

R. 97004

11 21751



CUADERNOS DE LA FUNDACION

Nº 47

**EL FRAUDE EN EL SEGURO DEL AUTOMÓVIL:
CÓMO DETECTARLO**

Autor: Mercedes Ayuso Gutiérrez

Enero, 1999

LISTA DE CUADERNOS DE LA FUNDACION MAPFRE ESTUDIOS EDITADOS:

1. Filosofía Empresarial
 2. Resultados de la Encuesta sobre "Altos Profesionales de Seguros" (A.P.S.)
 3. Dirección y Gestión de la Seguridad
 4. Los Seguros en una Europa cambiante: 1990-1995 (No disponible)
 5. La Distribución Comercial del Seguro: Sus Estrategias y Riesgos
 6. Elementos de Dirección Estratégica de la Empresa
 7. Los Seguros de Responsabilidad Civil y su Obligatoriedad de Aseguramiento
 8. La Implantación de un Sistema de Controlling Estratégico en la Empresa
 9. Técnicas de Trabajo Intelectual
 10. Desarrollo Directivo: Una Inversión Estratégica
 11. El Concepto de Seguridad en la Ciencia y la Ciencia de la Seguridad
 12. Los Seguros de Salud y la Sanidad Privada
 13. Calidad Total y Seguridad
 14. El Reaseguro de Exceso de Pérdidas
 15. El Coste de los Riesgos en la Empresa Española 1991
 16. La Legislación Española de Seguros y su Adaptación a la Normativa Comunitaria
- Número Especial: Informe sobre el Mercado de Seguros 1993

17. Medio Ambiente Seguro: Desarrollo Futuro
18. El Seguro de Crédito a la Exportación en los países de la OCDE (Evaluación de los resultados de los aseguradores públicos)
19. Una Teoría de la Educación
20. El Reaseguro en los Procesos de Integración Económica

Número Especial: Informe sobre el Mercado de Seguros 1994

21. La Nueva Regulación de las Provisiones Técnicas en la Directiva de Cuentas de la C.E.E. Provisiones Técnicas de Seguros de Vida en las Directivas Comunitarias
22. Rentabilidad y Productividad de Entidades Aseguradoras
23. Análisis de la Demanda de Seguro Sanitario Privado
24. El Seguro: Expresión de Solidaridad desde la Perspectiva del Derecho
25. El Reaseguro Financiero
26. El Coste de los Riesgos en la Empresa Española 1993
27. La Calidad Total como Factor para elevar la Cuota de Mercado en Empresas de Seguros
28. La Naturaleza Jurídica del Seguro de Responsabilidad Civil
29. Ruina y Seguro de Responsabilidad Civil Decenal

Número Especial: Informe sobre el Mercado de Seguros 1995

30. El Tiempo del Directivo

31. Tipos Estratégicos, Orientación al Mercado y Resultados Económicos: Análisis Empírico del Sector Asegurador Español
32. Decisiones Racionales en Reaseguro
33. La función del Derecho en la Economía
34. El Coste de los Riesgos en la Empresa Española 1995
35. El Control de Riesgos en Fraudes Informáticos
36. Cláusulas Limitativas de los Derechos de los Asegurados y Cláusulas Delimitadoras del Riesgo Cubierto. Las Cláusulas de Limitación Temporal de la Cobertura en el Seguro de Responsabilidad Civil

Número Especial: Informe sobre el Mercado de Seguros 1996

37. La Responsabilidad Civil por Accidente de Circulación. Puntual Comparación de los Derechos Francés y Español
38. Legislación y Estadísticas del Mercado de Seguros en la Comunidad Iberoamericana
39. Perspectiva Histórica de los Documentos Estadístico-Contables del Órgano de Control: Aspectos Jurídicos, Formalización y Explotación
40. Resultados de la Encuesta sobre la Organización y Gestión de la Seguridad en la Empresa (1996)
41. De Maastricht a Amsterdam: Un paso más en la integración europea

Número Especial: Informe sobre el Mercado de Seguros 1996

42. La Responsabilidad Civil por contaminación del entorno y su aseguramiento

43. Resultados de la Encuesta sobre Disponibilidad de Instalaciones de Protección contra Incendios en la Empresa 1997”
44. Resultados de la Encuesta sobre Implantación en la Empresa de la Ley de Prevención de Riesgos Laborales
45. Los Impuestos en una Economía Global
46. Evolución y Predicción de las Tablas de Mortalidad Dinámicas para la Población Española
47. El Fraude en el Seguro del Automóvil: Cómo detectarlo

Copyright: F.M.E.

Prohibida la reproducción total o parcial de este trabajo sin el permiso escrito del autor o de la FUNDACION MAPFRE ESTUDIOS.

EL FRAUDE EN EL SEGURO DE AUTOMÓVIL: CÓMO DETECTARLO

Autor: Mercedes Ayuso Gutiérrez

Departamento de Econometría,
Estadística y Economía Española

Universidad de Barcelona

Trabajo resultante de una Beca Riesgo y Seguro 1996/97, concedida a la autora por la Fundación MAPFRE Estudios.

MERCEDES AYUSO GUTIÉRREZ

*Miembro del Departamento de Econometría, Estadística y Economía Española
Universidad de Barcelona*

**EL FRAUDE EN EL SEGURO DEL AUTOMÓVIL:
CÓMO DETECTARLO**

Trabajo realizado dentro del marco BECAS RIESGO Y SEGURO 96/97
Título del proyecto: *Modelos de Detección y Control del Fraude en el Seguro del
Automóvil.*

Fundación **MAPFRE** Estudios

Este trabajo es un extracto de la Tesis Doctoral "*Modelos Económicos para la Detección del Fraude en el Seguro del Automóvil*" presentada por Mercedes Ayuso para la obtención del Título de Doctora en Ciencias Económicas y Empresariales, bajo la dirección de la Dra. Montserrat Guillén.

La autora agradece sinceramente el apoyo y la ayuda, siempre incondicionales, recibidos de su directora así como del Dr. Manuel Artís. Los consejos y la experiencia de ambos han sido de vital importancia en el desarrollo de éste y el resto de trabajos que ha realizado. Destacar, asimismo, la ayuda recibida del resto de miembros del Departamento de Econometría, Estadística y Economía Española de la Universidad de Barcelona y, en especial, de la Dra. Manuela Alcañiz y del Dr. Ramón Alemany por los comentarios e ideas sugeridas en relación a la investigación realizada.

El agradecimiento se extiende también a la Fundación MAPFRE Estudios. Su colaboración en el desarrollo de éste y otros estudios llevados a cabo en el mundo universitario demuestra la confianza que han depositado en un hecho importante: la Universidad puede generar trabajos de carácter aplicado y ponerlos a disposición del mundo empresarial. Saber que el estudio realizado responde a uno de los principales problemas con los que se encuentra actualmente el mercado asegurador español ha sido, sin duda, un factor alentador en su desarrollo.

Para terminar, agradecer al Departamento de Investigación y Análisis de Siniestros de MAPFRE Mutualidad la ayuda prestada en el diseño de la base de datos y su buena predisposición en todas las colaboraciones solicitadas.

Montserrat Guillén Estany es Doctora en Ciencias Económicas y Empresariales por la Universidad de Barcelona, Licenciada en Ciencias Matemáticas por la Universidad de Barcelona, Master en Análisis de Datos en Ciencias Sociales por la Universidad de Essex y Postgrado en Análisis de Datos por la Universidad de Essex.

En la actualidad es jefe de estudios de la Diplomatura en Estadística de la Universidad de Barcelona, es profesora en dicha Diplomatura y en la Licenciatura en Ciencias Actuariales y Financieras (Universidad de Barcelona) y colaboradora en la Licenciatura en Dirección y Administración de Empresas (Universitat Oberta de Catalunya).

Es autora de numerosas publicaciones, artículos y estudios relacionados con el campo financiero y asegurador. Participa, de manera habitual, en congresos nacionales e internacionales y colabora en revistas económicas. Es directora de numerosas tesis doctorales e investigadora responsable, junto al Dr. Manuel Artís, del equipo de investigación *Modelos Económicos del Riesgo*.

Miembro de la American Statistical Association, de la Econometric Society, de la Sociedad Española de Estadística e Investigación Operativa y de la Sociedad Catalana de Matemáticas, ha colaborado con instituciones públicas y privadas del mercado nacional e internacional.

e-mail: guillen@eco.ub.es
Tlfn: (34) 93.4021824

Mercedes Ayuso Gutiérrez es Doctora en Ciencias Económicas y Empresariales por la Universidad de Barcelona, Licenciada en Ciencias Económicas y Empresariales por la Universidad de Barcelona y Actuario de Seguros.

En la actualidad es profesora en la Licenciatura en Ciencias Actuariales y Financieras de la Universidad de Barcelona y es vicesecretaria del Departamento de Econometría, Estadística y Economía de dicha Universidad.

Su investigación se ha ceñido, fundamentalmente, al campo de los seguros no-vida aunque es autora de publicaciones relacionadas con la Biometría. Los artículos que ha escrito responden, principalmente, a un objetivo concreto: la aplicación de la Estadística y la Econometría al diseño de herramientas para detectar el fraude en el seguro.

