

ANÁLISIS GEOGRÁFICO DE LA SINIESTRALIDAD EN EL SEGURO DEL AUTOMÓVIL EN ESPAÑA

María del Carmen Melgar Hiraldo

Flor María Guerrero Casas

Departamento de Economía, Métodos Cuantitativos e Historia Económica

Universidad Pablo de Olavide

e-mails: mcmelhir@upo.es, fguecas@upo.es

Resumen

La importancia que tiene el ramo del automóvil en el mercado del seguro conlleva que, tanto la estimación del número de siniestros que los conductores sufren y declaran a sus compañías, como la determinación de los factores que influyen en dicha siniestralidad, se conviertan en temas de sumo interés para las aseguradoras.

El objetivo esencial del trabajo que se presenta es estudiar si, en el caso de España, la Comunidad Autónoma de residencia del asegurado constituye uno de esos factores significativos. Se utilizan, para tal fin, datos cedidos por una entidad aseguradora privada que opera en el mercado español. Sobre dichos datos, se aplica un modelo econométrico de tipo recuento o *count data*. Estos modelos son los que se usan habitualmente para abordar cometidos de esta índole. Nos referimos a algunos ya tradicionales como el de regresión de Poisson o el binomial negativo, así como a los más recientes modelos *inflados de ceros*, que se suelen adaptar mejor al exceso de ceros que normalmente muestran los datos en estas situaciones.

Palabras clave: Seguro del automóvil, siniestralidad, modelos *count data*, análisis territorial.

Área temática: Métodos Cuantitativos.

1. Introducción

El seguro del automóvil es uno de los principales ramos de la actividad aseguradora en la mayoría de los países desarrollados. El número creciente de vehículos que se matriculan cada año, así como la obligación y/o necesidad de protegerse frente a los gastos y responsabilidades que acarrearán los accidentes que pueden producirse, son dos de las causas fundamentales de la importancia que ha adquirido este sector en los últimos años. Debemos tener en cuenta además que, en los años venideros, el seguro del automóvil podría convertirse en un elemento más para luchar contra el elevado número de accidentes con víctimas que se producen anualmente. La “amenaza” de una prima elevada puede ser una incitación a la prudencia para el conductor y esto, combinado con otras propuestas actuales como la entrada en vigor del carné de conducir por puntos o el endurecimiento de las multas por infracciones a las normas de tráfico, debería ayudar a rebajar esa cifra de cerca de 100.000 accidentes con víctimas que se producen cada año (DGT, 2004).

En España, en el año 1999, se alcanzaron prácticamente 2 millones de nuevas matriculaciones, siguiendo la tendencia al alza iniciada al principio de la década de los 90. A partir de entonces, las cifras se han mantenido estables, a excepción del año 2002 en el que solo se cuentan 1.750.000 vehículos matriculados por primera vez, según los datos de la Dirección General de Tráfico (DGT, 2004).

Son cerca de 20 millones las pólizas de seguro de automóvil emitidas en la actualidad. En el año 2004, más del 43% de las primas recaudadas en el seguro no-vida procedían del ramo del automóvil; sobre el total del sector, el porcentaje sigue siendo elevado, sobrepasando el 25% (ICEA, 2004). De los más de 11.000 millones de euros recaudados, alrededor del 60% se debe a la garantía de responsabilidad civil, obligatoria en nuestro país. Las garantías optativas (robo, incendio, rotura de lunas, daños propios del vehículo, etc.) suponen únicamente el 40% restante.

Un ajuste correcto de las primas resulta de vital importancia para que las compañías puedan hacer frente a las coberturas elegidas por sus clientes. Para esta tarea, es primordial conocer con exactitud cuáles son los factores que influyen en una mayor o menor propensión a los siniestros por parte del conductor asegurado. Las compañías

usan datos estadísticos al respecto y hacen depender la cuantía de la prima de características observables como el tipo de vehículo, el uso que se le da, la edad y sexo del conductor, así como su experiencia en la conducción, medida generalmente por los años que hace que obtuvo el carné de conducir, o el lugar por el que suele circular habitualmente, por ejemplo.

El objetivo del trabajo que proponemos aquí es determinar cuáles de estas variables intervienen realmente en la siniestralidad y en qué sentido. Pondremos especial atención al papel que juega la Comunidad Autónoma en la que reside el asegurado, para determinar de ese modo si la región de circulación habitual es verdaderamente significativa cuando se estima el número de siniestros que puede sufrir un conductor, así como para analizar las diferencias que existen en cuanto a la siniestralidad, si es el caso, dentro del territorio español.

Los modelos econométricos que se utilizan habitualmente en estudios de este tipo son el de regresión de Poisson y el modelo binomial negativo. Destacamos, entre otros, los análisis de Richaudeau (1999) y Cohen (2002) que emplean datos de Francia e Israel, respectivamente, para estudiar la posible correlación entre la siniestralidad de los asegurados y la cobertura de la que disfrutaban y con ello la posible existencia de información asimétrica entre asegurado y asegurador.

