ESTUDIOS DE ECONOMÍA APLICADA VOL. 23-1, 2005. PÁGS. 355-375

Los siniestros en el seguro del automóvil: un análisis econométrico aplicado

MELGAR HIRALDO, M.C.(*) Y GUERRERO CASAS, F.M.
Departamento de Economía, Métodos Cuantitativos e Historia Económica.
Universidad Pablo de Olavide.
Ctra. de Utrera, Km. 1; 41013-SEVILLA. Tfno.: 954348548. (*)E-mail: mcmelhir@upo.es

RESUMEN

Este trabajo tiene como objetivo determinar cuáles son los principales factores determinantes del número de siniestros en el seguro del automóvil, a la vez que comprobar si, como prevé la teoría, las coberturas más elevadas están correlacionadas positivamente con un mayor número de siniestros. Utilizamos en este análisis empírico datos cedidos por una entidad aseguradora privada y aplicamos, para estimar la siniestralidad, un planteamiento reciente de los modelos econométricos tipo *count data* poco explotado en el seguro del automóvil: el modelo binomial negativo inflado de ceros. De este modo, se consigue distinguir entre las observaciones nulas que corresponden a conductores que no declaran los siniestros que sufren y las que realmente significan que no ocurre ningún accidente. Se concluye qué tipos de vehículos sufren un mayor número de siniestros, a qué uso se destinan y cómo están relacionadas con la siniestralidad variables como la edad, la experiencia en la conducción del asegurado, la prima pagada o el tipo de cobertura elegida.

Palabras clave: siniestro, seguro del automóvil, modelos *count data*, modelo binomial negativo inflado de ceros.

The accidents in the automobile insurance: an applied econometric analysis

ABSTRACT

The objective of this paper is to determine what the most important factors for the number of accidents in the automobile insurance are. We try to verify if the theory is right and the highest levels of coverage are positively correlated with a greater number of accidents. Data used in this empirical analysis have been provided by a private insurance company. In order to estimate the number of accidents, we apply a recent approach for considering count data situations, not very used in the automobile insurance: a zero-inflated negative binomial model. So we can make a distinction between the null observations from drivers that do not declare any accidents and the ones recorded when truly no accident happens. We finally show what types of vehicles suffer a greater number of accidents, what use they are devoted to, and how much related are accident variables and others: the age, the policyholders' driving experience, the premiums paid, or the different levels of insurance coverage.

Keywords: accident, automobile insurance, count data models, zero-inflated negative binomial model.

Clasificación JEL: C52, G22.

Artículo recibido en junio de 2004. y aceptado para su publicación en marzo de 2005.
La referencia electrónica de este artículo en la página www.revista-eea.net, ref.: e-23110

ISSN 1697-5731 (online) – ISSN 1133-3197 (print)

1. INTRODUCCIÓN

Desde que se inició, en los años 70, el análisis teórico de las relaciones contractuales en las que existe información asimétrica entre las partes, han sido muchos los estudios que han surgido al respecto. Sin embargo, la estimación empírica de los modelos de seguro con selección adversa o riesgo moral es mucho más reciente. Uno de los principales objetivos en el tratamiento empírico de la asimetría de información ha sido la comprobación de que las coberturas más elevadas están positivamente correlacionadas con un mayor número de accidentes. El mercado del seguro del automóvil ha resultado ser uno de los más propicios para la realización de estos estudios, aunque también se encuentran algunos análisis similares en otros mercados.

Entre las primeras aportaciones en este sentido, podemos mencionar a Dahlby (1983) y Boyer y Dionne (1989), ambas centradas en el mercado del seguro del automóvil. En estos dos trabajos, se obtienen resultados que no permiten descartar la existencia de información asimétrica. Sin embargo, los datos que usa Dahlby (1983) son datos agregados y no está claro que considerando datos individuales se siguieran dando los mismos resultados. Dahlby (1992) también trabaja con datos agregados del seguro del automóvil de Canadá.

Puelz y Snow (1994) fueron los primeros en utilizar datos individuales; en este caso, de un asegurador de Georgia. Entre los estudiosos del tema, se considera éste como punto de partida para el desarrollo de este campo de investigación. Puelz y Snow construyen un modelo de dos ecuaciones. La primera representa la política de precios de la compañía aseguradora. Mediante la estimación de la prima como función de la franquicia elegida y de otras variables específicas del individuo, comprueban que las primas más elevadas están asociadas a franquicias más bajas. En la segunda ecuación, describen la franquicia elegida a partir de la prima, de las características individuales del asegurado y de una variable ficticia que toma el valor 1 si el individuo ha sufrido accidentes y 0 en caso contrario. Se obtiene en la estimación un coeficiente negativo para esta variable *dummy*, lo que muestra que quienes sufren siniestros eligen franquicias más bajas y, por tanto, coberturas más elevadas.

Las principales críticas al trabajo de Puelz y Snow hacen referencia al uso de modelos lineales en la estimación de las ecuaciones y a la no inclusión de variables que afectan al riesgo, concretamente los años de conducción del asegurado.

Chiappori y Salanié (1997) proponen una aproximación muy general que, potencialmente, podría ser aplicable a cualquier situación con información asimétrica. La idea consiste en estimar simultáneamente dos ecuaciones no lineales: una sobre la elección de la franquicia en función únicamente de las variables exógenas específicas del individuo, sin incluir, a diferencia de Puelz y Snow, la ficticia sobre los siniestros ni la prima, y una segunda en la que la variable artificial que indica si ocurre o no siniestro se toma como variable dependiente. Estimando ambas ecuaciones simultáneamente, se determina si la siniestralidad está asociada o no a una mayor cobertura.

En relación con esto, Richaudeau (1999) hace una importante aportación, estimando el número de siniestros a través de un modelo de tipo recuento o *count data* (concretamente el modelo binomial negativo) en lugar de considerar únicamente la existencia o no de siniestro.

Los trabajos de Chiappori y Salanié (1997, 2000b) y Richaudeau (1999), entre otros, todos ellos con datos del seguro del automóvil en Francia, no muestran correlación entre siniestros y cobertura. Es importante resaltar, sin embargo, que Chiappori y Salanié basan su estudio únicamente en individuos con pocos años de experiencia en la conducción. La no existencia de correlación cobertura-accidentes para estos conductores, que no han tenido por tanto mucha oportunidad de adquirir una ventaja informativa respecto al asegurador sobre su tipo de riesgo, no implica necesariamente que dicha correlación no exista cuando la antigüedad del carné de conducir es mayor.

