \cdot \log(n) \quad (3.8)$$

Siendo n el número de observaciones. Su interpretación tiene el mismo sentido que el criterio AIC, cuanto menor valor de BIC mejor será el modelo estimado.

Todas las medidas propuestas para evaluar nuestro modelo representan criterios estadísticos objetivos, que evalúan su precisión y calidad. Sin embargo, es conveniente complementar el estudio utilizando también información cualitativa, que proporciona la propia experiencia y el conocimiento adquirido del negocio para dotar nuestro análisis con una completa valoración.

3.3 Modelos lineales generalizados

Los modelos lineales generalizados (GLMs) se enmarcan dentro de las técnicas estadísticas de análisis predictivo basados en los modelos de regresión múltiple. Estos tratan de evaluar y cuantificar la relación existente entre una variable respuesta (aquella que se quiere predecir) y las variables explicativas o predictores. Esta relación, que se establece a través de una función de distribución, proporciona las estimaciones de la variable respuesta y de los parámetros de la distribución a través de la optimización de las medidas de bondad de ajuste.

Los GLMs suponen una generalización del modelo de regresión lineal ordinario en dos aspectos fundamentales. Por un lado, y tal y como Nelder y Wedderburn ya demostraron (Brisard 2014; Ohlsson & Johansson, 2010), la distribución de probabilidad de la variable respuesta puede ser cualquier función de la familia

exponencial (ya sea para el caso discreto, el caso continuo y también de tipo mixtas, como es el caso de la distribución Tweedie), sin necesidad de ceñirse a la distribución normal. Y por otro lado, se establece una relación lineal entre una transformación de la media de la variable respuesta y las variables explicativas (De Jong & Heller, 2008; Denuit et al., 2007; Ohlsson & Johansson, 2010).

La aplicación de estos modelos multiplicativos para la tarificación de los seguros no vida es relativamente reciente, y comenzó a extenderse a partir de mediados de los años 90, debido en cierta forma a la desregulación del mercado asegurador en muchos países. Sin duda, estos modelos presentan una serie de ventajas respecto a los métodos tradicionales de tarificación. En primer lugar, se fundamentan en una teoría estadística general bien definida, que permite realizar estimaciones y obtener medidas estadísticas para evaluar la selección de modelos. En segundo lugar, acapara una amplia gama de áreas de aplicación, lo que respalda aún más su utilización. Y por último, existen ya unas cuantas aplicaciones informáticas que cuentan con el potencial suficiente para realizar esta técnica de análisis, como son los paquetes estadísticos SAS y R, incluso también existen otras herramientas comerciales especializadas en la tarificación, como es el caso de *Emblem* (Ohlsson & Johansson, 2010).

Una vez presentada la metodología general, corresponde ahora conocer las diferentes etapas en que consiste la elaboración de un modelo GLM. Más adelante, se profundizará también en la descripción matemática del modelo lineal general basado en la distribución Tweedie, que será la metodología que se va a utilizar en la presente investigación.

3.3.1. Guía de elaboración de GLMs

La construcción de un modelo GLM comprende el recorrido formado por una serie de etapas, con sus correspondientes objetivos que deben ir cumpliéndose para garantizar la obtención de un modelo óptimo. Estos modelos abarcan diferentes aplicaciones, que van desde actualizar los coeficientes multiplicativos o relatividades de una tarifa ya existente, y en tal caso no será necesario pasar ya por todas las etapas del proceso. O también puede tratarse del diseño de un producto nuevo, en cuyo caso sí que se debe realizar el proceso completo para elaborar la tarifa del producto.

Varios autores, (De Jong & Heller 2008; Goldburd et al., 2016), coinciden en que el proceso de elaboración de un modelo lineal generalizado debe pasar por las siguientes fases:

1. Obtención, tratamiento de datos y análisis exploratorio

El primer paso en cualquier análisis consiste en la obtención de información y su posterior tratamiento. La fase de recogida de información suele consumir bastante tiempo y en numerosas ocasiones es difícil de predecir con exactitud, esto se acrecienta aún más al tratarse del estudio de microseguros. Una vez que se dispone de los datos necesarios para el análisis, se procede a su consecuente tratamiento. Este acaba siendo un proceso iterativo que pasa, por limpiar los datos de posibles errores y por completar la información que esté incompleta con los datos que estén a nuestro alcance, todo con el fin de obtener una base de datos lo más completa posible. Es en este punto donde se plantea una limitación del modelo en cuanto a la plena credibilidad que se le otorga a la estimación, según la base de datos utilizada, sin considerar o no si los datos están perfectamente depurados (Goldburd et al., 2016). Posteriormente, cuando ya se dispone de la base de datos perfectamente depurada, se realiza un primer análisis exploratorio para conocer en detalle cómo es la información que vamos a manejar, para poder así después determinar cuáles van a ser las variables del modelo.

2. Establecer las hipótesis del modelo

Las hipótesis sientan las bases para la realización de cualquier modelo GLM, que comprende toda la parte de selección de las variables principales, la distribución de probabilidad y la función de enlace. En primer lugar, se definen las diferentes variables, la variable objetivo, que consiste en lo que se quiere predecir, esta va a condicionar cuál será la variable respuesta de acuerdo a la Tabla 3.1. También, se seleccionan los factores de riesgo según los datos disponibles que representarán las variables explicativas. En muchos casos, será necesario transformar y analizar dichas variables para su incorporación al modelo. En segundo lugar, se debe elegir la distribución de probabilidad de la familia exponencial que va a permitir modelizar el comportamiento de la variable respuesta. Y por último, la elección de la función de enlace, fácilmente deducible de la Tabla 3.3, que establece la forma funcional que relaciona la variable respuesta y las variables explicativas.

3. Ajuste del modelo lineal generalizado

El paso siguiente, una vez fijadas las hipótesis, consiste en calibrar el modelo predictivo obteniendo los resultados de los estimadores β , los parámetros y las medidas de bondad de ajuste del modelo. Con toda esta información se analizarán las medidas de ajuste del modelo y la significación estadística de los estimadores de cada predictor, para poder entonces disponer de diferentes alternativas de modelos. Todo ello se convierte en un proceso iterativo, en el que las variables explicativas juegan un papel principal ya que se irán transformando e incluso eliminando, en caso de que no resulten significativas, hasta lograr un buen ajuste.

4. Selección y validación del modelo

La selección del modelo óptimo se lleva a cabo comparando los diferentes modelos alternativos que se obtienen del análisis, por medio de las medidas estadísticas de bondad de ajuste. Además, es importante que la decisión tenga en cuenta no solo los criterios objetivos derivados de sus resultados, sino también un enfoque práctico que valore la mejor predicción a futuro de la variable objetivo. Por otra parte, también se debe realizar la correspondiente validación del modelo. Esta contempla varios aspectos que van, desde la verificación del modelo con diferentes muestras de datos para garantizar la eficacia de los resultados, hasta el análisis del equilibrio financiero del modelo para asegurar los mejores resultados en la práctica.

5. Traslado de los resultados

El resultado final del modelo va a dar lugar al plan de tarifas del producto analizado. Si el modelo se ha realizado correctamente, las primas resultantes deben cumplir los principios básicos de equidad y suficiencia de primas. Por este motivo, es interesante analizar aquí también el equilibrio financiero del modelo y compararlo con el que actualmente se aplica. El plan de tarifas calculado debe ser claro, preciso y adecuado a la clasificación de riesgos propuesta. En este punto, suele ser habitual plantearse la cuestión sobre la existencia de variables latentes que anteriormente no habían sido contempladas y que sí podrían afectar a la tarificación. En tal caso, es apropiado prestar atención tanto a las correlaciones entre las diferentes variables, como a valorar nuevos factores con la posibilidad de incluirlos dentro del modelo.

6. Mantenimiento y actualización del modelo

La capacidad predictiva del modelo suele disminuir con el paso del tiempo debido a que los datos utilizados pueden quedar obsoletos y no reflejar la situación real. Por este motivo, es importante mantener los datos bien actualizados para ir reconstruyendo de nuevo el modelo que maximice su capacidad predictiva. Este paso es crucial para realizar una tarificación más flexible, dinámica y que contenga la experiencia más reciente.

En general, todas las fases que se han descrito siguen un orden lógico y secuencial que en la práctica hacen que se convierta en un proceso iterativo idóneo para realizar un buen análisis predictivo capaz de proporcionar el modelo que mejor se ajuste a la distribución de riesgo, el modelo óptimo.

3.3.2. Modelo Tweedie

Tal y como se ha venido apuntando anteriormente, los modelos lineales generalizados presentan la gran ventaja, respecto al modelo lineal clásico, de poder emplear toda una gama de distribuciones de probabilidad pertenecientes a la familia exponencial para modelizar la variable respuesta.

Entre las diferentes distribuciones, vamos a prestar especial atención a la distribución Poisson compuesta, que es la suma de una serie de variables aleatorias independientes e idénticamente distribuidas, en las que el número se distribuye según una función primaria Poisson. Esta clase de funciones se caracterizan por ser de tipo mixto y son posibles de combinar con otras distribuciones secundarias para modelizar la variabilidad de la variable respuesta. En especial, la distribución Poisson compuesta con una Gamma, se trata de una distribución formada por una función primaria Poisson, discreta con una probabilidad positiva en el punto masa cero, y otra secundaria Gamma, que es de tipo continuo para el resto de números reales positivos.

Esta distribución Poisson compuesta con Gamma se enmarca dentro de los conocidos modelos Tweedie que presentan la siguiente forma:

$$f_s(s) = \begin{cases} e^{-\lambda}, & s = 0 \\ e^{-\beta s} e^{-\lambda} \sum_{n=1}^{\infty} \frac{(\lambda \beta^\alpha)^n}{n! \Gamma(n\alpha)} s^{n\alpha-1}, & s > 0 \end{cases} \quad (3.9)$$

Profundizando en las dos distribuciones de origen, el modelo de Poisson tiene un único parámetro λ que representa la media y la varianza. Y la distribución Gamma tiene a su vez dos parámetros, el de forma α y el de escala β .

Con la idea de demostrar que la distribución Tweedie pertenece a la familia exponencial de distribuciones, su función de densidad o de cuantía debe expresarse como:

$$f_{\theta}(x) = \exp\left\{\frac{x\theta - b\theta}{a(\phi)} + c(x, \phi)\right\} \quad (3.10)$$

Siendo a , b y c funciones arbitrarias, θ es el parámetro canónico de la distribución y ϕ el parámetro de escala.

Siguiendo con el modelo Tweedie, este se caracteriza por tener tres parámetros μ , ϕ y p que tienen una relación directa con las distribuciones subyacentes Poisson y Gamma:

$$\mu = \lambda \frac{\alpha}{\beta}, \quad p = \frac{\alpha + 2}{\alpha + 1}, \quad \phi = \frac{\lambda^{1-p}}{2-p} \left(\frac{\alpha}{\beta}\right)^{2-p} \quad (3.11)$$

Estos se pueden reparametrizar considerando ahora (λ, α, β) :

$$\lambda = \frac{\mu^{2-p}}{\phi(2-p)}, \quad \alpha = \frac{2-p}{p-1}, \quad \frac{1}{\beta} = \phi(p-1)\mu^{p-1} \quad (3.12)$$

Ahora bien, se demuestra que la suma depende de s , ϕ pero no de μ . Por tanto, se puede definir a modo de función $c(s, \phi)$ y $s > 0$:

$$c(s, \phi) = \log \left[\sum_{n=1}^{\infty} \frac{(\lambda\beta^{\alpha})^n}{n! \Gamma(n\alpha)} s^{n\alpha-1} \right], \quad c(0, \phi) = 1 \quad (3.13)$$

$$f_s(s) = \exp \left\{ -\beta s - \lambda + \log \left[\sum_{n=1}^{\infty} \frac{(\lambda\beta^{\alpha})^n}{n! \Gamma(n\alpha)} s^{n\alpha-1} \right] \right\} =$$

$$= \exp\{-\beta s - \lambda + c(s, \phi)\} = \exp\left\{\frac{-\beta\phi s - \lambda\phi}{\phi} + c(s, \phi)\right\}$$

Para conseguir igualdad con (3.10) y partiendo de $\theta = -\beta\phi$:

$$\theta(\mu) = \frac{-1}{(p-1)\mu^{p-1}} \quad (3.14)$$

$$\mu(\theta) = [-(p-1)\theta]^{-\frac{1}{p-1}} \quad (3.15)$$

La función acumulada sería,

$$b(\theta) = \lambda\phi = \frac{\mu^{2-p}}{2-p} = \frac{1}{2-p} [-(p-1)\theta]^{-\frac{p-2}{p-1}} \quad (3.16)$$

Y por otra parte, la función de varianza se calcularía como la segunda derivada:

$$V(\mu) = b''([b']^{-1}(\mu)) = b''(\theta(\mu)) \quad (3.17)$$

La derivada $b'(\theta)$ para el modelo de Poisson ($p = 1$) es $b'(\theta) = e^\theta$, para el modelo Gamma ($p = 2$) se sabe que $b'(\theta) = -\log(\theta)$ y para el caso de estudio:

$$b'(\theta) = [-(p-1)\theta]^{-\frac{1}{p-1}} \quad p > 1 \quad (3.18)$$

$$b''(\theta) = [-(p-1)\theta]^{-\frac{p}{p-1}} \quad p > 1 \quad (3.19)$$

Sustituyendo ahora $\theta(\mu)$ se obtiene así su correspondiente función de varianza:

$$V(\mu) = \mu^p \quad (3.20)$$

Se confirma que la distribución Tweedie pertenece a la familia exponencial, y como tal se caracteriza únicamente por su función de varianza sin necesidad de especificar la distribución completa.

Adicionalmente, la varianza tiene cierta relevancia respecto a las distribuciones de escala invariable. Se dice que una variable aleatoria Y , que pertenece a una determinada familia de distribuciones, adquiere esta propiedad si dada una constante $c > 0$, la familia de la distribución cY sigue perteneciendo a la misma familia de Y (Ohlsson & Johansson, 2010). Por supuesto, este aspecto es deseable en la tarificación de seguros no vida en el sentido de que el resultado debería ser independiente a las unidades de medida. Por este motivo, es recomendable trabajar con familias de distribuciones cerradas a transformaciones de escala o también llamadas de escala invariante. Dentro de la familia exponencial de distribuciones, el único modelo que cumple esta propiedad de escala invariable es el modelo Tweedie (Goldburd et al., 2016; Ohlsson & Johansson, 2010).

El modelo Tweedie contempla una serie de distribuciones de la familia exponencial, ya sean del tipo discreto, continuo o mixto. Esta versatilidad se debe a que, aparte de los parámetros estándar utilizados en la familia exponencial de distribuciones, la distribución Tweedie incorpora un tercer parámetro p llamado el parámetro de potencia. De este modo, y según el valor que tome, se pueden derivar estas otras distribuciones de la familia exponencial. A continuación se presenta un resumen de los diferentes modelos Tweedie clasificados según el valor del parámetro potencia p y la variable objetivo que se pretenda calcular.

Tabla 3.2. Resumen de los modelos Tweedie.

p	Distribución	Tipo de variable	Variable objetivo
$p=0$	Normal	Continua	–
$0 < p < 1$	–	Inexistente	–
$p = 1$	Poisson	Discreta	Frecuencia siniestro
$1 < p < 2$	Poisson compuesta	Mixta, no negativa	Prima pura
$p = 2$	Gamma	Continua, positiva	Cuantía siniestro
$p = 3$	Normal inversa	Continua, positiva	Cuantía siniestro

Fuente: elaboración propia a partir de Ohlsson E., & Johansson, B. (2010). *Non-life insurance pricing with generalized linear models*, vol. 21. Springer. Berlin.

En la Tabla 3.2 se resalta en color gris el caso particular de la distribución Poisson compuesta con Gamma cuando $1 < p < 2$, y que representa todos los casos intermedios entre las distribuciones Poisson y Gamma. Este caso es el que resulta de

mayor interés ya que permite modelizar la prima pura directamente sin necesidad de hacer un cálculo separado para la frecuencia y el coste medio. Tal y como demostraron Jørgensen y De Souza (1994), cuando los siniestros ocurren siguiendo un proceso de Poisson y cada evento genera un coste aleatorio que sigue una distribución Gamma, se demuestra que la cuantía total de todos los siniestros sigue entonces una distribución Tweedie.

Con todo lo dicho, queda claro entonces que la distribución Tweedie pertenece a la familia exponencial de distribuciones que son aquellas que permiten la aplicación de los modelos lineales generalizados tan importantes en la tarificación de seguros (Ohlsson & Johansson, 2010). Esta modelización predictiva consiste en evaluar y cuantificar la relación existente entre una variable respuesta y_i y las variables explicativas. Y se expresa a través de la siguiente distribución de probabilidad exponencial, que sigue una función de densidad para el caso continuo y una función de cuantía para el caso discreto (Ohlsson & Johansson, 2010):

$$f_{Y_i}(y_i; \theta_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{\phi/w_i} + c(y_i, \phi, w_i) \right\} \quad (3.21)$$

En sentido amplio, en todos los modelos GLM se distinguen tres componentes principales, el aleatorio de la variable respuesta, el sistemático por el predictor lineal y la función de enlace que establece la relación entre las dos anteriores.

En primer lugar, **la componente aleatoria**, que se corresponde con la variable respuesta y_i que sigue una distribución de la familia exponencial. θ_i es el parámetro de escala, que varía según las características de las i observaciones, el parámetro de dispersión $\phi > 0$ que se mantiene constante y w_i como el peso asociado a cada registro. Se asume la independencia entre las i observaciones del modelo sobre la variable respuesta.

La elección de la llamada función acumulada $b(\theta_i)$ pertenecerá a la familia exponencial de distribuciones de probabilidad. Esta vendrá determinada por los parámetros θ_i y ϕ y estará relacionada con la esperanza de la variable y_i . Dentro de la gama exponencial de distribuciones se encuentran la distribución Normal, Binomial, Poisson y Gamma entre otras. Y por tanto f_{Y_i} es la función de densidad en el caso continuo y la función de probabilidad en el caso discreto.

Una característica propia de las distribuciones de la familia exponencial es que permiten especificar un modelo mediante la relación entre su media y varianza, siendo la función de varianza:

$$V(\mu_i) = \frac{b''(\theta_i)}{w_i} \quad (3.22)$$

En segundo lugar, **la componente sistemática** es el predictor lineal η_i , que especifica las variables explicativas x_{ij} . El objetivo del análisis de la tarificación consiste en investigar cómo la variable respuesta y_i se ve influida por las r variables explicativas x_1, x_2, \dots, x_r , siendo r el nivel base. Esta combinación lineal se denomina predictor lineal:

$$g(\mu_i) = \beta_0 + \beta_1 x_1 + \dots + \beta_{r-1} x_{r-1} \quad (3.23)$$

Este último componente $g(\mu_i)$ es la **función de enlace** o también llamada función *link*, que establece la relación entre los dos componentes aleatorio y sistemático. El objetivo de esta función es relacionar linealmente la variable respuesta y las variables explicativas x_{ij} mediante la transformación de la variable respuesta. A diferencia del modelo lineal clásico, donde $\mu_i \equiv \eta_i$, en un modelo GLM la transformación de la media permite elegir diferentes funciones de enlace y se expresa de la siguiente forma:

$$g(\mu_i) = \eta_i = \sum_{j=1}^r x_{ij} \beta_j, \quad i = 1, 2, \dots, n \quad (3.24)$$

En un modelo GLM esta función de enlace $g(\mu_i)$ forma parte de la especificación del modelo, y su elección depende que sea una función diferenciable y estrictamente monótona. Cada distribución de la familia exponencial lleva asociada su correspondiente función de enlace, tal y como se puede observar en la siguiente Tabla.

Tabla 3.3. Funciones de enlace de las distribuciones de la familia exponencial.

Distribución	Función de enlace	$\eta_i = g(\mu_i)$	$\mu_i = g^{-1}(\eta_i)$
Normal	Identidad	μ_i	η_i
Poisson/Tweedie	Logarítmica	$\ln(\mu_i)$	e^{η_i}
Binomial	Logit	$\ln(\mu_i/(1-\mu_i))$	$e^{\eta_i}/(1+e^{\eta_i})$
Gamma	Recíproca	$1/\mu_i$	$1/\eta_i$

Fuente: elaboración propia a partir de De Jong, P., & Heller, G. Z. (2008). *Generalized linear models for insurance data*, vol. 10. Cambridge: Cambridge University Press.

En el campo de tarificación de seguros no vida, los modelos multiplicativos presentan una función de enlace que suele ser de tipo logarítmico. De hecho, para el presente estudio de investigación basado en el modelo Tweedie, y tal como se muestra en la Tabla 3.3, se utilizará la función de enlace logarítmica.

CAPÍTULO 4. TRATAMIENTO DE DATOS

El presente capítulo pretende ofrecer una visión general de las características de nuestro microseguro a partir de la información disponible. En este primer estudio trataremos de conseguir una base de datos lo más completa y fiable posible, que permita después realizar todo el análisis descriptivo, previo a la aplicación de la metodología de tarificación. Este planteamiento refuerza la importancia que se debe conceder a todo el ciclo de tratamiento y conservación de datos dentro de las compañías.

Para el desempeño de este trabajo vamos a seguir el siguiente recorrido. Primero, se empieza por describir de forma general el mercado objetivo característico de estos seguros inclusivos, hasta llegar a nuestro caso particular, de donde se obtendrá la valiosa información para el presente estudio. En segundo lugar, se explica el proceso llevado a cabo para el tratamiento y la depuración de los datos, hasta llegar a disponer de una buena fuente de información. Tercero, se evaluarán los factores como variables integrantes de nuestro futuro modelo predictivo y posteriormente, se realizará el análisis descriptivo de dichas magnitudes. En cuarto lugar, comenzaremos a segmentar las variables explicativas para adaptarlas al futuro modelo predictivo. Y por último, se realizará un estudio de dependencia e interacción de las variables explicativas, con el objeto de confirmar su idoneidad como factores de riesgo integrantes de la tarifa de nuestro microseguro.

4.1 Origen y descripción de la información

El punto inicial de partida comienza con una búsqueda exhaustiva de microseguros con cierta experiencia en el mercado y localizado en países emergentes. Un aspecto a resaltar ha sido la dificultad añadida en acceder a este particular mercado, ya que por un lado, se trata de un mercado nuevo en el que no todas las compañías trabajan, y por otro, debido a la distancia, al estar localizado en países emergentes. La región de estudio elegida se centra en la zona del Sudeste Asiático, que cuenta aproximadamente con un 70% de población mundial con bajos recursos y representa actualmente el mercado más grande de microseguros, y en particular Filipinas ha sido el país elegido para el presente estudio.

Es interesante destacar el progreso de Filipinas en esta materia por haber desarrollado considerables avances en los últimos años en materia de extensión, oferta y en la implantación de medidas regulatorias. De hecho, desde el año 2006 Filipinas cuenta con un marco regulatorio específico de microseguros (GIZ & RFPI, 2015). Esta regulación no solo suaviza los requerimientos a los proveedores, sino que además favorece la creación de nuevas clases de agentes que facilitan el acceso a los consumidores de bajos ingresos (Zimmerman & Magnoni, 2014). Los esfuerzos por parte de la Comisión de Seguros por impulsar medidas dirigidas a promover la participación de las compañías han favorecido el desarrollo de este mercado y su formalización, reduciendo los productos que eran ofrecidos de forma informal. El crecimiento en este mercado es indiscutible, partiendo del 2009 con 2,9 millones de asegurados a 25,2 millones en el año 2014 (GIZ & RFPI, 2015).

La información utilizada para esta investigación ha sido facilitada por una compañía de seguros con experiencia en este tipo de mercado en Filipinas. El producto objeto de análisis se trata de un microseguro combinado, es decir que cuenta con diferentes coberturas tanto del ramo de vida como de no vida, y cuyas sumas aseguradas se expresan en la moneda oficial de Filipinas, el peso filipino (su abreviatura Php¹⁷), moneda que será de referencia en todo el trabajo de investigación.

¹⁷ La conversión del peso filipino al euro supone 1€=0,02Php. Cambio obtenido el día 20 Marzo del 2017 a través de: <https://themoneyconverter.com/ES/PHP/EUR.aspx>.

En concreto, el producto engloba las tres siguientes coberturas:

- *Indemnización diaria por hospitalización.* Consiste en una paga de 500Php por cada día que el asegurado esté hospitalizado, a causa de lesiones sufridas por accidente y con un límite no superior a 20 días. La indemnización total no podrá superar los 10.000Php.
- *Indemnización por eventos de la naturaleza.* El asegurado recibirá hasta un máximo de 5.000Php de indemnización en caso de siniestro o por daños en la vivienda habitual (contenido y continente), siempre que sea por causas relacionadas con eventos de la naturaleza, como son los tifones, terremotos, deslizamientos de tierra, inundaciones e incendios.
- *Gastos funerarios.* El asegurador proporciona asistencia por gastos funerarios en caso de fallecimiento del asegurado, ya sea por causas naturales o por enfermedad, y con un importe máximo de 20.000Php.

La información del microseguro objeto de estudio se corresponde con uno de los colectivos con mayor experiencia y mayor número de asegurados de la compañía. Se dispone de varios años de experiencia del colectivo que comprenden el período que va desde su inicio, en el año 2012, hasta el momento más reciente, junio del 2016.

4.2 Tratamiento y depuración de datos

La información de origen viene separada en dos bases de datos diferentes. Por un lado, los datos de producción correspondientes a características de los asegurados, y por otro, los registros de siniestralidad del colectivo recogidos por la compañía. Ambas fuentes recogen la experiencia para cada uno de los años comprendidos en el período entre el 2012 y el 2016. El proceso de tratamiento y depuración tiene como objetivo obtener una base de datos lo más completa y fiable posible. Entre las limitaciones encontradas durante esta etapa, la más reseñable ha sido la dificultad de relacionar la base de datos de producción con la de siniestralidad, debido a que no existía ningún campo clave de identificación. Algunos autores, como Goldburd et al. (2016), atribuyen esta dificultad a que la mayoría de las compañías no almacenan la información en el mismo lugar. Así, la base de datos de primas a nivel de póliza suele gestionarse a través del área de suscripción, mientras que la correspondiente a siniestros corresponde al área de prestaciones. El problema surge cuando estas áreas se han ido desarrollado de forma separada e independiente, de forma que a la hora de

relacionar ambas bases de datos se convierte en una tarea bastante compleja. Por tanto, la solución práctica en nuestro caso ha consistido en asociar a cada asegurado un campo clave de identificación que permita la trazabilidad entre ambas muestras. De esta manera, cada registro a nivel de póliza y asegurado lleva asociado su correspondiente experiencia de siniestralidad.