Existen sin embargo otros modelos más complejos, como los inflados de ceros, que pueden ser más adecuados para estimar variables de recuento de datos cuando existe una gran cantidad de observaciones nulas y éstas pueden tener distintas interpretaciones. Estos modelos, poco utilizados en el seguro del automóvil hasta ahora, sí se han desarrollado recientemente en otros ámbitos de la ciencia. Böhning *et al.* (1999), que estudian la prevención de caries en epidemiología dental, o Yau, Wang y Lee (2003), que analizan la duración de la hospitalización de enfermos de páncreas, son ejemplos de estas aplicaciones. En cuanto al uso de modelos inflados de ceros en el seguro del automóvil, Lee *et al.* (2002) analizan los factores que influyen en la siniestralidad de los conductores noveles de Australia durante el año siguiente a la obtención del carné, mientras que, con datos españoles, podemos nombrar los trabajos de Melgar, Ordaz y Guerrero (2004) y Melgar y Guerrero

(2005), en los que se determinan también las variables que se muestran significativas al estimar dicha siniestralidad.

En el caso que nos ocupa, es muy elevado el porcentaje de conductores que no sufre siniestros, por lo que hay muchos valores nulos en las observaciones. Además, un conductor tiene la posibilidad de declarar o no sus siniestros, lo que otorga al valor 0 dos interpretaciones: ningún siniestro o siniestros no declarados. Los modelos inflados de ceros recogen estos matices, lo que no ocurre con los modelos más tradicionales; éste es el motivo que nos lleva a plantearnos su validez.

El trabajo que presentamos se organiza del modo siguiente: tras introducir el tema y fijar los objetivos que nos planteamos, incluimos en el Apartado 2 los métodos usados en el análisis. A continuación, presentamos los resultados empíricos del estudio en el Apartado 3, tanto los derivados de un análisis descriptivo de los datos como los que se deducen de la aplicación del modelo econométrico que elegimos para la estimación. Concluimos los principales resultados en el Apartado 4, para terminar con la bibliografía y un anexo final que incluye una breve descripción de las variables usadas en el análisis.

2. Metodología

Los estudios econométricos que tratan de estimar variables discretas no negativas se llevan a cabo habitualmente mediante modelos de tipo recuento o *count data*. El modelo de regresión de Poisson y el modelo binomial negativo son los más tradicionales y se encuentran desde hace años en la literatura especializada. Más recientemente se han formulado variantes más complejas de estos modelos, con la ventaja de disponer de un mayor ámbito de aplicación. Entre dichas variantes, los modelos inflados de ceros se adaptan especialmente a los datos de los que disponemos, como veremos más adelante.

Se exponen sucintamente a continuación las características de los modelos *count data* referidos. Pueden encontrarse más detalles en Winkelmann (2003).

2.1. Modelos tradicionales

Se supone un conjunto de individuos para cada uno de los cuales se desea estimar el valor de la variable aleatoria discreta no negativa Y .

Según el modelo de Poisson, para cada individuo i se tiene que:

$$P(Y = y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad (1)$$

siendo $y_i = 0, 1, 2, \dots$, $\lambda_i = e^{X_i \beta}$, X_i el vector que recoge los valores que toman las variables explicativas para el individuo i y β el de los coeficientes a estimar.

Para que la aplicación de este modelo sea correcta, es necesario que se dé equi-dispersión en los datos. En efecto, a partir de la formulación anterior se llega a que $E(Y) = \text{Var}(Y) = \lambda_i$. Es una hipótesis muy fuerte que, para el tipo de variable que nos interesa, se incumplirá en muchas ocasiones. En tal caso, se usa frecuentemente el modelo binomial negativo, como método alternativo.

En este nuevo modelo, se supone que la variable Y sigue una ley de Poisson con parámetro μ_i relacionado con el anterior λ_i a través de la igualdad $\ln \mu_i = \ln \lambda_i + \varepsilon_i$, donde e^{ε_i} se distribuye según una gamma con valor esperado igual a 1 y varianza igual a $\frac{1}{\nu}$. Se tiene entonces que:

$$P(Y = y_i) = \frac{\Gamma(y_i + \nu)}{\Gamma(y_i + 1) \cdot \Gamma(\nu)} \cdot \left(\frac{\nu}{\nu + \lambda_i} \right)^\nu \left(\frac{\lambda_i}{\nu + \lambda_i} \right)^{y_i} \quad (2)$$

siendo $\Gamma(\cdot)$ la función gamma.

Usando, como en Jones (2001), el parámetro de precisión $\alpha > 0$, dado por $\alpha = \frac{1}{\nu}$, se deduce que la varianza de Y es una función cuadrática de su media. En concreto, $E(Y) = \lambda_i$ y $\text{Var}(Y) = \lambda_i + \alpha \lambda_i^2$. Así, el parámetro de precisión puede considerarse como una medida del grado de sobre-dispersión que presentan los datos. Ambos modelos (Poisson y binomial negativo) coinciden cuando $\alpha \rightarrow 0$ y por tanto el

contraste del valor nulo de este parámetro es el que lleva a la correcta elección entre ellos.