Cohen (2002) dispone de una base de datos muy rica procedente de Israel, relativa a asegurados con experiencia diversa en la conducción. Confirma la conclusión de Chiappori y Salanié sobre la ausencia de correlación cobertura-accidentes en conductores noveles, pero sí encuentra que dicha correlación existe para los asegurados con más de dos años de experiencia. El estudio que hace Cohen es más completo que el de Chiappori y Salanié, ya que tiene en cuenta también el aprendizaje por parte del asegurador en cuanto al tipo de riesgo del asegurado, conseguido a través de una relación duradera entre ambos. Sin embargo, Cohen aplica un modelo Poisson que, debido al elevado porcentaje de asegurados que no sufren siniestros, no resulta el más adecuado.

Al igual que los estudios referidos, el trabajo que presentamos se enmarca en el seguro del automóvil. El objetivo que nos planteamos no es únicamente el de analizar la posible correlación entre la siniestralidad de los asegurados y la cobertura que eligen, sino determinar además las variables explicativas que se muestran significativas en la estimación del número de siniestros que los asegurados declaran a su compañía. No vamos a considerar la posibilidad de fraude, que puede existir y que falsearía estas cifras. El principal referente en el estudio de este aspecto en el seguro del automóvil en España es Artís *et al.* (1999), que usa modelos de elección discreta, en concreto modelos logísticos multinomiales y anidados, para cuantificar la probabilidad de aparición de diferentes tipos de fraude.

En nuestro trabajo, aplicaremos un modelo *count data* inflado de ceros, para combatir los problemas de sobre-dispersión que ocasiona el elevado número de ceros de la variable endógena, así como para incluir todos los matices de ese valor nulo. Como en gran parte de los trabajos en los que se usan estas modelizaciones, haremos el desarrollo desde un punto de vista frecuentista. Cabe reseñar, sin embargo, que estos modelos pueden presentar múltiples modas, lo que puede dificultar su estimación. La aproximación bayesiana del problema que desarrollan Scollnik (1998) y Angers y Biswas (2003), entre otros, y el uso de una distribución a priori informativa, pueden conseguir una reducción de los pesos de la moda de la verosimilitud, lo que supondría una ventaja computacional, y permitirían asimismo obtener predicciones para años posteriores que no se consiguen con el modelo frecuentista.

Los modelos inflados de ceros han tenido en los últimos años un gran auge. En ámbitos como el biométrico y de la salud, se encuentran a menudo datos que contienen un número de ceros demasiado elevado para que pueda usarse de modo conveniente la distribución tradicional de Poisson. El modelo de Poisson inflado de ceros es una alternativa que ha recibido una considerable atención en este campo. Pero su inflexibilidad, al suponer que los datos no nulos siguen una distribución de Poisson truncada en el cero, hace que la alternativa del modelo binomial negativo inflado de ceros pueda resultar más apropiada. En este sentido, Ridout *et al.* (2001) proponen un test para contrastar la validez de este último modelo frente al primero.

Entre las aplicaciones de los modelos inflados de ceros, destacamos, entre otras, la de Shankar *et al.* (1997), que se inclinan por la aplicación del modelo de Poisson para combatir la sobre-dispersión que provoca el elevado número de ceros que aparecen al analizar la frecuencia de accidentes en las calzadas, ya sea por la existencia de tramos de carreteras muy seguros o por no haberse registrado ningún accidente durante el periodo en estudio en un tramo inseguro; Böhning *et al.* (1999) usan ese mismo modelo en el estudio de la prevención de caries en epidemiología dental; Lee *et al.* (2002) aplican además el modelo binomial negativo, obteniendo en ambos casos resultados bastantes similares, para observar los factores que influyen en los accidentes que sufren los conductores durante los 12 meses siguientes a la obtención del carné de conducir; y, en el ámbito de la salud, Yau, Wang y Lee (2003) introducen efectos aleatorios en el modelo binomial negativo inflado de ceros, para analizar la duración de la hospitalización de un grupo de enfermos de páncreas.

En el trabajo que presentamos, describimos brevemente en el apartado 2, y tras esta introducción, la base de datos con la que hemos trabajado. Seguidamente, señalamos en el apartado 3 las principales características de los modelos *count data* que se utilizan habitualmente en estimaciones de este tipo, justificando a nivel teórico la elección del modelo binomial negativo inflado de ceros en nuestro caso. A continuación, los resultados de la estimación se muestran en el apartado 4, tras lo cual se incluyen unas breves conclusiones en el apartado 5. El trabajo finaliza con una selección de reseñas bibliográficas sobre el tema.

2. ESTUDIO EMPÍRICO

La base de datos que se emplea en la aplicación empírica que recoge el apartado 4 procede del ramo del automóvil de una aseguradora privada española. Se trata de una entidad que opera a nivel nacional y que tiene asegurados en todas las provincias, aunque en el conjunto de los datos disponibles, se observa una mayor representación de Sevilla y Córdoba, con participaciones del 12,1% y 10,2% respectivamente, seguidas de Murcia, La Coruña y Badajoz, con algo más de un 6% cada una. Las provincias que destacan por una menor presencia son Teruel y Soria (con un número prácticamente insignificantes de pólizas en ambos casos) así como Segovia, Palencia, Ceuta, Lérida, Huesca y Cuenca, con 1% de asegurados cada una aproximadamente, sobre el total.

Son 63.900 los registros de los que disponemos en principio, si bien tras depurar los datos nos quedamos finalmente con los correspondientes a 60.000 pólizas de vehículos asegurados entre el 16 de junio de 2002 y el 15 de junio de 2003. La información se ha clasificado en cuatro categorías diferenciadas:[1] características del vehículo asegurado, características del asegurado, características de la póliza de seguro y características de los siniestros acontecidos y declarados, definiéndose las variables que aparecen en el anexo final del trabajo.

Se muestra a continuación de modo resumido la composición de la base de datos.[2]

Con respecto al tipo de vehículo, más del 80% de los vehículos pertenecen al grupo *Turismo o furgoneta*; también destacan sobre las demás las categorías *Ciclomotor o moto* y *Tractor o maquinaria agrícola*, con alrededor de un 7% cada una. Las restantes categorías apenas superan en conjunto el 5% de los asegurados.