La fase previa comienza con una observación global de la información disponible de origen. En ella, se han establecido una serie de filtros iniciales hasta llegar a obtener la base de datos final con la que trabajaremos. Entre las consideraciones asumidas se describen las siguientes:

- *Período de análisis.* Comprende desde el año 2012 (primer año de comercialización del producto) hasta junio del 2016 (año incompleto). Analizando la información correspondiente a cada uno de los años disponibles (véase Anexo A.2), se ha decidido utilizar para el análisis únicamente el año 2014 por varios motivos. En primer lugar, el año 2014 es el período donde mayor experiencia en siniestralidad se registra para el microseguro analizado. En segundo lugar, el año 2014 es el período con la información más completa respecto al resto de los años. Aunque el resto de años se han descartado para realizar el análisis, sí que han sido de gran utilidad para realizar varias tareas de tratamiento y depuración de datos. Entre ellas, nos han servido como fuente de información para completar el año 2014, rellenando los registros ausentes de asegurados que contaban con más de un año de antigüedad. Además, gracias a este intervalo temporal disponible, se ha creado una nueva variable explicativa, la antigüedad de la póliza, en función de la permanencia de los asegurados durante dicho período. Por tanto, los datos que se utilizan se corresponden a todos los asegurados registrados en el año 2014 con independencia del tiempo de exposición durante ese año.
- *Selección de coberturas del producto.* De inicio, la base de datos recoge toda la experiencia en siniestralidad de todas las coberturas del microseguro combinado (véase Anexo A.2). Debido a que la cobertura de gastos funerarios apenas ofrece actividad durante el período considerado, se ha decidido excluir esta cobertura del análisis. Por tanto, el estudio se va a centrar exclusivamente en las garantías de indemnización diaria por hospitalización e indemnización por eventos de la naturaleza.

- *Alcance de la siniestralidad.* Se recogen todos los siniestros cuya ocurrencia sea el ejercicio 2014, incluyendo también los declarados con retraso o IBNR, es decir todos los siniestros con independencia de su fecha de comunicación. Por otro lado, en base al objetivo de análisis que aquí se plantea, la estimación de la prima pura, solo se van a considerar los siniestros con un coste último mayor que cero, de esta forma evitamos los posibles errores en la comunicación. Esta tarea de eliminación de dichos registros se ha realizado fácilmente porque aparecen identificados en el campo *estado del siniestro* como *NO LOSS*. Además, una ventaja adicional es que todos nuestros siniestros están liquidados, lo que implica que su coste último coincide con el importe total pagado.

Después, entramos en la fase de tratamiento y depuración de datos que ha consistido en revisar minuciosamente la información hasta ir completando la base de datos final. Muchos de los procesos seguidos en esta etapa se han tenido que realizar manualmente para garantizar la credibilidad y trazabilidad de los datos. En concreto, los principales pasos seguidos en esta etapa se explican a continuación:

- *Creación de un campo clave de identificación.* Inicialmente, las dos bases de datos disponibles, correspondientes a datos de asegurados y siniestralidad, estaban inconexas. Por este motivo, ha sido necesario crear un campo de identificación denominado *ID*, partiendo de la base de datos de producción. Este campo clave está formado por la unión del número de póliza, el número de suplemento y un número ordinal que se asigna a cada asegurado. Con este nuevo campo se logra asignar una referencia de identificación a todos los asegurados en la información de producción. El paso siguiente ha sido identificar, igualmente en la base de siniestralidad, cada registro con su correspondiente asegurado y clave de identificación. Para ello, el único campo común era el nombre y apellidos del asegurado, lo que ha permitido asociar a cada siniestro con su campo clave de identificación. Este proceso de asignación del campo clave se ha realizado manualmente para minimizar los posibles errores causados en el cruce de información. Este punto ha sido crucial para poder disponer de una única base de datos que garantizara la trazabilidad entre el perfil de los asegurados y su respectiva experiencia de siniestralidad.
- *Depuración y creación de variables.* En inicio se ha realizado un primer filtro con todas las variables explicativas potenciales que representaban el perfil de riesgo del colectivo asegurado. Entre estas se distinguen el estado civil, la provincia y el

nombre con los apellidos del asegurado. Para realizar dicha tarea ha sido necesario recurrir a la información disponible de otros años para completar y corregir buena parte de los datos.

Además de las variables ya mencionadas, se han creado otras variables como son el sexo, la antigüedad de póliza, la exposición, la región y la densidad de población, aplicando aproximaciones y recurriendo a otras fuentes, tanto internas como externas.

- *Depuración de datos.* Esta etapa ha tratado de verificar la existencia de registros únicos, comprobar la exactitud de los campos disponibles y la trazabilidad de los datos. Una de ellas, ha sido la eliminación de registros repetidos (dos, tres y hasta cuatro veces), utilizando como criterio de selección el asegurado con la fecha de efecto más antigua.

Todo este proceso de depuración y tratamiento de datos es de gran importancia para garantizar la calidad y fiabilidad de la información, y en ello se ha consumido una gran parte del tiempo hasta lograr disponer de una base de datos fiable (véase Anexo A.1). En base a esta experiencia es importante destacar el mayor esfuerzo y trabajo que deben dedicar las compañías para mejorar la recogida, el tratamiento y el mantenimiento de las bases de datos. Esto se traducirá después en un menor consumo de tiempo en tareas de depuración de datos, repercutiendo a favor de una mayor dedicación de tiempo en las tareas realmente importantes, como son el análisis de datos y las posteriores estimaciones.

4.3 Factores

Los factores hacen referencia a todas aquellas variables que van a configurar el modelo lineal generalizado y que engloban, la variable objetivo, la variable respuesta y las variables explicativas.

4.3.1. Variable objetivo

Representa el objetivo de estudio en nuestro modelo, y que será resuelto gracias a la aplicación de la metodología predictiva de los modelos lineales generalizados. En nuestro caso, la finalidad consiste en estimar las primas puras de riesgo de un microseguro en concreto.

4.3.2. Variable respuesta

Una vez establecido el objetivo de estudio mediante la variable anterior, el paso siguiente es identificar cuál es la variable respuesta que mejor permitirá estimar la prima pura de riesgo. Siguiendo las indicaciones de la Tabla 3.1 del capítulo anterior, es fácil deducir que esta variable Y_i es el coste último total por expuesto¹⁸, que representa todos los pagos incurridos por la compañía para hacer frente al siniestro ponderado por el tiempo de exposición del asegurado. Esta definición se traduce en la siguiente fórmula:

$$Y_i = \frac{X_i}{w_i} \quad (4.1)$$

Siendo X_i el importe de los pagos realizados por la compañía y w_i es la exposición del asegurado en la póliza, que se define a continuación.

4.3.3. Variables explicativas

La información disponible del microseguro nos proporciona la experiencia de la siniestralidad, que conforma el perfil de riesgo asociado al colectivo asegurado. De esta forma, todas las características observables y registradas en la base de datos representan los factores de riesgo, que se traducen en nuestro modelo a través de las variables explicativas o predictores.

Para nuestro caso de estudio se cuenta con varios predictores que será preciso estudiar en detalle. Tras su observación le seguirá una correspondiente selección de variables explicativas que cuente con la suficiente capacidad para predecir el comportamiento de riesgo futuro de nuestro colectivo. A continuación, se pasa a describir cada una de ellas separadas en dos categorías, variables a priori y variables a posteriori.

Variables a priori

- *Identificación (ID)*, es el campo de identificación del asegurado en la póliza correspondiente. Se trata de un registro único para cada asegurado y año que

¹⁸ Conocido en la literatura actuarial inglesa con el término *burning cost* (Wrede et al., 2016).

permite la trazabilidad de los datos de producción y siniestros como campo clave de la base de datos.

- *Exposición (exposure)*, se trata de una variable de escala, que representa la cantidad de tiempo en años que el asegurado mantiene su póliza en vigor. Esta variable es muy importante en el modelo porque influye directamente sobre la variable respuesta, además se utilizará como una medida de peso. También es de utilidad para el cálculo de probabilidades y frecuencias.
- *Edad del asegurado (age)*, se corresponde con la edad actuarial. Esta se calcula tomando de referencia la fecha de cumpleaños más próxima respecto a la fecha de efecto de la póliza.
- *Densidad de población (density)*, es una variable exógena estimada a partir de una fuente externa¹⁹ que mide la distribución relativa de la población asegurada en Filipinas. Se estima como el número total de habitantes de cada población sobre el total de personas que habitan en Filipinas.
- *Estado civil (civil_status)*, representa la condición del asegurado según el registro civil sobre si tiene pareja legal o no. Se asignan dos posibles categorías, soltero o casado (*single* o *married*).
- *Sexo (sex)*, se trata de la condición orgánica del asegurado con dos posibles valores hombre y mujer (*male* y *female*).
- *Provincia (province)*, se identifica con el área de residencia habitual del asegurado. Engloba varios términos, desde la provincia, ciudad o distrito de residencia. Cuenta con 95 diferentes niveles.
- *Región (region)*, es la agrupación de diferentes provincias por su localización geográfica. La autoridad de Filipinas señala que existen 18 regiones que serán los niveles establecidos para esta variable.
- *Antigüedad de la póliza (age_policy)*, hace referencia al período en años de permanencia como asegurado en el microseguro analizado. Se crea esta variable por el cruce de los asegurados en los diferentes años comprendidos entre 2012 y 2014. De esta forma se establecen tres posibles niveles numéricos en función del tiempo de permanencia como asegurados del producto.

¹⁹ Datos de la población filipina registrados en Junio del 2016 (Philippine Statistics Authority, 2016).

Variables a posteriori.

Obtenidas a partir de la experiencia acumulada del producto. Aunque se dispone de la información separada por cobertura, las variables que se manejarán corresponden al total coberturas analizadas del producto, indemnización por gastos de hospitalización e indemnización por eventos de la naturaleza. A continuación se pasa a describir cada una de ellas:

- *Número de siniestros (total_claims)*, indica el número de eventos registrados por la compañía por asegurado y que han ocurrido durante el período de cobertura.
- *Coste último total (total_paid)*, es la cuantía de la prestación realizada por el asegurador al asegurado a consecuencia de la ocurrencia del riesgo cubierto. En nuestro caso, como todos los siniestros están liquidados, el coste último coincide con el total de pagos realizados por la compañía.
- *Suma asegurada (total_tsi)*, corresponde al total de la suma asegurada establecida por la compañía ante la ocurrencia del siniestro y por cada asegurado.

4.4 Análisis descriptivo: datos y variables

Una vez definidas las variables objeto de estudio, es conveniente realizar un primer análisis descriptivo de la información. En primer lugar, se dará una visión general de la base de datos completa y posteriormente, se analizarán en detalle cada una de las variables explicativas que se contemplan hasta el momento.

4.4.1. Datos: visión general

La base de datos engloba el perfil de riesgo y experiencia del colectivo asegurado en el período de estudio 2014 para las dos coberturas de estudio, indemnización diaria por hospitalización e indemnización por eventos de la naturaleza. Contiene 9.475 asegurados con un total de 243 siniestros registrados. Cada observación representa a cada asegurado con su correspondiente perfil de riesgo y su experiencia de siniestralidad. Los campos de siniestralidad están desglosados por tipo de garantía y causa del siniestro en diferentes columnas para facilitar la identificación ante la ocurrencia de varios siniestros por asegurado.

Tabla 4.1. Resumen general de la base de datos.

	Exposición	Número de siniestros	Coste último total	Suma asegurada	Coste último total por expuesto
Suma	5.569,63	243	864.750	1.615.000	1.369.983
Media	0,59	0,03	91	170	145
Mínimo	0,07	0	0	0	0
Máximo	0,99	2	10.000	20.000	21.808

Fuente de elaboración propia

La Tabla 4.1 describe un resumen estadístico de las principales variables de análisis. Como se observa, la exposición pone de manifiesto que del total de 9.475 asegurados 5.569,63 son expuestos, lo cual se explica porque la exposición media del asegurado está en torno a algo más de medio año, concretamente en un 0,59. El número de siniestros asciende a 243 y la media por asegurado es de 0,03 siniestros, lo que demuestra el elevado número de ceros que registra la experiencia de siniestralidad acumulada de nuestro microseguro. El coste último total, que coincide con los pagos registrados, está muy por debajo de la suma asegurada. Y la variable respuesta del coste último total ponderado por la exposición, se sitúa entre medias del coste último y la suma asegurada. Lógicamente, esta variable es superior al coste último total debido a que la exposición media es inferior al año.

Aunque el análisis del modelo se va a realizar para el total de coberturas, es interesante ver la influencia que tiene cada cobertura en el producto. Por esto, se muestra a continuación un resumen similar desglosado por las dos coberturas objeto de estudio.

Tabla 4.2. Resumen de la base de datos por cobertura.

	Coberturas de indemnización	Número de siniestros	Coste último siniestros	Suma asegurada	Coste último total por expuesto
Suma	Eventos de la naturaleza	194	752.500	1.035.000	1.183.752
Media		0,021	79	109	125
Suma	Hospitalización	49	112.250	580.000	186.231
Media		0,005	12	61	20

Fuente de elaboración propia

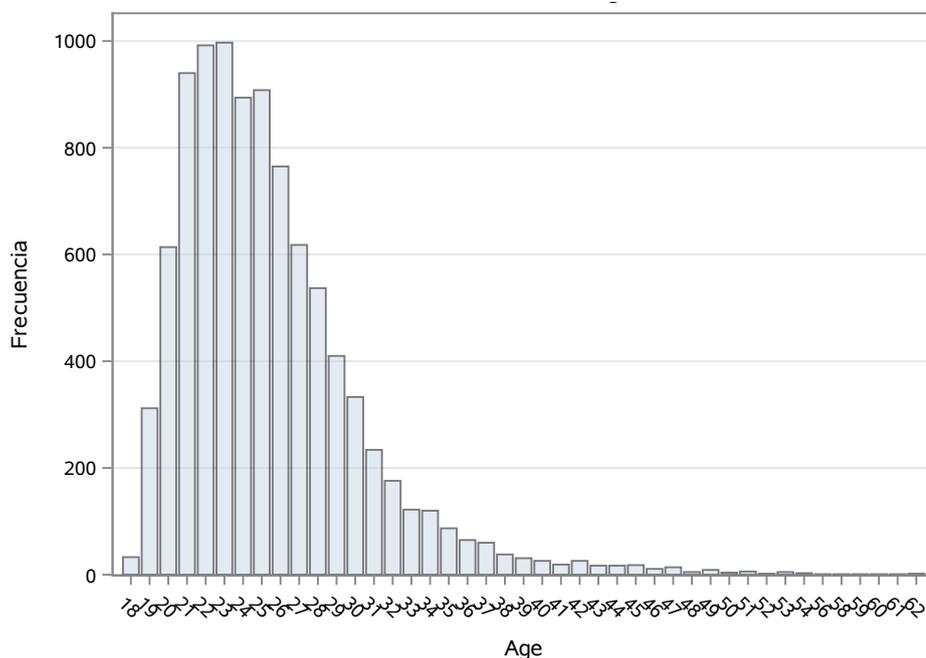
A la vista de los datos se puede comprobar como la cobertura de eventos de la naturaleza va a ejercer mayor peso en las estimaciones del microseguro, ya que cuenta con mayor siniestralidad. Ambas garantías presentan una baja siniestralidad, con un número medio de siniestros por asegurado de 0,021 y 0,005 siniestros para las coberturas de eventos de la naturaleza y hospitalización respectivamente. Esta última tan solo cuenta con 49 siniestros, pocos datos para poder realizar un análisis predictivo separado por cobertura. Por este motivo, y con el fin de buscar mayor consistencia en las estimaciones del modelo, se decide realizar el análisis de forma conjunta para el total coberturas. Este aspecto, que actualmente está condicionado por la limitación de información, sería lógicamente mejorable a futuro al acumular mayor experiencia y poder realizar el análisis de forma separada por cada garantía.

4.4.2. Variables explicativas

Características del asegurado

- La *edad (age)* se corresponde con la edad actuarial del asegurado a la fecha de efecto de la póliza. Los valores observados están comprendidos entre los 18 y los 62 años, aunque el producto ofrece cobertura desde los 18 hasta los 65 años. Con la idea de tener una visión sobre el perfil de edad del colectivo de nuestro microseguro se presenta el siguiente histograma de frecuencia para las diferentes edades observadas.

Gráfico 4.1. Histograma de frecuencia para la variable *edad*.



En el Gráfico 4.1 se muestra el histograma de frecuencia de la edad del asegurado y la moda se sitúa en los 23 años. De hecho, la mayoría de la exposición de la cartera se concentra en el intervalo de edad comprendido entre los 20 y los 30 años, lo que supone que la mayoría de los siniestros también están concentrados en dicho intervalo.

Tal y como se observa, existen asegurados prácticamente en todas las edades, en concreto con un total de 43 niveles, aunque no en todos ellos existe suficiente masa para representar de forma adecuada la edad del asegurado. Por tanto, es recomendable realizar una segmentación de la variable *edad*, agrupando diferentes niveles que tengan masa suficiente para realizar el análisis del modelo. Este proceso será planteado y desarrollado posteriormente.

- El sexo (*sex*). Esta variable se ha creado aplicando criterios generales a partir de otros campos disponibles, como son el nombre y la relación de parentesco con el beneficiario. Se distinguen dos posibles valores de la variable, hombre y mujer (*male*, *female*).

Tabla 4.3. Resumen descriptivo para la variable sexo.

Sexo	Frecuencia absoluta	Frecuencia relativa	Número de siniestros	Exposición
Hombre	5.195	55%	113	5.190
Mujer	4.280	45%	130	4.276
Total	9.475	100%	243	9.466

Fuente de elaboración propia

En la Tabla 4.3 se observa como la cartera está muy repartida entre ambos sexos, aún así los hombres representan la mayoría con un 55% sobre el total de asegurados. Es de notar, que las mujeres con una cuota menor de cartera tienen una mayor siniestralidad, tal y como indican los 130 siniestros respecto a los 113 que presentan los hombres.

- El *estado civil* (*civil_status*), presenta dos niveles según las dos posibles situaciones, soltero y casado (*single, married*).

Tabla 4.4. Resumen descriptivo para la variable *estado civil*.

Estado civil	Frecuencia absoluta	Frecuencia relativa	Número de siniestros
Soltero	8.604	91%	203
Casado	862	9%	40
Ausentes	9	0%	0
Total	9.475	100%	243

En la Tabla 4.4 se describe la composición de los asegurados según el estado civil, además existe un mínimo de observaciones ausentes, en concreto 9 registros, que no representa un peso significativo sobre el total de la cartera. El estado civil soltero es el predominante, con un 91% sobre el total de la cartera y lógicamente se corresponde con el de mayor número de siniestros registrados.

Características de la póliza

- *Antigüedad de la póliza* (*age_policy*). Esta variable se ha obtenido gracias a la información disponible de los años 2012 y 2013, por medio del cruce con los asegurados del año 2014 analizado. Los valores asignados a esta variable tienen un rango comprendido entre 0 y 2: el valor 0 se corresponde a los nuevos asegurados en el ejercicio 2014 que llevan menos de un año en la póliza; el valor 1, para los que llevan desde el año 2013, con al menos un año de antigüedad; y el valor 2, para los que llevan asegurados desde el año 2012, es decir con al menos dos años consecutivos en la póliza.

Tabla 4.5. Resumen descriptivo para la variable *antigüedad de la póliza*.

Antigüedad póliza	Frecuencia absoluta	Frecuencia relativa	Número de siniestros
0	4.801	51%	53
1	2.876	30%	79
2	1.798	19%	111
Total	9.475	100%	243

Fuente de elaboración propia

A la vista de los datos recogidos en la Tabla 4.5 se aprecia como casi la mitad de la cartera, un 49% de los asegurados, cuentan con más de un año de antigüedad en la

póliza. Un dato interesante es el comportamiento de la siniestralidad respecto a esta variable, los datos confirman que el número de siniestros aumenta con la antigüedad de los asegurados en la póliza. De forma, que los asegurados con más de dos años son los que presentan mayor siniestralidad de la cartera con un total de 111 siniestros. Este hecho se debe a dos razones, por un lado, cuanto mayor tiempo de exposición mayor conocimiento del producto y mayor propensión de uso. Y por otro lado, se pone de manifiesto la selección adversa en el producto, donde los riesgos peores con mayor siniestralidad son los que permanecen en la cartera, mientras que los riesgos buenos que presentan una baja siniestralidad acaban anulando la póliza.

Ambos aspectos deben ser controlados para evitar el uso fraudulento del producto por los clientes más antiguos, así como establecer medidas de fidelización destinadas a retener a los buenos clientes y evitar así que anulen la póliza. Los datos confirman la relevancia de la variable antigüedad de la póliza, ya que puede ser un buen factor de riesgo en la tarifa y que además puede servir como una medida evasiva para controlar la siniestralidad de los asegurados con mayor antigüedad en la cartera.

Zona geográfica

La zona geográfica corresponde al lugar de residencia del asegurado. La base de datos ofrece varios campos sobre la residencia del asegurado, relacionados con la distribución geográfica en Filipinas, como son la provincia, el *barangay* que se corresponde al término equivalente a pueblo o distrito y la dirección. De todas estas variables se plantea el estudio de las dos siguientes:

- *Provincia (province)*. Esta variable cuenta con 86 diferentes categorías que engloban las provincias, las ciudades y los cuatro distritos donde residen nuestros asegurados en Filipinas.

Dado el número elevado de niveles, es conveniente plantear otras alternativas más apropiadas que contemplen información sobre la residencia del asegurado. Se plantean a continuación dos posibles opciones para definir este factor de riesgo, la variable región y la variable densidad de población.

- *Región (region)*, es una variable que agrupa las diferentes zonas de residencia del asegurado por su cercanía geográfica. Partiendo de los 86 niveles de la variable provincia, se agrupan en un total de 18 regiones atendiendo al código geográfico estándar de Filipinas (Philippine Statistics Authority, 2016).

Tabla 4.6. Resumen descriptivo para la variable *región*.

Región	Nombre región	Frecuencia absoluta	Frecuencia relativa	Número de siniestros
1	Region I - Ilocos	163	2%	1
2	Region II - Cagayan Valley	70	1%	2
3	Region III - Central Luzon	225	2%	4
4	Region IV.A - Calabarzon	6.952	73%	216
5	Region IV.B - Mimaropa	183	2%	1
6	Region V - Bicol	357	4%	4
7	Region VI - Western Visayas	77	1%	3
8	Region VII - Central Visayas	38	0%	0
9	Region VIII - Eastern Visayas	116	1%	1
10	Region IX - Zamboanga peninsula	675	7%	5
11	Region X - Northern mindanao	31	0%	0
12	Region XI - Davao	23	0%	0
13	Region XII - Soccsksargen	24	0%	0
14	NCR - National Capital Region	425	4%	5
15	CAR - Cordillera Administrative Region	9	0%	0
16	ARMM - Autonomous Region Muslim Mindanao	10	0%	0
17	Region XIII - Caraga	43	0%	0
18	NIR - Negros Island Region	54	1%	1
Total		9.475	100%	243

Fuente de elaboración propia

La Tabla 4.6 proporciona la distribución de la cartera entre las diferentes regiones. Prácticamente la mayoría de la cartera está centrada en una sola región, la región 4 que cuenta con el 73% de los asegurados. Mientras que el porcentaje restante está repartido entre la región 10, 14 y 6 principalmente. Igualmente, el número de siniestros está también concentrado en dichas regiones donde se registra mayor exposición. Esta clasificación no es homogénea puesto que la cartera está concentrada principalmente en tres niveles. En cambio, el resto de regiones cuentan con muy poca exposición, lo que no resultaría adecuado a la hora de aplicar el modelo. La solución entonces vendría dada por realizar una oportuna segmentación que permita agrupar las provincias con una exposición más o menos homogénea.

- *Densidad de población (density)*. Se trata de una variable exógena, que considera la distribución de la población por área geográfica. Su utilización se justifica en el sentido de que existen poblaciones con mayor número de habitantes que otras, lo que repercute directamente en una mayor o menor siniestralidad según la zona donde resida el asegurado. Para el cálculo de la densidad de población se ha utilizado información externa (Philippine Statistics Authority, 2016), que proporciona el número de habitantes existente en cada una de las provincias. Los valores de esta variable se calculan como el cociente entre el número de habitantes en cada una de las provincias, respecto a la población total de Filipinas. Por tanto, de cada valor de la

variable provincia obtendremos un valor de densidad, con tantos niveles como zonas tenga la variable provincia, es decir un total de 86 niveles. Este número es muy elevado y tampoco sería adecuado a la hora de realizar el modelo, por lo que será necesario plantear una adecuada segmentación que permita agrupar las diferentes densidades en un número manejable de niveles. En los apartados siguientes se procederá a realizar este proceso de segmentación.

4.4.3. Variables a posteriori

Al igual que para las variables explicativas, es importante realizar un análisis descriptivo de las dos variables a posteriori que se utilizarán en el modelo, el número de siniestros y el coste último total.

- *Número de siniestros (total_claims)*. Esta variable representa el número de eventos que han sido registrados por la compañía durante el período de cobertura. A continuación se muestra el siguiente análisis descriptivo de la variable.

Tabla 4.7. Resumen descriptivo para el número de siniestros.