Su participación en congresos, tanto nacionales como internacionales, es habitual y es miembro colegiado del Instituto de Actuarios Españoles, del Col·legi d'Actuaris de Catalunya y de la International Actuarial Association (AFIR y ASTIN). Ha colaborado con entidades públicas y privadas del mercado nacional y a nivel internacional trabaja en contacto con prestigiosos organismos en la lucha contra el fraude como el Insurance Fraud Bureau de Massachusetts.

e-mail: ayuso@eco.ub.es
Tlfn: (34) 93.4021824

PRÓLOGO

La primera pregunta que surge al iniciar la lectura de una obra como la que aquí se presenta es saber si aportará elementos prácticos de cara a la lucha contra el fraude en el seguro del automóvil. En mi opinión, ese es precisamente el elemento más destacado de la misma.

El libro realiza un recorrido amplísimo sobre el fraude en el mercado asegurador. Desde una perspectiva histórica se revisan cuáles han sido hasta el momento, las principales actuaciones en la lucha contra el fraude. Seguidamente, y para el caso del mercado español, se presenta y se analiza el enorme impacto que la existencia de comportamientos deshonestos tiene en el balance final del negocio. Sin dejar de lado las consideraciones de la teoría económica, se plantea el problema de la detección del fraude y se justifica la necesidad de un control del mismo como herramienta de disuasión.

Uno de los apartados más interesantes lo constituye el dedicado a presentar cómo se implementan en la práctica los sistemas de detección de fraudes en el seguro del automóvil. En este punto, destaca el abanico de posibilidades técnicas, y hoy en día de fácil aplicabilidad, que pueden ayudar al tramitador en la gestión e identificación de los siniestros con indicios de fraude. La autora realiza una imparcial presentación de los métodos estadísticos que permiten conseguir los denominados indicadores de fraude. Además, revisa con extensión las aproximaciones propuestas en los diversos países, sus singularidades y lo que cada una de ellas aporta. Se trata de herramientas de soporte en la toma de decisiones, que en este caso, ayudan a determinar la conveniencia o no de investigar más a fondo un siniestro. Finalmente, se presenta un sistema novedoso y adaptado a las circunstancias del seguro de automóviles en España.

No sería justo llenar estas páginas de prólogo avanzando al lector lo que ya encontrará, con mayor rigor y extensión, en las páginas siguientes. Si se me permite, dedicaré algunas líneas a esbozar cómo se gestó este trabajo y cómo el tesón de la autora ha contribuido a su elaboración.

Fue en 1992 cuando Georges Dionne, uno de los más grandes científicos dedicados a la investigación en el campo de la economía del seguro, me preguntó qué se hacía en España para eliminar el fraude. A medida que avanzaba nuestra conversación, nos dimos cuenta de que en este país ya existía una verdadera concienciación de que el elevado volumen de siniestros y la agilidad en la tramitación de los mismos exigía disponer de modelos cuantitativos para ayudar a la detección automática de aquellos casos en los que confluían una serie de circunstancias, cuanto menos sospechosas, de fraude. Aunque la inspección manual de las reclamaciones de siniestros no parecía poder obviarse,

sí se ponía de manifiesto que era imprescindible diseñar sistemas de indicadores de fraude que, al menos, orientaran sobre la probabilidad de existencia del mismo. Por parte de las entidades de seguros, ya estaba latente la preocupación por el control del fraude y eso era un valor añadido muy importante.

Al establecer nuestra colaboración con MAPFRE Mutualidad no sólo intuimos que existía la necesidad de realizar este trabajo, sino que además hallamos un marco inigualable en el que desarrollar la investigación. Gracias a la inestimable ayuda de D. Miguel Iturgoyen al frente del Departamento de Investigación y Análisis de Siniestros, a su elevado grado de profesionalidad, tuvimos ocasión de aunar esfuerzos y encaminar correctamente nuestros pasos. El apoyo de MAPFRE, la principal entidad de seguros de automóviles en España, fue incondicional desde el primer momento. Ese estímulo, el saber que los resultados no sólo tenían un interés para la comunidad científica sino que además auguraban una inminente aplicabilidad práctica, ha constituido el hilo conductor de un trabajo que culmina en esta obra.

Dadas estas circunstancias, es lógico afirmar que el entorno de trabajo en el que D^a. Mercedes Ayuso pudo desarrollar su tesis doctoral fue inmejorable desde una perspectiva académica y que esa obra constituye el paradigma de lo que se considera una investigación excelente en sus múltiples facetas. Por lo tanto, los esfuerzos de tantos años, de los que puedo dar testimonio personal, proporcionan en esta obra, un trabajo completo, riguroso y de la más alta calidad científica. Tras estas páginas se conjuga el esfuerzo personal de la autora y una dedicación ejemplar que únicamente puede entenderse si se tiene en cuenta el talante optimista y tenaz de la misma. Se dedicó mucho tiempo a la revisión de trabajos, al contacto personal con investigadores de otros países y a la revisión de la información estadística, no siempre evidente, que se recoge en las declaraciones de siniestro. Además, para entender el funcionamiento de sistemas aseguradores de países distintos hace falta tener una formación actuarial competente y, a la vez, mantener una amplitud de miras que permita centrarse en el contexto general y no en las particularidades del seguro para cada marco legislativo particular.

Toda la dedicación de la autora a esta obra se vio reforzada por los consejos de una persona que merece ser destacada. Se trata del actual decano de la Facultad de Ciencias Económicas y Empresariales, D. Manuel Artís, que en su condición de catedrático y actuario, perfiló exquisitamente el interés que este trabajo tiene para el sector asegurador y la sociedad en general. Sus sugerencias abrieron perspectivas más amplias en la investigación, estimulando enormemente las distintas fases de su elaboración.

Finalmente, en tanto que persona que dirigió este trabajo desde su inicio, me honra destacar que la aportación que realiza D^a. Mercedes Ayuso a la lucha contra el fraude constituye un nuevo método, que sin duda aporta una gran ayuda en la agilización del proceso de identificación de fraudes. Al mismo tiempo, analiza con profundidad y contrasta si existe evidencia estadística de la correlación de determinadas circunstancias con la presencia de comportamientos

deshonestos.

Sin querer predisponer al lector, deseo destacar que la obra tiene una utilidad práctica para todo aquél que quiera saber cómo se puede mejorar la detección del fraude. Para responder a esta pregunta, basta leer este libro, pionero en España por sus características y en el mundo entero por sugerir avances técnicos y metodológicos que avalan los foros científicos de reconocido prestigio en los que se ha discutido.

Por lo tanto, me honra recomendar la lectura de esta obra sabiendo que será útil y que contribuirá a difundir una idea importante: que los grupos de investigación españoles, tanto en esta materia como en otros campos de la ciencia, tienen un nivel de competencia que constituye ya un referente a nivel internacional.

Dra. Montserrat Guillén
Profesora Titular de la Universidad de Barcelona
Jefe de Estudios de Estadística

ÍNDICE

INTRODUCCIÓN.....	15
1. EL FRAUDE EN EL MERCADO ASEGURADOR	21
1.1 Definición y antecedentes	21
1.2 La percepción del fraude en el seguro: un cambio de comportamiento	27
1.3 El posicionamiento del mercado asegurador español frente al fraude.....	32
1.4 La lucha contra el fraude en el seguro: una actuación sectorial que adolece la falta de tratamiento metodológico	37
2. EL FRAUDE EN EL SEGURO DEL AUTOMÓVIL.....	43
2.1 Aspectos preliminares	43
2.2 El seguro del automóvil en EE.UU.: tipología de fraude.....	47
2.3 El seguro del automóvil en España y los tipos de fraude más frecuentes dentro del marco europeo.....	52
2.4 Estadísticas de fraude en el seguro del automóvil.....	56
2.5 Aspectos prácticos del tratamiento del fraude en el seguro del automóvil.....	60
2.6 Categorías de fraude en el seguro del automóvil en España	69
3. ENFOQUE ECONÓMICO DEL FRAUDE EN EL SEGURO DEL AUTOMÓVIL Y SU TRATAMIENTO ECONOMETRICO.....	77
3.1 Consideraciones previas	77
3.2 La Teoría Clásica del Consumidor: ausencia de incertidumbre	79
3.3 Fundamentación teórica de la decisión de defraudar: Teoría de la Utilidad Esperada	82
3.4 Una nueva aproximación a la aplicación de la Teoría de la Utilidad Esperada en el tratamiento del fraude	87
3.5 Métodos cuantitativos de detección y control del fraude en el seguro del automóvil.....	91