En cuanto al uso del vehículo, sobresale el uso *Particular* al que se destina casi el 80% de los vehículos asegurados, seguido del uso *Transporte de mercancías*, casi el 10% de los vehículos, y el uso *Agrícola propio*, con un 8%.

Las características del asegurado que resultan de interés para las aseguradoras españolas son, fundamentalmente, la edad, el sexo, la antigüedad del carné de conducir y la zona de circulación habitual. La edad media entre los conductores es aproximadamente de 48 años; por tramos, solo el 2,26% son menores de 25 años, el 5,59% tienen al menos 25 pero menos de 30 años y el 92,15% son asegurados con 30 o más años. En la gran mayoría de los casos, la experiencia en la conducción es de 2 años al menos. Menos del 1% sacó el carné de conducir hace menos de 2 años. Y en lo que se refiere al sexo, alrededor del 85% de los asegurados son hombres, frente al 15% restante de mujeres.

Se han considerado 3 zonas de circulación del vehículo, en función de la "peligrosidad" contrastada experimentalmente de la provincia de residencia del asegurado. Finalmente, observamos que sobre el 31% de los vehículos asegurados circulan por la zona con menor probabilidad de accidente, el 43% por la zona intermedia y el

[1] Dionne, Gouriéroux y Vanasse (1999) y Cohen (2002), entre otros, usan una clasificación similar. Hay que señalar, no obstante, que la información que recopilan las aseguradoras en otros países sobre los conductores y vehículos es mucho más completa que la que tienen las aseguradoras españolas.

[2] Para un análisis descriptivo más detallado, consultar Melgar (2004).

resto, aproximadamente el 26%, por la zona con mayor propensión a los siniestros.

En relación con la prima, podemos decir que casi la mitad oscilan entre los 200 y 400 € anuales. Las primas más bajas son las menos frecuentes (menos del 10%) y, por encima de los 300 €, a medida que aumenta la prima va disminuyendo el porcentaje de asegurados que la pagan.

Con respecto a la cobertura, hemos definido cuatro posibles grados o niveles, en función de las garantías contratadas en cada póliza, según se observa en la Tabla 1 y la Tabla 2. Globalmente, más de la mitad de los asegurados disfrutan del grado más bajo de cobertura. A medida que ésta aumenta, disminuye la proporción de conductores que la contratan. Frente al 27,55% que tienen un nivel de cobertura medio-bajo, solo un 10,85% elige el grado medio-alto y un 7,71% el nivel más elevado de cobertura.

Tabla 1. Grado de cobertura para turismos y furgonetas

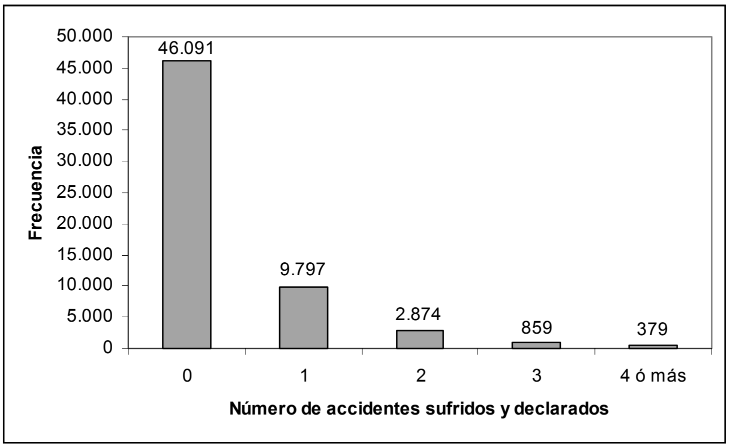
Grado de cobertura	Garantías
Bajo (1)	responsabilidad civil obligatoria, responsabilidad civil suplementaria, defensa y reclamaciones, muerte, invalidez, asistencia
Medio-bajo (2)	grado 1 + incendio y/o rotura de lunas
Medio-alto (3)	grado 1 + robo y/o retirada permiso de conducir
Alto (4)	grado 1 + daños propios, daños por colisión o pérdida total

Tabla 2. Grado de cobertura para el resto de categorías

Grado de cobertura	Garantías
Bajo (1)	responsabilidad civil obligatoria, responsabilidad civil suplementaria, defensa y reclamaciones
Medio-bajo (2)	grado 1 + muerte, invalidez, y/o asistencia
Medio-alto (3)	grado 1 + robo, incendio, roturas y/o retirada permiso de conducir
Alto (4)	grado 1 + daños propios, daños por colisión o pérdida total

Como se ha indicado, el objetivo principal de este trabajo es el de determinar las variables que resultan significativas al estimar el número de siniestros sufridos y declarados a la compañía por los asegurados. Se muestra en la Figura 1 la distribución de dicha variable. Destacamos, como dato más relevante, que para el 76,82% de los asegurados no consta ningún siniestro a lo largo del periodo que hemos tomado como referencia.

Figura 1. Distribución del número de siniestros



3. MÉTODOS

Los modelos econométricos más adecuados para estimar variables discretas con valores no negativos son los de tipo recuento o *count data*. El modelo de regresión de Poisson y el modelo binomial negativo son los más utilizados tradicionalmente. Existen sin embargo otros modelos a través de los que es posible estimar también variables que cuentan número de sucesos, como son, por ejemplo, los modelos inflados de ceros. Se lleva a cabo a continuación un breve repaso de las principales características de cada uno de ellos[3], haciendo especial hincapié en los aspectos diferenciadores para justificar la elección del que pensamos resulta ser el más adecuado para la estimación de la variable que nos ocupa: el número de siniestros declarados por los conductores a su aseguradora.

3.1. Modelo de regresión de Poisson

Se pretende estimar una variable aleatoria discreta Y para cada uno de los individuos de una muestra, en función K de variables explicativas (incluido el término independiente), que denotamos X_κ , $\kappa = 1,2,... K$.

El modelo de Poisson estima la probabilidad de que la variable aleatoria Y tome el valor y_i para el individuo i de modo que:

[3] Para un análisis más detallado, ver Jones (2001) o Winkelmann (2003).