Número de siniestros	Frecuencia absoluta	Frecuencia relativa
0	9.234	97,46%
1	239	2,52%
2	2	0,02%
Total	9.475	100,00%

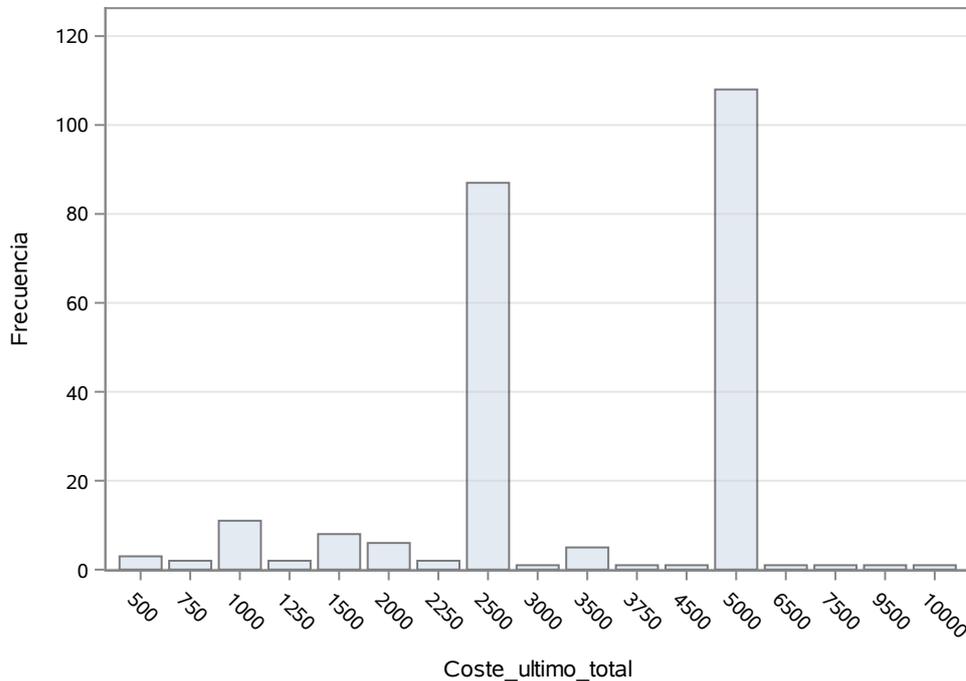
Fuente de elaboración propia

Tal y como se observa en la Tabla 4.7, la experiencia registrada muestra que existe un número elevado de ceros con un 97,46%, lo que indica una baja siniestralidad del microseguro ya que solo el 2,54% restante presenta algún siniestro. La distribución del número de siniestros para el total de coberturas analizadas toma valores comprendidos entre cero y dos, y con solo 242 siniestros registrados durante el período de observación.

- *Coste último total (total_paid)*, se corresponde con la prestación que realiza el asegurador para dar cobertura a los siniestros ocurridos. Durante el período observado se han registrado 241 siniestros liquidados por lo que la indemnización correspondiente coincide con el total de pagos registrados por la compañía. A continuación, se realiza un análisis descriptivo de la variable pagos, centrándonos solo

en los casos donde haya habido al menos un siniestro para poder conocer así la distribución de esta variable.

Gráfico 4.2. Distribución de la frecuencia del coste último total.



Observando la distribución de los pagos realizados se observa un rango bastante acotado entre los 500Php y los 10.000Php. Del total de 241 siniestros registrados, la cuantía de la prestación se concentra principalmente en dos valores, un importe de 5.000Php, que representa un 45%, y de 2.500Php con un 36%. Esto indica que la distribución de la cuantía es prácticamente discreta, al existir un rango tan limitado además de que la mayor parte de siniestros están concentrados en estos dos valores.

4.5 Segmentación de las variables explicativas

Después de haber analizado la estadística descriptiva de nuestras variables explicativas, en algunas de ellas se planteaba la necesidad de transformarlas con la idea de disponer de unos predictores que contengan un número representativo de niveles homogéneo para la exposición de nuestro colectivo. Este proceso se lleva a cabo a través del procedimiento de segmentación que ahora veremos, y que se ha descrito previamente en la sección 3.2.2 del capítulo anterior.

Recapitulando las variables predictoras disponibles para nuestro análisis, se resumen a continuación todas ellas, indicando cuáles precisan o no transformación. Acto seguido, analizaremos cuáles de ellas son o no apropiadas para incluir en nuestro modelo.

Tabla 4.8. Resumen de las variables explicativas disponibles.

Características	Variable	Niveles	Requiere segmentación
Asegurado	Edad	43	✓
	Estado civil	2	–
	Sexo	2	–
Póliza	Antigüedad póliza	3	–
Zona geográfica	Región	8	✓
	Densidad	86	✓

Fuente de elaboración propia

En la Tabla 4.8 se aprecia claramente que aquellas variables con un número elevado de niveles no son adecuadas para incluir directamente al modelo, lo que hace necesaria su correspondiente segmentación. De hecho, si se incluyesen directamente en el modelo sin un tratamiento previo, los resultados no serían coherentes hasta el punto de que el modelo pudiese no converger (Goldburd et al., 2016).

En nuestro estudio particular, los predictores que sí van a requerir realizar la correspondiente segmentación son la edad, como característica del asegurado, y dentro de la zona geográfica, la región y la densidad. En el resto de variables, como son el estado civil, el sexo y la antigüedad de la póliza se disponen de pocos niveles, lo que posibilita utilizarlas directamente en el modelo.

Características del asegurado

La edad del asegurado cuenta con un rango inicial de 43 niveles para las edades comprendidas entre los 18 y los 62 años. El proceso de segmentación seguido se ha realizado en 4 etapas hasta obtener la variable *age5* con 4 niveles, tal y como se muestra en el siguiente resumen.

Tabla 4.9. Segmentación de la variable explicativa *edad*.

Paso	Variable segmentada	Niveles	Intervalos	AIC	Desviación escalada	Máxima Pr(ChiSq)
0	age	43	x=18,..., 54,56,...,62	1.464	1.854	–
1	age1	10	[(18-20); (21-22); 23; 24; (25-26); 27; (28-30); (31-36); (37-44); (45-62)]	1.416	1.858	0,3348
2	age2	6	[(18-22); (23-24); (25-27); (28-36); (37-44); (45-62)]	1.411	1.860	0,2308
3	age3	5	[(18-22); (23-25); (26-30); (31-44); (45-62)]	1.410	1.867	0,2303
4	age5	4	[(18-24); (25-30); (31-36); (37-62)]	1.413	1.859	0,0004

Fuente de elaboración propia

En la Tabla 4.9 se muestran los principales resultados de la segmentación de la variable *edad*, con los intervalos elegidos en cada clase y marcando en rojo aquellos niveles de la variable que no son significativos. En la última columna, aparece como referencia la máxima probabilidad del test Chi-cuadrado en el nivel que presenta menor significación. Como medidas de referencia, se han utilizado los estadísticos AIC y la desviación escalada obtenidos con el modelo. Es fácil comprobar como, a medida que se va agrupando la variable en cada etapa, el valor AIC va también disminuyendo hasta llegar al paso 4 donde su valor vuelve a repuntar. De forma inversa, en el caso de la desviación escalada, esta va aumentando a medida que segmentamos debido a que se va perdiendo información al ir agrupando. Sin embargo, esta tendencia parece que se revierte cuando realizamos el último paso 4 de la segmentación.

Cabe decir que no existe un único criterio de selección a la hora de elegir la mejor variable segmentada, porque no solo hay que valorar los resultados obtenidos de las medidas de ajuste, sino que además debe tenerse en cuenta que la mayoría de los niveles de la variable sean significativos. Se hace especial hincapié en este aspecto debido a la diferencia que existe entre el criterio empleado para la segmentación, basado en el modelo Poisson con la variable número de siniestros, y el planteamiento de estimación de la prima pura que utiliza el modelo Tweedie. Por este motivo, se da mayor prioridad a lograr la mayor significación de los diferentes niveles de la variable.

Dicho esto, para la variable *edad* se van a proponer dos posibles alternativas para incluir en el modelo de prima pura. La variable *age2* porque presenta buenas medidas AIC y desviación, aunque cuenta con un nivel no significativo. Y por otro lado, la segmentación correspondiente a la variable *age5*, que presenta 4 niveles con unas medidas similares a la anterior y sobre todo cuenta con todos los niveles de la variable

significativos. Cualquiera de estas dos variables segmentadas pueden representar una buena alternativa como factor de riesgo *edad* en el modelo Poisson. No obstante, el modelo Tweedie, enfocado en estimar la prima pura, será el que nos permita decidir cuál de las dos variables proporcionará las mejores estimaciones.

Zona geográfica

En el ámbito geográfico se disponen también, como vimos anteriormente, de dos posibles opciones. La variable *región*, que se corresponde con una agrupación estándar por cercanía geográfica de las diferentes provincias donde residen los asegurados. Y, la variable exógena *densidad*, que recoge el efecto de la distribución de población en cada una de las provincias.

En principio, la primera de ellas, la variable *región*, parecía ser una buena alternativa para agrupar las diferentes provincias atendiendo a su localización geográfica. Esta agrupación obedece a un criterio general que se corresponde con el código geográfico estándar de Filipinas (Philippine Statistics Authority, 2016), accesible y conocido por el mercado nacional, lo que favorecería su implementación. Sin embargo, al aplicar este criterio a los datos disponibles del colectivo asegurado (véase Tabla 4.6), se comprueba que la mayoría de la población asegurada, concretamente un 73% de la exposición, estaba concentrada en una sola región, con lo que cualquier segmentación posterior no va a ofrecer mejores resultados.

Por otro lado, la segunda alternativa consiste en la *densidad*. Esta presenta una doble ventaja respecto a *región*, puesto que no solo recoge la información de la propia provincia, sino que además tiene en cuenta cómo está distribuida la población en cada una de ellas. De hecho, al considerar esta variable en el modelo hace que el coste total del siniestro por expuesto no solo vaya a depender de la propia provincia, sino de si está más o menos poblada en relación al resto de provincias.

A la vista de estas observaciones, se ha decidido finalmente trabajar directamente con la *densidad* como variable identificativa de la zona geográfica del asegurado por su mayor representatividad.

La *densidad* es una variable continua que presenta los mismos niveles de inicio que presenta la variable provincia, exactamente 86 niveles (véase Anexo B.1). El proceso de segmentación sigue los mismos pasos que en el caso anterior para la variable

edad. En este caso, el proceso de agrupación se ha realizado en 5 pasos hasta llegar a la variable más segmentada, *den5*.

Tabla 4.10. Segmentación de la variable explicativa *densidad*.

Paso	Variable segmentada	Niveles	Intervalos	AIC	Desviación escalada	Máxima Pr(ChiSq)
0	density	86	d=0,0002,...,0,0461	1.493	1.888	–
1	den1	12	[0,0118; 0,0208; 0,0299; 0,0358; 0,0364; 0,0401; 0,0523; 0,0562; 0,0672; 0,0977; 0,2211; 0,2515]	1.345	1.903	0,9996
2	den2	7	[0,0208; 0,0364; 0,082; 0,092; 0,0977; 0,1195; 0,4727]	1.338	1.874	0,9992
3	den3	4	[0,0208; 0,036; 0,2718; 0,5922]	1.351	1.752	0,0004
4	den5	3	[0,0364; 0,2928; 0,5922]	1.354	1.760	<0,0001

Fuente de elaboración propia

A la vista del resumen de la Tabla 4.10 se observan las diferentes etapas seguidas en la segmentación de la variable *densidad*. El proceso se inicia con la variable *density* que parte de 86 niveles hasta llegar al nivel más bajo, con 3 niveles para la variable *den5*. A medida que se va segmentando se van reduciendo el número de niveles de la variable, aunque en mayor medida en los primeros pasos. Las medidas de ajuste AIC y la desviación escalada también van mejorando al disminuir sus valores a medida que segmentamos. Aunque al llegar al paso 3, el AIC repunta de nuevo, lo que indica que una posterior agrupación lleva implícita una pérdida en el detalle de la información. A pesar de que las variables *den3* y *den5* presentan un AIC superior, no ocurre lo mismo con la desviación, cuyo mejor resultado lo obtiene la variable *den3*. Además, es importante señalar que ambas variables son significativas en todos sus niveles, y en especial con una mejor significación, p-valor < 0,0001, con la variable *den5*.

Por tanto, a través de la segmentación realizada se confirman dos posibles alternativas para representar la zona geográfica, las variables *den3* y *den5*, que son adecuadas para el modelo de Poisson basado en el número de siniestros y por extensión para el modelo de prima pura. Sin embargo, la decisión final sobre cuál será la variable que mejor represente la zona geográfica dependerá de los resultados que se obtengan en el modelo de prima pura que se pretende realizar posteriormente.

4.6 Estudio de influencia de las variables explicativas

Es lógico pensar que pueden existir relaciones entre las variables explicativas propuestas para nuestro modelo. Por ello, antes de entrar en el análisis predictivo y ver la influencia de las variables explicativas sobre la variable respuesta, se va a realizar en primer lugar, un análisis estadístico sobre la posible relación o asociación entre ellas para conocer su grado de correlación. En segundo lugar, se va a plantear si la combinación entre las diferentes variables puede tener un efecto sobre la variable respuesta, lo que se denomina interacción. En resumen, este estudio permite conocer el grado de relación que existe entre las potenciales variables seleccionadas con el propósito de conseguir la mayor capacidad explicativa en las estimaciones del modelo.

4.6.1. Análisis de dependencia

El estudio de dependencia sirve para conocer el comportamiento de las variables explicativas que intervendrán en el modelo e identificar cuáles pueden o no ocasionar algún problema. La herramienta SAS ofrece diferentes opciones para comprobar el grado de asociación entre variables a través de diferentes procedimientos.

- Al realizar el propio análisis GLM con el procedimiento "PROC GENMOD". Este permite ajustar modelos lineales generales siempre y cuando la correlación entre las variables no sea muy grande. Por este motivo, siempre hay que comprobar que existe convergencia en el modelo.
- Análisis de colinealidad. Este mide la correlación que existe entre dos o más variables. Se dice que hay una alta colinealidad cuando existe una correlación alta. En este caso, la información que proporcionan ambos predictores es muy similar o puede existir cierto solapamiento, obteniendo así unos estimadores imprecisos.

El procedimiento utilizado en SAS es "PROC REG" que permite realizar diagnósticos de colinealidad a través de las opciones de *VIF*, *TOL*, *COLLIN* y *COLLINOINT*. Probablemente la opción más utilizada es la inflación de la varianza (*VIF*). Se trata de una medida relativa del incremento de la varianza debido a la colinealidad de una variable con las otras. Un *VIF* superior a 10 indica que existe colinealidad entre las variables. Sin embargo, esta medida no facilita información sobre cuáles de las variables son colineales con cada una de ellas. Otra medida es la tolerancia (*TOL*), la inversa de *VIF*, que determina que existe una alta colinealidad cuando su valor no es superior a 0,1. Las dos restantes opciones son *COLLIN* y *COLLINOINT*, que se basan

en autovalores y tienen la ventaja de indicar qué grupos de variables pueden ser colineales. La primera de ellas, *COLLIN*, sí considera el término independiente, mientras que *COLLINOINT* lo ajusta por fuera. Ambas medidas proporcionan el llamado índice de la condición. Este indicador mide el grado de dependencia entre las variables, si su valor se acerca a 10 puede surgir una ligera dependencia que podría afectar a la regresión, y si este número fuese superior a 100 es muy probable que existan errores en el modelo de regresión. En el caso de analizar la multicolinealidad, es decir cuando la asociación se produce entre más de dos variables, el índice de la condición toma de referencia valores superiores a 30 para indicar que existe multicolinealidad entre las variables del modelo (SAS Institute, 2008).

Profundizando en nuestro estudio, este análisis de colinealidad se aplica a través de un modelo de regresión lineal, considerando como variable respuesta el coste último total por expuesto y como variables explicativas todos los posibles factores de riesgo. Debido a que se dispone de dos alternativas para los factores edad y zona geográfica, se probarán de forma separada dos posibles modelos que combinan las dos alternativas existentes para cada variable.

Tabla 4.11. Análisis de colinealidad (variables segmentadas *age2* y *den3*).

Variable	Estimador	Error estándar	Tolerancia (TOL)	Inflación de la varianza (VIF)
Intercept	42,05	53,00	—	0
age2	33,79	8,98	0,83	1,21
Antigüedad póliza	127,78	13,56	0,95	1,06
Sexo	-78,18	20,55	0,99	1,01
Estado civil	54,97	37,86	0,87	1,14
den3	-31,69	51,87	0,99	1,01

Fuente de elaboración propia

Número	Collin		Collinooint	
	Autovalor	Índice de la condición	Autovalor	Índice de la condición
1	4,895	1,000	1,46	1,00
2	0,504	3,115	1,04	1,18
3	0,363	3,673	0,99	1,21
4	0,143	5,843	0,92	1,26
5	0,069	8,402	0,60	1,56
6	0,025	13,908		

Fuente de elaboración propia

La Tabla 4.11 corresponde al análisis de colinealidad teniendo en cuenta las variables explicativas segmentadas *age2* y *den3*. En la parte superior, aparecen los valores de los estimadores del modelo, comenzando por el término independiente (*intercept*), seguido de todas las variables explicativas. Cada uno de ellos lleva asociado su índice de tolerancia e inflación de la varianza, los cuales confirman que no existe colinealidad entre las variables. El estadístico *VIF* es inferior a 10 para todas las variables, y en consecuencia el valor *TOL* está por encima de 0,1.

En la parte inferior, se describe el diagnóstico de colinealidad con las dos opciones *COLLIN* y *COLLINOINT*. En ambos casos los valores del índice de la condición, que miden el grado de asociación, son inferiores a 10. Existe una excepción en el primer análisis, con un índice es de 13,908, lo que nos puede indicar que es posible que exista una débil dependencia en la última variable *den3*.

Tabla 4.12. Análisis de colinealidad (variables segmentadas *age5* y *den5*).

Variable	Estimador	Error estándar	Tolerancia (TOL)	Inflación de la varianza (VIF)
Intercept	28,15	53,23	—	0
age5	46,88	14,59	0,81	1,23
Antigüedad póliza	131,95	13,39	0,97	1,03
Sexo	-74,83	20,54	0,99	1,01
Estado civil	55,85	38,70	0,84	1,19
den5	16,17	51,54	0,99	1,01

Fuente de elaboración propia

Número	Collin		Collinooint	
	Autovalor	Índice de la condición	Autovalor	Índice de la condición
1	4,919	1,000	1,46	1,00
2	0,504	3,124	1,05	1,18
3	0,352	3,737	1,00	1,21
4	0,136	6,016	0,91	1,27
5	0,064	8,758	0,58	1,59
6	0,025	13,975		

Fuente de elaboración propia

En la Tabla 4.12 los resultados corresponden al modelo que incluye las variables *age5* y *den5*. Los resultados son muy similares al anterior modelo, en la parte superior, para

todas las variables consideradas, sus índices *VIF* y *TOL* confirman que no existe colinealidad. En la sección inferior, en ambos diagnósticos de colinealidad *COLLIN* y *COLLINOINT*, los índices de la condición están por debajo de 10 salvo en el primer análisis en el que también la última variable *den5* presenta un índice de 13,975. Al igual que en el modelo anterior, esto nos podría advertir de una débil dependencia en esta última variable. Es de notar, como a medida que los autovalores se van reduciendo los índices correspondientes van aumentando.

Comparando los dos modelos analizados, se comprueba que no existe colinealidad entre las variables explicativas, tan solo se advierte una posible ligera dependencia asociada al factor de riesgo zona geográfica. Esta observación per se, no demuestra una clara dependencia por lo que simplemente se debe tener en cuenta a la hora de realizar el modelo de prima pura.

- Coeficiente V de Cramer. Consiste en una medida basada en el estadístico χ^2 de Pearson y proporciona la intensidad en la asociación entre dos o más variables (Anderson et al., 2007; Towers Watson, 2010). Su fórmula de cálculo es sencilla y se define como:

$$V = \sqrt{\frac{\chi^2}{nt}} \quad (4.2)$$

Siendo n el tamaño de la muestra y t el mínimo valor entre el número de filas r menos uno o columnas c menos uno respecto de las dos variables analizadas. De forma que:

$$t = \min(r - 1, c - 1) \quad (4.3)$$

Los valores posibles del coeficiente V de Cramer están comprendidos en el intervalo $0 < V < 1$, cuando $V = 0$ indica independencia absoluta entre las variables y $V = 1$ muestra la dependencia perfecta entre las variables. A la hora de interpretar los valores de este estadístico, una regla de oro consiste en que aquellos valores inferiores al 5% suponen una débil dependencia, en cambio los que estén por encima del 30% implican una fuerte dependencia y deben prestarse especial atención (Towers Watson, 2010).

Aplicando el coeficiente V de Cramer a las variables explicativas potenciales para el modelo de prima pura se obtiene la siguiente matriz de coeficientes V de Cramer.

Tabla 4.13. Matriz de coeficientes V de Cramer.

	age2	age5	estado civil	sexo	antigüedad póliza	den3	den5
age2			0,414	0,102	0,212	0,046	0,055
age5			0,431	0,080	0,187	0,047	0,057
estado civil	0,414	0,431		0,029	0,052	0,055	0,055
sexo	0,102	0,080	0,029		0,054	0,128	0,127
antigüedad póliza	0,212	0,187	0,052	0,054		0,121	0,120
den3	0,046	0,047	0,055	0,128	0,121		
den5	0,055	0,057	0,055	0,127	0,120		

Fuente de elaboración propia

En la Tabla 4.13 de la matriz de coeficientes V de Cramer se observa, que prácticamente la mayoría de las variables tienen valores bajos, lo que indica que no hay dependencia entre ellas. Analizando los coeficientes, el *estado civil* presenta los coeficientes más elevados con 0,414 y 0,431 respecto a las dos variables de *edad* (*age2* y *age5* respectivamente). En realidad, tiene su lógica pensar que el estado civil del asegurado esté relacionado con la edad. Y a pesar de que el coeficiente obtenido con nuestra base de datos no representa una alta dependencia, sí es conveniente tenerlo en cuenta para el posterior análisis. La variable *sexo* presenta sus valores más altos en relación al factor de riesgo zona geográfica. Esto hace pensar que el sexo de los asegurados condiciona en cierta medida el lugar de residencia de los asegurados. Por último, la variable *antigüedad de la póliza* igualmente parece tener sus valores más altos con los factores de riesgo edad y zona geográfica. Los datos nos confirman que los asegurados con más antigüedad en la póliza son los más adultos y que residen en las provincias de mayor índice de población.

En definitiva, los datos proporcionados por la matriz de coeficientes V de Cramer nos confirman que no existe una fuerte dependencia entre ninguna de las variables según la información disponible. Sí resulta práctico realizar este análisis para conocer qué grado de asociación existe entre las variables e interpretar las posibles causas de dependencia. Todo este estudio nos servirá como herramienta en el posterior análisis, para valorar qué predictores deben o no incluirse dentro de nuestro modelo de prima pura.

4.6.2. Análisis de interacción

Las interacciones entre las variables explicativas, que permiten incorporar el análisis multivariante, tienen la ventaja de que pueden mejorar los resultados estimados por el modelo aunque a su vez esto también pueda complicar su interpretación. Dicho esto, y según el procedimiento previamente explicado en la sección 3.2.2 del capítulo anterior, se plantea analizar las posibles interacciones entre los predictores antes de proseguir con el análisis predictivo del modelo.

Al realizar este análisis se comprueba que, prácticamente para todas las posibles combinaciones de nuestras variables explicativas, los resultados del modelo no son favorables. La comprobación se observa a través del estadístico AIC, que tiende a aumentar al incluir el término interacción, incluso en algunos casos su incorporación provoca la no convergencia del modelo y por ende la no significación de las variables. Tan solo, en un caso concreto, se distingue que la interacción entre las variables *age5* y la *antigüedad de la póliza* provoca una ligera mejoría en el AIC, pero repercute en la falta de significación de la variable *age5*. A la vista de estos resultados, se confirma que la interacción de variables en nuestro estudio particular no produce mejores resultados y por tanto, se descarta la incorporación de interacciones en nuestro modelo.

CAPÍTULO 5. ANÁLISIS PREDICTIVO

Este último capítulo empírico se centra en la aplicación de la metodología de tarificación del modelo lineal generalizado que sigue la distribución Tweedie, al caso particular del microseguro objeto de estudio. La finalidad consiste en demostrar que es posible obtener una tarifa multivariante ajustada al perfil de riesgo individual y que a su vez contribuya con una mejor gestión del riesgo asumido por la compañía.

Para lograr este cometido se ha organizado el capítulo de la siguiente forma. En primer lugar, partiendo de la propuesta metodológica del modelo Tweedie, explicaremos cómo realizaremos su implementación en la herramienta de análisis SAS. En segundo lugar, se plantearán las hipótesis de partida del modelo en relación a los predictores y a la validación de la muestra empleada. En tercer lugar, y como núcleo de la investigación, se ejecutará el modelo de forma iterativa introduciendo uno a uno los diferentes predictores hasta lograr obtener los modelos predictivos más robustos. Entre las alternativas planteadas se escogerá un modelo óptimo que será el que mejor represente la previsión futura del riesgo asegurable. Finalmente, se va a realizar un estudio comparativo de nuestro modelo óptimo y el modelo real que actualmente está utilizando la compañía, con el fin de demostrar las ventajas que supone la aplicación de una metodología más sofisticada, como es el modelo Tweedie para el cálculo de la tarifa del microseguro.

5.1 Metodología del caso de estudio: el modelo Tweedie

El análisis predictivo que se va a plantear en este estudio utiliza un solo modelo lineal general para nuestro objetivo, el modelo Tweedie, que consiste en estimar las primas de riesgo del microseguro mediante la variable respuesta, coste último total por expuesto. Ahora bien, tal y como se describía en el capítulo anterior sobre la metodología GLM, la distribución Tweedie cuenta con tres parámetros μ , ϕ y p , y en función de los valores del parámetro p se van a estimar diferentes miembros de la familia Tweedie y de la familia exponencial (véase Tabla 3.2).