4. PRESENTACIÓN DEL MODELO DE ELECCIÓN DE FRAUDE CON MÚLTIPLES ALTERNATIVAS EN EL SEGURO DEL AUTOMÓVIL.....	109
4.1 Modelo de elección de fraude: una decisión jerárquica entre alternativas	109
4.2 El Modelo Lógit Multinomial.....	113
4.3 El Modelo Lógit Anidado.....	120
4.4 Optimización del punto de corte en modelos con múltiples alternativas.....	123
5. APLICACIÓN EMPÍRICA A LA CUANTIFICACIÓN Y DETECCIÓN DE FRAUDE: DESCRIPCIÓN DE LA BASE DE DATOS.....	133
5.1 Comentarios generales.....	133
5.2 Descripción de la base de datos utilizada en la modelización	135
5.2.1 Naturaleza de los datos	135
5.2.2 Descripción de variables.....	137
5.3 Estadísticos descriptivos básicos	144
6. APLICACIÓN EMPÍRICA : MODELIZACIÓN LOGÍSTICA MULTINOMIAL Y ANIDADA.....	147
6.1 Introducción	147
6.2 Criterio de selección muestral. Determinación de las ponderaciones a introducir en los modelos	150
6.3 Estimación logística multinomial de la probabilidad de existencia de <i>fraude a favor del asegurado o a favor de un tercero</i> (MODELO 1).....	157
6.3.1 Análisis de la significación individual y global del modelo	157
6.3.2 Independencia de Alternativas Irrelevantes.....	166
6.3.3 Análisis de la calidad del ajuste	167
6.3.4 Estimación de la utilidad esperada por el individuo al elegir entre las alternativas	173
6.3.5 Cuantificación del coste esperado por siniestro tras la aplicación de un modelo de control de fraude	177
6.4 Estimación logística multinomial de la probabilidad de existencia de fraude <i>para conseguir una indemnización o para incrementar una indemnización</i> (MODELO 2.)	182
6.4.1 Análisis de la significación individual y global del modelo	182
6.4.2 Independencia de Alternativas Irrelevantes.....	191
6.4.3 Análisis de la calidad del ajuste	192

6.4.4 Estimación de la utilidad esperada por el individuo al elegir entre las alternativas	195
6.4.5 Cuantificación del coste esperado por siniestro tras la aplicación de un modelo de control de fraude	197
6.5 Modelización logística anidada.....	199
7. CONCLUSIONES	205
8. BIBLIOGRAFÍA.....	221

INTRODUCCIÓN

Que el fraude es un hecho inherente a la propia dinámica del seguro es algo que todos los aseguradores saben. Que es imposible erradicarlo en su totalidad es un hecho que también conocen. Que es posible prevenirlo y diseñar herramientas que ayuden a actuar frente al mismo es algo que queremos hacerles saber.

La lucha contra el fraude en el seguro, especialmente en el ramo del automóvil, ha despertado el interés de las aseguradoras, nacional e internacionalmente hablando. ¿Cómo obviar la importancia que la existencia de comportamientos fraudulentos puede tener en el número de siniestros declarados y en el coste medio de los mismos?.

Hasta hace poco tiempo, el fraude era considerado un factor ineludible de riesgo. Las entidades eran conscientes de su existencia pero no se habían detenido a considerar la influencia que podía tener en sus resultados. Su preocupación por mantener una imagen social adecuada (¿cómo cuestionar el comportamiento del asegurado?), la delimitación de breves plazos de indemnización y las consecuencias que de ello se derivan a la hora de realizar una adecuada investigación de los siniestros, eran sin duda, factores alentadores de la existencia de comportamientos fraudulentos.

Sin embargo, algo está cambiando. Durante los últimos años, las entidades de nuestro país parecen haber abandonado su actitud pasiva para hacer frente, por fin, al problema. ¿Por qué permitir que los asegurados honestos tengan que pagar las consecuencias de los deshonestos? ¿por qué no intentar tomar

medidas contra comportamientos que pueden distorsionar notablemente los resultados esperados de la entidad? ¿cómo permanecer impasivos ante un llamamiento realizado desde el propio Comité Europeo de Seguros?.

El problema es, realmente, muy importante. El análisis de las estadísticas publicadas por I.C.E.A (1998) pone de manifiesto cifras sorprendentes. Si gracias a su actuación, las 27 entidades que participaron en el 4º concurso sectorial organizado por esta asociación evitaron un volumen de fraude, en el seguro del automóvil, cuantificado en 6.657 millones de pesetas siendo la capacidad media de detección de fraudes de las entidades de Automóviles tan sólo del 0.5% de los siniestros aperturados en un ejercicio, ¿a qué cantidad ascienden las magnitudes defraudadas a nuestras aseguradoras?.

Las instituciones y entidades involucradas en la lucha contra el fraude han hecho eco de su preocupación. Artículos en revistas especializadas del sector, la publicación de manuales como *El Fraude en el Seguro de Automóviles* (C.E.S, 1992), el *Manual de Investigación de Siniestros y Lucha contra el Fraude en el Seguro de Automóviles* (Cobo,1993) o *Le Guide de l'anti-fraude à l'assurance en Europe* (C.E.A, 1996), en los que se señalan los comportamientos fraudulentos más típicos y las pautas a seguir para actuar frente a ellos y, una mayor profesionalización del personal de las compañías, son algunas de las herramientas utilizadas.

Ahora bien, ¿por qué no utilizar el gran volumen de información de que disponen las compañías para diseñar técnicas cuantitativas de control y detección del fraude?, ¿por qué no facilitar el trabajo a los tramitadores?, ¿por qué no planificar de forma estructurada la investigación de los siniestros atendiendo a su posible contenido en fraude?.

En los últimos años, la participación de la comunidad científica en el estudio cuantitativo del fraude está ganando terreno. Así, hoy en día es posible hablar ya de la existencia de tres grupos de investigación que, a nivel mundial,

modelizan el comportamiento deshonesto dentro de las reclamaciones de siniestros. Los estudios realizados en el Insurance Fraud Bureau de Massachusetts, de la mano del Dr. Richard Derrig; los llevados a cabo en la École des Hautes Études Commerciales afiliada a la Universidad de Montreal, bajo la dirección del profesor Georges Dionne y los desarrollados en la Universidad de Barcelona respecto a España, han permitido extraer conclusiones importantes a tener en cuenta en el diseño de una adecuada política de control de fraude. El ámbito de estudio de estas instituciones ha sido, fundamentalmente, el del seguro del automóvil, que suele estar caracterizado por resultados técnicos negativos. Aprovechamos este momento para agradecer a R. Derrig y a G. Dionne sus valiosos comentarios en relación al trabajo que presentamos.

A nivel español, es más, a nivel europeo, no se conocen hasta el momento estudios cuantitativos de modelización del fraude en el seguro del automóvil (tampoco, en ningún otro ramo). De esta forma, el presente trabajo es totalmente novedoso e intenta cubrir el vacío existente en la determinación de las variables (relacionadas con el propio siniestro, con las partes que intervienen, con la póliza y el vehículo,...) que resultan significativas a la hora de explicar los diferentes tipos de comportamientos deshonestos.

El tratamiento estadístico y econométrico de la información que rodea a la póliza de seguros, al siniestro y a su declaración y a las partes involucradas en el accidente deriva en la consecución de dos hechos de gran importancia: el diseño del *perfil del defraudador* en una cartera de asegurados (teniendo en cuenta los diferentes tipos de fraude que pueden cometerse) y la generación de *señales de alerta* que advierten de la posible existencia de comportamiento fraudulento. La determinación de indicadores de fraude validados estadísticamente es uno de los principales objetivos perseguidos en el presente libro. Señalar cuáles son las variables sobre las que la entidad ha de prestar una especial atención puede llegar a constituirse, sin ninguna duda, en una pieza clave en la investigación de siniestros. Como resultado del trabajo

realizado, el diseño de una herramienta capaz de cuantificar la probabilidad de aparición de fraude en un siniestro (avanzando incluso en la cuantificación de la probabilidad de aparición de diferentes tipos de comportamientos fraudulentos), dotada de la sofisticación adecuada hasta el punto de ser capaz de orientar adecuadamente la investigación a realizar por la entidad, responde a una nueva aproximación al tratamiento del fraude mediante la utilización de técnicas cuantitativas.

La estructuración del libro atiende a tres partes básicas.

En la primera (Capítulos 1 y 2), de contenido meramente teórico, se realiza un análisis completo de los antecedentes existentes en relación al tratamiento del fraude en el campo asegurador. Es aquí donde detallamos las principales aportaciones realizadas por otros autores, considerando no sólo la vertiente estadística y econométrica sino, también, la asociada a un enfoque más conceptual (posibles tipos de fraude, recomendaciones sobre la forma de actuar, etc.).

En la segunda parte (Capítulos 3 y 4), de contenido metodológico, presentamos la técnica econométrica utilizada en nuestro estudio. El objetivo no se limita a cuantificar la probabilidad de aparición de los diferentes tipos de fraude, sino que lleva, de forma más importante, a determinar las variables que influyen en cada uno de ellos. Por tanto, los modelos aplicados incorporan en su definición el proceso de decisión seguido por el individuo al elegir entre alternativas. Se propone el uso de modelos lógit multinomiales y anidados, teniendo en cuenta esquemas de elección unietápicos y polietápicos, con métodos de estimación adecuados a las características de la muestra disponible.

En la última parte (Capítulos 5 y 6), de contenido empírico, se han detallado los resultados obtenidos al aplicar el modelo de control y detección de fraude a una muestra de expedientes de siniestros facilitada por la principal entidad del

mercado español en seguro de automóviles. La coherencia entre nuestros resultados y los hallados por otros autores que trabajan con muestras de muy distinta índole y en sistemas bastante diferentes, nos hace pensar que, al menos a grandes rasgos, nuestras conclusiones son extrapolables, en general, a toda la población de asegurados en el ramo de automóviles en España.

Sirva el presente trabajo para justificar el peso que la Estadística y la Econometría poseen dentro del campo asegurador.

“Aunque desde los principios del Seguro el potencial para el fraude y el azar moral ha aumentado, sólo recientemente se ha comprendido su alcance” (Hoyt, 1990).