2.2. Modelos inflados de ceros

En ocasiones, el hecho de que la variable endógena tome el valor 0 puede interpretarse de dos modos distintos. Cuando es así, los modelos inflados de ceros pueden proporcionar mejores resultados que los modelos de Poisson y/o binomial negativo, ya que éstos no tienen en cuenta en la estimación esos posibles aspectos diferenciadores.

Con respecto a los modelos tradicionales, estos desarrollos más recientes de los modelos de tipo *count data* suponen que la variable Y es el producto de una ley binaria y una ley de Poisson o binomial negativa, $Y = ZY^*$, de manera que Z es un proceso modelizado como un logit que toma valores $z_i \in \{0,1\}$ e Y^* determina el valor de Y , sujeto a la condición $z_i = 1$.

La idea en la que se basan estos modelos más complejos es el hecho de que la población consta de dos tipos de individuos: para una parte de ellos, la observación es siempre nula mientras que, para el resto, los resultados se generan según la distribución de Poisson o binomial negativa, según se considere. Subyace así una decisión inicial de participación de cada individuo, representada por la variable Z . Sin participación ($z_i = 0$), Y vale 0; cuando hay participación ($z_i = 1$), Y se conduce según la ley elegida, pudiendo también, en particular, anularse. La posibilidad de que se obtengan valores nulos para la variable dependiente por estas dos vías, conlleva que el porcentaje de ceros observados sea elevado. Se tiene así que:

$$\begin{aligned} P(Y = 0) &= P(Z = 0) + P(Z = 1, Y^* = 0) = q_i + (1 - q_i)f(0) \\ P(Y = y_i > 0) &= P(Z = 1, Y^* = y_i) = (1 - q_i)f(y_i) \end{aligned} \tag{3}$$

donde q_i es la probabilidad que se asigna a la no participación, $q_i = P(z_i = 0)$, y $f(y_i)$ es la ley de Poisson, según la ecuación (1), o la ley binomial negativa, según la ecuación (2).

En el primer caso, se obtiene el modelo de Poisson inflado de ceros (*zero-inflated Poisson* o ZIP) y, en el segundo, el modelo binomial negativo inflado de ceros (*zero-inflated negative binomial* o ZINB). Evidentemente, si todos los individuos participaran ($q_i = 0$), se volvería al modelo tradicional correspondiente.

El modelo ZIP se expresa como:

$$\begin{aligned} P(Y_i = 0) &= q_i + (1 - q_i)e^{-\lambda_i} \\ P(Y = y_i > 0) &= (1 - q_i) \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \end{aligned} \quad (4)$$

donde $\lambda_i = e^{X_i\beta}$, como en los modelos anteriores.

La media y la varianza vienen dados ahora por $E(Y) = (1 - q_i)\lambda_i$ y $Var(Y) = \lambda_i(1 - q_i)(1 + \lambda_i q_i)$ y se observa que coinciden con las del modelo de Poisson si $q_i = 0$.

En cuanto al modelo ZINB, la distribución de probabilidades se formula como:

$$\begin{aligned} P(Y_i = 0) &= q_i + (1 - q_i) \left(\frac{\nu}{\nu + \lambda_i} \right)^\nu \\ P(Y = y_i > 0) &= (1 - q_i) \frac{\Gamma(y_i + \nu)}{\Gamma(y_i + 1) \cdot \Gamma(\nu)} \cdot \left(\frac{\nu}{\nu + \lambda_i} \right)^\nu \left(\frac{\lambda_i}{\nu + \lambda_i} \right)^{y_i} \end{aligned} \quad (5)$$

y los momentos $E(Y) = (1 - q_i)\lambda_i$ y $Var(Y) = (1 - q_i)(1 + \alpha\lambda_i + q_i\lambda_i)\lambda_i$. En caso de participación, estas expresiones se reducen a las del modelo binomial negativo; si $\alpha \rightarrow 0$, a las del Poisson inflado de ceros y si se dan ambas condiciones simultáneamente, se tendría el modelo de regresión de Poisson.

Existen distintas variantes de los modelos inflados de ceros, según la ley que se elija para q_i . El software con el que trabajamos, *Limdep* v.7.0, se basa en el supuesto de que se trata de una distribución logística, en función de un nuevo parámetro $\tau \in R$: $q_i = \Lambda(\tau X_i\beta)$ (Greene, 1995). Con esta formulación, el modelo tradicional y el inflado de ceros correspondiente no están anidados, lo que hace más complicada la elección entre uno u otro modelo.

2.3. Método de estimación y bondad del ajuste

El método de estimación que se utiliza para obtener los valores estimados de los parámetros es el de máxima verosimilitud, que consiste en adoptar como valores para los parámetros aquéllos para los que la función de verosimilitud $(\prod_{i=1}^N P(Y = y_i))$ o, de modo equivalente, su logaritmo $(\sum_{i=1}^N \ln P(Y = y_i))$ alcanza su valor máximo.