En nuestro caso de estudio, la aplicación del modelo GLM basado en la distribución Tweedie se justifica perfectamente por lo siguiente. El producto, objeto de tarificación, consiste en un microseguro, y a la vista de los datos disponibles es evidente que hay una limitación de información que impide realizar dos análisis GLM de forma separada (uno para la frecuencia y otro para el coste medio). Esto nos conduce a la aplicación de un único modelo para el coste último total por expuesto basado en la distribución Tweedie. Además, no se debe olvidar las características propias del tipo de producto que se quiere estimar la tarifa, el microseguro. En su mayoría, los microseguros se conciben como productos básicos y sencillos, con una o varias coberturas pero con una suma asegurada especialmente reducida como también lo deben ser sus primas correspondientes. Por lo cual, la justificación de aplicar una metodología de tarificación más sofisticada que permita captar mejor el comportamiento de los riesgos, sí debe estar acompañada de un criterio de sencillez y simplicidad que favorece en este caso la aplicación de un único modelo.

El modelo Tweedie que se va a utilizar en la presente tesis, pretende modelizar la prima pura basándonos en el caso particular de $1 < p < 2$, que se corresponde con la distribución conocida como Poisson compuesta con Gamma y permite estimar la prima pura²⁰. Para nuestro caso de estudio, la estimación de parámetros (μ, p, ϕ) y los consecuentes resultados se obtendrán directamente aplicando la metodología de cálculo planteada por Shi (2010). En su estudio implementa una macro en SAS basada en la propiedad ortogonal que existe entre el parámetro μ y los parámetros (p, ϕ) para realizar sus optimizaciones de forma separada. La macro consiste en un proceso iterativo que combina dos pasos, un primer procedimiento que estima el

²⁰ Demostración realizada por Jørgensen y De Souza (1994).

parámetro μ a través del “PROC GENMOD” y por otro lado, la estimación de los parámetros (p, ϕ) con el procedimiento “PROC NLMIXED”²¹.

- El procedimiento “GENMOD”, ampliamente utilizado para el análisis GLM, permite fácilmente lograr la convergencia del modelo para las distribuciones preestablecidas. En el caso de la distribución Tweedie, este procedimiento no contempla el cálculo automático de algunas medidas, como son la desviación y la varianza, aunque sí permite programarlo manualmente para obtener sus valores correspondientes. Este procedimiento utiliza la función de enlace logarítmica para relacionar la variable respuesta con las diferentes variables explicativas. Por lo cual, a la hora de obtener las relatividades de la prima pura se deberá realizar la oportuna transformación tomando las exponenciales de los parámetros estimados.
- Por otro lado, el procedimiento “NLMIXED” se caracteriza por su flexibilidad y permite especificar la función de verosimilitud correspondiente a la distribución Poisson compuesta perteneciente a la distribución Tweedie (Yan et al., 2009).

La conjunción de ambos procedimientos permite obtener un modelo óptimo que converge bajo la optimización de sus tres parámetros de la distribución Tweedie (Shi, 2010).

Profundizando en nuestro caso de estudio, se han realizado las oportunas modificaciones del código SAS para adaptarlo convenientemente (véase Anexo C). El siguiente paso ahora consiste en asignar las diferentes variables macro del programa SAS con nuestras variables disponibles, tal y como se define a continuación.

- La variable respuesta (en el programa como *resp*) se corresponde con el coste último total por expuesto (*b_cost_T*) y coincide con el total de pagos registrados del siniestro sobre la exposición.
- La variable peso (definida como variable macro *wght*) será la exposición de cada asegurado (*exposure*).

²¹ El procedimiento “NLMIXED” fue desarrollado en SAS originalmente por Yan et al. (2009).

- La categorización de las variables explicativas no va a seguir el criterio establecido en la macro²², sino que se realiza manualmente con el propósito de indicar qué nivel base corresponde a cada variable discreta. El criterio elegido para designar el nivel de referencia para nuestros predictores es aquel que presente la mayor exposición. SAS ofrece la opción de elegir el orden según la frecuencia de las observaciones aplicando la opción *DESC ORDER=FREQ*, y que en nuestro caso coincide con el de mayor exposición.
- Las variables explicativas del modelo (en la variable macro *pred*) contendrán todos los predictores seleccionados y previamente segmentados. Todos ellos aparecen descritos en la siguiente Tabla 5.1, teniendo en cuenta que se deberá seleccionar un único predictor para los factores de riesgo edad y zona geográfica.
- El número de siniestros (*total_claims*) será utilizado en el procedimiento “NLMIXED” (como variable macro *clmct*).

Con toda esta definición pormenorizada de variables para nuestro caso de estudio, quedan identificados los elementos necesarios para el correcto funcionamiento del modelo Tweedie en esta macro de SAS. Mediante este código (véase Anexo C) se podrá realizar el cálculo de los correspondientes parámetros y estimadores del modelo.

5.2 Hipótesis del modelo

Una vez que se ha explicado cómo se va a implementar el modelo Tweedie en la herramienta SAS junto con la correspondiente definición de variables para nuestro microseguro corresponde avanzar en el estudio, concretando las hipótesis de partida necesarias para desarrollar el análisis predictivo. Las hipótesis de partida comprenden dos partes diferenciadas. Por una parte, el criterio de selección de las variables explicativas, que representan los factores de riesgo de la tarifa. Y por otro, el planteamiento adoptado para validar los resultados con la muestra de datos disponible.

²² Se elimina la variable macro *vclass* definida para la clasificación de los niveles de cada variable.

5.2.1. Variables explicativas

Son los predictores del modelo y los que mejor van a representar el comportamiento de nuestra variable respuesta, el coste último total por expuesto. Todas ellas representan los factores de riesgo del colectivo que formarán parte de la tarifa del microseguro. Como paso previo, en el capítulo anterior ya se realizó el correspondiente análisis estadístico de estas variables, su segmentación, el análisis de dependencia e interacción, hasta llegar a seleccionar los diferentes predictores en función de las características del asegurado, de la póliza y de la zona geográfica. A continuación, se resumen las 7 variables explicativas potenciales según sus características y que junto con el término independiente formarán parte del modelo.

Tabla 5.1. Resumen de las variables explicativas potenciales del modelo.

Características	Variable	Niveles
Asegurado	age2	6
	age5	4
	Estado civil	2
	Sexo	2
Póliza	Antigüedad póliza	3
Zona geográfica	den3	4
	den5	3

Fuente de elaboración propia

Cabe decir, que tanto para el factor de riesgo edad como para la zona geográfica se disponen de dos variables alternativas y excluyentes, cuya selección dependerá de los resultados obtenidos en el posterior análisis predictivo. Por tanto, será necesario calcular diferentes modelos que combinen las diferentes variables para seleccionar aquel que proporcione los mejores resultados.

El procedimiento de selección de variables consiste en replicar el modelo con las diferentes combinaciones de variables, comparar sus resultados y elegir aquellos que proporcionen las mejores estimaciones para nuestro modelo. Este proceso deberá tener en cuenta las siguientes consideraciones que se pasa a comentar a continuación:

- Analizar la significación individual de cada variable. Esto se traduce en la ejecución del modelo para cada una de las variables explicativas, comprobando su significación con un nivel de confianza del 95%. En la práctica, con la

herramienta SAS, este análisis se determina aplicando el test *TYPE3* dentro del procedimiento "PROC GENMOD". Como medidas de ajuste y referencia se van a utilizar la desviación residual y los criterios AIC y BIC. El objetivo de esta fase consiste en seleccionar qué variables se van a utilizar como predictores y el orden en el que deben ser incluidas en el modelo predictivo.

- Método de selección hacia adelante²³. Este enfoque consiste en la elección del mejor modelo estimado de forma secuencial incluyendo una sola variable en cada paso. El proceso comienza incluyendo la variable que presente mayor significación respecto a la variable respuesta. Seguidamente, se van añadiendo una a una todas las variables con el criterio de mayor significación y bajo un efecto de anidación, en el que una vez las variables se han añadido deben ya permanecer en el modelo. El objetivo final es ir incorporando progresivamente las variables al modelo hasta alcanzar los mejores resultados, mediante la mayor significación y unas buenas medidas de ajuste que representen la mejor explicación del modelo.

Las medidas de ajuste utilizadas son la desviación residual, los criterios AIC y BIC, así como la reducción de la desviación residual del modelo. Para esta última medida será necesario calcular el modelo nulo²⁴, como referencia para cuantificar la reducción de la variabilidad del modelo según las variables consideradas.

Es importante decir que no existe un único modelo que sea perfecto, puesto que la combinación de las diferentes variables puede dar lugar a varios modelos con buenos resultados y totalmente válidos para nuestro caso de estudio. En consecuencia, la decisión final debe ir encaminada a elegir aquel modelo que presente una tarifa bien segmentada, de acuerdo a los factores de riesgo y que proporcione la mejor predicción del comportamiento futuro de nuestro colectivo asegurado.

5.2.2. Partición de la base de datos

El análisis del modelo predictivo se va a realizar a partir del planteamiento definido en el punto 3.2.3 del capítulo anterior, que separa la base de datos en dos muestras

²³ Conocido en la literatura anglosajona como *forward-stepwise regression* (Hastie et al., 2008).

²⁴ El modelo nulo considera solo el *intercept*, dejando los coeficientes a cero de las variables explicativas (Hastie et al., 2008).

diferentes, una de entrenamiento y otra de test. La primera de ellas, se va a utilizar para calcular los estimadores de la tarifa del microseguro y la segunda, nos servirá para poder comprobar posteriormente que dicha tarifa estimada es adecuada para otra muestra de datos diferente. Bajo este criterio de validación, se resumen a continuación los datos más representativos de cada una de las muestras que se van a utilizar.

Tabla 5.2. Resumen de muestras para la validación del modelo.

	Muestra de datos	Exposición	Número de siniestros	Coste último total	Suma asegurada	Coste último total por expuesto
N	Entrenamiento	6.633	6.633	6.633	6.633	6.633
Suma	(70%)	3.906,82	178	632.750	1.155.000	1.010.047
N	Test (30%)	2.842	2.842	2.842	2.842	2.842
Suma		1.662,80	65	232.000	460.000	359.936

Fuente de elaboración propia

En la Tabla 5.2 se describe el reparto de la base de datos en las dos muestras que se van a utilizar en el análisis predictivo. La muestra de entrenamiento, con el 70% de los datos, contiene 6.633 asegurados y 178 siniestros, mientras que la muestra de test cuenta con 2.842 asegurados y 65 siniestros. En ambos casos el comportamiento es similar. A la vista de los datos, se puede comprobar como el coste último total por expuesto supera al coste último total, la razón se debe a que el primero de ellos está ponderado por la exposición, que como ya vimos al ser inferior al año repercute en un mayor coste por expuesto. Con vistas a realizar el modelo, se utilizará la muestra de entrenamiento que es la que contiene la mayor parte de información.

5.3 Modelización de la prima pura

El presente análisis de modelos predictivos con GLMs se basa en el modelo Tweedie, ya descrito el punto 5.1, y se realiza siguiendo el planteamiento de validación con la muestra de entrenamiento del 70%. El procedimiento se inicia con un análisis univariante, seguidamente en las siguientes etapas se van a ir incorporando una a una las variables explicativas con la idea de ir mejorando la explicación del modelo.

Paso 1 - Análisis univariante.

Este primer paso comienza aplicando el modelo Tweedie individualmente para cada una de las variables explicativas de la Tabla 5.1. Para seleccionar los predictores

definitivos del modelo se establece el criterio de mayor significación de variables, fijando un umbral máximo de 0,02 para cada una de ellas. Este umbral es superior al nivel de confianza del 95% establecido para el modelo, la idea radica en seleccionar de inicio aquellos predictores con mayor capacidad predictiva con la idea de reforzar la significación de nuestro modelo de partida. De esta forma, todos aquellos predictores que superen el umbral establecido serán excluidos directamente del análisis.

Tabla 5.3. Análisis univariante de los predictores.

Características	Variables	Desviación escalada	AIC	BIC	Significatividad (p-valor)
Asegurado	age2	911	923	964	<0,0001
	age5	912	920	947	<0,0001
	estado civil	922	926	940	0,0921
	sexo	921	925	939	0,0249
Póliza	antigüedad póliza	903	909	930	<0,0001
Zona geográfica	den3	880	888	915	<0,0001
	den5	881	887	907	<0,0001

Fuente de elaboración propia

A la vista de los resultados de la Tabla 5.3, se observa lo siguiente. Las variables *estado civil* con 0,0921 y *sexo* con 0,0249 superan el umbral de significación, que se ha establecido en 0,02, lo que implica su directa eliminación como variables explicativas del modelo predictivo. Respecto al resto de predictores, todos ofrecen un nivel significación inferior a 0,0001, lo que conlleva su consideración como variables explicativas dentro del modelo.

Con vistas a la selección de las variables alternativas de la edad y zona geográfica, se aprecia que tanto *age5* como *den5* obtienen menores valores en las medidas AIC y BIC frente a *age2* y *den3* respectivamente. Por el contrario, la desviación sí es ligeramente mayor en ambos casos. De acuerdo con los resultados de este análisis univariante, aún no es posible decidir cuál de ambas opciones correspondientes a los factores de edad y zona geográfica van a proporcionar mejores modelos. Por tanto, la decisión será que ambas variables permanezcan en las posteriores etapas del análisis predictivo.

Paso 2. Análisis bivariante.

En esta siguiente etapa se va a incorporar una segunda variable explicativa que proporcionará las diferentes combinaciones de modelos. El criterio de incorporación de

variables sigue un orden lógico, en función de la mayor significación de los predictores resultante del anterior análisis univariante. A continuación, se analizan los resultados obtenidos en las diferentes combinaciones de variables explicativas.

Tabla 5.4. Análisis bivariante de los predictores.

Variablen	Desviación escalada	AIC	BIC	D ²	Variable1 (p-valor)	Variable2 (p-valor)
–	923	925	932	–	–	–
den5 antig_poliza	854	864	898	7,4%	<0,0001	<0,0001
den5 age5	864	876	917	6,3%	<0,0001	<0,0001
den5 age2	864	880	934	6,4%	<0,0001	<0,0001
den3 antig_poliza	853	865	905	7,6%	<0,0001	<0,0001
den3 age5	864	878	925	6,4%	<0,0001	<0,0001
den3 age2	862	880	942	6,5%	<0,0001	<0,0001
antig_poliza age5	898	910	950	2,7%	<0,0001	0,0003
antig_poliza age2	898	914	968	2,7%	<0,0001	0,0031

Fuente de elaboración propia

La Tabla 5.4 muestra los resultados de las diferentes pares de combinaciones con sus correspondientes medidas de ajuste, AIC, BIC, D^2 así como el p-valor que analiza la significación estadística de las variables. En la primera fila, aparecen los resultados correspondientes al modelo nulo, sin incluir ninguna variable, y que nos sirve de referencia para estimar la medida D^2 que estima la reducción de variabilidad del modelo.

La mejor combinación corresponde a las variables *antigüedad de la póliza* y al factor de riesgo zona geográfica, representado tanto por *den3* como *den5*. Esta combinación presenta los mínimos valores de desviación, AIC, BIC y D^2 que suponen un 7,6% y un 7,4% respectivamente. En cuanto al resto de modelos, los resultados muestran que la combinación entre los factores de riesgo zona geográfica y edad también proporcionan buenos resultados. A la cola estaría la combinación entre la *antigüedad de la póliza* y el factor de riesgo edad, con sus dos posibles variables alternativas, *age5* o *age2*, por la menor reducción de variabilidad respecto al modelo nulo con un 2,7% y los valores más altos de las medidas AIC y BIC.

En cuanto a la selección de los factores de riesgo edad y zona geográfica, ambas alternativas proporcionan resultados similares. Mientras *age5* y *den5* presentan menores medidas de AIC y BIC, *age2* y *den3* son las que presentan mejores

resultados relacionados con la desviación y la medida D^2 . En vista de estas observaciones, ambas variables van a continuar en el análisis y seguirán combinándose en las siguientes fases del modelo hasta tener argumentos suficientes para una correcta selección.

Paso 3. Análisis multivariante.

Finalmente, en esta etapa se añade la tercera y última variable explicativa a los mejores modelos obtenidos en el análisis bivariante. Después, se pasa a analizar los resultados obtenidos al realizar las diferentes combinaciones de modelos multivariantes.

Tabla 5.5. Análisis multivariante de los predictores.

Variables	Desviación escalada	AIC	BIC	D^2	Variable1 (p-valor)	Variable2 (p-valor)	Variable3 (p-valor)
den5 antig_poliza age5	845	861	916	8,4%	<0,0001	<0,0001	0,0006
den5 antig_poliza age2	844	864	932	8,5%	<0,0001	<0,0001	0,0027
den3 antig_poliza age5	844	862	923	8,6%	<0,0001	<0,0001	0,0007
den3 antig_poliza age2	843	865	940	8,6%	<0,0001	<0,0001	0,0033

Fuente de elaboración propia

La Tabla 5.5 recoge cuatro posibles modelos con tres factores de riesgo disponibles, la zona geográfica, la *antigüedad de la póliza* y la edad. En todos ellos, las tres variables son significativas como se aprecia en las tres últimas columnas. Aunque las dos posibles variables del factor edad, *age2* y *age5*, presentan una menor representatividad como puede comprobarse con sus p-valores más altos en la última columna. Esto puede ser un indicio de que no exista significación en todos los niveles de la variable edad, lo que denota que para dichos niveles los estimadores estén fuera del intervalo de confianza.

Desde el punto de vista de las medidas de ajuste, la desviación, el AIC, y D^2 son bastante similares en los distintos modelos. El criterio de Akaike presenta sus valores más bajos cuando la variable edad es *age5*. Por otra parte, la medida D^2 es ligeramente superior al considerar como factor geográfico la variable *den3*. En cambio, la medida BIC presenta más variabilidad en los resultados, decantándose por las variables *age5* y *den5* frente a *age2* y *den3*.

En vista de los resultados y las observaciones realizadas se puede confirmar que, en base a la información disponible, los cuatro modelos presentan buenos resultados y

explican bien el comportamiento de la variable respuesta. Ahora bien, partiendo de este análisis multivariante es posible seleccionar cuál/cuáles de los modelos presentan mejor el comportamiento futuro de los siniestros del microseguro analizado. De todos estos comentarios descritos en este análisis se deducen varios argumentos que confirman la menor representatividad de la variable *age2* como factor de riesgo edad. Por tanto, a partir de ahora se excluirán los dos modelos que contienen esta variable *age2* con el fin de ir acotando la selección del modelo óptimo.

Después de realizar todo este análisis predictivo y teniendo en cuenta la limitación de información de nuestro caso de estudio, se ha creído oportuno realizar nuevas pruebas para detectar la existencia de posibles variables omitidas en el modelo u otras ya existentes pero que puedan presentar una forma funcional incorrecta o recoger efectos no contemplados en las variables ya analizadas. Estas pruebas han consistido en utilizar y transformar los predictores ya seleccionados, con la idea de confirmar que los modelos obtenidos representan la mejor opción posible en base a la información disponible. Las opciones contempladas se describen a continuación:

- Modificación del tipo de variable: consiste en considerar los factores de riesgo edad y zona geográfica como variables continuas (*age* y *density*).
- Función polinómica: se incluye la variable continua *densidad* como una función polinómica de término cuadrático.

En ambos casos, los resultados obtenidos han sido buenos aunque no alcanzaban a mejorar los modelos ya analizados previamente. Se concluye por tanto, que en base a la información disponible, los predictores seleccionados no solo representan la mejor alternativa para los factores de riesgo, sino que además son capaces de explicar el comportamiento de nuestra variable respuesta.

5.4 Comparación de los modelos seleccionados

A partir del análisis predictivo realizado basado en el modelo Tweedie, se han obtenido finalmente dos modelos multivariantes que representan la mejor explicación del comportamiento de la variable respuesta. Profundizando en estos dos modelos, se procede ahora a completar su información con la estimación de los parámetros de la distribución Tweedie y los correspondientes estimadores de cada modelo.

Tabla 5.6. Análisis comparativo de modelos: parámetros y medidas.

Modelos	Variabes	Parámetro potencia β	Parámetro dispersión ϕ	Parámetro escala	Desviación escalada	AIC
1	den5 antig_poliza age5	1,1765	1.590	39,88	845	861
2	den3 antig_poliza age5	1,1758	1.593	39,91	844	862

Fuente de elaboración propia

Según se observa en la Tabla 5.6, ambos modelos comparten los mismos factores de riesgo para la edad (*age5*) y la *antigüedad de la póliza*, salvo en la zona geográfica, que según la segmentación realizada en el punto 4.5 se disponen de dos posibles alternativas (*den3* y *den5*). Es de notar, que los dos modelos presentan resultados muy similares, tanto en sus parámetros como en sus medidas de bondad de ajuste, lo que hace más difícil la elección del modelo óptimo. Está claro entonces que cualquiera de ellos podría utilizarse como modelo predictivo para estimar la prima futura de nuestro microseguro, aunque la idea es tratar de realizar una serie de cálculos que faciliten la toma de decisión a la hora de elegir el modelo óptimo.

Por otra parte, los resultados del modelo GLM proporcionan los estimadores β de los parámetros, asignando las relatividades correspondientes a cada una de las variables que permiten estimar las primas de riesgo del microseguro. A continuación, se muestran los resultados para cada uno de los dos modelos seleccionados.

Tabla 5.7. Resultados del modelo 1: análisis de los estimadores.

Parámetros	Nivel	Estimador β	Relatividad e^β	Límites de confianza (95%)		Pr>ChiSq
Intercept		4,5285	92,62	4,1353	4,9217	<0,0001
den5	0,036	-1,8087	0,16	-2,2539	-1,3635	<0,0001
den5	0,293	0,0000	1,00			
den5	0,592	-2,5151	0,08	-3,3293	-1,7009	<0,0001
antig_poliza	0	0,0000	1,00			
antig_poliza	1	0,8509	2,34	0,4194	1,2824	<0,0001
antig_poliza	2	1,4270	4,17	1,0107	1,8434	0,0001
age5	1	0,0000	1,00			
age5	2	0,8070	2,24	0,4197	1,1943	<0,0001
age5	3	0,5357	1,71	-0,0807	1,1521	0,0885
age5	4	0,5745	1,78	-0,2936	1,4425	0,1946

Fuente de elaboración propia

Tabla 5.8. Resultados del modelo 2: análisis de los estimadores.

Parámetros	Nivel	Estimador β	Relatividad e^{β}	Límites de confianza (95%)		Pr>ChiSq
Intercept		4,5271	92,49	4,1346	4,9197	<0,0001
den3	0,021	1,2243	3,40	0,0438	2,4049	0,0421
den3	0,036	-1,7864	0,17	-2,2322	-1,3406	<0,0001
den3	0,272	0,0000	1,00			
den3	0,592	-2,4925	0,08	-3,3073	-1,6777	<0,0001
antig_poliza	0	0,0000	1,00			
antig_poliza	1	0,8258	2,28	0,3930	1,2587	0,0002
antig_poliza	2	1,4141	4,11	0,9978	1,8305	0,0001
age5	1	0,0000	1,00			
age5	2	0,7966	2,22	0,4097	1,1836	<0,0001
age5	3	0,5251	1,69	-0,0900	1,1402	0,0943
age5	4	0,5764	1,78	-0,2894	1,4423	0,1920

Fuente de elaboración propia

Cada una de las salidas de los modelos GLM recoge en cada columna: los estimadores β para cada predictor, sus relatividades asociadas al hacer la correspondiente transformación exponencial, sus intervalos de confianza al 95% y el p-valor asociado que mide la significación estadística de cada estimador. Los niveles base de los modelos se identifican con los niveles de cada variable que presentan mayor exposición y se identifican claramente cuando su $\beta = 0$. En concreto, para los factores de riesgo de los dos modelos seleccionados, corresponde a la edad con *age5* el nivel 1, a la *antigüedad de la póliza* el nivel 0. Para el caso de la variable densidad, el modelo 1 con *den5* cuenta con el nivel base de 0,293 y el modelo 2 con *den3* tiene un nivel base de 0,272.

La diferencia entre ambos modelos viene determinada por el predictor que describe el factor de riesgo zona geográfica, *den5* para el modelo 1 y *den3* para el modelo 2. Aunque en los dos modelos la variable densidad es estadísticamente significativa, en el modelo 1 todos los niveles presentan un p-valor<0,0001, mientras que en el modelo 2 uno de los niveles presenta menor significatividad acercándose al umbral de 0,05. Respecto a la variable *age5*, y tal como ya vimos, se consideraba un predictor significativo, pero esto no sucede en todos sus niveles. En ambos casos, la variable edad presenta dos niveles con un p-valor por encima del nivel de significación del 0,05, lo que implica que el valor cero está incluido dentro del intervalo de confianza cuando este debería quedar fuera del mismo.

El paso siguiente, consiste en calcular la prima de riesgo correspondiente a cada modelo multiplicativo combinando los diferentes estimadores β de los predictores y según los factores de riesgo de cada asegurado. Se define μ_i como el predictor lineal de una observación i , aplicando la siguiente fórmula:

$$\mu_i = \log(E[C]) = \beta_0 + \sum_{j=1}^3 \beta_{1j} \cdot d_j + \sum_{j=1}^3 \beta_{2j} \cdot a_j + \sum_{j=1}^4 \beta_{3j} \cdot n_j \quad (5.1)$$

Siendo C la variable respuesta, coste último total por expuesto. β_0 es el término independiente, β_{1j} los estimadores de la densidad representados por cualquiera de las dos variables de los modelos correspondientes a la zona geográfica, *den5* o *den3*. β_{2j} los estimadores de la variable *antigüedad de la póliza* y β_{3j} los correspondientes a la edad del asegurado. Los coeficientes $d_j, a_j, y n_j$ pueden tomar los valores de (0,1) en función del perfil de riesgo del asegurado asociado al nivel de cada variable explicativa. El logaritmo de la variable respuesta, determinado por la función de enlace, es igual al producto de las relatividades asociadas a los estimadores de las variables explicativas de cada asegurado.