1. EL FRAUDE EN EL MERCADO ASEGURADOR

1.1 Definición y antecedentes

La existencia de comportamientos fraudulentos en el mercado asegurador no es algo nuevo, propio de los años recientes, sino que de acuerdo con una opinión prácticamente generalizada a todos los aseguradores, puede remontarse a los orígenes de la misma empresa aseguradora. Sin embargo, la importancia que el tratamiento del fraude ha ganado en algunos campos económicos y sociales no ha sido tan marcada dentro del mundo del seguro, que lo ha considerado durante muchos años un factor ineludible de riesgo.

A modo de ejemplo, el tratamiento del fraude fiscal ha sido objeto, desde hace años, de profundas investigaciones. Allingham y Sandmo (1972) analizan el proceso de decisión seguido por el contribuyente a la hora de evadir el pago de impuestos. Análogamente, y desde otro punto de vista, la existencia de fraude en el ámbito empresarial¹ ha suscitado una gran preocupación en el campo de la auditoría interna, que incluye dentro de su declaración de normas de actuación, la necesidad de *“disuadir, detectar, investigar e informar del fraude”* (Institute of Internal Auditors, 1985).

Sin embargo, la actitud de las compañías aseguradoras no ha seguido el mismo camino. A pesar de que la década de los noventa ha supuesto un cambio de comportamiento en las entidades hacia una mayor preocupación por la detección, control y cuantificación del fraude, aún queda mucho por hacer.

¹ En el libro *“Cómo Luchar contra el Fraude en la Empresa”* (Hevia y Lafuente, 1992) aparece una detallada exposición de todas aquellas actuaciones fraudulentas que pueden ir en contra del funcionamiento de la empresa.

Pero, ¿qué entendemos por fraude en seguros?. Las definiciones que se presentan en la literatura sobre el tema son variadas. Hoyt (1990), presentando un enfoque fundamentalmente económico, define el fraude como *"un abuso del mecanismo asegurador para obtener ganancias económicas"*, interpretándolo, básicamente, como un determinado tipo de azar moral². En esta misma línea, Derrig y Ostaszewski (1994b) definen el fraude como *"la actuación del demandante y/o de otros partícipes del mecanismo asegurador (abogados, médicos,...) que, manipulando el proceso normal de un siniestro, pretenden obtener ganancias en beneficio propio"*. Una definición análoga puede encontrarse en Cummins y Tennyson (1992).

Según el C.E.S.³ (1992) se entiende por fraude en el seguro *"toda actuación de mala fe llevada a cabo por una persona con el objeto de obtener para sí misma, o en beneficio de un tercero, un enriquecimiento injusto e ilícito a expensas de una compañía de seguros, mediante la utilización de un artificio o engaño"*. Para la reaseguradora Münchener Rück (1988) se define como *"el comportamiento del asegurado o de un tercero con el objeto de obtener un amparo de seguro que normalmente no puede contratarse, de pagar una prima más baja o de reclamar el pago de una indemnización que, por su motivo o cuantía, no está justificada"*.

El contrato de seguro se fundamenta en la honestidad de las partes, apareciendo como necesario para la cobertura de los numerosos riesgos⁴ a los que el individuo está expuesto: *"el contrato de seguro es aquél por el que el asegurador se obliga, mediante el cobro de una prima y para el caso de que se produzca el evento cuyo riesgo es objeto de cobertura, a indemnizar, dentro de los límites pactados, el daño producido al asegurado o a satisfacer un capital, una renta u otras prestaciones convenidas"* (Art. 1 de la Ley 50/1980, de 8 de octubre, de Contrato de Seguro). De

² Bajo el término de *"moral hazard"*, Cummins y Tennyson (1996) definen todas aquellas acciones realizadas expresamente por el asegurado y que afectan a la probabilidad de siniestro o a la cantidad de pérdidas resultantes de un accidente.

³ Centro de Estudios del Seguro.

⁴ Siguiendo a Garrigues (1974) podemos definir Riesgo como *"la posibilidad de que por azar ocurra un hecho que produzca una necesidad patrimonial"*. Podríamos, por tanto, definirlo también como *"la probabilidad de que ocurra un hecho con consecuencias económicas negativas"*.

esta forma, cuando el asegurado comete fraude actúa en contra del principio de buena fe.

El conjunto de definiciones presentadas, gozando de una notable variedad, poseen en realidad un contenido, si no común, bastante similar. En todas ellas se pone de manifiesto una situación en la que el asegurado pretende obtener un beneficio ilícito de la entidad aseguradora. Aunque las formas de actuar por parte del defraudador pueden ser variadas, tal y como veremos más adelante, la primera cuestión a tratar se centra en la mayor o menor importancia que las entidades aseguradoras han otorgado a la detección y tratamiento del fraude en los últimos años.

Durante mucho tiempo, los aseguradores, aun siendo conscientes de la existencia del fraude, lo han considerado un factor ineludible de riesgo ignorando los grandes problemas que lleva asociado. Su actitud se ha limitado a subsanar sus efectos negativos mediante incrementos en las primas aplicadas, derivando de esta forma el problema en el conjunto de asegurados. La referencia a los elevados costes que supondría la instalación de mecanismos de control y detección ha sido habitualmente presentada como la principal justificación ante su falta de actuación. Sin embargo, las razones son variadas, tal y como veremos a continuación.

La competencia en el establecimiento de las primas, la preocupación por la aparición de nuevos riesgos o el aumento desmesurado en el importe de las indemnizaciones derivadas de las sentencias judiciales⁵, han llevado a considerar el fraude como un riesgo secundario. En las últimas décadas, el asegurador ha visto crecer de forma notable su cartera de clientes y ha diversificado ampliamente el número y la sofisticación de las coberturas ofertadas. De este modo, ha entrado en una guerra de competencia, en la que, el deseo de ganar mercado ha derivado en una simplificación de los requisitos necesarios para acceder al seguro. La menor

⁵ El hecho de que la aplicación de un baremo de indemnizaciones en el marco del seguro del automóvil no haya sido obligatoria hasta finales de 1995 (en base a la Ley 30/1995 de Ordenación y Supervisión de los Seguros Privados) ha generado durante muchos años la concesión al asegurado, por parte del sistema judicial, de sumas indemnizatorias muy elevadas.

información requerida en la póliza, la rápida aceptación de riesgos y la velocidad en la tramitación de siniestros sirven como indicadores de lo señalado. Lógicamente, todos estos puntos son favorables a la aparición de comportamientos fraudulentos.

La escasa acción social frente a la existencia de fraude implica que éste sea considerado como un delito de poca relevancia. Las entidades aseguradoras son consideradas por parte de los asegurados, empresas grandes y ricas, que aplican primas para hacer frente a los riesgos asumidos. Éstas son para la opinión pública particularmente elevadas, sobre todo, desde el punto de vista de aquellos asegurados que, no teniendo ningún siniestro, consideran que gran parte de la cantidad abonada derivará directamente en un aumento de los beneficios de la entidad. De esta forma, defraudar al seguro se considera por algunas personas como la forma de recuperar, al menos, una parte de las primas pagadas, sin tener en cuenta, en ningún momento, los costes que la empresa aseguradora ha tenido que afrontar en la aceptación del riesgo.

Se echa de menos la falta de concienciación pública mediante la difusión de campañas publicitarias semejantes a las que se han llevado a cabo para asuntos fiscales. Sin embargo, sería difícil que ésta pudiese lograr su objetivo sin poner en entredicho los propios mecanismos de control de la compañía, e incluso podría producirse el efecto contrario al buscado. De esta forma, podría ocurrir que los tomadores que solían actuar de buena fe decidieran cambiar su actitud para recuperar ilícitamente la prima, siguiendo el mal ejemplo de otros asegurados. El diseño de una correcta política de marketing debe partir de una premisa básica: la entidad debe hacer pública su preocupación frente al problema, alegando siempre la defensa del asegurado honesto.

Desde el punto de vista legal, la situación no es más favorable.

La jurisprudencia penal y civil no han desarrollado focos importantes de acción contra las actuaciones fraudulentas en el campo asegurador. De hecho, el nuevo Código Penal, promulgado mediante Ley Orgánica 10/95, de 23 de noviembre, y

que entró en vigor en mayo de 1996, suprime la referencia expresa a la internacionalmente denominada “estafa del seguro” o “fraude del seguro”. El anterior Código mostraba una mayor preocupación del legislador ante esta realidad económica y social, recogiendo como circunstancia agravante del delito de estafa la destrucción, daño u ocultación de cosa propia, la agravación de lesiones sufridas o autolesión, para defraudar al asegurador. El nuevo Código Penal, sin variar sustancialmente el concepto básico de “estafa”, ha eliminado la referencia expresa al seguro, basándose para ello en el escaso uso de dicho precepto durante su vigencia y en la insatisfacción derivada de su aplicación en algunos casos. Sin duda, este hecho contrasta con la actitud seguida en la redacción de los nuevos Códigos Penales de otros países. Por ejemplo, Portugal en 1993 y Francia en 1994 optaron por incluir en su articulado una referencia expresa a la “estafa en el seguro”.

Quizá la explicación a lo anterior se deba a un hecho concreto: la defensa del individuo que, dentro de la sociedad de consumo, aparece como perjudicado. Es cierto que, en el marco asegurador, el asegurado suele comparecer en un litigio como la parte menos favorecida, frente a la entidad aseguradora, a la que caracterizan atributos de poder en la mayoría de los casos.