La significatividad de cada variable explicativa se contrasta a través del parámetro correspondiente, utilizando el estadístico z que sigue una distribución normal tipificada. Para un nivel de significación del 5%, valores de z mayores en valor absoluto que 1,96 indicarán significatividad de la variable asociada al parámetro en la estimación.

Para determinar la validez del modelo inflado de ceros frente al correspondiente modelo tradicional, se utiliza el estadístico que define Vuong (1989) como:

$$V = \frac{\sqrt{N} \left[\frac{1}{N} \sum_{i=1}^N m_i \right]}{\sqrt{\frac{1}{N} \sum_{i=1}^N (m_i - \bar{m})^2}}$$

donde $m_i = \ln \frac{P_1(Y = y_i)}{P_2(Y = y_i)}$, $P_1(Y = y_i)$ y $P_2(Y = y_i)$ son las funciones de probabilidad para los modelos inflado de ceros y tradicional, respectivamente, y $\bar{m}$ la media de m_i , $i = 1, \dots, N$.

Vuong demuestra que este estadístico es bidireccional y sigue una distribución asintótica normal tipificada. Valores de $|V|$ menores que 1,96 no permiten que nos decantemos por ninguno de los modelos, mientras que valores de V mayores que 1,96 son evidencias a favor del modelo inflado de ceros y valores menores que -1,96 favorecen el modelo tradicional.

3. Resultados del análisis

Como se ha mencionado anteriormente, el propósito del estudio empírico que realizamos es el de determinar si el comportamiento de los conductores difiere de una Comunidad Autónoma a otra, en lo que se refiere al número de siniestros que sufren y declaran a su compañía aseguradora. Además de llevar a cabo este análisis geográfico de la siniestralidad, prestaremos también atención al resto de variables que se muestran significativas en la estimación.

Los datos que se utilizan en el trabajo proceden de una entidad aseguradora privada que opera a nivel nacional en España. Por motivos computacionales, el estudio se ha llevado a cabo a partir de una muestra de 15.000 asegurados.

Exponemos a continuación algunos resultados del análisis descriptivo para pasar posteriormente a comentar los que se deducen de la aplicación del modelo elegido.

3.1. Análisis descriptivo

Las variables con las que trabajamos hacen referencia a características del vehículo asegurado (categoría y uso), características del asegurado (edad, sexo, experiencia en la conducción y región de residencia), características de la póliza del seguro (prima anual pagada y cobertura elegida) así como al número de siniestros que el conductor ha sufrido y ha declarado a su compañía aseguradora. El periodo del estudio abarca desde el 16 de junio de 2002 hasta el 15 de junio de 2003. En el anexo final del trabajo se incluye una breve descripción de todas las variables empleadas.

Nos detendremos aquí básicamente en lo relacionado con la región de residencia y el número de siniestros, que son las dos variables que nos interesan especialmente en este análisis. Puede verse un estudio descriptivo más pormenorizado en Melgar, Ordaz y Guerrero (2004).

La Figura 1 muestra la distribución del número de siniestros. Destaca la gran cantidad de ceros que presenta esta variable. En concreto, 11.558 conductores de los 15.000 que conforman la muestra no han sufrido y/o declarado siniestro alguno durante el periodo considerado, lo que supone un 77,05 % del total. Han sido

exactamente 5.015 los siniestros declarados, lo que representa una media de 1,46 entre los 3.442 conductores con siniestros.

Con respecto al lugar de residencia y circulación habitual del asegurado, hemos utilizado la clasificación por Regiones o nivel NUTS-1. Se trata de una división geográfica consensuada a nivel europeo por Eurostat y que, en el caso español, se corresponde con la agregación de varias Comunidades Autónomas. Así, son 7 en principio las Regiones tenidas en cuenta: Noroeste (Cantabria, Galicia y Principado de Asturias), Noreste (Aragón, Comunidad Foral de Navarra, La Rioja y País Vasco), Madrid (Comunidad de Madrid), Centro (Castilla y León, Castilla-La Mancha y Extremadura), Este (Cataluña, Comunidad Valenciana e Islas Baleares), Sur (Andalucía y Región de Murcia) y Canarias (Canarias). Hemos añadido sin embargo una región más, que no se tiene en cuenta en la clasificación de las NUTS-1, y que está formada por las ciudades autónomas de Ceuta y Melilla.

Según se observa en la Figura 2, el mayor peso dentro de la muestra lo soportan los asegurados de la Región Sur, con más del 45% del total; le siguen bastante emparejadas las Regiones Centro, Noroeste y Este, con porcentajes del 16,81%, 15,43% y 12,07%, respectivamente; Canarias y el Noreste representan solo un 4,38% la primera y un 3,21% del total, la segunda; por último, Madrid y Ceuta y Melilla son las Regiones menos representadas en la muestra, con porcentajes de 1,35% en el primer caso y 0,43%, únicamente, en el segundo.