5.4.1. Análisis de equilibrio financiero

Hasta ahora, todo el análisis predictivo de modelos se ha basado en los resultados de las medidas de bondad de ajuste. El siguiente paso va a consistir en complementar el estudio con nuestra herramienta de decisión que nos permitirá seleccionar el modelo óptimo. Este estudio trata de realizar un análisis de equilibrio financiero de los modelos, que comprende los siguientes cálculos: el análisis de la variabilidad de primas, la estimación del error cuadrático medio y el análisis del ratio de siniestralidad. Además, como parte del proceso de validación ya comentado, todo este estudio se va a realizar con la muestra test que corresponde con el 30% de los datos.

1. Análisis de variabilidad de primas

Lo primero de todo, es necesario calcular las primas μ_i a partir de los estimadores β correspondientes de cada modelo, tal y como se describía anteriormente en la fórmula (5.1). Una vez calculadas las primas μ_i , se obtiene la variable objeto de análisis en este cálculo que será la prima ponderada por la suma asegurada A , es decir μ_i/A .

Esta variable supone el coste unitario del producto por cada unidad de suma asegurada, y tiene la ventaja de ser una medida objetiva que puede ser comparada con otros productos similares. La suma asegurada A del producto es constante y común para todos los asegurados con un valor de 15.000Php. A continuación se resume un análisis descriptivo de esta nueva variable para conocer su variabilidad y forma.

Tabla 5.9. Análisis comparativo de modelos: variabilidad de primas.

Modelo	Variable	N	Media	Desviación	Varianza	Coefficiente de variación	Curtosis	Asimetría
1	u_i/A	2.842	0,01030	0,01380	0,00019	134,03	4,20	2,17
2	u_i/A	2.842	0,01016	0,01358	0,00018	133,64	4,80	2,23

Fuente de elaboración propia

En la Tabla 5.9 se describen las principales medidas de dispersión y de forma de la variable μ_i/A , que resultan muy similares en ambos modelos. Las medidas de dispersión indican que existe una alta variabilidad en las primas ponderadas por la suma asegurada, al contar con unos coeficientes de variación superiores al 100%. Si se comparan los dos modelos, el segundo de ellos es el que presenta una ligera menor variabilidad debido a sus valores inferiores. Las medidas de forma permiten identificar las características de la distribución de la variable, que en ambos casos presentan una asimetría positiva, lo que denota la presencia de colas a la derecha, siendo superior para el modelo 2. El coeficiente de curtosis, que informa sobre el apuntamiento de la distribución, también es positivo con un valor superior a 3 en ambos casos. Este hecho indica que la distribución de μ_i/A es leptocúrtica, con la presencia de una cola gruesa en ambos modelos, aunque para el modelo 2 los valores son ligeramente superiores a los del modelo 1.

Por tanto, a la vista de estas observaciones se puede afirmar que ambos modelos presentan resultados similares, con una alta variabilidad superior en el modelo 1 y con una asimetría positiva que indica la existencia de una cola gruesa a la derecha, aunque sí más pronunciada para el modelo 2. No se debe olvidar que este análisis se está realizando con la muestra test del 30% de los datos, lo que implica que los resultados se deben ceñir a la validación del modelo pero no hacerlo extensible al total de la cartera.

2. Estimación del error cuadrático medio

En cualquier modelo predictivo donde se realiza la correspondiente estimación de primas, es necesario analizar la diferencia entre el valor estimado y el valor observado a través del error cuadrático medio (ECM). Esta medida nos servirá para conocer la variabilidad asociada a las estimaciones realizadas en los dos modelos propuestos. El cálculo del ECM se hará siguiendo el criterio previamente descrito en el capítulo anterior (véase sección 4.3).

Tabla 5.10. Análisis comparativo de modelos: ECM.

Modelo	ECM
1	0,00173
2	0,00171

Fuente de elaboración propia

En la Tabla 5.10 se recogen los resultados de los dos modelos que presentan un reducido ECM y muy similar para ambos casos. Aparentemente, el modelo 2 presenta un menor error de estimación, lo que indica una ligera mayor eficiencia en su aplicación. Al igual que en el análisis anterior, estos resultados se basan en la muestra test del 30% de los datos, lo que implica que los resultados pueden diferir al hacerlo extensible a la totalidad de la cartera.

3. Análisis del ratio de siniestralidad

El último cálculo que comprende este análisis de equilibrio financiero trata de estudiar el comportamiento derivado del perfil de riesgo y la siniestralidad de la cartera después de obtener las primas de riesgo estimadas por los modelos. El ratio de siniestralidad es una medida objetiva que relaciona la siniestralidad total de la cartera respecto al total de sus primas. Según esto, se presenta a continuación un análisis descriptivo de las primas obtenidas en cada uno de los modelos, calculando su correspondiente ratio de siniestralidad.

Tabla 5.11. Análisis comparativo de modelos: ratio de siniestralidad.

Modelo	Prima pura de riesgo estimada					Coste último siniestros	Ratio siniestralidad
	Media	Suma	Mínimo	Máximo	Rango		
1	154,43	438.889,62	7,49	864,83	857,34	232.000	52,9%
2	152,38	433.061,18	7,65	1.593,92	1.586,27		53,6%

Fuente de elaboración propia

La Tabla 5.11 muestra como las magnitudes descriptivas, relativas a las primas de riesgo estimadas, evidencian la existencia de variabilidad entre sus valores mínimos y máximos. Los dos modelos destacan por tener un amplio rango en las primas estimadas, en las que el modelo 2 casi alcanza duplicar el rango del modelo 1. Esto justifica los resultados obtenidos respecto a la asimetría positiva y a la curtosis vistos anteriormente, que recordamos presentaba valores superiores para el caso 2 (véase Tabla 5.9). También se puede ver como la prima media es superior en el modelo 1, lo que va a suponer una mayor prima total de cartera y en consecuencia esto repercutirá en un menor ratio de siniestralidad, del 52,9% por debajo del 53,6% del modelo 2. Aunque este ratio de siniestralidad presenta valores muy similares en ambos modelos, sí es cierto que un menor valor indica que, para la muestra observada del 30%, la suficiencia de primas es mayor en el caso 1 que en el caso 2.

Finalmente, este estudio comparativo confirma que los dos modelos seleccionados presentan resultados muy parecidos y que cualquiera de ellos puede modelizar perfectamente nuestro microseguro analizado. Ahora bien, habida cuenta de que nuestro análisis presenta una menor variabilidad y un mayor grado de eficiencia estimado por el ECM a favor del modelo 2, sí es cierto que las primas asociadas presentan un rango muy amplio que se verifica por una mayor asimetría y una cola más gruesa de la distribución. Por este motivo, el modelo 1 se convierte así en el mejor modelo para nuestro estudio, justificado tanto por un rango bastante más acotado en las primas estimadas, como por presentar un ratio de siniestralidad inferior. En definitiva, el modelo 1 es la mejor elección para predecir el comportamiento futuro del colectivo asegurado en el microseguro analizado.

5.5 Modelo óptimo

Una vez realizado todo el análisis predictivo de los modelos y comparar las dos posibles alternativas, se ha elegido el modelo que mejor representa el comportamiento futuro de nuestro colectivo asegurado. Dicho análisis ha seguido el planteamiento previamente estipulado en la validación: inicialmente el modelo se ha calculado con la muestra de entrenamiento del 70% de los datos y posteriormente, se han realizado los análisis utilizando la muestra del 30% restante. A partir de aquí, con el modelo ya elegido, es conveniente reconstruirlo de nuevo con toda la base de datos disponible para dotar a las estimaciones con la mayor credibilidad y robustez posible.

Siguiendo esta consideración, se construye de nuevo el modelo con toda la información y bajo las hipótesis inicialmente asumidas. A continuación, vamos a describir todos los resultados obtenidos del modelo óptimo que comprenden: las medidas de bondad de ajuste, los parámetros del modelo lineal general con la distribución Tweedie y los estimadores para determinar la tarifa del microseguro.

Tabla 5.12. Modelo óptimo: medidas de bondad del ajuste.

Variabes	Desviación escalada	AIC	BIC	D ²	Variable1 (p-valor)	Variable2 (p-valor)	Variable3 (p-valor)
–	1.282	1.284	1.291	–	–	–	–
den5 antig_poliza age5	1.192	1.208	1.266	7,0%	0,0001	<0,0001	<0,0001

Fuente de elaboración propia

En la Tabla 5.12 se incluyen los resultados del modelo óptimo, incluyendo también los del modelo nulo (situados en la primera fila) necesarios para estimar la medida de reducción de variabilidad de la desviación. En general, se aprecia como todas las medidas de ajuste calculadas, la desviación, el AIC y el BIC se incrementan respecto a los resultados obtenidos con la muestra de entrenamiento. Por otro lado, D^2 cuantifica en un 7% la variabilidad explicada por el modelo óptimo. Si comparamos esta cifra con los resultados de la muestra anterior, el modelo óptimo presenta una reducción de algo más de un punto al considerar toda la base de datos.

En relación a los parámetros y medidas de nuestro modelo óptimo, se resumen a continuación las principales magnitudes.

Tabla 5.13. Modelo óptimo: parámetros y medidas.

Variabes	Parámetro potencia p	Parámetro dispersión ϕ	Parámetro escala	Desviación escalada	AIC
den5 antig_poliza age5	1,1720	1.659	40,73	1.192	1.208

Fuente de elaboración propia

En la Tabla 5.13 se recogen los resultados del modelo óptimo. Si los comparamos con la muestra de entrenamiento resultan muy similares, salvo por un ligero incremento en los parámetros de dispersión y de escala, así como una pequeña reducción del parámetro potencia p . Este último, se identifica con el tipo de distribución asociada al modelo Tweedie, en este caso es de 1,172 que al situarse entre $1 < p < 2$ se asocia a

la distribución Poisson compuesta con Gamma, caso que permite la estimación de la prima pura de riesgo (véase Tabla 3.2).

El siguiente paso, va a consistir en analizar los estimadores obtenidos en el modelo óptimo. Estos valores van a proporcionar las relatividades que finalmente conformarán la tarifa de nuestro microseguro.

Tabla 5.14. Modelo óptimo: estimadores.

Parámetros	Nivel	Estimador β	Relatividad e^{β}	Límites de confianza (95%)		Pr>ChiSq
Intercept		4,5439	94,06	4,2138	4,8740	<0,0001
den5	0,036	-1,4839	0,23	-1,8336	-1,1342	<0,0001
den5	0,293	0,0000	1,00			
den5	0,592	-2,4563	0,09	-3,1518	-1,7607	<0,0001
antig_poliza	0	0,0000	1,00			
antig_poliza	1	0,7554	2,13	0,3913	1,1194	<0,0001
antig_poliza	2	1,3538	3,87	1,0110	1,7065	0,0001
age5	1	0,0000	1,00			
age5	2	0,7209	2,06	0,3927	1,0490	<0,0001
age5	3	0,5868	1,80	0,0784	1,0951	0,0237
age5	4	0,7924	2,21	0,0964	1,4883	0,0256

Fuente de elaboración propia

Una ventaja importante añadida a los resultados que se obtienen del modelo óptimo, cuando este se hace extensivo a toda la base de datos, es que todos los niveles de las diferentes variables alcanzan el nivel de significación previsto, a diferencia del análisis previo donde los niveles 3 y 4 de la variable *age5* contaban con un p-valor>0,05.

Los valores positivos de los estimadores β , que se observan tanto en la *antigüedad de la póliza* como en la edad (*age5*), indican un incremento de la tarifa respecto a su nivel base, con unas relatividades por encima de uno. En cambio, los valores negativos, como es el caso de la variable densidad (*den5*), producen una reducción en la tarifa respecto al nivel base, con relatividades por debajo de uno.

5.5.1. Análisis de residuos y de valores atípicos

Una vez se ha elaborado y analizado los nuevos resultados del modelo óptimo, sí es necesario realizar otras comprobaciones que permitan garantizar la validez de nuestro modelo. El siguiente estudio tiene como finalidad asegurar que nuestro modelo óptimo responde a la mejor explicación del comportamiento de riesgo de nuestro colectivo en base a la información disponible. Este comprende un doble análisis: por un lado, examinando los residuos y por otro, un estudio de los valores atípicos que pueden llegar a distorsionar las estimaciones del modelo.

Análisis de residuos

Se trata de un criterio útil para conocer en qué grado se ajusta el modelo a los datos analizados, a su vez sirve también para conocer aquella información que puede quedar fuera de la explicación del modelo. La forma de residuo más sencilla es la diferencia entre el valor observado y el valor estimado, $y_i - \hat{\mu}_i$. En consecuencia, cuanto más cercano a cero sea su valor, mejor explicado estará nuestro modelo.

Para realizar el análisis de modelos GLM suele ser más adecuado utilizar los residuos estandarizados, que bajo unas correctas hipótesis del modelo, su varianza debe ser constante. Y bajo estas consideraciones, se han calculado tanto los residuos estandarizados de Pearson, como los de la desviación residual del modelo que suelen ser más eficaces al trabajar con datos reales (Ohlsson & Johansson, 2010).

La herramienta SAS nos facilita la obtención de este tipo de residuos a través de unas sencillas sentencias. Bastaría entonces con incorporar, en las opciones de salida del modelo, la opción *STDRESCHI* para los residuos estandarizados de Pearson, y *STDRESDEV* para los residuos de la desviación residual. El cálculo de estos residuos se estandarizan para tener su varianza asintótica unitaria que se realiza a través de la siguiente formulación (SAS Institute, 2008):

- Los residuos estandarizados de Pearson:

$$r_{pi} = \frac{y_i - \hat{\mu}_i}{\sqrt{v_i(1 - h_i)}} \quad (5.2)$$

Siendo v_i la varianza de la variable respuesta *i-ésima* y h_i se define como el punto palanca.

- Los residuos estandarizados de la desviación residual:

$$r_{Di} = \frac{\text{sign}(y_i - \hat{\mu}_i)\sqrt{d_i}}{\sqrt{\phi(1 - h_i)}} \quad (5.3)$$

Donde d_i es la contribución a la desviación total de la observación i . En el numerador, $\text{sign}(y_i - \hat{\mu}_i)$ toma el valor 1 siempre que $y_i - \hat{\mu}_i$ sea positiva, y el valor -1 en el caso de que sea negativa.

Incluyendo estas opciones dentro de nuestro modelo óptimo se obtiene el siguiente análisis descriptivo de residuos.

Tabla 5.15. Modelo óptimo: análisis descriptivo de residuos estandarizados.

Residuos estandarizados	Media	Varianza	Mínimo	Máximo	Asimetría
Pearson	-0,002	1,263	-0,372	59,862	22,884
Desviación residual	-0,150	0,103	-0,578	5,220	6,140

Fuente de elaboración propia

En la Tabla 5.15 se recogen las dos clases de residuos analizados. Ambos tipos presentan unos resultados parecidos, con unos valores medios ligeramente por debajo del cero, lo que indica que en media las estimaciones están por encima de los valores observados. La varianza indica cierta variabilidad de los residuos, con la presencia de valores atípicos que corroboran tanto sus valores máximos como la asimetría positiva de los residuos. Comparando entre las dos clases de residuos, se observa como los que utilizan la desviación residual minimizan la variabilidad producida por la existencia de los valores atípicos.

Análisis de valores atípicos

En algunos seguros, la existencia de siniestros graves puede repercutir en un incremento sustancial sobre el coste del siniestro. En el caso de los microseguros, aunque este aspecto está bastante más acotado debido a las menores indemnizaciones, es conveniente de igual modo realizar el correspondiente estudio para tener controlado la posible influencia de las observaciones atípicas que puedan alterar de forma significativa los resultados del modelo.

Una magnitud adecuada para medir la posible influencia de estos valores en los modelos lineales generalizados es la distancia de Cook. Esta medida nos dice cómo cambia el vector de estimadores $\hat{\beta}$ cuando se elimina cada observación, así un valor elevado de la distancia de Cook indica un nivel con gran influencia en el modelo (Goldburd et al., 2016). El proceso es sencillo, se ordenan de forma descendente todos los valores de la distancia de Cook, identificando los primeros registros con los valores más altos, que son los que mayor influencia ejercen sobre los resultados del modelo.

En nuestro modelo Tweedie planteado, la macro de SAS que se utiliza nos permite incluir la opción *COOKSD* para calcular la distancia de Cook estandarizada para evaluar la influencia de cada una de las observaciones sobre el ajuste del modelo (SAS Institute, 2008).

De acuerdo con dicho planteamiento, se ha aplicado todo este proceso para nuestro modelo óptimo, estableciendo diferentes escenarios con varios umbrales para identificar los valores atípicos. Con cada escenario se ha generado un posible nuevo modelo. En todos ellos se repetía el mismo patrón, un efecto de mejora sobre los resultados, gracias a la exclusión de los valores atípicos, pero que iba en detrimento de la significación de los niveles de las variables explicativas. Ante esta situación, se considera prioritario disponer de la mayor significación de todos los predictores y en todos sus niveles. Por tanto, la decisión ha sido mantener el modelo óptimo con todas las observaciones, sin excluir ningún valor atípico.

5.6 Análisis comparativo de primas: modelo óptimo vs modelo real

El modelo óptimo obtenido con el análisis GLM basado en el modelo Tweedie nos proporciona las primas puras de riesgo de nuestro microseguro, en función de los tres factores de riesgo, edad, zona geográfica y antigüedad de la póliza. Sin embargo, nuestro interés es ir un paso adelante y comprobar que el modelo óptimo puede realmente aplicarse en la práctica actuarial del negocio microasegurador. En este sentido, se plantea realizar un estudio comparativo entre las primas de riesgo estimadas por nuestro modelo y el plan de tarifas que utiliza actualmente la compañía de seguros. De esta forma, se podrán evaluar los beneficios que realmente aporta la aplicación de nuestro modelo. Este estudio va a consistir en dos etapas: la primera,

será analizar la estabilidad del modelo, y la segunda se evaluará la viabilidad del modelo predictivo frente al modelo actual.

Lo primero de todo, es indispensable conocer el enfoque real que está aplicando la compañía de seguros en la actualidad. El modelo actual aplica un criterio de tarificación sencillo, sin diferenciar por factores de riesgo y utilizando la experiencia de siniestralidad junto con los gastos incurridos de los tres últimos años para realizar así las convenientes estimaciones. En base a estos criterios, se calcula entonces la prima neta y unos recargos, obteniendo una prima recargada única para todos los asegurados. A continuación, se muestra un resumen del plan de tarifas de la compañía para el microseguro combinado, desglosado por cobertura y con los importes en la moneda oficial, pesos filipinos (PhP).

Tabla 5.16. Modelo real: plan de tarifas del microseguro.

Cobertura	Suma asegurada	Prima neta	Gastos	Prima recargada
Hospitalización	10.000	35	15	50
Eventos de la naturaleza	5.000	70	40	110
Total	15.000	105	55	160

Fuente de elaboración propia

En la Tabla 5.16 se recoge el desglose de las primas del microseguro para las dos coberturas analizadas. El término de prima recargada recoge la prima pura más los gastos que contempla la compañía, que asciende a 160PhP para el total de coberturas. Este concepto será el que se tome como referencia para el modelo real, por el hecho de considerar ese margen de seguridad que está contemplando la compañía para garantizar la solvencia del producto. En este sentido, nuestro enfoque metodológico deberá contemplar ese margen de prudencia dentro de las primas estimadas.

5.6.1. Análisis de equilibrio financiero

Al igual que anteriormente se analizó la estabilidad de los diferentes modelos alternativos propuestos (véase punto 5.4.1), es necesario justificar también aquí el equilibrio financiero del modelo óptimo. Siguiendo esta línea, se va a realizar este análisis comparando los resultados obtenidos del modelo óptimo junto con los del

modelo real. El estudio comprende las siguientes partes, el análisis de variabilidad de primas, la estimación del ECM y el análisis del ratio de siniestralidad.

1. Análisis de variabilidad de primas

Consiste en un análisis descriptivo de la prima de riesgo unitaria por suma asegurada, es decir la estimación ponderada por la suma asegurada μ_i/A . En este caso, solo es posible analizar la variabilidad del modelo óptimo, puesto que para el caso real no procede al ser una prima constante y única para todos los asegurados.

Tabla 5.17. Modelo óptimo: análisis de variabilidad de primas.

Variable	N	Media	Desviación	Varianza	Coefficiente de variación	Curtosis	Asimetría
u_i/A	9.475	0,01007	0,01241	0,00015	123,21	3,88	2,11

Fuente de elaboración propia

En la Tabla 5.17 se describen los resultados del modelo óptimo con la muestra total de los 9.475 asegurados. Como es lógico, al aumentar el tamaño muestral la variabilidad se reduce, tal y como muestra el descenso de la varianza que pasa de 0,00019 a 0,00015. Aún así, las medidas de dispersión siguen confirmando una alta variabilidad de μ_i/A con un coeficiente de variación por encima del 100%, concretamente del 123%. De igual manera, las medidas de forma se reducen respecto a la muestra test del 30%, aunque siguen manteniendo su asimetría positiva e igualmente el coeficiente de curtosis continúa con un valor por encima de 3, lo que confirma la existencia de una cola gruesa a la derecha debido a la existencia de primas elevadas.

2. Estimación del error cuadrático medio

El error cuadrático medio mide la variabilidad de μ_i/A respecto a P_i/A ponderado por la exposición (véase fórmula 3.3). Este será utilizado como medida de ajuste en la estimación de primas respecto a la siniestralidad registrada. El ECM se calcula, tanto para el modelo óptimo como para el modelo real de la compañía para poder así comparar sus resultados de acuerdo a la metodología utilizada.

Tabla 5.18. Modelo óptimo vs modelo real: ECM.

Modelo	ECM
Óptimo	0,00182
Real	0,00187

Fuente de elaboración propia

En la Tabla 5.18 se observa que el ECM es bastante similar en ambos casos. Aún así, nuestro modelo óptimo presenta un error ligeramente inferior al plan actual de tarifas de la compañía. Esta diferencia favorece la aplicación de una tarifa basada en modelos multivariantes GLM, al obtener un mejor ajuste de la prima respecto al coste real del siniestro, frente a una prima constante y única para todo el colectivo asegurado.

3. Análisis del ratio de siniestralidad

El último cálculo de este estudio de estabilidad se centra en el cálculo del ratio de siniestralidad. Se trata de una medida de gran utilidad para las compañías porque permite analizar la suficiencia de las primas y otros aspectos relacionados con el comportamiento de la siniestralidad de la cartera (Werner & Modlin, 2016). En particular, aplicado al análisis comparativo que aquí se plantea, el ratio de siniestralidad nos va a servir de referencia para medir el grado de eficiencia del negocio según la metodología empleada en cada uno de los modelos planteados.

Tabla 5.19. Modelo óptimo vs real: análisis del ratio de siniestralidad.

Modelo	Prima pura riesgo estimada					Coste último	Ratio siniestralidad
	Media	Suma	Mínimo	Máximo	Rango		
Óptimo	151	1.431.695	8	804	796	864.750	60,4%
Real	160	1.516.000	160	160	0		57,0%

Fuente de elaboración propia

La Tabla 5.19 recoge un resumen de las primas de riesgo, la siniestralidad final así como el ratio de siniestralidad correspondiente a los dos planteamientos. Nuestro modelo óptimo presenta una prima media de 151Php, inferior a la prima recargada actual del producto de 160Php. Es fácil distinguir en los resultados, como la técnica predictiva de GLMs genera mayor variabilidad en las primas, por sus valores mínimo máximo y su correspondiente rango, respecto al criterio de tarificación actual que muestra una prima plana. En el conjunto de toda la cartera, una prima media inferior

en nuestro modelo supone un encaje total de primas por debajo del obtenido para la compañía. En consecuencia, al calcular el ratio de siniestralidad del modelo propuesto, este se incrementa situándose en un 60,4%, por encima del 57% que presenta la compañía, es decir que existe una diferencia positiva de 3 puntos porcentuales. Para calibrar la importancia de esta diferencia en el negocio microasegurador se acude a fuentes externas del mercado de microseguros.

El panorama de microseguros correspondiente al mercado de Filipinas puede perfectamente equipararse a la información procedente del mercado de microseguros asiático²⁵. Esta información nos proporciona los ratios de siniestralidad para los principales tipos de microseguro en el mercado de Asia y Oceanía.

Tabla 5.20. Mercado de microseguros asiático: ratios de siniestralidad.

Tipo de producto	Ratios de siniestralidad
Vida	90%
Accidentes	121%
Salud	79%
Propiedad	81%
Agrícolas	59%
Total	79%

Fuente: Elaboración propia a partir del estudio de Fundación Munich Re (2013) *Landscape of Microinsurance in Asia and Oceania*. Con el apoyo de GIZ-RFPI y Microinsurance Network.

Los ratios de siniestralidad del mercado asiático, descritos en la Tabla 5.20, se corresponden con los principales productos microaseguradores. A la vista, se observa como todos presentan unos ratios elevados, en su mayoría por encima del 70% y dejando previsiblemente el 30% restante a cubrir los gastos comerciales y de adquisición. Esta situación hace que para algunas modalidades sea difícil su sostenibilidad a largo plazo en el mercado microasegurador (Fundación Munich Re, 2013).

Si nos centramos en nuestro estudio particular, el microseguro que estamos utilizando combina dos coberturas de indemnización, por hospitalización y por daños derivados

²⁵ Informe procedente de Fundación Munich Re (2013) *Landcape of microinsurance in Asia and Oceania*. Microinsurance Network.

de actos de la naturaleza. Ambas se pueden asemejar a dos tipos diferentes de productos que se resaltan en gris en la Tabla 5.20, los relacionados con la salud y la propiedad respectivamente. Para estos productos, sus ratios de siniestralidad son prácticamente similares y en media se situarían en torno al 80%.

Si se contrasta ahora estos datos de mercado con el ratio de siniestralidad estimado en nuestro microseguro, se comprueba que el 60,4% está muy por debajo de la media del sector. Esto concede a la compañía una mayor tranquilidad, al disponer de un margen más amplio para cubrir los gastos correspondientes. Respecto a la diferencia de 3 puntos porcentuales entre los ratios de ambos modelos, lógicamente se confirma que esta no resulta significativa y entra dentro del rango de lo razonable, al disponer de unos ratios de siniestralidad que favorecen la sostenibilidad del microseguro analizado.