El establecimiento de un régimen extremo en cuanto a la aportación de pruebas nos permite hablar, asimismo, de una regulación de derecho civil, que actúa otorgando grandes prerrogativas a los asegurados. De esta forma, la entidad aseguradora no luchará legalmente contra el fraude salvo que sus pruebas sean concluyentes, pues de lo contrario sólo conseguirá deteriorar su imagen social. Además, los pequeños fraudes y robos suelen llevar asociados un cierto automatismo en el comportamiento policial, que la mayoría de la veces deriva en un abandono de los casos sin haber obtenido pruebas fehacientes para instruirlos. Todo ello propicia cadenas recursivas de fraude, ante el convencimiento de la falta de actuación, tanto de la entidad aseguradora como de la autoridad legal. Mientras el asegurado honesto no sea claramente diferenciado del defraudador, de forma

que la condición de este último sea conocida públicamente, la tendencia a actuar fraudulentamente se verá acrecentada.

La visión económica del problema nos lleva señalar un hecho importante: la demanda para muchas líneas de seguro es relativamente inelástica⁶ (Hoyt, 1990) y, por tanto, poco sensible a cambios en las primas. Aunque las razones son variadas, quizá la obligatoriedad que caracteriza a algunos tipos de seguros (por ejemplo, para España la cobertura de Responsabilidad Civil Obligatoria en el seguro del automóvil), sea una de las más significativas. Ante una función de demanda de estas características, los incrementos en el coste del seguro (incluido el incremento en costes provocado por el fraude) pueden ser trasladados a los asegurados sin descensos significativos en la compra, en este caso suscripción, del producto. Sin embargo, y aunque quizá sea ésta la idea principal para comprender la falta de actuación de las entidades, adicionalmente al desconocimiento de las ventajas que pueden obtenerse de la creación de mecanismos de detección de fraude⁷, las consecuencias finales no son sólo negativas para el asegurado sino también para la propia compañía.

El asegurado se verá afectado, ante el incremento en primas, por un aumento en el gasto total que destina a seguros. Ante una disminución de la cantidad de dinero disponible para gastos alternativos o para ahorro, puede optar por reducir las coberturas contratadas con carácter voluntario, lo que, sin duda, incidirá en el mercado asegurador. La compañía, que ante ello puede ver peligrar el equilibrio técnico en algunos ramos, se encontrará en una situación no deseada, que puede

⁶ La elasticidad demanda-precio se define como el ratio que mide el cambio porcentual en la cantidad demandada de un bien ante un cambio porcentual en el precio del mismo.

⁷ Según Hoyt (1990), las compañías temen que el coste añadido de crear un mecanismo de detección supere el beneficio derivado del mismo. A pesar de la dificultad que supone cuantificar el verdadero coste del fraude en seguros, el autor manifiesta la existencia de una relación coste/beneficio positiva (al menos, para algunos niveles de control de fraude). La compañía Münchener Rück (1988) afirma que las experiencias recogidas en los países que, con el fin de luchar contra el fraude han creado organizaciones propias, indican que el coste financiero asociado a la inversión redundará en un beneficio de tres a seis veces el coste desembolsado. Derrig, Weisberg y Chen (1994) analizan como incide la toma de determinadas medidas en la tramitación de siniestros (con interés manifiesto en la detección de fraude) sobre el total de indemnizaciones pagadas por daños corporales, dentro del seguro del automóvil.

empeorar si los incrementos en precios se traducen en una reducción en competitividad.

En definitiva, el asegurador que no combate el fraude (y que, por lo tanto, no realiza una adecuada política de control de costes) podrá obtener, a corto plazo, ventajas competitivas; sin embargo, el incremento a largo plazo en el coste medio por siniestro derivado de la existencia de acciones fraudulentas, probablemente desvirtuará los efectos positivos que en un principio se pudieron conseguir. Una falta de actuación puede poner en peligro el equilibrio del mercado asegurador (Picard, 1996).

Ante todo ello, la creación de medidas de detección y control del fraude, no sólo dentro de cada entidad en particular, sino también dentro del entorno asegurador en general, es cada vez más necesaria. Desde este punto de vista, el campo de investigación que se nos ofrece es muy amplio: el tratamiento metodológico y científico existente es escaso para un problema con la connotación de ser de gran interés, no sólo para la empresa aseguradora, sino también para el asegurado y la sociedad en general.

1.2 La percepción del fraude en el seguro: un cambio de comportamiento

Los últimos años han supuesto un cambio en la mentalidad de los aseguradores de numerosos países frente al fraude. Lógicamente, la reacción frente al problema no ha sido homogénea, ni en todos los países, ni en todas las compañías, pero, sin duda, la toma de conciencia en relación a su presencia es ya un hecho. En la actualidad, el control de las acciones fraudulentas aparece como una gran preocupación para las entidades, que establecen mayores restricciones a la asegurabilidad de determinados tipos de riesgo. La situación comentada anteriormente, en la que los aseguradores permanecían ajenos a la evolución del problema, compensándolo intrínsecamente con una elevación de primas, está perdiendo vigencia.

La empresa aseguradora buscará una eficiente gestión de riesgos, en la que el control del fraude no puede quedar en lugar secundario. Las razones, aunque variadas, se pueden concretar en dos. Por un lado, la elevada dimensión alcanzada por las acciones fraudulentas impide que pueda ser compensada a través del incremento en las primas aplicadas. Por otro, los aseguradores han comprendido que la falta de medidas para contrarrestar el fraude derivará, directa o indirectamente, en un incremento de éste.

Las magnitudes que valoran la presencia de fraude en el seguro, genéricamente considerado, son realmente sorprendentes. Hoyt (1990) señala que un 10% de los siniestros soportados por los aseguradores de Estados Unidos contienen algún tipo de fraude⁸ y Brockett, Xia y Derrig (1995) indican que el coste anual del fraude en el mercado asegurador estadounidense es aproximadamente de 20000 millones de dólares⁹. Por otro lado, el Comité Europeo de Seguros (1996) estima que el fraude en la industria aseguradora europea implica unas pérdidas de 8000 millones de ECUS, lo que supone aproximadamente el 2% del total de las primas anuales recaudadas de todos los ramos. En España las cifras son también muy elevadas. El C.E.S. (1992) señala que, técnicamente, el fraude tiene incidencia entre el 10% y el 15% de los siniestros declarados en nuestro país, mientras que el departamento de investigación de I.N.E.S.E. (1996a) lo cuantifica en más de 100000 millones de pesetas anuales.

Ahora bien, ¿cuál ha sido la posición del mercado asegurador frente al problema estudiado en los diferentes momentos del tiempo?

Clarke (1990), en su artículo "The Control of Insurance Fraud: A Comparative View", indica que el proceso de actuación frente a la presencia de comportamientos fraudulentos en el seguro puede identificarse con el establecimiento de cuatro etapas. En ellas quedan recogidas las variaciones en los niveles de respuesta que

⁸ Estimación realizada por The United States Chamber of Commerce.

⁹ Estimación realizada por The National Insurance Crime Bureau (N.I.C.B.) en 1994.

los aseguradores han ido ofreciendo, con el paso del tiempo, a la existencia de fraude.

En una primera etapa, situada por el autor en la mitad de los años sesenta y referenciada textualmente como *“la edad de oro”*, el fraude no aparece como un problema importante, y aunque las entidades son conscientes de su existencia, lo aceptan como un riesgo más. Al definirlo, se habla de la actuación de algunos asegurados que obran de mala fe o con el objetivo de solucionar problemas económicos. En todo caso, estos últimos suelen tener pequeña importancia, admitiéndose sólo en casos excepcionales la posible presencia de grandes escalas de fraude ligadas a la actividad de defraudadores profesionales. La mayoría de los asegurados son clasificados como honestos, y por tanto, no se considera necesaria la actuación frente al problema. Mientras tanto, el aumento en las primas aplicadas es el necesario para cubrir, en la medida de lo posible, los siniestros acaecidos, sean éstos fraudulentos o no. La escasa actuación contra el fraude en esta etapa da pie a la generalización de dicho comportamiento entre un número, cada vez mayor, de asegurados.

Las consecuencias de lo anterior se reflejan en una segunda etapa, en la que las acciones fraudulentas son más numerosas. Sin embargo, y aunque continúan siendo percibidas como la actuación de grupos minoritarios de individuos con una escasa relevancia en costes, las entidades aseguradoras comienzan ya a plantearse la aplicación de medidas que, a nivel interno, frenen el desarrollo de este comportamiento evaluando la propensión al fraude en las diferentes áreas de riesgo. A pesar de ello la respuesta queda limitada a acciones individuales dentro de cada compañía, sin que se establezcan contactos formales entre entidades de cara a homogeneizar los datos disponibles. Los cambios de información son ocasionales (llamadas de teléfono, cartas a otros aseguradores,...) y en la mayoría de los casos aparecen totalmente encubiertos ante el temor de que la publicidad sobre el fraude pueda alentarlos. Además todas las actuaciones se someten a las restricciones que las leyes sobre protección de datos y sobre libertades civiles generan.

Los defraudadores pueden moverse entre entidades para evitar los perjuicios derivados de una detección de fraude en su comportamiento. La cooperación con las autoridades públicas es prácticamente inexistente. Sólo en aquellos casos en los que la evidencia de una acción fraudulenta es clara, con pruebas para demostrarla, la entidad aseguradora se pone en contacto con el poder público, con el objetivo de procesar dicha actuación.