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

donde $y_i = 0,1,2,\ldots$ para cada individuo, X_i es el vector de regresores, β el de los coeficientes a estimar y el parámetro λ_i verifica $\ln \lambda_i = X_i \beta$ o, de modo equivalente, $\lambda_i = e^{X_i \beta}$.

El supuesto implícito de igualdad entre la media y la varianza condicionales de la variable Y es la característica principal de este modelo ($E(Y|_{X_i}) = \lambda_i = Var(Y|_{X_i})$). Es un supuesto fuerte que supone una seria restricción para la aplicación del método. Es muy normal que no se dé equi-dispersión en el tipo de variable que estamos considerando; en concreto, suele tenerse sobre-dispersión, es decir, el valor de la varianza de la variable endógena es mayor que el de su media. El elevado número de valores nulos que presentan los datos suele ser motivo de ello. Este hecho provoca sesgos, por lo que es preciso comprobar si la hipótesis de equi-dispersión es aceptable. En caso de rechazarse, será necesario emplear una distribución de probabilidad más flexible, recurriéndose habitualmente a la binomial negativa.

3.2. Modelo binomial negativo

El modelo binomial negativo representa una alternativa común al modelo de regresión de Poisson que permite abordar situaciones de recuento de datos cuando éstos muestran una gran heterogeneidad, pudiendo presentar, en particular, un elevado número de ceros, como es el caso de la variable que queremos estimar.

Una posible explicación de que exista sobre-dispersión y exceso de valores cero en determinadas variables dependientes es lo que Mullahy (1997) denomina *heterogeneidad inobservable*. La heterogeneidad adicional entre los individuos, que no reflejan las diferencias en las variables explicativas, conducen a una mayor dispersión en la distribución. El modelo binomial negativo contempla esta heterogeneidad inobservable y por ello se convierte en el más apropiado para estos casos.

Se supone en este modelo que la variable $Y|_{X_i}$ sigue una distribución de Poisson con parámetro μ_i tal que $\ln \mu_i = \ln \lambda_i + \varepsilon_i$ donde, de nuevo, $\lambda_i = e^{X_i \beta}$ y e^{ε_i} sigue una distribución gamma de parámetros ν, ν con lo que $E(e^{\varepsilon_i}) = 1$ y $Var(e^{\varepsilon_i}) = \frac{1}{\nu}$. Así, la probabilidad de que la variable Y tome el valor y_i para el individuo i puede calcularse según la expresión:

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

donde $\Gamma(.)$ es la función gamma.[4]

Si introducimos el llamado parámetro de precisión[5], dado por $\alpha = \frac{1}{\nu}$, se puede demostrar que $E(Y|_{X_i}) = \lambda_i$ y $Var(Y|_{X_i}) = \lambda_i + \alpha\lambda_i^2$. Observamos así cómo, en este modelo, la varianza condicional de es una función cuadrática de su media condicional.

Además, si $\alpha \to 0$ la media y la varianza de la variable dependiente coinciden, con lo que se obtiene el modelo de Poisson. Por tanto, puede considerarse como una medida del grado de sobre-dispersión que presentan los datos y contrastando su significatividad, se determina la validez del modelo de Poisson o, por el contrario, la necesidad de elegir un modelo alternativo.

3.3. Modelo de Poisson inflado de ceros

Un planteamiento más reciente que los expuestos hasta ahora, para abordar el análisis de datos de recuento con una gran cantidad de ceros, consiste en usar distribuciones *infladas de ceros*. Éstas se basan en la hipótesis de que la población consta de dos tipos de individuos: para uno de ellos, el resultado de la observación siempre es "0"; para el otro, los datos se generan según la distribución tradicional que se considere (Poisson en este apartado, binomial negativo en el siguiente), pudiendo por tanto tomar también el valor "0". Se puede decir que, de algún modo, subyace una decisión inicial de participación en el proceso. La posibilidad de obtener observaciones nulas por estas dos vías conlleva, en estos casos, que el porcentaje de ceros observados sea elevado.

Si q_i es la probabilidad que se asigna a los no participantes (para los que la observación siempre es "0") y $1 - q_i$ la de los participantes (para los cuáles la probabilidad correspondiente se obtiene según la ley de Poisson), el modelo de Poisson inflado de ceros (ZIP) puede expresarse como:

$$P(Y_i = 0) = q_i + (1 - q_i)e^{-\lambda_i}$$
$$P(Y = y_i > 0) = (1 - q_i)\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \qquad [3]$$

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

---

[4] Según la terminología de Cameron y Trivedi (1986) este modelo es el binomial negativo II y es un caso particular (para $k = 0$) del modelo más general que se obtiene tomando $\nu_i = \frac{1}{\alpha}\left(e^{X_i\beta}\right)^k$. Para $k = 1$, se tendría el modelo binomial negativo I, en el que la media y la varianza condicionales de Y son proporcionales.

[5] Ver Jones (2001).

A partir de esta formulación, se obtienen las igualdades $E(Y|_{X_i}) = (1-q_i)\lambda_i$ y $Var(Y|_{X_i}) = \lambda_i(1-q_i)(1+\lambda_i q_i)$, que coinciden con los momentos dados para el modelo de Poisson si $q_i = 0$.

Dependiendo de la ley que se elija para q_i, surgen distintas variantes del modelo ZIP. El software con el que trabajamos usa la conocida como ZIP(τ), basada en el supuesto de que q_i sigue una distribución logística en función de un nuevo parámetro $\tau \in R$: $q_i = \Lambda(\tau X_i \beta)$ (Greene, 1995). Esta formulación tiene el inconveniente de que el modelo tradicional y el inflado de ceros no están anidados, con lo cual resulta más complejo determinar si la distribución es realmente Poisson o Poisson inflada de ceros, de lo que lo sería si únicamente hubiera que contrastar la nulidad de un parámetro.

3.4. Modelo binomial negativo inflado de ceros

Siguiendo un razonamiento similar al detallado en el apartado anterior y tomando ahora como referencia el modelo binomial negativo, obtenemos la distribución de probabilidades para el modelo binomial negativo inflado de ceros (ZINB), dada por:

$$
\begin{aligned}
P(Y_i = 0) &= q_i + (1-q_i)\left(\frac{\nu}{\nu+\lambda_i}\right)^{\nu} \\
P(Y = y_i > 0) &= (1-q_i)\frac{\Gamma(y_i+\nu)}{\Gamma(y_i+1)\cdot\Gamma(\nu)}\cdot\left(\frac{\nu}{\nu+\lambda_i}\right)^{\nu}\left(\frac{\lambda_i}{\nu+\lambda_i}\right)^{y_i}
\end{aligned}
\qquad [4]
$$

donde, una vez más, $\lambda_i = e^{X_i\beta}$ y q_i es la probabilidad de no participación.