Figura 1. Número de siniestros

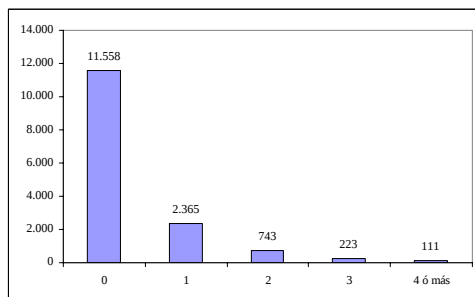
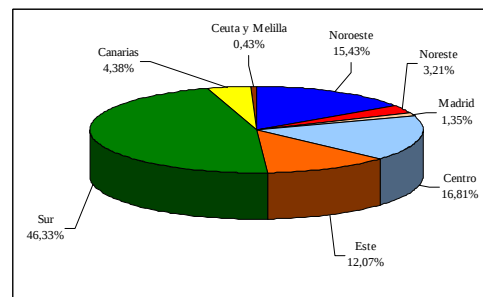


Figura 2. Región de residencia



En lo que se refiere a la relación entre Regiones y Número de siniestros, la Tabla 1 nos muestra los resultados desde el punto de vista descriptivo. Se observa que el porcentaje de no declaraciones oscila entre el 71,29% de la Región de Madrid y el

81,04% de la Región Centro. Por debajo de la media nacional, aunque ligeramente, y por tanto con una siniestralidad más elevada, se encuentran, además de la Región de Madrid, la Región Noreste, la Región Este, la Región Sur y la Región Ceuta y Melilla. El Noroeste, con un 77,28%, coincide prácticamente con lo que ocurre a nivel nacional (77,05% de valores nulos para la variable que indica el número de siniestros). Finalmente, con un mayor porcentaje de ceros que en el conjunto español y por tanto con menor siniestralidad, observamos la Región Canarias, además de la ya mencionada Región Centro.

Tabla 1. Distribución del número de siniestros según región de residencia

			NÚMERO DE SINIESTROS					
			0	1	2	3	4 ó más	TOTAL
REGIÓN	NOROESTE	Recuento	1789	360	110	37	19	2315
		% de Región	77,28%	15,55%	4,75%	1,60%	0,82%	100%
	NORESTE	Recuento	364	87	20	7	3	481
		% de REGIÓN	75,68%	18,09%	4,16%	1,46%	0,62%	100%
	MADRID	Recuento	144	44	12	2	0	202
		% de REGIÓN	71,29%	21,78%	5,94%	0,99%	0,00%	100%
	CENTRO	Recuento	2043	347	100	27	4	2521
		% de REGIÓN	81,04%	13,76%	3,97%	1,07%	0,16%	100%
	ESTE	Recuento	1369	291	99	35	17	1811
		% de REGIÓN	75,59%	16,07%	5,47%	1,93%	0,94%	100%
	SUR	Recuento	5284	1124	366	108	67	6949
		% de REGIÓN	76,04%	16,17%	5,27%	1,55%	0,96%	100%
	CANARIAS	Recuento	517	101	32	6	1	657
		% de REGIÓN	78,69%	15,37%	4,87%	0,91%	0,15%	100%
	CEUTA Y MELILLA	Recuento	48	11	4	1	0	64
		% de REGIÓN	75,00%	17,19%	6,25%	1,56%	0,00%	100%
Total		Recuento	11558	2365	743	223	111	15000
		% de REGIÓN	77,05%	15,77%	4,95%	1,49%	0,74%	100%

3.2. Estimación del modelo

En la tarea de identificar las variables que intervienen de manera significativa cuando se estima la siniestralidad en el seguro del automóvil en España, hemos aplicado el modelo binomial negativo inflado de ceros, realizando las estimaciones a través del software *Limdep v.7.0*.

Las aplicaciones de los modelos inflados de ceros al seguro del automóvil no son frecuentes. En los últimos años, Lee *et al.* (2002) han estudiado la siniestralidad de los conductores noveles australianos durante su primer año de conducción, aplicando el modelo ZIP. Con datos españoles, Melgar, Ordaz y Guerrero (2004) y Melgar y

Guerrero (2005) aplican el modelo ZINB para analizar los factores que influyen en la siniestralidad; con esos mismos datos, Melgar, Ordaz y Guerrero (2005) comparan las estimaciones obtenidas al aplicar tanto los modelos tradicionales como el modelo ZINB.

En el caso que nos ocupa ahora y desde el punto de vista teórico, nos decantamos por el modelo binomial negativo inflado de ceros por varios motivos. El primero de ellos es la sobre-dispersión que presentan los datos con los que trabajamos, y que se debe probablemente al elevado número de observaciones nulas que encontramos. La segunda razón de la elección del modelo señalado es la doble interpretación que podemos dar al valor nulo de la variable endógena. El número de siniestros sufridos y declarados por los asegurados será igual a 0 si el conductor no ha tenido ningún accidente, pero también puede serlo cuando, habiendo tenido alguno, ha decidido no dar conocimiento de ello a su compañía. Las diferencias cualitativas que conlleva el comportamiento más cuidadoso, por lo general, de los conductores que no sufren siniestros es otro de los motivos que nos lleva a decantarnos por un modelo inflado de ceros, ya que éstos tienen en cuenta todos estos matices. En la práctica, el valor del parámetro de sobre-dispersión Alpha y del estadístico de Vuong (Tabla 2) confirman la idoneidad de este modelo frente a los otros modelos *count data* descritos en el Apartado 2.