La escasa evolución en la detección y control de las acciones fraudulentas da paso a la tercera etapa en la que Clarke hace referencia al *"fraude como deporte"*. En esta fase, los aseguradores son vistos como agentes que se enriquecen a partir de los asegurados. De esta forma, el fraude es concebido como un delito sin importancia sobre el que existe un amplio conocimiento público. El fraude cometido por aficionados continúa siendo el más frecuente, mientras que el profesional va ganando terreno, ante las grandes posibilidades que el marco asegurador ofrece.

Desde el punto de vista de las compañías, el cambio de actitud es notable, y comienzan a valorar objetivamente la importancia del fraude. La cooperación entre entidades aparece como una medida necesaria para realizar un control efectivo. Se plantea la creación de registros comunes, con el objetivo de poder verificar la identidad de los asegurados y su historial de siniestros, intentando tener localizados en todo momento a los posibles infractores. Se mejora la cooperación con los organismos públicos. En concreto, los contactos con la policía, con el poder judicial y con diferentes estamentos, han de ayudar, sin duda, a obtener buenos resultados. Crece la preocupación por realizar una correcta formación de personal especializado dentro de las entidades y por mejorar la tramitación de siniestros. Asimismo, el deseo de conocer el perfil del defraudador y las circunstancias que rodean la ocurrencia del siniestro fraudulento anticipa ya, en esta tercera etapa, la importancia que se concederá a la aplicación de mecanismos de detección y control del fraude.

En la cuarta y última etapa, la presencia de fraude en el seguro aparece reconocida universalmente, así como su impacto en el aumento en primas. No obstante, los

asegurados siguen sin considerarlo un delito grave, tal vez por la falta de acciones legales severas, y aparece, de forma más notable, el fraude profesional. Frente a todo ello, los aseguradores exigen del Estado una respuesta más rigurosa, acentuándose, asimismo, la acción colectiva entre entidades y potenciándose la formación de personal especializado. A pesar de todo, la actitud de los asegurados apenas sufre variaciones: continúa la hostilidad hacia las primas elevadas y el convencimiento de la presencia de elevados beneficios para la empresa aseguradora. En esta última fase, por tanto, Clarke (1990) asume la imposibilidad de cambiar la mentalidad del asegurado frente a las compañías, pero identifica a esta últimas dentro de un marco claramente diferenciado de lucha contra la existencia de actuaciones deshonestas.

El esquema presentado resulta suficientemente indicativo de cuál ha sido el posicionamiento de las entidades aseguradoras con el paso del tiempo. Lógicamente, si tuviésemos que destacar alguna característica del mismo sería su conceptualización desde un punto de vista genérico, siendo muy difícil señalar con exactitud en qué posición se encuentra cada país (concebido como un marco asegurador concreto) en cada momento del tiempo. Dicha localización se complica cuando tenemos en cuenta que aspectos típicos de alguna de las etapas pueden aparecer en otras, de forma que los límites entre ellas no están claramente definidos.

Es correcto indicar que existen países, como EE.UU. que se encuentran en condiciones de situarse en la cuarta etapa. El desarrollo de grandes esfuerzos coordinados a nivel nacional, la creación de departamentos especializados en el tratamiento y procesamiento de los comportamientos fraudulentos, la formación de personal especializado, el desarrollo de campañas publicitarias o la cooperación activa con los representantes legales sirven como muestra de lo señalado. Otros países, como Holanda y Francia, son localizados por Clarke en la segunda o en la tercera etapa, mientras que la situación descrita en la primera fase prácticamente puede considerarse inexistente en la actualidad. ¿En qué posición estaría el mercado asegurador español?. Veámoslo detenidamente.

1.3 El posicionamiento del mercado asegurador español frente al fraude

España, en base a las características descritas en el punto anterior para cada una de las fases evolutivas y desde nuestro punto de vista, se situaría aproximadamente en la segunda etapa, aunque las actuaciones desarrolladas en los últimos años han sido muy significativas.

La comparación con otros países, tanto de su entorno (Francia, Los Países Bajos,...) como más lejanos (EE.UU., Canadá,...), en donde la creación de organismos y el desarrollo de actividades frente al fraude han sido mucho más marcados, impide situar a nuestro país en un nivel superior dentro del esquema presentado. Es cierto, sin embargo, que la dificultad antes manifestada para delimitar cada una de las etapas podría dar lugar a diferentes puntos de vista en relación a esta afirmación.

Al igual que en el resto de campos socio-económicos, la integración de España en el mercado europeo ha tenido un gran impacto en el ámbito asegurador. De esta forma, y particularizando al tratamiento del fraude, cada vez se hace más notable la necesidad de una actuación sectorial coordinada, no sólo a nivel nacional, sino también a nivel comunitario, hecho que se ha venido reiterando últimamente desde el Comité Europeo de Seguros. Para este organismo¹⁰, la actuación de los defraudadores es similar en todos los países con lo que las acciones de las compañías deberán también ser parecidas. Dado que, tal y como venimos manifestando, el posicionamiento frente al fraude varía de unos países a otros, este órgano se ha propuesto como objetivo intensificar los intercambios de información y de experiencia, en lo que ha venido denominando la creación de una red europea anti-fraude.

¹⁰ Una explicación de los principales aspectos relacionados con la prevención, detección y sanción de las actividades fraudulentas a nivel europeo, así como, una estimación del coste asociado al fraude en Europa, pueden encontrarse en el manual *Le Guide de l'anti-fraude à l'assurance en Europe* elaborado por el organismo mencionado en 1996.

Pero, ¿cuál es el tratamiento institucional y sectorial otorgado actualmente al fraude en el mercado asegurador español?

Los porcentajes estimados de fraude en la siniestralidad declarada en España y, el estudio del ahorro derivado de la detección de comportamientos fraudulentos, han hecho que, si bien hasta hace poco tiempo, era inusual oír la palabra fraude entre los aseguradores españoles, la tendencia actual sea totalmente distinta. Los cambios se están dejando sentir tanto a nivel legal, como a nivel sectorial y de las propias entidades individualmente consideradas. No obstante aún queda mucho por hacer.

Por un lado, y desde una óptica institucional, la nueva Ley 30/1995, de 8 de noviembre, de Ordenación y Supervisión de los Seguros Privados, ha provisto de naturaleza a los ficheros de datos personales para la prevención del fraude en la selección de riesgos y en la liquidación de siniestros. A nivel sectorial, la creación de una comisión (en fase inicial) para el tratamiento del fraude en el seno de U.N.E.S.P.A.¹¹ y la actuación de I.C.E.A.¹² en la elaboración de bases de datos (fichero de Siniestros Importantes -S.I.- e implantación de un concurso sectorial anual de detección de fraudes¹³) y, en el desarrollo de una amplia campaña publicitaria¹⁴, son muestras de una mayor coordinación dentro del sector asegurador.

Por otro lado, la actuación individual de las empresas aseguradoras ha ganado especial relevancia. En base a los resultados extraídos de una encuesta realizada por I.N.E.S.E. en 1995, entre las principales aseguradoras del mercado español (muestra formada por 24 entidades), cabe destacar:

¹¹ Unión Española de Entidades Aseguradoras y Reaseguradoras.

¹² Investigación Cooperativa entre Entidades Aseguradoras.

¹³ Según I.C.E.A (1995) el objetivo de los concursos sectoriales de detección de fraudes, es doble: por un lado, potenciar y premiar la investigación y, por otro, crear una base de datos sobre casos de fraude.

¹⁴ Bajo el slogan "Unidos contra el Fraude", la campaña publicitaria pretende no sólo recordar al personal de las entidades la existencia del problema, sino también informar al público (asegurados) de una acción sectorial frente al mismo.

- en un 56% de las entidades de la muestra existe departamento de lucha contra el fraude aunque éste no actúa de forma autónoma (aproximadamente, en la mitad de los casos se encuadra dentro del departamento de siniestros, y en la otra mitad dentro de otros departamentos),
- en un 68% de los casos dicho departamento se ocupa de todos los siniestros fraudulentos (en el 32% restante sólo lo hace sobre los que afectan al ramo de automóviles),
- el número de personas que forman parte de la unidad de lucha contra el fraude no es elevado (un promedio aproximado de 4 personas¹⁵); son individuos con un nivel de estudios medio-superior, que no se dedican de forma exclusiva a la investigación del fraude y que han recibido, en la mayoría de los casos, formación específica sobre el tema.

Asimismo, en relación a la colaboración entre entidades aseguradoras, la encuesta pone de manifiesto la existencia de intercambios de información.

Desde un punto de vista más conceptual, diferentes artículos publicados en revistas especializadas del sector revelan hechos o circunstancias ya comentados por autores de otros países. El comportamiento fraudulento por parte de un asegurado dependerá de varias premisas. Su cultura, su nivel social, su situación familiar o incluso sus características étnico-religiosas pueden aparecer como condicionantes. Sin embargo, y aunque las entidades saben que determinadas características del asegurado, y/o del contrario, así como de la póliza y del siniestro, pueden incidir en la probabilidad de aparición de fraude, no disponen hasta la fecha de mecanismos que les ayuden a dirigir la investigación. No existen tratamientos metodológicos exhaustivos que determinen las variables significativas en la explicación de los comportamientos deshonestos. Las compañías no realizan una adecuada recogida de información (es frecuente encontrar partes de declaración de siniestros con

¹⁵ No obstante, en relación a este dato, se observa una inferior participación en las respuestas (60% del total de encuestados).

muchas preguntas sin contestar) y obvian que el fraude puede aparecer en cualquier momento de la vida del seguro.