Los momentos vienen dados en este caso por las expresiones $E(Y|_{X_i}) = (1-q_i)\lambda_i$ y $Var(Y|_{X_i}) = \lambda_i(1-q_i)(1+\alpha\lambda_i+\lambda_i q_i)$, siendo de nuevo α el parámetro de precisión ya definido como $\alpha = \frac{1}{\nu}$. Si $\alpha \to 0$, dichas expresiones son similares a los del modelo de Poisson inflado de ceros; mientras que si $q_i = 0$, coinciden con las del modelo binomial negativo.

Por tanto, el contraste del valor nulo de los parámetros α y q_i sería el que determinaría, en la práctica, cuál es el modelo más adecuado para nuestros datos. Sin embargo, si consideramos, como en el caso del modelo ZIP(τ), que q_i sigue una distribución logística, $q_i = \Lambda(\tau X_i \beta)$, $\tau \in R$, resulta de nuevo que el modelo binomial negativo y el correspondiente inflado de ceros no están anidados.

No se encuentran habitualmente aplicaciones de estos modelos en el seguro del automóvil. Sin embargo, los modelos binomiales negativos inflados de ceros son los que parecen más apropiados para la estimación que queremos realizar. En nuestro caso, el valor 0 de la variable endógena *NUMSIN* puede deberse a que el conductor realmente no ha sufrido siniestros o a que, habiendo tenido alguno, no ha dado conocimiento de ello a su compañía de seguros. Además, el grupo de conductores sin siniestros suele mostrar, por lo general, un comportamiento más cuidadoso que el resto, lo que conlleva diferencias cualitativas con respecto a los que sí los sufren. Esto, junto con el alto porcentaje de ceros que, como hemos visto en el apartado anterior, presenta la variable dependiente, nos lleva a la elección desde el punto de vista teórico del modelo binomial negativo inflado de ceros para llevar a cabo el análisis econométrico señalado.

3.5. Método de estimación

Para obtener los valores estimados de los parámetros, se utilizará el método de máxima verosimilitud, que proporciona estimadores consistentes y asintóticamente eficientes.

Este método consiste en adoptar como valores para los parámetros aquéllos para los que se maximiza la función de verosimilitud o, equivalentemente, su logaritmo. La función de verosimilitud no es más que la función de probabilidad conjunta que se deduce de cada modelo y que, de forma general, podemos expresar como $\prod_{i=1}^{N} P(Y = y_i)$. Su logaritmo viene dado por $\sum_{i=1}^{N} \ln P(Y = y_i)$.

En el caso del modelo de Poisson, el estimador de máxima verosimilitud para el parámetro β se obtiene maximizando la función $L(\beta) = \sum_{i=1}^{N} y_i X_i\beta - e^{X_i\beta} - \ln(y_i!)$.

Para el modelo binomial negativo, hay que estimar además el parámetro α. En ese caso, el logaritmo de la función de verosimilitud cuya maximización habría que abordar viene dado por:

$$L(\alpha,\beta) = \sum_{i=1}^{N} y_i \ln \alpha + y_i X_i\beta - (y_i + \frac{1}{\alpha})\ln(1 + \alpha e^{X_i\beta}) + \ln\Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln(y_i!) - \ln\Gamma\left(\frac{1}{\alpha}\right)$$

En el modelo de Poisson inflado de ceros, la formulación sería:

$$L(\beta,\tau) = \sum_{y_i=0} \ln(e^{\tau X_i\beta} + e^{-e^{X_i\beta}}) + \sum_{y_i>0} y_i X_i\beta - e^{X_i\beta} - \ln(y_i!) - \sum_{i=1}^{N} \ln(1 + e^{\tau X_i\beta})$$

siendo β y τ los valores a calcular.

Por último, si nos centramos en el modelo binomial negativo inflado de ceros, la maximización de la función logaritmo de verosimilitud, que se expresa de la forma

$$L(\beta,\alpha,\tau) = \sum_{y_i=0} \ln\left(e^{\tau X_i\beta} + \left(\frac{\nu}{\nu + e^{X_i\beta}}\right)^{\nu}\right) - \ln\left(1 + e^{\tau X_i\beta}\right) +$$

$$+\sum_{y_i>0} y_i \ln\alpha + y_i X_i\beta - (y_i + \frac{1}{\alpha})\ln(1+\alpha e^{X_i\beta}) + \ln\Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln(y_i!) - \ln\Gamma\left(\frac{1}{\alpha}\right)$$

llevará a la determinación de los valores estimados de los parámetros β, α, τ.

Para contrastar la significatividad individual de cada parámetro y, en consecuencia, de la variable explicativa correspondiente, se utiliza el estadístico z que sigue una distribución normal tipificada, de modo que $P(-N_{\alpha/2} < z < N_{\alpha/2}) = 1-\alpha$, siendo z el cociente entre el valor estimado de cada coeficiente y el error estándar y α el nivel de significación elegido. Para un nivel de significación del 5%, como el que tomaremos, valores de z mayores en valor absoluto que 1,96 indicarán significatividad de la variable asociada en la estimación.

Como hemos indicado anteriormente, los modelos *inflados de ceros* formulados en función del parámetro τ no se reducen a los modelos tradicionales cuando este parámetro toma el valor 0, con lo que las verosimilitudes no son directamente comparables. Vuong (1989) define un estadístico que permite decidir entre ambos modelos y que viene dado por:

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

siendo $m_i = \ln\frac{P_1(Y=y_i)}{P_2(Y=y_i)}$, $P_1(Y=y_i)$ y $P_2(Y=y_i)$ las funciones de probabilidad para el modelo inflado de ceros y el tradicional, respectivamente, y $\overline{m}$ la media de m_i, $i=1,\dots,N$.