La relación de variables significativas, con un nivel de significación del 5%, cuando se aplica el modelo ZINB se incluye en la Tabla 2.

Si nos centramos en la variable que indica la Región de residencia del asegurado, observamos que únicamente las Regiones Noroeste (que incluye Cantabria, Galicia y Principado de Asturias) y Centro (Castilla y León, Castilla-La Mancha y Extremadura) aparecen entre los factores que muestran significatividad en el análisis llevado a cabo. Tanto en una como en otra, el coeficiente estimado indica que se espera que para los asegurados de dichas regiones el número de siniestros sufridos y declarados sea menor que para los conductores de la Región Sur, que hemos tomado como categoría base, y los de todas las demás regiones que no han resultado significativas. Además, en el caso de la región Centro, es mucho menor la siniestralidad que para la región Noroeste.

Este último resultado concuerda con lo que se observaba en el análisis descriptivo, en el que quedó de manifiesto que la Región Centro era la que presentaba un mayor porcentaje de observaciones nulas de la variable endógena y, por tanto, una menor siniestralidad. En cuanto a la Región Noroeste, estaba prácticamente en la media nacional, y solo muy ligeramente por debajo de ésta, en lo que se refiere al porcentaje de siniestros observados.

Entre las demás variables con significatividad en nuestro análisis, encontramos algunas categorías de vehículos que se comportan de modo distinto al resto, mostrando en concreto un menor número de siniestros. Son, ordenadas de menor a mayor siniestralidad, los remolques, los ciclomotores y motos y los camiones.

En función del uso al que se destinan, los vehículos diferenciados de los demás son los que se dedican al transporte escolar (con siniestralidad más alta) así como los que tienen uso industrial y los de uso agrícola (ambos con un número de accidentes inferior al del resto de vehículos).

Tabla 2. Estimación a través del modelo ZINB

Variable	Coefficiente estimado	Estadístico z
Constante	-1,220	-7,025
CAMION	-0,456	-3,527
REMOLQUE	-1,441	-3,637
CICL_MOT	-0,966	-8,049
USO_INDU	-0,666	-2,903
USO_TESC	0,600	2,966
USO_AGRI	-0,943	-4,977
EDAD	-0,002	-1,977
ANTIG_2A	0,320	2,094
NOROESTE	-0,168	-4,146
CENTRO	-0,885	-1,986
P200_300	0,724	4,818
P300_400	0,948	6,091
P400_500	1,085	6,696
P500_750	1,257	7,568
P750_	1,384	8,041
GR_MBAJO	0,175	4,770
GR_MALTO	0,153	3,342
GR_ALTO	0,222	3,633
Parámetros		
Alpha (sobredispersión)	0,047	2,180
Tau	-0,462	-3,197
Log-Máxima verosimilitud		-10.592,434
Estadístico de Vuong		9,056
Número de observaciones		15.000

La edad y la antigüedad en el carné de conducir tienen el comportamiento esperado: a mayor edad, menor siniestralidad; por otro lado, si hace menos de 2 años que el conductor obtuvo el carné, su siniestralidad será mayor que si tuviera más experiencia.

En cuanto a la prima, a medida que aumenta también lo hace el número de accidentes esperados; y, por último, en lo que se refiere al grado de cobertura del que disfruta el asegurado, se observa que el más elevado es el que conlleva una mayor siniestralidad.

4. Conclusiones

El gran número de vehículos nuevos matriculados cada año unido a la obligatoriedad de contratar un seguro, que cubra al menos la responsabilidad civil, son los dos factores fundamentales que convierten sin duda al seguro del automóvil en el ramo más importante del seguro no-vida y en uno de los principales en el conjunto del sector asegurador, tanto en España como en la mayoría de los países con economías desarrolladas.

Para los aseguradores, un buen ajuste de las primas resulta esencial para que puedan cubrirse sin problema los costes asociados a los siniestros declarados por los clientes. Es de gran importancia para conseguir un ajuste adecuado conocer una buena estimación del número de accidentes que pueden tener lugar, así como identificar los factores que intervienen en la siniestralidad.

En el trabajo que hemos presentado, hemos aplicado el modelo binomial negativo inflado de ceros para determinar, en particular, si la Comunidad Autónoma de residencia y circulación del asegurado es una de las variables que se muestran significativas en dicha estimación. Este modelo permite distinguir las dos posibles interpretaciones de la observación 0 de la variable endógena, el número de siniestros sufridos y declarados a la compañía aseguradora. En efecto, este valor puede, por un lado, indicar simplemente que el asegurado no ha sufrido siniestros durante el periodo analizado mientras que, por otro lado, también puede indicar que el conductor no ha declarado nada a su compañía, incluso si ha tenido uno o varios siniestros. Esto junto con la sobre-dispersión que presentan los datos, a consecuencia

del gran porcentaje de valores nulos observados, son los motivos que nos hacen pensar que, desde el punto de vista teórico, este modelo sea el más adecuado para nuestro propósito.