El asegurador ha de comenzar la fase de detección del fraude en el momento de suscripción de la póliza. La experiencia pone de manifiesto que la avidez de las entidades por formalizar un mayor número de pólizas, la actuación social de los agentes (en muchos casos más favorable para el asegurado que para la compañía) y, la falta de personal especializado, son factores que inciden en que el riesgo no sea correctamente examinado en la formalización del contrato.

En todos los ramos del ámbito asegurador, la realización de una correcta apreciación del riesgo debe suponer una inspección profunda de éste. La utilización de un cuestionario de solicitud detallado, en el que se consideren más aspectos de los que se tienen en cuenta actualmente (como puede ser, la situación laboral y familiar del solicitante¹⁶), la actuación del personal de la compañía en el momento de recoger los partes de siniestros (llevando a cabo una revisión de los mismos fundamentalmente dirigida a completar todos los datos requeridos) y en el momento de tramitarlos¹⁷, son actuaciones que pueden aumentar las posibilidades de detectar el fraude.

Los datos obtenidos en la inspección del riesgo han de ser reflejados dentro de la estructura del contrato y en la tarificación. El establecimiento de cláusulas asociadas, por ejemplo, a límites de responsabilidad, exclusiones en la póliza, franquicias y/o carencias puede frenar la conducta deshonesto del asegurado.

La elaboración de ficheros en los que aparezcan datos sobre los defraudadores puede constituir una buena herramienta en la lucha contra el fraude y ello, no sólo

¹⁶La consideración de la situación laboral (si el asegurado posee o no trabajo, o si la fecha en la que comenzó a trabajar es cercana a la fecha de contratación de la póliza y/o de ocurrencia del siniestro) y de la situación familiar (si el asegurado es miembro o no de una familia) son algunas de las variables analizadas por Weisberg y Derrig (1993) en la creación de un sistema de *indicadores de fraude* en el marco del seguro del automóvil.

¹⁷ El "historial de siniestros del asegurado o del demandante" o la "duración del tratamiento médico derivado del siniestro" son algunas de las variables, directamente asociadas a la tramitación del siniestro, consideradas por Weisberg y Derrig (1993).

para las compañías, sino también, y de forma muy importante, para la comunidad científico-investigadora. Lógicamente, para las entidades, el conocimiento a la hora de contratar un seguro de los datos propios de los asegurados deshonestos evitaría que el problema fuese pasando de una a otra compañía sin una solución adecuada. Sin embargo, en este aspecto entra en juego la ley de protección de datos y, como materia de regulación legal, no consideramos apropiado establecer matizaciones o conclusiones definitivas.

Ahora bien, desde el punto de vista investigador no es necesario conocer los datos personales de los individuos involucrados en el siniestro para diseñar un perfil del asegurado defraudador frente al honesto. El objetivo es determinar cuáles son las características generales de la póliza y del propio siniestro que muestran una mayor relación con la existencia de fraude, sin que intervenga para nada la identidad de los implicados. Indicar cuáles son las "señales de fraude" sobre las que la compañía ha de realizar una investigación detallada de cara a controlar el problema, es, sin duda, un paso necesario en el diseño de una adecuada política de control. A pesar de que la detección y el tratamiento del fraude llevan asociadas connotaciones importantes de subjetividad (análisis detallado de los partes de siniestro por parte del personal de la compañía, contactos con la entidad contraria,...), el uso de modelos estadísticos y econométricos ha de permitir, tal y como demostraremos en partes posteriores de este libro, no sólo cuantificar la probabilidad de aparición de fraude, sino también determinar qué variables son significativas en la explicación del mismo.

La existencia de una estrecha colaboración entre el sector asegurador y la comunidad científica puede ser decisiva a la hora de obtener una solución para el problema que estamos tratando.

1.4 La lucha contra el fraude en el seguro: una actuación sectorial que adolece la falta de tratamiento metodológico

La progresiva actuación que las compañías aseguradoras han realizado durante los últimos años frente al fraude ha derivado, como se ha puesto de manifiesto en páginas anteriores, en una mayor implicación del sector. Las entidades, bien de forma individual, bien a nivel de cooperación, presentan una posición activa frente al problema, siendo varios los países en los que la organización de la industria aseguradora y la política social existente han desembocado en la creación de instituciones específicas¹⁸ de índole suprarregional para combatir el fraude en el seguro.

La actuación eficiente de estos órganos (ver Clarke, 1990, para una revisión extensa) viene determinada por una disponibilidad actualizada de datos, así como por una estricta colaboración con las entidades aseguradoras. Dichos organismos trabajan con los medios de comunicación haciendo publicidad de las principales actuaciones fraudulentas y de las sentencias judiciales correspondientes, con fines, principalmente, disuasorios. Además se encargan de crear un banco de datos de elevada calidad informática, que ayude a las empresas de seguros miembros en la suscripción de riesgos y en la tramitación de siniestros. Finalmente, facilitan una adecuada formación profesional con el objetivo de mejorar los conocimientos que los empleados de las entidades han de poseer sobre el fraude en seguros. La formación, asimismo, de un conjunto cualificado de colaboradores, que tengan amplios conocimientos en materia criminalista y que posean la capacidad técnica necesaria, así como los contactos apropiados con las autoridades civiles y penales, contribuye al esclarecimiento de las causas de los siniestros. La coordinación de las

¹⁸ Por ejemplo, el Insurance Crime Prevention Institute (I.C.P.I.), el Insurance Fraud Bureau (I.F.B.) o los SIUs (unidades propias de investigación de cada compañía) estadounidenses, el Insurance Crime Prevention Bureau (I.C.P.B.) canadiense, el Central Information System for Risk Insurance (C.I.S.) holandés o la A.P.S.A.I.R.D. (Assemblée Plénière des Sociétés d'Assurance contre l'Incendie et les Risques Divers) francesa, son algunas de las múltiples instituciones que, a nivel internacional, luchan contra la existencia de comportamientos fraudulentos en el mercado asegurador.

actividades de investigación y la ayuda a los peritos de las entidades en la aclaración de los motivos de los siniestros, son otras de sus numerosas funciones.

Ahora bien, si la actuación de estas instituciones se podría calificar de eficiente en la mayoría de los casos, hay un aspecto sobre el que el comportamiento seguido no puede calificarse de suficiente, siendo en muchos casos inexistente. La idea de que la posición de las compañías no ha de basarse únicamente en una actitud subjetiva (en el sentido de control del fraude en base a la preparación del personal de la compañía, pero incorporando un elevado contenido de intuición personal e incluso de suerte) está ganando terreno. De esta forma, la necesidad manifestada por las entidades aseguradoras de poseer algún sistema de detección del fraude es, cada día, más notable. El interés de que la actuación del sector vaya acompañada, e incluso dirigida, por un "modelo de comportamiento" fundamentado en criterios económicos, estadísticos y econométricos es creciente, sobre todo teniendo en cuenta la reducción de costes (temporales y monetarios) que posiblemente implicará. La compañía concentrará sus esfuerzos investigadores en el estudio de aquellos datos que realmente pueden ayudar a detectar el fraude.

Sin embargo, y aunque la idea expuesta en el párrafo anterior ha sido reconocida durante los últimos años en la literatura sobre el tema, no existe un tratamiento metodológico exhaustivo que fundamente un análisis estadístico de las reclamaciones fraudulentas. Quizá sea la dificultad en la obtención de bases de datos o la escasa motivación demostrada por las aseguradoras las razones que justifican esta pobre actuación científica, siendo poco numerosos los estudios relacionados con la modelización teórica del fraude en el mercado asegurador en general¹⁹.

Para Hoyt (1990), el fraude en seguros puede manifestarse bajo diferentes formas. En primer lugar, el siniestro puede ser creado de forma deliberada. En este caso, se simula la ocurrencia del accidente para cobrar la indemnización correspondiente.

¹⁹ Fundamentalmente, y aunque tampoco son numerosos, los estudios realizados se refieren al tratamiento del fraude en el marco de los seguros del automóvil, de incendios, de enfermedad y de vida.

En segundo lugar, y tras la ocurrencia real del siniestro, los daños derivados pueden ser intencionadamente inflados para obtener una mayor compensación monetaria. El siniestro puede ser usado, en tercer lugar, para reparar, mediante las coberturas de la póliza, daños producidos anteriormente. Por último, la información facilitada por el asegurado puede no ser verdadera, en orden a conseguir una determinada póliza de seguros o una menor prima (selección adversa). Una clasificación similar, particularizada al seguro del automóvil, puede encontrarse en Weisberg y Derrig (1993).

Una vez considerados los diferentes tipos de fraude, Hoyt (1990) analiza su impacto desde un punto de vista económico (estimación del coste asociado, influencia en la elasticidad demanda-precio del seguro y determinación de la relación coste-beneficio en la creación de un mecanismo de control por parte de las entidades aseguradoras) y genera un modelo sencillo de control basado en la utilidad esperada del comportamiento fraudulento. Sin embargo, ¿cuál es la principal crítica que puede hacerse al planteamiento del autor?