Este autor demuestra que V es un estadístico bidireccional con una distribución asintótica normal tipificada. Cuando $|V|$ es menor que 1,96 el contraste no permite decantarse por uno u otro modelo; en caso contrario, un valor positivo es una evidencia a favor del modelo inflado de ceros, mientras que un valor negativo de V favorece el modelo tradicional.

4. RESULTADOS DE LA ESTIMACIÓN

A partir de la información disponible sobre los asegurados, procedemos a determinar las variables que resultan estadísticamente significativas en la estimación del número de siniestros sufridos y declarados a la compañía. Prestaremos especial atención, en la línea de algunos de los trabajos mencionados en la introducción, a observar si el disfrute de un grado de cobertura más elevado se traduce necesariamente en un aumento del número de siniestros, en cuyo caso se ratificaría la previsión teórica para los mercados de seguro con información asimétrica.

El hecho de que la variable endógena *NUMSIN* tome valores discretos no negativos hace que lo más apropiado sea utilizar un modelo de tipo *count data*. Según las características expuestas en el apartado 3.1, el modelo de regresión de Poisson no parece el más adecuado para la estimación que queremos llevar a cabo, puesto que no se cumple el supuesto de equi-dispersión. El elevado número de ceros de la variable *NUMSIN* hace más aconsejable el modelo binomial negativo. En efecto, como se ha mostrado en el apartado 2, casi el 77% de los asegurados no ha sufrido ningún siniestro durante el periodo de referencia. Más aún, si consideramos que existen diferencias cualitativas entre los asegurados que tienen siniestros y los que no (aquéllos suelen tener una actitud más descuidada en la conducción, en el mantenimiento del vehículo y/o en la elección del lugar de aparcamiento, por ejemplo) y que la posibilidad de que el asegurado no dé a conocer a la compañía todos los siniestros que sufre otorga al valor cero dos posibles interpretaciones, debemos elegir, para una estimación más exacta, un modelo inflado de ceros.

El software econométrico que hemos utilizado en la modelización es *Limdep v. 7.0*. La estimación de los parámetros se lleva a cabo mediante el método de máxima verosimilitud. Los problemas de optimización no lineal se resuelven aplicando el método de los gradientes y el algoritmo de Davidon, Fletcher y Powel o alguna de sus variantes (Greene, 1995). Por problemas computacionales de limitación de memoria, tomamos para el análisis empírico una muestra aleatoria de 15.000 asegurados, número suficiente para garantizar su representatividad.

Mostramos en primer lugar, en la Figura 2, las frecuencias observadas para el número de siniestros en la muestra tomada, así como las frecuencias estimadas según el modelo ZINB. En la Tabla 3, hemos incluido los principales resultados de los ajustes efectuados con los distintos modelos. Hay que tener en cuenta que, como ya se ha indicado, las verosimilitudes de los modelos inflados de ceros no son directamente comparables con las de los modelos tradicionales. El valor del estadístico de Vuong indica que debemos decantarnos por un modelo inflado de ceros frente al correspondiente modelo tradicional; por otro lado, la significatividad de Alpha en el modelo ZINB junto con los motivos teóricos que ya hemos expuesto, nos hacen pensar que éste es más adecuado para la modelización que proponemos que el modelo ZIP.

Figura 2. Distribuciones empírica y estimada con el modelo ZINB del número de siniestros

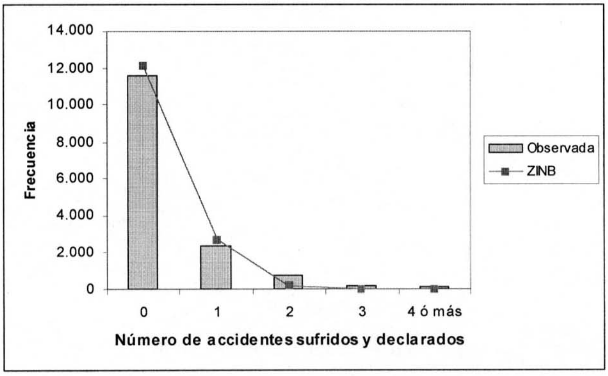


Tabla 3. Resumen de los ajustes

	Poisson	Binomial Negativo	ZIP	ZINB
Logaritmo de verosimilitud	-10.975,404	-10.556,327	-10.605,433	-10.602,787
Alpha (significatividad)	-	1,187 (<0,001)	-	0,052 (0,019)
Tau (significatividad)	-	-	-0,567 (<0,001)	-0,440 (0,003)
Estadístico de Vuong	-	-	34,826	9,076

Aplicamos entonces el modelo ZINB partiendo de las variables que se definen en el anexo. Tras varias simulaciones, llegamos al resultado que se muestra en la Tabla 4. Son 17 (incluido el término independiente) las variables explicativas que se muestran significativas, con un nivel de confianza del 95%.

Los *camiones*, *remolques* y *ciclomotores o motos* tienen un comportamiento significativamente distinto al de los *turismos y furgonetas*. En todos los casos el número de siniestros que se estima es menor que en la categoría base.

En lo que respecta a los usos, en el *agrícola propio* se observa una menor siniestralidad que en el uso base (el *particular*) y que en el uso *industrial*, aunque éste también es menos propenso a los siniestros que el uso *particular*. El uso de *transporte escolar* tiene sin embargo un comportamiento contrario: el modelo señala que hay una mayor siniestralidad en este uso que en el base.

Tabla 4. Estimación a través del modelo binomial negativo inflado de ceros

Variable	Coeficiente estimado	Estadístico *z*
Constante	-1,284	-7,275
CAMION	-0,446	-3,431
REMOLQUE	-1,447	-3,594
CICL_MOT	-0,961	-7,948
USO_INDU	-0,629	-2,718
USO_TESC	0,549	2,602
USO_AGRI	-0,967	-5,038
EDAD	-0,002	-1,985
ANTIG_2A	0,348	2,254
P200_300	0,753	4,951
P300_400	0,978	6,210
P400_500	1,110	6,779
P500_750	1,281	7,630
P750_	1,400	8,058
GR_MBAJA	0,164	4,516
GR_MALTA	0,171	3,688
GR_ALTA	0,219	3,563
Parámetros		
Alpha (dispersión)	0,052	2,349
Tau	-0,440	-3,024
Log-Máxima verosimilitud		-10.602,787
Estadístico de Vuong		9,076
Número de observaciones		15.000

La edad, variable que presenta una menor significatividad, resulta beneficiosa en el sentido de que es menor el número de siniestros a medida que aquélla avanza, como refleja la Tabla 4.