5.6.2. Viabilidad del modelo

Otro tema interesante es el estudio de la viabilidad del modelo, que se centra en analizar los resultados desde dos planos diferentes. Por un lado, la habilidad del modelo en adaptarse al perfil de riesgo de la población asegurada y por otro lado, la contribución del asegurado a la estabilidad del producto. Este estudio contribuye a conocer cómo es el perfil de riesgo de la cartera, así como prevenir uno de los mayores riesgos detectados en el negocio microasegurador, la selección adversa.

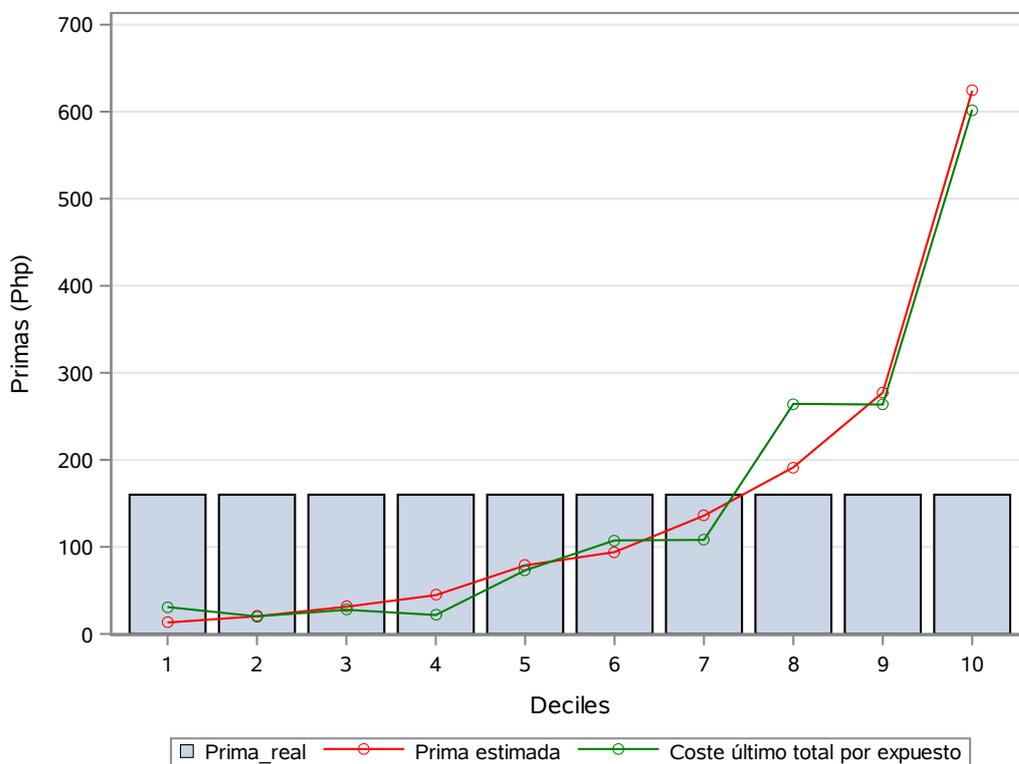
La viabilidad del modelo se va a materializar en la elaboración y análisis de varios gráficos. En primer lugar, el gráfico de elevación, que permite visualizar la capacidad de cada enfoque para proporcionar una tarifa ajustada al riesgo de cada asegurado. Y en segundo lugar, la curva de Lorenz con el índice de Gini para visualizar la distribución de las primas en la cartera.

En primer lugar, el **gráfico de elevación**²⁶ que permite comparar las estimaciones de prima pura con nuestro modelo predictivo Tweedie respecto a la variable observada, el coste último total por expuesto. Como referencia, se tiene en cuenta la tarifa real que aplica la compañía para comprobar las ventajas que ofrece la aplicación de la técnica predictiva basada en modelos GLM.

²⁶ Más conocido en la literatura inglesa como *Lift Chart* (Goldburd et al., 2016).

La construcción del gráfico se realiza de la siguiente forma, primero se ordena de forma ascendente las primas estimadas por nuestro modelo multivariante, y después se agrupan todas estas observaciones ordenadas en deciles, formando 10 grupos homogéneos. A cada grupo de asegurados se le asocian sus correspondientes primas estimadas y su siniestralidad. Gráficamente, los puntos de cada cuantil representan los valores de las dos variables analizadas: la línea trazada en rojo se corresponde con las primas estimadas por nuestro modelo, y la línea verde representa nuestra variable respuesta, el coste último total por expuesto. Las barras grises inferiores reproducen la prima constante y común para toda la cartera que aplica la compañía con su actual modelo.

Gráfico 5.1. Gráfico de elevación de las primas puras del modelo Tweedie.



Visualmente, es fácil comprobar en el Gráfico 5.1, por la proximidad entre las dos líneas, como las estimaciones de prima pura derivadas del modelo Tweedie se ajustan bastante bien a la siniestralidad de la cartera. Además, el gráfico permite identificar la distribución del perfil de riesgo del colectivo, a la izquierda en los primeros deciles se sitúan los riesgos buenos con una siniestralidad baja, y hacia la derecha en los últimos

deciles se encuentran los riesgos malos con una elevada siniestralidad. Este hecho pone de manifiesto la existencia de un rango amplio de primas, que se incrementa de forma considerable a partir del 8º decil por la presencia de valores extremos.

Desde el plan de tarifas real, que viene representado por las barras grises, la prima de 160Php se mantiene constante para todos los perfiles de riesgo. Si se contrasta con la línea verde de la siniestralidad, se observa que para los riesgos buenos la prima actual es más que suficiente para hacer frente a los siniestros. En cambio, para los riesgos malos, a partir del octavo decil, la situación es a la inversa y la tarifa actual se sitúa muy por debajo del coste real del siniestro. Esta situación pone de manifiesto una clara insuficiencia en la tarifa para un perfil de riesgo malo, lo que pondría en peligro la sostenibilidad financiera del microseguro.

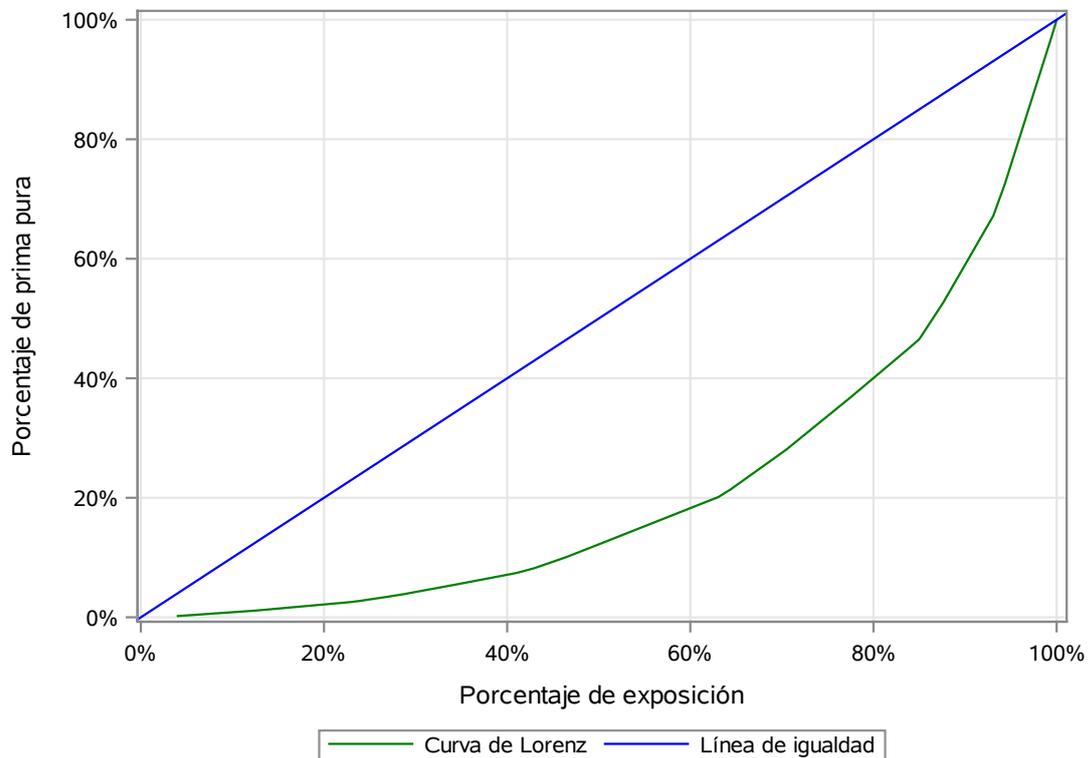
Por todo lo dicho, se confirma que la técnica predictiva del modelo GLM, aquí propuesta, realiza una mejor predicción de las primas de riesgo respecto a la siniestralidad registrada en los diferentes perfiles de riesgo. El modelo actual, basado en una prima constante para todo el colectivo, está utilizando la estrategia de compensar los riesgos buenos y malos para alcanzar la suficiencia del negocio. Sin embargo, este criterio es sostenible siempre y cuando el perfil de la cartera esté concentrado en los riesgos buenos, obviando la posibilidad de un trasvase de asegurados con un peor perfil de riesgo que sin duda pondría en peligro la suficiencia del negocio. Además, esta situación está propiciando claramente la selección adversa, en el que los asegurados buenos acaban anulando la póliza por una prima elevada y los malos permanecen en la cartera. Este aspecto se ha hecho patente en este microseguro analizado, tal y como corrobora el factor de riesgo *antigüedad de la póliza* (véase Tabla 4.5), donde los asegurados con mayor antigüedad de la póliza evidencian una mayor siniestralidad registrada.

El segundo análisis se basa en **el índice de Gini y la curva de Lorenz**. Ambos son indicadores que tratan de reflejar el grado de equidistribución del total de valores de una determinada variable. Se aplican fundamentalmente a variables de tipo económico, como es el caso particular del grado de desigualdad de ingresos de un país (Goldburd et al., 2006). El índice de Gini es la forma porcentual del coeficiente de Gini que se sitúa en el intervalo (0,1): el valor cero corresponde a la igualdad perfecta, cuando todo el mundo tiene los mismos ingresos; y el valor 1 cuando existe plena desigualdad, en el que todos los ingresos pertenecen a una sola persona y el resto no

dispone de ingreso alguno. Este índice cobra gran utilidad cuando se compara con diversos países, otros productos o se analiza su evolución en el tiempo.

A partir de este índice, se realiza el gráfico de la curva de Lorenz. Esta representación consiste en la concentración de la distribución de la variable objeto de estudio, que para nuestro caso particular va a ser la contribución que realizan los expuestos por cada nivel de prima estimada, en términos porcentuales. Esta curva, representada en color verde, se compara respecto a la llamada línea de igualdad, representada en azul, en la que todos los asegurados aportan la misma prima.

Gráfico 5.2. Curva de Lorenz: modelo óptimo vs modelo real.



En el Gráfico 5.2 se distingue por un lado, la curva de Lorenz, que representa la distribución de la prima estimada por nuestro modelo Tweedie respecto a la exposición de la cartera. Y por otro lado, la línea de igualdad, que se identifica claramente con el modelo actual de la compañía, en el que todos los asegurados aportan la misma prima de 160Php. Atendiendo al Gráfico, nuestro modelo explica que el 60% de la exposición contribuye con casi el 20% de la prima total de la cartera. Esto indica que en su mayoría los asegurados aporten unas primas bajas y por debajo de la prima real. Este

hecho está condicionado por la baja siniestralidad del colectivo, que hace que nuestro modelo estime unas primas menores para la mayoría de la cartera. Sin embargo, el modelo actual de la compañía, representado por la línea de igualdad azul, hace que todos los asegurados contribuyan con la misma prima sin diferenciar por su perfil de riesgo.

Al comparar el índice de Gini para los dos modelos, se obtiene que para nuestro modelo óptimo este valor es de un 56,76%, ya que la prima varía según el perfil de riesgo del asegurado, mientras que para el modelo actual es de cero al no discriminar por perfil de riesgo.

En resumen, debe quedar claro que ambos indicadores no se utilizan para cuantificar la rentabilidad del producto, sino que sirven para medir la capacidad y habilidad del modelo en gestionar los diferentes perfiles de riesgos. De esta forma, se pone a disposición de la compañía una herramienta adicional para realizar diversas funciones, como las relacionadas con la gestión de riesgos, la adecuación del plan de tarifas, así como aplicar determinadas políticas de suscripción que garanticen la rentabilidad y sostenibilidad del negocio.

CONCLUSIONES

Llegados al final de nuestro recorrido expositivo, es el momento de resumir en breve las conclusiones que se desprenden de la investigación realizada. Para mayor claridad, estas se han organizado atendiendo a la misma secuencia de la exposición y con especial referencia a los objetivos propuestos. Con ellas se pretende describir los principales resultados obtenidos del trabajo como una estimable contribución a la investigación de los microseguros, sin dejar de señalar también aquellas limitaciones detectadas en el camino. Limitaciones que, dicho sea de paso, nos indican que es un campo donde todavía existen muchas zonas por explorar. Y para dar continuidad, a partir de las aportaciones de esta tesis se abre la posibilidad de iniciar nuevas líneas de investigación que pueden ser el punto de partida para futuros trabajos.

1. Respecto al marco teórico bajo un enfoque social

- Una primera conclusión se refiere al papel de los microseguros como objetivo de desarrollo sostenible -

El enfoque social y humano toma cada vez más fuerza en la sociedad debido al impulso dado por los organismos internacionales. El más relevante, el de Naciones Unidas, que establece entre sus objetivos de desarrollo la promoción de un crecimiento económico sostenible e inclusivo centrado en la población más vulnerable. Lo que significa una apremiante llamada a las instituciones financieras para que asuman su responsabilidad desarrollando productos que amplíen el servicio y el acceso a este nicho de mercado. Y justamente aquí aparece con especial protagonismo el microseguro como producto inclusivo, que responde al objetivo social y humano de integrar al segmento de población más vulnerable poniendo a su alcance la protección que necesitan.

Más aún, entre las características del microseguro está la de contribuir al desarrollo económico y cultural de los países emergentes. Unas veces, fomentando su educación financiera que, aparte de otros beneficios, les capacita para conocer mejor el producto a la vez que incentiva su decisión de compra. Otras veces, identificando las necesidades y preferencias del colectivo con medidas encaminadas a su bienestar tanto familiar como social, cuya aplicación contribuirá a reducir la desigualdad endémica que se observa en los colectivos más vulnerables como son las mujeres y los niños. Por todo ello, está claro que el microseguro se perfila como un producto

inclusivo que, además de cumplir con el objetivo social de integrar en el mercado a la población de bajos ingresos, aporta el valor añadido de proporcionar un mayor bienestar entre la población más desfavorecida.

- Una segunda conclusión responde a la cuestión: ¿cómo de diferente es el microseguro respecto al seguro tradicional? -

La principal diferencia del microseguro con respecto al seguro tradicional viene determinada por el segmento de población objetivo al que va dirigido, que es la población de bajos ingresos. Por tanto, la cobertura debe identificarse con las necesidades de esta población vulnerable que cuenta con un perfil de riesgo diferente al del mercado convencional. Por eso mismo, en el producto debe prevalecer el criterio de sencillez y flexibilidad con el fin de facilitar a este segmento de población el acceso a dicho producto, mientras que en el seguro tradicional los productos son más complejos y con coberturas más amplias.

Pese a estas diferencias, tampoco son pocas las similitudes que existen entre ambos productos. Desde el punto de vista regulador, el microseguro debe respetar también los principios que rigen el seguro y cumplir con la normativa vigente que por defecto se equipara a la del seguro tradicional. Respecto a la tipología de productos, el microseguro sigue líneas de negocio paralelas a las existentes en el seguro tradicional. No obstante, hay aspectos diferenciadores que deben mantenerse debido al público objetivo al que el producto va dirigido, lo que se traduce en coberturas más básicas y sencillas respecto a la mayor complejidad del seguro convencional.

En definitiva, el microseguro comparte muchas de las características del seguro tradicional pero su finalidad de protección social enfocada a las personas más vulnerables le otorga un enfoque social e inclusivo no existente en el seguro tradicional. Ahora bien, partiendo de las similitudes existentes en ambos y con la finalidad de aprovechar el conocimiento y la experiencia del seguro tradicional, es lógico plantear la utilización de las sinergias propias del seguro para favorecer todas aquellas áreas susceptibles de mejora en el microseguro.

2. Respecto al marco teórico bajo un enfoque asegurador

- La tercera conclusión se centra en la creación de un marco regulatorio del seguro inclusivo en aras de un crecimiento del negocio microasegurador -

El desarrollo de cualquier actividad de negocio, y en particular la de los microseguros, está supeditada a la legislación vigente del país. Pero ocurre que la situación en los países emergentes, donde se desarrolla principalmente este negocio, se caracteriza por una generalizada falta de recursos, lo que supone una seria dificultad a la hora de intentar acceder a estos segmentos de población tan vulnerable. Y en consecuencia, también se hace más difícil la creación de una regulación específica para este negocio, y se opta por la fácil solución de mantener vigente un modelo único igual para todos, que no lleva sino a desincentivar el desarrollo del mercado de microseguros. Es de notar, que bajo este entorno, existe un número significativo de organizaciones informales que operan fuera del ámbito legal, motivado en parte por las exigencias regulatorias de estos países. Pese a todo, es alentador el hecho de que, dentro de este desfavorable contexto, se cuenta con la iniciativa de la Asociación Internacional de Supervisores de Seguros (IAIS), que se encarga de promover de forma responsable un marco de regulación y supervisión estable y fuerte para este seguro inclusivo, con el fin de favorecer el crecimiento y desarrollo sostenible del mercado microasegurador.

- La cuarta conclusión se orienta a la mejora del diseño del microseguro desde el punto de vista de la tarificación -

En general un producto bien diseñado es garantía de éxito tanto para el cliente como para la compañía. Es entonces cuando se logra un perfecto equilibrio en las necesidades detectadas en la demanda y las identificadas en las coberturas del producto ofertado. Un equilibrio que no siempre se consigue, como se ve por los frecuentes problemas observados en este mercado (Biener, 2013; Biener & Eling, 2012; Purcal, 2012), que ponen de manifiesto la necesidad de una mayor investigación que asegure la estabilidad del negocio microasegurador. Bajo esta premisa, se distinguen dentro del ámbito actuarial diferentes áreas de mejora en la creación y desarrollo de los microseguros, tanto en la gestión y control de riesgos como en la búsqueda de una tarificación que contribuya a la sostenibilidad financiera del producto, que es precisamente la piedra angular de esta tesis.

Es en el área de gestión y control de riesgos donde se encuentra el problema de la asimetría de información. Así tenemos que la selección adversa y el riesgo moral ponen al descubierto los múltiples casos de uso excesivo del producto, que se acentúa especialmente entre la población de bajos recursos. Sin olvidar el tan extendido uso fraudulento del producto. Este punto requiere, por parte de las compañías, un especial seguimiento para evitar que el sobre coste asociado a esos riesgos repercuta de forma directa, bien que proporcional, en un aumento de las primas.

En el área de la tarificación, se aglutinan diversas limitaciones que obstaculizan el buen ejercicio actuarial en la estimación de primas. La principal restricción viene por la escasez de información que rodea a este negocio, que está marcada por la propia novedad del producto, todavía con una corta experiencia y una consecuente baja exposición en un mercado donde ya de por sí los recursos son escasos. Otra barrera a considerar es el tipo de población objetivo al que va dirigido el microseguro, que condiciona en gran medida aspectos como el perfil de riesgo, la capacidad de pago y la forma de realizarlo, así como otras características en el diseño del producto.

Por todo lo dicho, es comprensible que en la práctica la actual tarificación de microseguros sea muy básica y esté alejada del perfil de riesgo del colectivo. Con la opción inevitable y nada deseable de suplir esa falta de conocimiento e incertidumbre del negocio con unos elevados márgenes de seguridad dentro de la tarifa. Obviamente, esta medida debiera entenderse como una solución temporal en espera de tener un mayor conocimiento del negocio para poderlo trasladar de forma adecuada a una correcta tarificación que contemple fielmente el riesgo del colectivo asegurado.

La contribución de esta investigación se encuadra dentro de una de las áreas de mejora mencionadas en el apartado sobre el diseño de los microseguros, tales como la de promover una tarificación más dinámica que se adapte a las necesidades reales del mercado y la de lograr así la sostenibilidad financiera del producto. Además, con ello se plantea dar solución al problema de incertidumbre asociado a la tarificación actual de microseguros, que en cierta medida obliga a establecer recargos para garantizar la suficiencia de las primas. Para materializar este objetivo, se ha propuesto aplicar una metodología más precisa, ya de uso extendido en el seguro tradicional, que proporcione una tarifa ajustada al perfil de riesgo asegurado y en consecuencia permita eliminar los recargos de seguridad derivados de la incertidumbre del negocio.

3. Respecto al objetivo general

- La quinta conclusión se centra en “conocer la viabilidad de la aplicación de los modelos lineales generalizados (GLMs) a un programa de microseguros, con la idea de obtener unas primas de riesgo más ajustadas a la verdadera distribución de riesgo, de manera que sean más efectivas en este mercado emergente”.-

A lo largo del presente trabajo se ha hecho patente la necesidad de ahondar en un mayor conocimiento sobre el comportamiento del riesgo asegurado en el negocio microasegurador, convirtiéndose en un auténtico reto el desarrollo de la tarificación de estos productos. La aplicación de una metodología más sofisticada proporciona la herramienta eficaz para alcanzar esta meta. En esta materia, se plantea aprovechar la experiencia acumulada en el seguro tradicional como fuente de conocimiento de las diversas metodologías que se utilizan en el cálculo de las primas de seguros y que, en cierta medida, son fácilmente trasladables al microseguro.

A partir de aquí, surge el planteamiento metodológico utilizado para esta investigación que se fundamenta en el análisis predictivo de los modelos lineales generalizados con la distribución Tweedie. Esta técnica proporciona un modelo predictivo capaz de evaluar y cuantificar la relación entre la variable respuesta, que representa nuestro objetivo en el modelo, y las variables explicativas que conforman los factores de riesgo de los asegurados. La ventaja de esta metodología es que permite realizar un estudio pormenorizado del perfil de riesgo del colectivo asegurado, seleccionando los factores de riesgo más influyentes y ajustando el modelo a una determinada distribución que refleje correctamente el comportamiento del riesgo asegurado. Los resultados obtenidos proporcionan las relatividades de una tarifa multivariante capaz de adaptarse fácilmente en función del perfil de riesgo individual de cada asegurado.

La investigación realizada demuestra cómo es posible aplicar esta técnica predictiva, basada en GLMs, para obtener una tarifa más ajustada al perfil de riesgo correspondiente a nuestro caso de estudio. Es de notar, que el estudio se ha realizado con un caso real de un microseguro combinado que se comercializa actualmente en el mercado de Filipinas. Basándonos en dicho producto nos hemos centrado en analizar las dos coberturas de indemnización, una por hospitalización y la otra por los daños derivados de eventos de la naturaleza. La selección de este producto viene marcada

por el hecho de ser el colectivo más numeroso y con mayor experiencia en la compañía, y además porque se trata de un producto combinado con diversas coberturas que suele ser habitual encontrar en el mercado microasegurador.

Entre las dificultades que ha habido que superar para la realización de este trabajo, las más complicadas han venido por el lado de la información utilizada. El proceso de tratamiento y depuración de datos ha sido largo y tedioso hasta lograr disponer de una base de datos con la suficiente calidad y claridad para poder aplicar los modelos predictivos. En cuanto a la estructura de datos, es importante destacar la falta de conexión inicial entre la información de los asegurados y la de la siniestralidad registrada, lo que ha obligado a crear un campo clave para su correcta identificación. Adicionalmente, la escasez de información disponible ha propiciado la creación de otras variables exógenas, recurriendo a otras fuentes, tanto internas como externas. En definitiva, se puede afirmar que, a pesar de todas las limitaciones de información encontradas, ha sido posible establecer de forma adecuada todas las hipótesis del modelo. Hipótesis que van desde elegir la distribución de la variable respuesta, pasando luego a seleccionar y transformar las variables explicativas, hasta conseguir ajustar el modelo de forma que proporcione los parámetros necesarios para estimar las primas del microseguro.

Con todo lo dicho, se demuestra que una metodología predictiva de uso extendido en el mercado asegurador, como son los modelos GLM, es lo suficientemente versátil como para poder aplicarse a un caso real, como es nuestro microseguro combinado, y que por extensión, podría ampliarse su aplicación a otros microseguros que actualmente están presentes en distintos países emergentes.

4. Respecto a los objetivos específicos

- La sexta conclusión se refiere al objetivo de "comprobar que el modelo predictivo Tweedie es la mejor alternativa de la metodología de tarificación con GLMs para nuestro caso de estudio". -

La tarificación a priori, sobre la que se fundamenta este trabajo de investigación, depende principalmente de la experiencia acumulada del colectivo y por tanto de la calidad y cantidad de datos disponibles. De hecho, es esta información la que va a condicionar la elección entre las posibles técnicas de tarificación, que básicamente se

engloban en dos grupos diferenciados. En el primero, se dispone de la suficiente desagregación de datos para realizar un cálculo separado aplicando dos modelos de regresión, un modelo Poisson con la variable número de siniestros, y otro modelo Gamma para la cuantía media del siniestro. Esta opción es la que prefieren utilizar las compañías de seguro tradicional ya que proporciona más información sobre el comportamiento de riesgo individual de cada variable. El segundo grupo atiende a una restricción de datos que dificulta la desagregación de la información y solo permite la aplicación de una única regresión para la variable siniestralidad total. Esta situación es la que viene produciéndose en el mercado de los microseguros.