Desde el punto de vista empírico, esta idoneidad queda también ratificada por los valores de los coeficientes de sobre-dispersión y por el valor del estadístico a través del que se mide la bondad del ajuste.

En cuanto a los resultados concretos que deseábamos analizar, se comprueba que los asegurados de las Regiones Noroeste y Centro tienen un comportamiento diferenciado del resto, en lo que se refiere a la siniestralidad. En concreto, se aprecia que es menor el número de siniestros que sufren y declaran estos asegurados que el del resto de asegurados españoles, sobre todo en el caso de la Región Centro.

Destacamos también que se observa una mayor siniestralidad para aquellos conductores que disfrutan de un mayor grado de cobertura. Esto es sin duda muestra de la presencia de asimetría de información en los datos. Puede existir una componente de selección adversa, ya que los conductores que se saben más arriesgados intentan protegerse con el mayor número de garantías posibles; pero por otro lado, tampoco se puede descartar una componente de riesgo moral, ya que un conductor asegurado contra muchas contingencias será menos prudente que aquél que solo tiene un seguro a terceros.

Bibliografía

1. Böhning, D.; Dietz, E.; Schlattmann, P.; Mendonça, L. y Kirchner, U. (1999): “The Zero-Inflated Poisson Model and the Decayed, Missing and Filled Teeth Index in Dental Epidemiology”, *Journal of Royal Statistical Society A*, **162**, 2, pp. 195-209.
2. Boyer, M. y Dionne, G. (1989): “An Empirical Analysis of Moral Hazard and Experience Rating”, *Review of Economics and Statistics*, **71**, pp. 128-134.
3. Chiappori, P.A. y Salanié, B. (1997): “Empirical Contract Theory: The Case of Insurance Data”, *European Economic Review*, **41**, pp. 943-950.

4. Chiappori, P.A. y Salanié, B. (2000): "Testing for Asymmetric Information in Insurance Markets", *Journal of Political Economy*, **108**, 1, pp. 56-78.
5. Cohen, A. (2002): "Asymmetric Information and Learning: Evidence from the Automobile Insurance Market", *Discussion Paper* nº371, Harvard Law School, Cambridge.
6. DGT (2004): Anuario Estadístico General 2003. Dirección General de Tráfico, Ministerio del Interior, Madrid.
7. Dionne, G. y Laberge-Nadeau, C. (1999): *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation*, Kluwer Academic Publishers.
8. Greene, W.H. (1995): *Limdep Version 7.0: User's Manual*, Econometric Software.
9. ICEA (2004): *Evolución del Mercado Asegurador. Estadística Año 2003*, Informe nº917. ICEA, Madrid.
10. Jones, A.M. (2001): *Applied Econometrics for Health Economists - A Practical Guide*, Office of Health Economics.
11. Lee, A.H.; Stevenson, M.R.; Wang, K. y Yau, K.K.W. (2002): "Modeling Young Driver Motor Vehicle Crashes: Data with Extra Zeros", *Accident Analysis and Prevention*, **34**, pp. 515-521.
12. Melgar, M.C.; Ordaz, J.A. y Guerrero, F.M. (2004): "The Main Determinants of the Number of Accidents in the Automobile Insurance: an Empirical Analysis", *Etudes et Dossiers*, **286**, pp. 45-56.
13. Melgar, M.C. y Guerrero, F.M. (2005): "Los Siniestros en el Seguro del Automóvil: un Análisis Econométrico Aplicado", *Estudios de Economía Aplicada*, pendiente de publicación.
14. Melgar, M.C.; Ordaz, J.A. y Guerrero, F.M. (2005): "Diverses Alternatives pour Déterminer les Facteurs Significatifs de la Fréquence d'Accidents dans l'Assurance Automobile", *Assurances et Gestion des Risques*, pendiente de publicación.

15. Puelz, R. y Snow, A. (1994): "Evidence on Adverse Selection: Equilibrium Signalling and Cross-Subsidization in the Insurance Market", *Journal of Political Economy*, **102**, 2, pp. 236-257.
16. Richaudeau, D. (1999): "Automobile Insurance Contracts and Risk of Accident: An Empirical Test Using French Individual Data", *Geneva Papers on Risk and Insurance Theory*, **24**, pp. 97-114.
17. Vuong, Q.H.: (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses", *Econometrica*, **57**, pp. 307-333.
18. Winkelmann, R. (2003): *Econometric Analysis of Count Data*, Springer.
19. Yau, K.K.W.; Wang, K. y Lee, A.H. (2003): "Zero-Inflated Negative Binomial Mixed Regression Modeling of Over-Dispersed Count Data with Extra Zeros", *Biometrical Journal*, **45**, 4, pp. 437-452.