La antigüedad en el permiso de conducir está fuertemente relacionada con el número de siniestros sufridos. Concretamente, se observa que la variable ANTIG_2A es altamente significativa, lo que pone de manifiesto que los asegurados con menos de 2 años de experiencia en la conducción tendrán un comportamiento distinto del de los que llevan más años conduciendo, traducido en este caso en un mayor número de siniestros.

En relación con la prima, la Tabla 4 muestra la plena significatividad de todas sus categorías, con coeficientes positivos y crecientes que apuntan un mayor número de siniestros a medida que aumenta la prima en relación con la categoría base: casi el doble para las primas superiores a 750 € con respecto a las primas entre 200 y 300 €; entre tramos consecutivos, el incremento se sitúa entre el 10 y el 20%, aproximadamente.

Los coeficientes asociados a los grados de cobertura siguen un orden creciente. Es decir, cuanto mayor es la cobertura elegida, mayor es la siniestralidad con respecto a la cobertura mínima. Concluimos así que cada uno de estos niveles tiene un comportamiento distinto al grado de cobertura bajo, que conlleva en todos los casos un mayor número de siniestros. Los grados intermedios están bastante parejos, pero el coeficiente del grado más elevado de cobertura es algo mayor. Así, con respecto a la cobertura baja, la cobertura alta es la que presenta una mayor diferencia, una siniestralidad más notable en nuestro caso.

La predicción teórica de que los mayores niveles de cobertura deben ser para quiénes tienen un tipo de riesgo más elevado queda comprobada en nuestro análisis. Esto concuerda con los resultados obtenidos por Dionne, Gouriéroux y Vanasse (1999) y Cohen (2002), aunque no con los que deducen Chiaporri y Salanié (1997, 2000b) en los estudios que realizan sobre asegurados que tienen poca experiencia en la conducción. En nuestro caso, el escaso porcentaje de éstos en la población (solo 489) ha impedido que pudiéramos aplicar de forma aceptable las técnicas econométricas para comprobar si el comportamiento de los conductores noveles era distinto al de los restantes conductores en ese sentido.

5. CONCLUSIONES

El trabajo que presentamos se ha centrado en el seguro del automóvil. Hemos trabajado con datos cedidos por una entidad aseguradora privada, con el objetivo de determinar los factores más relevantes en la estimación del número de siniestros sufridos y declarados por los asegurados. Además, en el sentido de los trabajos empíricos existentes en otros países, hemos comprobado que se cumple en nuestros datos la predicción teórica de que los individuos expuestos a mayores riesgos de siniestros son los que eligen mayores coberturas. A diferencia de gran parte de estos estudios, que usan como variable dependiente la existencia o no de siniestros, nos hemos decantado por utilizar en su lugar el número de siniestros ocurridos. Debemos destacar aquí el hecho de que la información que las aseguradoras españolas, en general, y la que nos ha facilitado los datos, en particular, solicitan a sus clientes es mucho menos detallada que la que tienen las aseguradoras de otros países. La disponibilidad de información constituye siempre una dificultad añadida en este tipo de estudios. Es posible que alguna variable desconocida pudiera influir en la siniestralidad.

El modelo que se ha aplicado es un binomial negativo inflado de ceros, descartando otros modelos *count data* como el modelo de regresión de Poisson, por el elevado número de ceros que presenta la variable dependiente, o el modelo binomial negativo por no distinguir entre los valores nulos de la variable dependiente y los demás. Las diferencias cualitativas entre los asegurados que sufren siniestros y los que no, así como los matices que encierra el valor cero, sí son tenidos en cuenta en el modelo binomial negativo inflado de ceros. El resultado de la modelización confirma que ésta es mejor que la que se habría obtenido con los modelos de regresión de Poisson, con el binomial negativo o con el Poisson inflado de ceros.

Entre los factores que resultan determinantes en la siniestralidad, encontramos algunas categorías de vehículos, como camiones, remolques y ciclomotores o motos: algunos usos, concretamente el agrícola propio, el industrial y el transporte escolar; la edad; la experiencia en la conducción; la prima pagada; y el grado de cobertura. Con respecto a esta última variable, se observa que a medida que aumenta la cobertura, mayor es la siniestralidad, lo que concuerda con la previsión para los mercados en los que existe información asimétrica.

REFERENCIAS BIBLIOGRÁFICAS

ANGERS, J.F. y BISWAS, A. (2003): A Bayesian Analysis of Zero-Inflated Generalized Poisson Model, *Computational Statistics and Data Analysis*, 42, pp. 37-46.

ABBRING, J.H.; CHIAPPORI, P.A.; HECKMAN, J.J. y PINQUET, J. (2003): Adverse Selection and Moral Hazard in Insurance: Can Dynamic Data Help to Distinguish? *Journal of the European Economic Association*, 1 (Papers and Proceedings), pp. 512-521.

ARTÍS, M.; AYUSO, M. y GUILLÉN, M. (1999): Modelling Different Types of Automobile Insurance Fraud Behaviour in the Spanish Market, *Insurance: Mathematics and Economics*, 24, pp. 67-81.

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.

BOYER, M. y DIONNE, G. (1989): An Empirical Analysis of Moral Hazard and Experience Rating, *Review of Economics and Statistics*, 71, pp. 128-134.

CAMERON, A. y TRIVEDI, P. (1986): Econometric Models Based on Count Data: Comparison and Applications of Some Estimators and Tests, *Journal of Applied Econometrics*, 1, pp. 29-54.

CHIAPPORI, P.A. (1999): Asymmetric Information in Automobile Insurance: an Overview, en *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation* Dionne, G. y C. Laberge-Nadeau (eds.), pp. 1-11.

CHIAPPORI, P.A. y SALANIÉ, B. (1997): Empirical Contract Theory: The Case of Insurance Data, *European Economic Review*, 41, pp. 943-950.

CHIAPPORI, P.A. y SALANIÉ, B. (2000a): Testing Contract Theory: a Survey of Some Recent Work, *World Congress of the Econometric Society*, Seattle.