Desde el punto de vista de nuestro microseguro, la información disponible presenta una serie de rasgos característicos que van a condicionar la metodología de aplicación. La mayor dificultad viene por el elevado porcentaje de ceros presentes en la muestra. En concreto, el 97% de los asegurados no registran ningún siniestro durante el período considerado. Atendiendo a la forma de nuestros datos, el coste último total por expuesto, que es nuestra variable respuesta, presenta una clara asimetría positiva que indica la existencia de colas a la derecha. Además, existe una falta de normalidad en la distribución de la variable respuesta que acota las posibilidades de aplicación de cualquier metodología.

Bajo estas limitaciones, las posibles técnicas de tarificación ofrecen dos posibles modelos predictivos para estimar las primas de riesgo de nuestro microseguro. El primero de ellos es el de la regresión censurada propia del modelo Tobit. Esta opción aunque sí considera la principal restricción de nuestro caso de estudio, que trata del elevado número de ceros de la variable dependiente, sin embargo, se ha descartado por la existencia de la asimetría positiva y por la falta de normalidad presente en nuestros datos, y que no contempla este modelo de regresión.

El segundo modelo consiste en el modelo GLM basado en la distribución Tweedie que pertenece a la familia exponencial de distribuciones. Si nos centramos en el caso particular de la distribución Poisson compuesta con Gamma cuando el valor del parámetro $p \in (1,2)$, nos permite modelizar directamente las primas de riesgo sin necesidad de hacer un cálculo separado. Se trata de una distribución mixta que tiene en cuenta todas las propiedades observadas en nuestros datos. Por un lado, la variable número de siniestros recoge el elevado número de ceros y sigue una distribución Poisson. Y por otro, cada evento genera un coste aleatorio que sigue una

distribución Gamma. A partir de aquí se deduce que la distribución de probabilidad de la variable coste último total por expuesto, que presenta una clara asimetría positiva y falta de normalidad, sigue entonces una distribución Tweedie.

Por tanto, la selección de la técnica predictiva basada en el modelo Tweedie está plenamente justificada por varias razones. No solo se adapta perfectamente a la limitación de información presente en nuestro microseguro, sino que denota simplicidad al modelizar directamente la prima de riesgo, sencillez que debe prevalecer también en la tarificación de este tipo de productos. Además, se trata de una metodología de uso extendido en el campo actuarial de tarificación para el seguro tradicional, lo que respalda la experiencia en su aplicación a los microseguros. Con todo, la aplicación del modelo Tweedie para estimar las primas de riesgo a un nuevo mercado de microseguros se convierte así en un planteamiento novedoso y con gran potencial de desarrollo.

- La séptima conclusión trata de “conseguir la certeza suficiente de que el modelo Tweedie proporciona la mejor estimación de las primas de riesgo de nuestro producto”. -

La aplicación del modelo Tweedie permite calcular, con mayor aproximación que otros modelos, la tarifa para nuestro ejemplo. Se trata de un enfoque que proporciona la mejor explicación de la variable respuesta, el coste último total por expuesto, respecto a las variables explicativas seleccionadas que representan los factores de riesgo característicos de nuestro colectivo, como son la edad del asegurado, la antigüedad de la póliza y la densidad de población. A partir de los estimadores de cada predictor se calculan las relatividades asociadas a los factores de riesgo que conforman la tarifa de nuestro microseguro.

El estudio realizado en el análisis predictivo con GLMs, de acuerdo a las hipótesis asumidas, nos ofrece dos posibles modelos multivariantes que, por una parte difieren únicamente en la segmentación de la variable densidad de población, y por otro lado, presentan resultados muy similares y totalmente válidos para modelizar nuestro microseguro. El objetivo final del análisis trata de decidir cuál es el modelo óptimo, teniendo en cuenta no solo los mejores resultados de la estimación de sus parámetros con sus medidas de ajuste y significación, sino que además ofrezca la mejor predicción del comportamiento futuro de nuestro colectivo a través del análisis de equilibrio financiero.

La decisión requiere, por tanto, la tarea previa de comparar sus resultados, que han sido obtenidos a partir de la muestra test que recoge el 30% de los datos, y en los que se destacan las siguientes características. En ambos modelos las primas presentan una alta variabilidad, que redundará en una alta asimetría positiva y la existencia de colas gruesas a la derecha. También, en los dos casos el error cuadrático medio es reducido y proporcionan ratios de siniestralidad similares en torno al 53%. La diferencia fundamental entre ambos radica en la amplitud del rango de primas estimadas, el cual prácticamente se duplica en el modelo 2, con el resultado de unas primas máximas bastante elevadas tal y como confirma la asimetría positiva de los modelos. A pesar de esto, la prima media en el modelo 1 asciende a 154,43Php, siendo ligeramente superior a los 152,38Php del modelo 2. Por tanto, la decisión del modelo óptimo va a consistir en dar prioridad a los resultados que presenten una menor variabilidad, y en consecuencia el modelo 1 se alza como la alternativa óptima que mejor va a representar el comportamiento del perfil de riesgo de nuestro colectivo.

El modelo óptimo debe realizarse bajo la hipótesis de validación del mismo con toda la muestra de datos, con el fin de dotar a las estimaciones de una mayor credibilidad. Entre las características más relevantes del modelo óptimo se mencionan las siguientes: el valor del parámetro p es de 1,172 que se sitúa en el caso particular $1 < p < 2$ asociado a la distribución Poisson compuesta con Gamma del modelo Tweedie para la estimación de las primas de riesgo. Las medidas de bondad de ajuste, la desviación, el AIC y el BIC mejoran al utilizar toda la muestra de datos, lo que nos da una garantía de que cuanto más experiencia se tenga más robusto será el modelo. En cuanto a los estimadores β , todos los niveles de las variables son significativos con un p-valor < 0,05 lo que garantiza obtener unas relatividades de prima adecuadas para la tarifa de nuestro producto.

En definitiva, se confirma que el análisis predictivo basado en la distribución Tweedie ofrece un buen modelo de regresión para la variable respuesta, capaz de proporcionar las primas de riesgo de nuestro microseguro ajustadas al perfil de riesgo de nuestro colectivo asegurado. Perfil que viene representado por tres factores de riesgo: la edad, la antigüedad de la póliza y la densidad de la población.

- La octava conclusión en cuanto a “llegar a valorar nuestra propuesta, promoviendo la aplicación de nuestro modelo óptimo a la práctica microaseguradora”. –

Técnicamente, no hay duda que esta investigación demuestra que es posible obtener un buen modelo multivariante, capaz de establecer una tarifa individualizada con factores de riesgo para nuestro microseguro. Pero yendo un paso más allá, es importante comprobar que este modelo se puede realmente extender en la práctica aseguradora, y de ese modo contribuir al desarrollo del campo actuarial de la tarificación en el negocio microasegurador. Con este fin, se plantea entonces la cuestión de comparar nuestro modelo óptimo con el que actualmente está utilizando la compañía para ver las ventajas que lleva consigo su implementación en este negocio.

Desde el punto de vista metodológico, ambos modelos utilizan una tarificación a priori basada en la experiencia de siniestralidad acumulada. En concreto, nuestro modelo proporciona una tarifa individualizada según el perfil de riesgo del asegurado, mientras que el modelo actual utiliza una tarificación colectiva con un enfoque sencillo que ofrece una prima plana común para todo el perfil del colectivo asegurado. La falta de definición sobre el comportamiento del riesgo del colectivo obliga a la compañía a complementar la prima pura con una serie de gastos y recargos que garanticen la solvencia del producto, obteniendo una prima total recargada que será la que se utilice como referencia para este estudio comparativo.

La propia tarificación lleva implícita una gestión de riesgos que, con nuestro modelo multivariante, permite incluir diferentes factores de riesgo que ayudan a solventar problemas derivados de la asimetría de información, tales como la selección adversa, el riesgo moral o el fraude, tan presentes en el negocio microasegurador. Un ejemplo visible aparece aquí con la variable *antigüedad de la póliza*, que del simple análisis descriptivo nos lleva a deducir que el asegurado con mayor antigüedad presenta una mayor siniestralidad. En nuestra tarifa se incluye esta variable como factor de riesgo, lo que servirá de medida para controlar el uso abusivo del producto, favoreciendo con una prima menor a los riesgos buenos y recargando la prima de los malos. En cambio, con una tarifa única para todos los asegurados, que es la que actualmente ofrece la compañía, no se aplica criterio alguno para aminorar este efecto. Por tanto, la consideración de este nuevo factor de riesgo en nuestra tarifa puede contribuir a una mejor gestión de riesgos y a la vez favorecer la permanencia de los riesgos buenos en la cartera.

Este estudio comparativo nos lleva a profundizar en el análisis desde dos nuevas perspectivas. Por un lado, mediante un análisis del equilibrio financiero del negocio; y por otro, observando gráficamente la viabilidad y sostenibilidad de ambos modelos:

- En *el análisis del equilibrio financiero* se destaca en primer lugar, la variabilidad de las primas de nuestro modelo, motivado lógicamente por una tarifa multivariante con la presencia de una cola gruesa a la derecha que evidencia la existencia de los riesgos malos, frente a la estabilidad de una prima plana que aplica actualmente la compañía. En segundo lugar, el cálculo del error cuadrático medio confirma que nuestras primas proporcionan un ligero mejor ajuste a la siniestralidad respecto a la tarifa real, lo que favorece la utilización del modelo Tweedie. Y por último, en nuestra propuesta la prima media de riesgo asciende a 151Php, por debajo de la prima recargada actual de 160Php, lo que se traduce en un encaje total de primas también inferior y, en lógica, en un mayor ratio de siniestralidad del 60,4% frente al 57% que presenta la compañía. En cuanto al mayor ratio, cabe decir que los datos del mercado asiático confirman que entra dentro de lo razonable para este tipo de microseguros.
- En *el análisis gráfico de la viabilidad y estabilidad*, se puede observar claramente como la tarifa multivariante realiza un buen ajuste y se adapta perfectamente a la siniestralidad, proporcionando primas más reducidas a los riesgos buenos y recargando a los riesgos malos por su peor siniestralidad. Sin embargo, la prima constante de 160Php no establece diferenciación entre los riesgos buenos y malos, siendo insuficiente para estos últimos donde el coste supera con creces la prima fijada por la compañía. Por otro lado, la curva de Lorenz permite comprobar como nuestro modelo proporciona en general primas inferiores y diferenciadas por el perfil de riesgo, mientras que el modelo actual se sitúa en la línea de igualdad donde todos los asegurados contribuyen de igual manera con independencia de su comportamiento frente al riesgo.

En definitiva, este estudio demuestra que la aplicación del modelo Tweedie puede ayudar a mejorar la tarificación del microseguro, proporcionando unas primas ajustadas al verdadero comportamiento de siniestralidad de los asegurados. Sí es cierto, que la flexibilidad de esta técnica predictiva lleva implícita una gran variabilidad en las primas que debe ser convenientemente tratada y adaptada por la compañía para evitar grandes saltos entre los diferentes perfiles de riesgo. Además, esta

metodología de tarificación se convierte así en una herramienta adicional para prevenir riesgos patentes como son la selección adversa y el uso excesivo del producto, problemas que se han detectado en nuestro microseguro.

5. Líneas abiertas y repercusión para futuras investigaciones

Una tesis solo abarca una pequeña área de estudio, que, sin embargo, al culminar su recorrido, logra situarnos en una posición privilegiada desde la que nos es posible divisar todo un campo de investigación que queda aún por explorar. A partir del estudio realizado y de las conclusiones derivadas de la investigación, consideramos que la presente tesis ofrece una serie de aportaciones que pueden servir como punto de referencia para desarrollar futuras líneas de investigación. Tales aportaciones se resumen a continuación:

- *Una revisión de literatura de los microseguros* bajo dos perspectivas, un enfoque social como producto asegurador inclusivo, y un enfoque de la actividad aseguradora en relación a las mejoras que contribuyen al desarrollo de este negocio emergente.
- *La propuesta para aplicar una metodología de tarificación* basada en los modelos lineales generalizados que sigue una distribución Tweedie para así estimar las primas de riesgo del microseguro más ajustadas según el perfil de riesgo individual de los asegurados.
- *La recomendación de hacer extensivo el uso de nuestro modelo a la práctica microaseguradora.* En el sentido de promover su aplicación en la práctica actuarial de tarificación de estos productos, que viene justificada por las ventajas comparativas que supone respecto al modelo tradicional que emplea hoy en día la compañía.

La principal contribución del presente trabajo de investigación se ha centrado en la mejora del diseño del producto desde el punto de vista actuarial de la tarificación, con el valor añadido de haber utilizado el caso real de un microseguro combinado. Esto constituye un avance en el desarrollo de estos productos. A lo que hay que añadir diferentes aspectos de mejora que se han detectado y apuntado a lo largo de la tesis y ahora sugerimos que puedan servir como punto de partida para iniciar nuevas líneas de trabajo.

Se describen a continuación como **sugerencias en relación al marco teórico**, las siguientes:

Fomentar la cultura financiera y aseguradora del microseguro en todos los niveles. Comenzando por impartir una mayor educación entre la población de bajos ingresos, así como una formación específica para el organismo supervisor, aprovechando incluso la experiencia adquirida del microcrédito. Y por último, transmitir el conocimiento entre el personal de las compañías aseguradoras.

Buscar nuevos canales de distribución con un enfoque predominantemente inclusivo. Con la idea de alcanzar a la mayor cantidad de personas desfavorecidas, que en tantas ocasiones resulta inaccesible. Y apostar, entre otras opciones, por la tecnología como una vía con gran potencial para acceder y comunicarse con este nicho de mercado.

Apoyar, desde las instituciones, la creación de un marco regulatorio para el sector de microseguros. Esta iniciativa reguladora impulsada por la IAIS y la Comisión de Servicios Financieros, debe ser respaldada con la colaboración de las distintas instituciones para crear un entorno legal que promueva el desarrollo de este negocio. En esta misma línea, se echa en falta una mínima regulación para los esquemas informales de microseguros, que actualmente ejercen su actividad fuera del ámbito legal, y que pone en desventaja a los esquemas formales que sí que cumplen con la regulación.

Explorar alternativas para la obtención, tratamiento y conservación de la información. Con la finalidad que se convierta en una auténtica fuente de conocimiento del negocio de microseguros para realizar posteriormente un correcto análisis y seguimiento del negocio. Entre las múltiples opciones que se plantean tenemos:

- Invertir mayor tiempo y recursos en el diseño, el tratamiento y la conservación de una buena base de datos interna de la compañía. Además de coordinar todos los requerimientos de información necesarios dentro de la compañía.
- Potenciar las fuentes externas colaborativas, fomentando el apoyo entre las diferentes instituciones, bien entre las propias compañías de seguros, o bien mediante asociaciones sectoriales encargadas de recopilar datos del mercado.
- Desarrollar estudios de mercado más flexibles y dinámicos, que permitan identificar con mayor exactitud las verdaderas necesidades y preferencias de la

población de bajos ingresos para diseñar así productos realmente acordes a la demanda del mercado.

Avanzar en la creación y desarrollo de medidas del rendimiento y la eficacia de los microseguros encaminadas a garantizar su sostenibilidad. En esta línea, aunque se cuenta con la creación de los indicadores clave de rendimiento (KPIs), sí que existen varias lagunas por cubrir, como son la diferenciación entre la clase de proveedores, así como integrar la función social realizada por las compañías.

En cuanto a las **sugerencias en relación al estudio empírico**, se destacan las siguientes:

Explorar nuevas vías de gestión y análisis de riesgos. Sobre todo, las enfocadas a tratar los problemas asociados a los microseguros, tales como reducir el uso abusivo del producto por el riesgo moral, control mayor del fraude, controlar la selección adversa con el fin de conservar una cartera dominada por riesgos buenos y reducir así la alta tasa de anulación en estos productos.

Evaluar y desarrollar otras metodologías de tarificación de microseguros. Y a ser posible con una doble función: proporcionar una tarifa ajustada al perfil de riesgo del colectivo, como también ayudar a controlar y gestionar los riesgos inherentes de este negocio.

Estudiar alternativas para mejorar las limitaciones encontradas en la aplicación de nuestro modelo Tweedie. En primer lugar, buscando la posibilidad de reducir la variabilidad de las primas asociada a los diferentes perfiles de riesgo para acotar el rango de primas. En segundo lugar, y ligado al anterior, investigando más a fondo sobre el tratamiento de los valores atípicos en microseguros. Y finalmente, valorando las opciones de incorporar nueva información al modelo, bien que permita realizar un cálculo separado por coberturas, o bien con la idea de actualizar las tarifas para años sucesivos.

Utilizar este estudio como guía de aplicación del modelo Tweedie en otro tipo de microseguros y para otros países, con el objetivo de comprobar la validez global de la metodología propuesta y contribuir al desarrollo de una mejor tarificación de los microseguros.

Plantear la posibilidad de implantar el microseguro en España, siguiendo la reciente recomendación de la OCDE²⁷ en potenciar su crecimiento inclusivo en una sociedad con cada vez menos empleos y salarios más bajos, en detrimento del bienestar social deseable. La creación de un nuevo producto, como es el microseguro, promueve la inversión de empresas innovadoras a la vez que impulsa el sector microfinanciero desde el mercado asegurador.

²⁷ Recomendación de los últimos Estudios Económicos de la OCDE: España, Marzo 2017 (OCDE, 2017).

BIBLIOGRAFÍA

- Albarrán, I., & Alonso González, P. (2007). Microseguro: fundamentos técnicos y ámbito de aplicación de un producto en expansión. *Revista Española de Seguros*, (131), 297-310.
- Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Shirmacher, E., & Thandi, N. (2007). A Practitioner's Guide to Generalized Linear Models. *Casualty Actuarial Society Discussion Paper Program*, tercera edición, 1-116.
- Azevedo, F.C., Oliveira, T.A., & Oliveira, A. (2016). Modelling non-life insurance price for risk without historical information. *Statistical Journal*, vol.14 (2), 171-192.
- Banco Mundial. (2013). Población pobre 2013. Agregación regional PovcalNet. Recuperado el 5 de Mayo del 2016, desde: <http://iresearch.worldbank.org/PovcalNet/povDuplicateWB.aspx>
- Bankable Frontier Associates (BFA). (2015). Can the digitalization of microinsurance make all the difference?
- Bauchet, J., Damon, A., & Hunter, B. (2016). Asymmetric Information in Microinsurance Markets: Experimental Evidence from Mexico.
- Barbin, E. A., Lomboy, C., & Soriano, E. (2002). A field study of microinsurance in the Philippines. ILO.
- Bejarano V.E. (2008). The base of the pyramid, the change of microinsurance. *Revista Trébol*, vol.49 (4), 6-9.
- Biener, C. (2013). Pricing in microinsurance markets. *World Development*, vol. 41, 132-144.
- Biener, C., & Eling, M. (2012). Insurability in microinsurance markets: An analysis of problems and potential solutions. *The Geneva Papers on Risk and Insurance Issues and Practice*, vol. 37 (1), 77-107.
- Biener C., & Eling M. (2014). Regulation in Microinsurance Markets: Principles, Practice, and Directions for Future Development. *World Development*, vol. 58, 21-40.
- Blacker, J., & Yang, M. (2015). *Actuaries in Microinsurance. Managing Risk for the Underserved*. Winsted, CT: ACTEX Publications.

- Botero, F., Churchill, C., McCord, M.J., & Qureshi, Z. (2006). The future of microinsurance. En Churchill, C.F., *Protecting the poor: a microinsurance compendium, vol.1*, 583-601. International Labour Organization.
- Breidert, C., Hahsler, M., & Reutterer, T. (2006). A review of methods for measuring willingness-to-pay. *Innovative Marketing*.
- Brisard, E. (2014). *Pricing of Car Insurance with Generalized Linear Models*. Trabajo fin de Máster. Universidad de Vrije. Ámsterdam, Bruselas.
- Burke, M., de Janvry, A., & Quintero, J. (2010). Providing Index-Based Agricultural Insurance to Smallholders: Recent Progress and Future Promise. Documento de Trabajo. CEGA, Universidad de California, Berkeley.
- BSLatAm. (2015). Microseguros en América Latina: Estudio del mercado y casos de éxito.
- Cachero, M. L. (1993). *Fundamentos y métodos de estadística*. Ediciones Pirámide. Madrid, España.
- Camargo, L.A., & Montoya, L.F. (2011). Microinsurance: Analysis of Outstanding Experiences in Latin America and the Caribbean". FIDES, FOMIN y Fundación MAPFRE. México.
- Carter, M., De Janvry, A., Sadoulet, E., & Sarris, A. (2014). Index-based weather insurance for developing countries: A review of evidence and a set of propositions for up-scaling. Documento de trabajo de políticas de desarrollo, (nº111).
- Clarke D. (2011). *Insurance Design for Developing Countries*. Tesis doctoral. Universidad de Oxford. Reino Unido.
- Clarke, D., Mahul, O., Rao, K.N., & Verma, N. (2012). *Weather Based Crop Insurance in India*. Documento de trabajo, (nº 5985). Banco Mundial. Recuperado el 17 de Enero del 2016, desde: <http://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-5985>.
- Cohen, M., & Sebstad, J. (2005). Reducing vulnerability: The demand for microinsurance. *Journal of International Development, vol.17* (3), 397-474.
- Chen, X., Ender, P.B., & Mitchell, M.N. (2003). *Regression Tricks of the Trade*. UCLA Academic Technology Services. Recuperado el 9 de Enero del 2017, desde: <http://www.ats.ucla.edu/stat/webbooks/rtt/rttch2.pdf>.
- Churchill, C.F. (2002). Trying to understand the demand for microinsurance. *Journal of International Development, vol.14* (3), 381-387.

- Churchill, C.F. (2006). *Protecting the poor: a microinsurance compendium, vol.1*. International Labour Organization.
- Churchill, C.F., & Matul, M. (2014). *Protegiendo a los pobres: un compendio sobre microseguros, vol.2*. International Labour Organization.
- Dalal, A. (2015). Learning from others' mistakes, n° 42. International Labour Organization.
- Dalal, A., & Morduch, J. (2010). *The Psychology of Microinsurance: Small Changes can make a surprising difference*. ILO.
- De Bock, O., & Gelade, W. (2012). The demand for microinsurance: a literature review. *ILO Microinsurance Innovation Facility Research Paper*, (n° 26).
- De Jong, P., & Heller, G. Z. (2008). *Generalized linear models for insurance data, vol. 10*. Cambridge: Cambridge University Press.
- Demirgüç-Kunt, A., Klapper, L.F., Singer, D., & Van Oudheusden, P. (2015). The global finindex database 2014: Measuring Financial inclusion around the World. Documento de trabajo, (n° 7255). Banco Mundial. Recuperado el 2 de Enero del 2017, desde: <http://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-7255>.
- Déniz, E.G., & Alegría, J.M.S. (2008). *Teoría de la credibilidad: desarrollo y aplicaciones en primas de seguros y riesgos operacionales*. Fundación MAPFRE.
- Denuit, M., Maréchal, X., Pitrebois, S., & Walhin, J. F. (2007). *Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems*. John Wiley & Sons.
- Dercon, S., Gunning, J. W., & Zeitlin, A. (2011). The demand for insurance under limited credibility: Evidence from Kenya. *International Development Conference, DIAL*.
- Dunn, P. K., & Smyth, G. K. (2008). Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Statistics and Computing*, vol.18 (1), 73-86.
- Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Chapman and Hall/ CRC press.
- Embrechts, P. (2000). Actuarial versus Financial Pricing of Insurance. *The Journal of Risk Finance*, vol. 1 (4), 17-26.
- England, P., & Verrall, R. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: mathematics and economics*, vol. 25 (3), 281-293.

Fundación Munich Re (2013). *The Landscape of Microinsurance in Asia and Oceania*. Con el apoyo de GIZ-RFPI y Microinsurance Network.

Garayoa, F.J. (2011). *La actividad microaseguradora sostenible en America Latina*. Tesis doctoral. Universidad Pontificia de Salamanca, España.

Garrido, J., Genest, C., & Schulz, J. (2016). Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics*, 70, 205-215.

Gil, J.A., Heras, A.J., & Vilar, J.L. (1999). *Matemática de los seguros de vida*. Fundación MAPFRE Estudios. Madrid, España.

GIZ, & RFPI. (2015). Regulatory Impact Assessment of Microinsurance Philippines. Study Report. Manila, Philippines. Recuperado el 6 de Septiembre del 2016, desde: <http://www.inclusiveinsuranceasia.com/docs/RIA-MI-PH-report.pdf>.

Goldburd, M., Khare, A., & Tevet, C.D. (2016). Generalized linear models for insurance rating. *Casualty Actuarial Society*.

Gradl, C., Herrndorf, M., Knobloch, C., & Sengupta, R. (2010). Learning to Insure the Poor. Microinsurance Report–Allianz Group.

GRET. (2014). Learning journey: Health Insurance Project for the garment sector in Cambodia. Con el apoyo de *Microinsurance Innovation Facility & ILO*.

Hastie, T., Tibshirani, R. & Friedman, J. (2008). *The Elements of Statistical Learning; Data Mining, Inference and Prediction*. Segunda edición. Springer, Berlin.

Hazell, P., Anderson, N., Balzer, A., Hastrup Clemmensen, U., & Rispoli, F. (2010). *El potencial para la ampliación y sostenibilidad de los seguros basados en índices climáticos para la agricultura y la subsistencia rural*. Con el apoyo de FIDA y PMA.

Huang, S. (2015). *Risk Assessment and Pricing for Group Health Claims*. Tesis doctoral. Universidad de Connecticut.

IAIS, & CGAP Working Group on Microinsurance. (2007): Issues in regulation and supervision of Microinsurance.

International Labour Organization (ILO). (2005). Fourth ítem on the Agenda: Social protection as a productive factor. *Comitte on Employment and Social Policy*.

International Labour Organization (ILO). (2006). Implementing the Global Employment Agenda: Employment strategies in support of decent work, vision document.

International Labour Organization (ILO). (2013). Mobile phones and microinsurance. Microinsurance paper, (nº 26).