ANEXO: VARIABLES UTILIZADAS EN EL ANÁLISIS ECONÓMétrICO

<i>Categoría del vehículo asegurado</i>
- TUR_FUR = 1 si el vehículo asegurado es un turismo o una furgoneta, 0 en caso contrario (categoría base) - CAMION = 1 si el vehículo asegurado es un camión, 0 en caso contrario - REMOLQUE = 1 si el vehículo asegurado es un remolque, 0 en caso contrario - AUTOCAR = 1 si el vehículo asegurado es un autocar, 0 en caso contrario - TRAC_MA = 1 si el vehículo asegurado es un tractor o una maquinaria agrícola, 0 en caso contrario - VEH_IND = 1 si el vehículo asegurado es un vehículo industrial, 0 en caso contrario - CICL_MOT = 1 si el vehículo asegurado es un ciclomotor o una moto, 0 en caso contrario
<i>Uso del vehículo asegurado</i>
- USO_PART = 1 si el uso del vehículo asegurado es el uso particular, 0 en caso contrario (categoría base) - USO_SP = 1 si el uso del vehículo asegurado es el servicio público, 0 en caso contrario - USO_ALQU = 1 si el uso del vehículo asegurado es el alquiler, 0 en caso contrario - USO_ESCU = 1 si el uso del vehículo asegurado es escuela de conductores, 0 en caso contrario - USO_COMP = 1 si el uso del vehículo asegurado es la compra-venta, 0 en caso contrario - USO_INDU = 1 si el uso del vehículo asegurado es el industrial, 0 en caso contrario - USO_TMER = 1 si el uso del vehículo asegurado es el transporte de mercancías, 0 en caso contrario - USO_TESC = 1 si el uso del vehículo asegurado es el transporte escolar, 0 en caso contrario - USO_TGV = 1 si el uso del vehículo asegurado es el transporte general de viajeros, 0 en caso contrario - USO_AGRI = 1 si el uso del vehículo asegurado es el agrícola propio, 0 en caso contrario - USO_RPC = 1 si el uso del vehículo asegurado es la retirada de permiso de conducir, 0 en caso contrario
<i>Edad del asegurado</i>
- EDAD = edad del asegurado a fecha 15 de diciembre de 2002
<i>Experiencia del asegurado como conductor</i>
- ANTIG_2A = 1 si el asegurado obtuvo el carné de conducir hace menos de 2 años (a fecha 15 de diciembre de 2002), 0 en caso contrario

<i>Sexo del asegurado</i>
- MUJER = 1 si el asegurado es mujer, 0 en caso contrario
<i>Región de residencia del asegurado</i>
- NOROESTE = 1 si el asegurado reside en la región Noroeste (Cantabria, Galicia, Principado de Asturias), 0 en caso contrario - NORESTE = 1 si el asegurado reside en la región Noreste (Aragón, Comunidad Foral de Navarra, La Rioja, País Vasco), 0 en caso contrario - MADRID = 1 si el asegurado reside en la región de Madrid (Comunidad de Madrid), 0 en caso contrario - CENTRO = 1 si el asegurado reside en la región Centro (Castilla y León, Castilla-La Mancha, Extremadura), 0 en caso contrario - ESTE = 1 si el asegurado reside en la región Este (Cataluña, Comunidad Valenciana, Islas Baleares), 0 en caso contrario - SUR = 1 si el asegurado reside en la región Sur (Andalucía, Región de Murcia), 0 en caso contrario (categoría base) - CANARIAS = 1 si el asegurado reside en la región de Canarias (Canarias), 0 en caso contrario - CEU_MELI = 1 si el asegurado reside en la región de Ceuta y Melilla (ciudades autónomas de Ceuta y Melilla), 0 en caso contrario
<i>Prima anual pagada por el asegurado</i>
- P0_200 = 1 si la prima anual pagada por el asegurado no supera los 200 €, 0 en caso contrario (categoría base) - P200_300 = 1 si la prima anual pagada por el asegurado es mayor de 200 € y no supera los 300 €, 0 en caso contrario - P300_400 = 1 si la prima anual pagada por el asegurado es mayor de 300 € y no supera los 400 €, 0 en caso contrario - P400_500 = 1 si la prima anual pagada por el asegurado es mayor de 400 € y no supera los 500 €, 0 en caso contrario - P500_750 = 1 si la prima anual pagada por el asegurado es mayor de 500 € y no supera los 750 €, 0 en caso contrario - P750_ = 1 si la prima anual pagada por el asegurado es mayor de 750 €, 0 en caso contrario
<i>Cobertura del seguro</i>
- GR_BAJO = 1 si el asegurado disfruta del grado de cobertura bajo, 0 en caso contrario (categoría base) - GR_MBAJO = 1 si el asegurado disfruta del grado de cobertura medio-bajo, 0 en caso contrario - GR_MALTO = 1 si el asegurado disfruta del grado de cobertura medio-alto, 0 en caso contrario - GR_ALTO = 1 si el asegurado disfruta del grado de cobertura alto, 0 en caso contrario

<i>Número de siniestros sufridos y declarados por el asegurado</i>
- NUMSIN = número de accidentes sufridos por el asegurado y declarados a la compañía de seguros entre el 16 de junio de 2002 y el 15 de junio de 2003 (variable dependiente)