CHIAPPORI, P.A. y SALANIÉ, B. (2000b): Testing for Asymmetric Information in Insurance Markets, *Journal of Political Economy*, 108, 1, pp. 56-78

COHEN, A. (2002): Asymmetric Information and Learning: Evidence from the Automobile Insurance Market, *Discussion Paper* nº371, Harvard Law School, Cambridge.

DAHLBY, B. (1983): Adverse Selection and Statistical Discrimination: An Analysis of Canadian Automobile Insurance Market, *Journal of Public Economics*, 20, pp. 121-131.

DAHLBY, B. (1992): Testing for Asymmetric Information in Canadian Automobile Insurance, *Contributions to Insurance Economics*, editado por Dionne, G., Kluwer Academic Publishers, pp. 423-443.

DIONNE, G. (1998): La Mesure Empirique des Problèmes d'Information, *Cahier de recherche*, nº 9833, U.F.R. de Science Économiques, Gestion, Mathématiques et Informatique, Paris.

DIONNE, G. (2000) (Ed.): *Handbook of Insurance*, Kluwer Academic Publishers.

DIONNE, G.; GOURIÉROUX, C. y VANASSE, C. (1999): Evidence of Adverse Selection in Automobile Insurance Markets, en *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation*, Dionne, G. y C. Laberge-Nadeau (eds.), pp. 13-46.

GOURIÉROUX, C. (1999): The Econometrics of Risk Classification in Insurance, *Geneva Papers on Risk and Insurance Theory*, 24, pp. 119-137.

GREENE, W.H. (1995): *Limdep Version 7.0: User's Manual*, Econometric Software.

GREENE, W.H. (2000): *Análisis Econométrico*, Prentice Hall.

JONES, A.M. (2001): *Applied Econometrics for Health Economists - A Practical Guide*, Office of Health Economics.

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.

MELGAR, M.C. (2004): *Análisis de las Componentes de la Demanda de Seguro. Aplicación al Seguro del Automóvil*. Tesis Doctoral. Universidad de Sevilla.

MULLAHY, J. (1997): Heterogeneity, excess Zeros, and the Structure of Count Data Models, *Journal of Applied Econometrics*, 12, pp. 337-350.

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.

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.

RIDOUT, M.; HINDE, J. y DEMÉTRIO, C.G.B. (2001): A Score Test for Testing a Zero-Inflated Poisson Regression Model against Zero-Inflated Negative Binomial Alternatives, *Biometrics*, 57, pp. 219-223.

SCOLLNIK, D.P.M. (1998): On the Analysis of the Truncated Generalized Poisson Distribution using a Bayesian Method, *ASTIN Bulletin*, 28, 1, pp. 135-152.

SHANKAR, V.; MILTON, J. y MANNERING, F. (1997): Modeling Accident Frequencies as Zero-Altered Probability Processes: an Empirical Inquiry, *Accident Analysis and Prevention*, 29, 6, pp. 829-837.

VUONG, Q.H.: (1989): Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, *Econometrica*, 57, pp. 307-333.

WINKELMANN, R. (2003): *Econometric Analysis of Count Data*, Springer.

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.

YAU, K.K.W; YIP, K.C.H. y YUEN, H.K. (2003): Modelling Repeated Insurance Claim Frequency Data Using the Generalized Linear Mixed Model, *Journal of Applied Statistics*, 30, 8, pp. 857-865.

ANEXO
DEFINICIÓN DE LAS VARIABLES UTILIZADAS EN EL ANÁLISIS ECONOMÉTRICO

- TUR_FUR* = 1, si el vehículo asegurado es un turismo o una furgoneta; 0, en caso contrario.
- 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.
- TRACT_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.
- USO_PART* = 1, si el uso al que se destina el vehículo asegurado es el uso particular; 0, en caso contrario.
- USO_SP = 1, si el uso al que se destina el vehículo asegurado es el servicio público; 0, en caso contrario.
- USO_ALQU = 1, si el uso al que se destina el vehículo asegurado es el alquiler; 0, en caso contrario.
- USO_ESCU = 1, si el uso al que se destina el vehículo asegurado es escuela de conductores; 0, en caso contrario.
- USO_COMP = 1, si el uso al que se destina el vehículo asegurado es la compraventa; 0, en caso contrario.
- USO_INDU = 1, si el uso al que se destina el vehículo asegurado es industrial; 0, en caso contrario.
- USO_TMER = 1, si el uso al que se destina el vehículo asegurado es el transporte de mercancías; 0, en caso contrario.
- USO_TESC = 1, si el uso al que se destina el vehículo asegurado es el transporte escolar; 0, en caso contrario.
- USO_TGV = 1, si el uso al que se destina el vehículo asegurado es el transporte general de viajeros; 0, en caso contrario.
- USO_AGRI = 1, si el uso al que se destina el vehículo asegurado es el agrícola propio; 0, en caso contrario.
- USO_RPC = 1, si el uso al que se destina el vehículo asegurado es la retirada de permiso de conducir; 0, en caso contrario.
- EDAD: edad del asegurado a fecha 15 de diciembre de 2002.
- ANTIG_2A = 1, si el asegurado sacó el carné de conducir hace menos de 2 años (a fecha 15 de diciembre de 2002); 0, si lo sacó hace al menos 2 años.

- MUJER = 1, si el asegurado es mujer; 0, si es hombre.
- ZONA_1* = 1, si la zona habitual de circulación del vehículo asegurado es la zona 1; 0, en caso contrario.
- ZONA_2 = 1, si la zona habitual de circulación del vehículo asegurado es la zona 2; 0, en caso contrario.
- ZONA_3 = 1, si la zona habitual de circulación del vehículo asegurado es la zona 3; 0, en caso contrario.
- P0_200* = 1, si la prima anual pagada por el asegurado no supera los 200 •; 0, en caso contrario.
- 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.
- GR_BAJA = 1, si el asegurado disfruta del grado de cobertura bajo; 0, en caso contrario.
- GR_MBAJA = 1, si el asegurado disfruta del grado de cobertura medio-bajo; 0, en caso contrario.
- GR_MALTA = 1, si el asegurado disfruta del grado de cobertura medio-alto; 0, en caso contrario.
- GR_ALTA = 1, si el asegurado disfruta del grado de cobertura alto; 0, en caso contrario.
- NUMSIN: número de siniestros sufridos por el asegurado y declarados a la compañía.

* CATEGORÍA BASE DE LA VARIABLE