Jacquier, C., Ramm, G., Marcadent, P., & Schmitt-Diabate, V. (2006). The social protection perspective on microinsurance. En Churchill, C.F., *Protecting the poor: a microinsurance compendium, vol.1*, 45-64. International Labour Organization.

Jørgensen B., & De Souza M. (1994). Fitting Tweedie's compound Poisson model to non-life insurance. *Scandinavian Actuarial Journal, vol.3*, 69-93.

Kadigi, R.M.J., Kaliba A.R., & Kingu P.M. (2013). "Rainfall Indexed Microinsurance Program as a Viable Risk Management Strategy to Secure Agricultural Production in Tanzania". *Time journals of agriculture and veterinary sciences, vol.1* (3), 31-46.

Kudryavtsev, A.A. (2009). Using quantile regression for rate-making. *Insurance: Mathematics and Economics, vol.45* (2), 296-304.

Lloyd's. (2009). Insurance in developing countries: exploring opportunities in microinsurance. Lloyd's 360° Risk Insight.

Martín, J.A. (2015). *Análisis e inclusión de variables exógenas en la tarificación de autos mediante modelización por GLM*. Trabajo fin de Máster. Universidad Carlos III. Madrid, España.

McCord, M.J. (2005). Microinsurance: Sustainable risk management for the low income market. *Financial Sector Development Conference, 23–24 Junio*.

McCord, M.J. (2011). SUAVE Checklist for Microinsurance Products: Enhancing the potential for success. Microinsurance Center.

McCord, M.J., Buczkowski, G., & Saksena, P. (2006). Premium collection: Minimizing transaction costs and maximizing customer service. En Churchill, C.F., *Protecting the poor: a microinsurance compendium, vol.1*, 197-215. International Labour Organization.

McCord, M., Opdebeeck, B., Biese, K., & Pandey, M. (2014). "If you can't measure it, you can't manage it: microinsurance by the numbers". *Enterprise Development and Microfinance, vol. 25* (1), 41-56.

McCord, M.J., Tatin-Jaleran, C. & Ingram, M. (2012). El panorama de los microseguros en América Latina y el Caribe: una nota informativa. BID y FOMIN.

Mechler, R., Linnerooth-Bayer, J., & Peppiatt, D. (2006). *Microinsurance for Natural Disasters in Developing Countries: Benefits, Limitations and Viability*. ProVention Consortium, Geneva.

Microinsurance Network. (2015a). *The Landscape of Microinsurance Africa*. Con el apoyo de la Fundación Munich Re y el Gobierno de Luxemburgo.

Microinsurance Network. (2015b). *The Landscape of Microinsurance in Latin America and the Caribbean. A changing market*. Con el apoyo de la Fundación Munich Re y el Gobierno de Luxemburgo.

Microinsurance Network, (2015c) [en línea]. World Map of Microinsurance. Recuperado el 9 de Septiembre del 2016, desde: <http://worldmapofmicroinsurance.org/#>.

Milliman. (2015). Health microinsurance instructional pricing tool. [Herramienta Excel]. Disponible en: <http://www.milliman.com/insight/2015/Health-microinsurance-instructional-pricing-tool/>

Moreno, T. y Ramos, L. (2003). *Aplicación de Modelos de Credibilidad para el Cálculo de Primas en el Seguro de Automóviles*. Comisión Nacional de Seguros y Fianzas. México.

Mosleh, A. & Ramm, G. (2006). Meeting the special needs of women and children. En Churchill, C.F., *Protecting the poor: a microinsurance compendium, vol.1*, 130-144. International Labour Organization.

Muganda, F. (2015). *A Generalized Linear Model for Rating the Premiums of a Micro Health Insurance Policy*. Trabajo de investigación. Universidad de Strathmore. Nairobi, Kenya.

Naciones Unidas. (2015). Resolución aprobada por la Asamblea General 70/1. Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible. Recuperado el 5 de Mayo del 2016, desde: <http://www.un.org/es/comun/docs/?symbol=A/RES/70/1>.

Norberg, R. (2002). *Basic Life Insurance Mathematics*. Version: September 2002. Notas de trabajo, Laboratorio de Matemática Actuarial, Universidad de Copenhague.

OCDE. (2017). *Estudios Económicos de la OCDE. España. Marzo 2017. Visión general*. Recuperado el 20 de Marzo del 2017, desde: <http://www.oecd.org/eco/surveys/Spain-2017-OECD-economic-survey-overview-spanish.pdf>

Ohlsson, E., & Johansson, B. (2010). *Non-life insurance pricing with generalized linear models*, vol. 21. Springer. Berlín.

Oxfam América. (2013). *Saving for Change: Financial inclusion and resilience for the world's poorest people*. Con apoyo de *Freedom from Hunger*.

Peña, I. (2006). *Métodos de tarificación con credibilidad en seguros colectivos*, (85), 71-94. ICEA.

Pettifer, A., & Pettifer, J. (2012). *A Practical Guide to Commercial Insurance Pricing*. Artículo presentado en el Actuaries Institute General 2012 Insurance Seminar. Sydney. Recuperado el 10 de Enero del 2017, desde: <http://www.actuaries.asn.au/Library/Events/GIS/2012/GIS2012PaperAlinaPettifer.pdf> .

Philippine Statistics Authority. (2016). *Philippine Standard Geographic Code (PGSC)*. Recuperado el 15 Julio del 2016, desde: <http://psa.gov.ph>.

Philippine Statistics Authority (PSA), & ICF Internacional. (2014). *Philippines National Demographic and Health Survey 2013*. Manila, Philippines y Rockville, Maryland, USA: PSA y ICF International.

Portula D. O., & Vergara R. (2013). *The Philippine experience on Microinsurance Market Development*. Training Program of Insurance Supervisors in Asia. Access to Insurance Initiativ, AITRI, GIZ-RFPI y Toronto Center. Manila, Filipinas.

Purcal S., Englmann, T., Pang-Rey, M., & Suan, I.D. (2012). *Microinsurance Pricing under the Microscope: Findings from the Philippines*.

Radermacher, R., Dror, I., & Noble, G. (2006). *Challenges and strategies to extend health insurance to the poor*. En Churchill, C.F., *Protecting the poor: a microinsurance compendium*, vol.1, 66-93. International Labour Organization.

Ramsey, J.B. (1970). *Models, specification error, and inference: A discussion of some problems in econometric methodology*. *Oxford Bulletin of Economics and Statistics*, vol. 32 (4), 301-318.

Reinhard, D., & Qureshi, Z. (2013). *How to provide sustainable insurance for low-income farmers*, (nº 3). Trendmonitor.

Richardson, E. (2011). *Application of Standard Actuarial Pricing Techniques for Health Microinsurance Schemes*. Trabajo fin de Master. City University. Londres, Reino Unido.

Roth, J., McCord, M. J., & Liber, D. (2007). The landscape of microinsurance in the world's 100 poorest countries.

Sandmark, T., Debary, J.C., & Tatin-Jaleran, C. (2014). Surgimiento y desarrollo de los microseguros agrícolas. MicroInsurance Network con el apoyo de ADA y el Gobierno de Luxemburgo.

SAS Institute. (2008). *SAS/STAT® 9.2 User's Guide. The GENMOD Procedure*. Cary, NC, USA.

Shi, S.G. (2010). Direct Analysis of Pre-Adjusted Loss Cost, Frequency or Severity in Tweedie Models. *Casualty Actuarial Society E-Forum*.

Siegel, P., Alwang, J., & Canagarajah, S. (2001). *Viewing microinsurance as a social risk management instrument, vol.116*, 17-25. Social Protection Discussion Paper Series.

Smyth, G. K., & Jørgensen, B. (2002). Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. *Astin Bulletin, vol. 32 (01)*, 143-157.

Swiss Re (2010). *Microinsurance - Risk protection for 4 billion people*. Sigma nº 6. Zurich, Suiza.

Towers Watson (2010). *Predictive Modeling for Life Insurance*.

Werner, G., & Modlin, C. (2016). Basic ratemaking. *Casualty Actuarial Society*.

Wiedmaier-Pfister, M., & Chatterjee, A. (2006). An enabling regulatory environment for microinsurance. En Churchill, C.F., *Protecting the poor: a microinsurance compendium, vol.1*, 488-507. International Labour Organization.

Wipf, J., & Garand, D. (2006). Pricing microinsurance products. En Churchill, C.F., *Protecting the poor: a microinsurance compendium, vol.1*, 238-253. International Labour Organization.

Wipf, J., & Garand, D. (2010). Performance indicators for microinsurance. Segunda edición. ADA con el Luxembourg Development Cooperation, BRS y Microinsurance Network.

World Bank Group (WBG). (2014a). Actuarial Basis. Manual 2. Con apoyo de *International Finance Corporation, Watts and Associates, & International Institute for Agricultural Risk Management*.

World Bank Group (WBG). (2014b). Data and information management. Manual 3. Con apoyo de *International Finance Corporation, Watts and Associates, & International Institute for Agricultural Risk Management*.

Wrede, P., Phily, C., Wipf, J. J., Tatin-Jaleran, C., & Garand, D. (2016). Pricing for Microinsurance: a Technical Guide. Impact Insurance & International Labour Organization.

Xacur, O. A. Q., & Garrido, J. (2015). Generalised linear models for aggregate claims: to Tweedie or not?. *European Actuarial Journal*, vol. 5 (1), 181-202.

Yan, J., Guszczka, J., Flynn, M., Wu, C.S. (2009). Applications of the Offset in Property–Casualty Predictive Modelling. *Casualty Actuarial Society E-Forum*.

Zimmerman, E., Magnoni, B. (2014). Consumer protection in microinsurance: Challenges and good practices from the Philippines and Colombia. *Enterprise Development and Microfinance*, vol. 25 (1), 25-40.

ANEXOS

Anexo A. Cuadros resumen de la base de datos del microseguro

Tabla A.1. Descripción de la base de datos de microseguros para el año 2014.

Base de datos	Nombre del campo	Descripción
Datos del asegurado	age	Edad
	civil status	Estado civil
	sex	Sexo
	province	Provincia
	region	Región
	id2	Clave de identificación
	age_policy	Antigüedad de la póliza
	exposure	Exposición
Datos de siniestralidad	total claims	Nº total de siniestros
	total closed	Nº total de siniestros liquidados
	total no loss	Nº total de siniestros a coste cero
	id earthquake	Nº siniestro por terremoto
	eq tsi amount	Suma asegurada por terremoto
	eq loss amount	Valoración siniestro por terremoto
	eq loss paid	Pagos del siniestro por terremoto
	eq claim status	Situación expediente por terremoto
	id fire assist	Nº siniestro por incendio
	fa tsi amount	Suma asegurada por incendio
	fa loss amount	Valoración siniestro por incendio
	fa loss paid	Pagos del siniestro por incendio
	fa claim status	Situación expediente por incendio
	id flood	Nº siniestro por inundación
	fl tsi amount	Suma asegurada por inundación
	fl loss amount	Valoración siniestro por inundación
	fl loss paid	Pagos del siniestro por inundación
	fl claim status	Situación expediente por inundación
	id landslide	Nº siniestro por deslizamiento
	ld tsi amount	Suma asegurada por deslizamiento
	ld loss amount	Valoración siniestro por deslizamiento
	ld loss paid	Pagos del siniestro por deslizamiento
	ld claim status	Situación expediente por deslizamiento
	id typhoon	Nº siniestro por tifón
	typ tsi amount	Suma asegurada por tifón
	typ loss amount	Valoración siniestro por tifón
	typ loss paid	Pagos del siniestro por tifón
	typ claim status	Situación expediente por tifón
	id actnature	Nº siniestro por total actos de la naturaleza
	an tsi amount	Suma asegurada por total actos de la naturaleza
an loss amount	Valoración siniestro por total actos de la naturaleza	
an loss paid	Pagos del siniestro por total actos de la naturaleza	
an claim status	Situación expediente por total actos de la naturaleza	
id hospital income	Nº siniestro por hospitalización	
hi tsi amount	Suma asegurada por hospitalización	
hi loss amount	Valoración siniestro por hospitalización	
hi loss paid	Pagos del siniestro por hospitalización	
hi claim status	Situación expediente por hospitalización	

Tabla A.2. Resumen descriptivo de la información por año/cobertura.

Año	Nº siniestros	Suma asegurada	Valoración siniestros	Pagos
Gastos funerarios				
2012	0	0	0	
2013	1	40.000	0	0
2014	2	40.000	0	0
2015	5	100.000	20.000	0
2016	0	0	0	
Hospitalización				
2012	4	40.000	4.500	4.500
2013	26	320.000	51.000	51.000
2014	52	630.000	94.250	94.250
2015	69	690.000	153.500	93.500
2016	22	240.000	55.000	30.500
Actos de la naturaleza^(*)				
2012	65	325.000	220.000	220.000
2013	131	1.195.000	340.000	340.000
2014	332	1.795.000	762.500	762.500
2015	4	15.000	10.000	7.500
2016	0	0	0	

Fuente de elaboración propia.

(*) Se incluyen todos los siniestros derivados por causas de terremoto, incendio, inundación, deslizamiento de tierras y tifón.

La Tabla A2 presenta un resumen de la base de datos de origen, que se corresponde con uno de los colectivos más numerosos de la compañía, desglosado para las tres coberturas del producto y para los diferentes años de experiencia acumulada por la compañía. Se observa claramente como la cobertura de gastos funerarios apenas registra siniestralidad, lo que sin duda nos conduce a excluir esta cobertura de nuestro análisis. En cuanto al horizonte temporal, los datos contienen información de los años que van desde el inicio del colectivo, en el año 2012, hasta el momento más reciente, junio del 2016. A la vista de los datos y bajo estas observaciones, se ha elegido el año 2014 por ser el período con mayor siniestralidad registrada para realizar nuestro caso de estudio particular.

Anexo B: Información demográfica nacional de Filipinas

Los siguientes datos recogen tanto la información sobre la población en las diferentes provincias, ciudades y distritos de Filipinas (Philippine Statistics Authority, 2016), como el nuevo factor de riesgo densidad de población elaborado a partir de dichos datos. La

variable densidad se estima como el cociente de la población registrada en cada una de las provincias sobre la población total de Filipinas.

Tabla A.3. Descripción de la población y densidad a nivel provincial.

Nombre	Nivel	Población	Densidad
Philippines	País	100.981.437	1,000
First District	Distrito	1.780.148	0,018
Second District	Distrito	4.650.613	0,046
Third District	Distrito	2.819.388	0,028
Fourth District	Distrito	3.627.104	0,036
Abra	Provincia	241.160	0,002
Agusan del Norte	Provincia	354.503	0,004
Agusan del Sur	Provincia	700.653	0,007
Aklan	Provincia	574.823	0,006
Albay	Provincia	1.314.826	0,013
Antique	Provincia	582.012	0,006
Apayao	Provincia	119.184	0,001
Aurora	Provincia	214.336	0,002
Basilan	Provincia	346.579	0,003
Bataan	Provincia	760.650	0,008
Batanes	Provincia	17.246	0,000
Batangas	Provincia	2.694.335	0,027
Benguet	Provincia	444.224	0,004
Biliran	Provincia	171.612	0,002
Bohol	Provincia	1.313.560	0,013
Bukidnon	Provincia	1.415.226	0,014
Bulacan	Provincia	3.292.071	0,033
Cagayan	Provincia	1.199.320	0,012
Camarines Norte	Provincia	583.313	0,006
Camarines Sur	Provincia	1.952.544	0,019
Camiguin	Provincia	88.478	0,001
Capiz	Provincia	761.384	0,008
Catanduanes	Provincia	260.964	0,003
Cavite	Provincia	3.678.301	0,036
Cebu	Provincia	2.938.982	0,029
Compostela Valley	Provincia	736.107	0,007
Davao del Norte	Provincia	1.016.332	0,010
Davao del Sur	Provincia	632.588	0,006
Davao Occidental	Provincia	316.342	0,003
Davao Oriental	Provincia	558.958	0,006
Dinagat Islands	Provincia	127.152	0,001
Eastern Samar	Provincia	467.160	0,005
Guimaras	Provincia	174.613	0,002
Ifugao	Provincia	202.802	0,002
Ilocos Norte	Provincia	593.081	0,006
Ilocos Sur	Provincia	689.668	0,007
Iloilo	Provincia	1.936.423	0,019
Isabela	Provincia	1.593.566	0,016
Kalinga	Provincia	212.680	0,002
La Union	Provincia	786.653	0,008
Laguna	Provincia	3.035.081	0,030
Lanao del Norte	Provincia	676.395	0,007
Lanao del Sur	Provincia	1.045.429	0,010

Nombre	Nivel	Población	Densidad
Leyte	Provincia	1.724.679	0,017
Maguindanao	Provincia	1.173.933	0,012
Marinduque	Provincia	234.521	0,002
Masbate	Provincia	892.393	0,009
Misamis Occidental	Provincia	602.126	0,006
Misamis Oriental	Provincia	888.509	0,009
Mountain	Provincia	154.590	0,002
Negros Occidental	Provincia	2.497.261	0,025
Negros Oriental	Provincia	1.354.995	0,013
North Cotabato	Provincia	1.379.747	0,014
Northern Samar	Provincia	632.379	0,006
Nueva Ecija	Provincia	2.151.461	0,021
Nueva Vizcaya	Provincia	452.287	0,004
Occidental Mindoro	Provincia	487.414	0,005
Oriental Mindoro	Provincia	844.059	0,008
Palawan	Provincia	849.469	0,008
Pampanga	Provincia	2.198.110	0,022
Pangasinan	Provincia	2.956.726	0,029
Quezon	Provincia	1.856.582	0,018
Quirino	Provincia	188.991	0,002
Rizal	Provincia	2.884.227	0,029
Romblon	Provincia	292.781	0,003
Sarangani	Provincia	544.261	0,005
Siquijor	Provincia	95.984	0,001
Sorsogon	Provincia	792.949	0,008
South Cotabato	Provincia	915.289	0,009
Southern Leyte	Provincia	421.750	0,004
Sultan Kudarat	Provincia	812.095	0,008
Sulu	Provincia	824.731	0,008
Surigao del Norte	Provincia	485.088	0,005
Surigao del Sur	Provincia	592.250	0,006
Tarlac	Provincia	1.366.027	0,014
Tawi-Tawi	Provincia	390.715	0,004
Western Samar	Provincia	780.481	0,008
Zambales	Provincia	590.848	0,006
Zamboanga del Norte	Provincia	1.011.393	0,010
Zamboanga del Sur	Provincia	1.010.674	0,010
Zamboanga Sibugay	Provincia	633.129	0,006
Bacolod	Ciudad	561.875	0,006
Cagayan de Oro	Ciudad	675.950	0,007
Calamba	Ciudad	454.486	0,005
Tabaco	Ciudad	133.868	0,001
Tanauan	Ciudad	173.366	0,002
Cotabato	Ciudad	299.438	0,003
Lipa	Ciudad	332.386	0,003
Lucena	Ciudad	266.248	0,003
San Pablo	Ciudad	266.068	0,003
Zamboanga	Ciudad	861.799	0,009

Fuente: elaboración propia a partir de los datos (PGSC) de la población filipina registrados en junio 2016 (Philippine Statistic Authority, 2016).

Anexo C: Código SAS del modelo Tweedie

La siguiente macro de SAS se basa en el código elaborado por Shi (2010) para aplicar el modelo GLM que sigue la distribución Tweedie. Este programa tiene la gran ventaja que nos permite realizar la convergencia de los diferentes parámetros (μ, p, ϕ) de la distribución Tweedie a través de dos procedimientos, "PROC GENMOD" y "PROC NLMIXED". Se trata de un proceso iterativo que va a ser capaz de proporcionarnos un modelo óptimo para la variable respuesta, coste último total por expuesto, en el caso particular de nuestro microseguro.

MACRO MODELO TWEEDIE

Parámetros:

```
dn -- nombre de la base de datos
wght -- variable peso
resp -- variable respuesta
pred -- predictores
clmcnt -- número de siniestros
offset -- variable offset
```

/* Inicio */

```
%macro tweedie(dn=,wght=,resp=,pred=,clmcnt=,offset=);
title3;
data Est_save_;
format p_ phi_ sigma_ p_lower p_upper phi_lower phi_upper
       sigma_lower sigma_upper 15.4 p_change sigma_change 15.4;
p_ = 1.5;
p_lower = .;
p_upper = .;
phi_ = 1;
phi_lower = .;
phi_upper = .;
sigma_ = 1;
sigma_lower = .;
sigma_upper = .;
p_change = .;
sigma_change = .;
call symput('p',trim(left(put(p_,15.4))));
call symput('phi',trim(left(put(phi_,15.4))));
call symput('sigma',trim(left(put(sigma_,15.4))));
call symput('p_lower',trim(left(put(p_lower,15.4))));
call symput('p_upper',trim(left(put(p_upper,15.4))));
call symput('phi_lower',trim(left(put(phi_lower,15.4))));
call symput('phi_upper',trim(left(put(phi_upper,15.4))));
call symput('sigma_lower',trim(left(put(sigma_lower,15.4))));
```

```
call symput('sigma_upper',trim(left(put(sigma_,15.4))));
run;

/* Estimación de máxima verosimilitud */

%let converge = 0;
%let i=1;
%do %until ((&converge eq 1) or (&i gt 10));
title3 'Optimization Step &i';
%optimize(&dn,&wght,&resp,&pred,&clmcnt,&offset,0);
%let i = %eval(&i+1);
  data Est_save_(drop=p_old sigma_old);
  set Est_save_ end=last;
  p_old = p_;
  sigma_old = sigma_;
  retain p_old sigma_old;
  output;
  if last then do;
    p_ = &p;
    p_lower = &p_lower;
    p_upper = &p_upper;
    phi_ = &phi;
    phi_lower = &phi_lower;
    phi_upper = &phi_upper;
    sigma_ = &sigma;
    sigma_lower = &sigma_lower;
    sigma_upper = &sigma_upper;
    p_change = abs(p_-p_old);
    sigma_change = abs(sigma_-sigma_old);
    p_old = p_;
    sigma_old = sigma_;
    if (p_change le 1e-5) and (p_change ne .) and (sigma_change
le 1e-5)
      and (sigma_change ne .) then call symput('converge','1');
  output; end;
run; %end;

/* Resultados*/

%if (&converge eq 1) %then %do;
title3 'Tweedie Model with Converged Parameter Estimates';
%optimize(&dn,&wght,&resp,&pred,&clmcnt,&offset,1);
%put Power parameter = &p with 95% C.I. (&p_lower, &p_upper);
%put Dispersion parameter = &phi with 95% C.I. (&phi_lower,
&phi_upper); %put SAS scale parameter = &sigma with 95% C.I.
(&sigma_lower,
&sigma_upper);
%end; %else %do;
```

```

%put Not converged after 10 iterations: ;
%put Power parameter = &p;
%put Dispersion parameter = &phi;
%put SAS scale parameter = &sigma;
%put at the end of 10th iteration.;
%end;
title3;

%mend tweedie;
%macro optimize(dn,wght,resp,pred,clmcnt,offset,flag);

ods graphics on;
proc genmod data=&dn plots=leverage;
class age5 (desc order=freq) den5 (desc order=freq) age_policy
(desc order=freq) /param=ref;
    p_ = &p;
    mu_ = _MEAN_;
    y_ = _RESP_;
    v_ = mu_**p_;
    if y_ gt 0 then d_ = 2*(y_*(y_**(1-p_)-mu_**(1-p_))/(1-p_)-
(y_**(2-p_)-mu_**(2-p_))/(2-p_));
    else d_ = 2*(mu_**(2-p_))/(2-p_);
    variance var = v_;
    deviance dev = d_;
    weight &wght;
    model &resp = &pred /link=log noscale scale=&sigma type3
%if %length(&offset) eq 0 %then ;
%else offset=&offset;;
    output out=Out_mu_pred=yhat_ cooks=dCook_ leverage=Lev_
stdreschi=rstP_resdev=rDev_stdresdev=rstDev_;
    ods output parameterestimates= parms;
run;

%if &flag ne 1 %then %do;
ods trace on;
ods output ParameterEstimates=Est_;

ods graphics off;

proc nlmixed data=Out_mu_;
    format p_ 15.4 phi_ 15.4;
    parms p_=&p phi_=&phi;
    bounds 1<p_<2, phi_>0;
    n_ = &clmcnt;
    w_ = &wght;
    y_ = &resp;
    mu_ = yhat_;

```

```
t_ = y_*mu_**(1-p_)/(1-p_)-mu_**(2-p_)/(2-p_);
a_ = (2-p_)/(p_-1);
if (n_ eq 0) then rll_ = (w_/phi_)*t_;
else rll_ = n_*((a_+1)*log(w_/phi_)+a_*log(y_)-a_*log(p_-1)-
log(2-p_)) -lgamma(n_+1)-lgamma(n_*a_)-log(y_)+(w_/phi_)*t_;

/* log likelihood of (p_,phi_) with mu_ known */
model y_ ~ general(rll_);
run;
ods trace off;
data _null_;
  set Est_;
  if Parameter eq 'p_' then do;
    call symput('p' ,trim(left(put(Estimate,15.4))));
    call symput('p_lower' ,trim(left(put(Lower,15.4))));
    call symput('p_upper' ,trim(left(put(Upper,15.4))));
  end; else if Parameter eq 'phi_' then do;
    call symput('phi' ,trim(left(put(Estimate,15.4))));
    call symput('phi_lower' ,trim(left(put(Lower,15.4))));
    call symput('phi_upper' ,trim(left(put(Upper,15.4))));
    call symput('sigma' ,trim(left(put(sqrt(Estimate),15.4)))); call
symput('sigma_lower',trim(left(put(sqrt(Lower),15.4)))); call
symput('sigma_upper',trim(left(put(sqrt(Upper),15.4))));
  end; run;

%end;
%mend optimize;
%tweedie(dn=phd.db_philf2, wght=exposure, resp=b_cost_T,
pred=age5 den5 age_policy, clmcnt=total_claims);
```

Fuente: elaboración propia a partir del código elaborado por Shi (2010). Direct Analysis of Pre-adjusted Loss Cost, Frequency or Severity in Tweedie Models. *Casualty Actuarial Society E-Forum*.