Desde nuestro punto de vista, hay un aspecto básico susceptible de mejora. El planteamiento teórico presentado por Hoyt lleva asociado un modelo genérico, en el sentido de que no incorpora las diferentes formas en que el fraude puede manifestarse y que él mismo comenta. De esta manera, determina la utilidad esperada de un comportamiento fraudulento, pero no distingue de qué tipo se trata. Por ejemplo, dada la clasificación presentada, el modelo se podría ver notablemente enriquecido al diferenciar, en su formulación, entre la intencionalidad *a priori*²⁰ y *a posteriori*²¹ de que se produzca el siniestro. De esta forma el mecanismo de control podría ser más potente al diferenciar variables estadísticamente significativas en la explicación de uno u otro tipo de fraude y al cuantificar la probabilidad de su aparición.

²⁰ El siniestro es provocado o planeado.

²¹ La ocurrencia del siniestro es real, pero el asegurado (de forma individual o en connivencia con terceras partes) distorsiona los daños derivados.

El objetivo, a lo largo del libro, consistirá en plantear una primera aproximación a la identificación de las reclamaciones fraudulentas. Para ello se recurrirá a la utilización de modelos de decisión jerárquicos que permitan el tratamiento unificado de los diferentes tipos de fraude. El uso de un modelo de elección individual como el que propondremos para el fraude en el seguro del automóvil, supondrá un análisis del comportamiento de cada asegurado, con el fin de hallar una mejor caracterización de los siniestros fraudulentos y, en consecuencia, de agilizar su detección. De esta forma tendremos en cuenta, no sólo la perspectiva económica de la utilidad esperada, sino también las diferentes situaciones o actuaciones que el asegurado puede seguir a la hora de decidir si actúa o no fraudulentamente. De modo indirecto, el modelo que propondremos actuará sobre el presentado por Hoyt (1990) y que analizaremos en páginas posteriores, al cuantificar la probabilidad de ocurrencia de fraude y la probabilidad de su detección. Por tanto, se dotará a las compañías de una herramienta de control basada en la información estadística.

Para Picard (1996), la existencia o no de un compromiso por parte de las entidades aseguradoras de realizar una política de control, deriva en diferentes consideraciones, todas ellas de elevado contenido económico, en relación a un aspecto fundamental: el alcance del equilibrio de mercado ante la presencia de información asimétrica. Las compañías aluden normalmente a los elevados costes de investigación como la causa de una falta de actuación frente al problema. Ante ello, el autor sugiere la conveniencia de aplicar una política de regulación de mercado y, propone la creación de una agencia común a todas a las aseguradoras que se encargue total o parcialmente de los costes derivados del control del fraude. De esta forma, aunque el siniestro es directamente auditado por cada compañía, una parte de los costes son asumidos por el organismo creado (financiado a partir de los derechos de participación pagados por cada entidad) y la investigación se ve potenciada.

Según Picard, el desarrollo de una correcta política antifraude puede ser objeto de tratamiento mediante la generación de sofisticados modelos económicos. Sin embargo, y al igual que ocurría en el planteamiento presentado por Hoyt, dicha

modelización adolece la falta de consideración de los diferentes tipos de fraude existentes. Parece lógico pensar, ya desde nuestro punto de vista, que los costes de investigación pueden oscilar notablemente en función del comportamiento fraudulento seguido por el asegurado oportunista y ello ha de quedar reflejado en el tratamiento teórico. La cuantificación de las probabilidades asociadas a los diferentes tipos de fraude (probabilidad de crear intencionadamente un siniestro, probabilidad de incrementar de forma deliberada los costes derivados del mismo, probabilidad de declarar datos falsos en la contratación de póliza,...) y la determinación de las variables estadísticamente significativas en la explicación de cada uno de ellos, ofrece un enfoque alternativo a los presentados por los autores mencionados.

Terminamos este punto señalando que la modelización teórica presentada por Hoyt (1990) y por Picard (1996) responde a un estudio del fraude considerando el mercado asegurador general, es decir, sin distinguir por coberturas o por ramos. Sin embargo y, aunque la modelización económica del fraude en el sector, globalmente considerado, está ganando terreno (Crocker y Morgan, 1998), dentro de la literatura existente es más frecuente encontrar estudios dirigidos al análisis de la presencia de fraude en ámbitos aseguradores concretos. Entre ellos y por su importancia, destacan aquéllos que particularizan el estudio al seguro del automóvil, que serán analizados con detalle en el siguiente apartado del libro, aunque también son de especial interés aquéllos que se concentran en el análisis del fraude en el seguro de enfermedad y de vida (Münchener Rück, 1994) o de incendios (Brotman y Fox, 1988; Cloninger, 1990).

2. EL FRAUDE EN EL SEGURO DEL AUTOMÓVIL

2.1 Aspectos preliminares

La presencia de fraude es un hecho cada vez más aceptado dentro del contexto asegurador, tanto nacional como internacional. La incidencia de los comportamientos fraudulentos se deja sentir en, prácticamente, la totalidad de los ramos aseguradores y, el estudio de la problemática que genera podría realizarse, tanto dentro del contexto de los seguros de vida como en el de los de no vida.

Ocultar datos relevantes en la contratación de la póliza, como puede ser la no declaración de una enfermedad en la suscripción de un seguro de vida (siendo éste el ejemplo más clásico de selección adversa), la provocación de un incendio o la declaración de un falso atraco, bajo la cobertura de un seguro multirriesgo del hogar, son algunos de los numerosos ejemplos que nos permiten verificar lo comentado en el párrafo anterior. Sin embargo, si hay un ramo en el que se ha concedido una mayor importancia a la detección del fraude, éste es el del automóvil, ya que las consecuencias derivadas de las acciones fraudulentas son muy importantes.

Las características del seguro del automóvil desde el punto de vista de su contratación, en muchos casos obligatoria (como ocurre en España), hace que las cifras manejadas sobre el volumen de primas recaudadas y la siniestralidad sean realmente sorprendentes¹. Además hay un hecho que normalmente se relaciona

¹ En Ayuso (1995) puede encontrarse un análisis detallado de la evolución de ambos conceptos en el marco del seguro del automóvil español.

con el seguro automovilístico: la presencia de resultados técnicos negativos, y ello a pesar de que las primas pagadas por los asegurados son crecientes.

Diversos trabajos han puesto de manifiesto la dificultad de desarrollar una adecuada política de control de costes fundamentada en el control de la siniestralidad. Cummins y Tennyson (1992) manifiestan que el incremento de los precios del seguro del automóvil en la década de los ochenta aparece asociado fundamentalmente al aumento en los costes de los factores, sobre todo de las indemnizaciones por daños personales. Weisberg y Derrig (1992a) plantean la necesidad de desarrollar nuevos sistemas de tarificación en lo que respecta a la cobertura de daños corporales, dada la existencia de un incremento en los costes por siniestro superior a la tasa de inflación (en la década de los ochenta). Por su parte, Derrig, Weisberg y Chen (1994) señalan el rápido crecimiento en el importe de los pagos por daños corporales como el principal problema del seguro del automóvil en la década de los noventa.

En España, el análisis de la trayectoria global seguida por el seguro automovilístico (Ayuso, 1995) revela la existencia de dos razones que causan un incremento en los costes del seguro: el aumento de la siniestralidad relativa (número de siniestros) y el del coste promedio de las reclamaciones por accidente. El porcentaje relativo de siniestros es el que muestra una progresión más acentuada.

¿Cuáles son las razones que inducen a un crecimiento en las indemnizaciones o en el número medio de siniestros? Lógicamente, pueden ser variadas, pero hay una que cada vez está ganando más peso en la explicación de comportamientos no esperados. Cummins y Tennyson (1996) destacan la importancia que tiene la presencia de azar moral en la declaración de siniestros, valorando tanto su influencia en la probabilidad de ocurrencia del accidente como en la determinación de las pérdidas resultantes del mismo². Dado que el asegurador no

² En definitiva, su repercusión sobre el número y coste medio de los siniestros.

puede controlar la acción seguida por el asegurado para prevenir riesgos, la frecuencia de materialización de los mismos aumenta.

La falta de información que el asegurador posee sobre el posible comportamiento fraudulento del asegurado es, por lo tanto, uno de los principales problemas a los que se enfrenta actualmente el seguro del automóvil; así pues, es necesario establecer una clara definición de fraude antes de realizar el diseño de medidas de prevención y control del mismo.

A pesar de las diferencias que pueden encontrarse a nivel internacional en la regulación legal del tipo de seguro que nos ocupa, la tipología de los fraudes cometidos bajo su cobertura es bastante similar en todos los países. Frente a otros tipos de seguros donde los capitales asegurados pueden ser de elevada cuantía, en el seguro del automóvil las cantidades defraudadas no tienen por qué alcanzar cifras desmesuradas. Sin embargo, la frecuencia con que se produce este tipo de comportamiento hace que las consecuencias finales sean más negativas que en el resto de modalidades aseguradoras (Cobo, 1993).

En relación a las coberturas más afectadas por la ocurrencia de fraude, los daños materiales se sitúan en España, en primer lugar, aunque las elevadas cuantías asociadas a los daños personales hacen que su montante económico final sea igualmente considerable e incluso más elevado. Las dificultades que comporta la investigación de siniestros se acentúan en el ramo de automóvil ya que la elevada contratación del mismo y la frecuencia en la declaración de siniestros impiden, en muchos casos, realizar con la profundidad deseada, el seguimiento de algunos de los accidentes producidos.

El estudio de las circunstancias relevantes en la explicación del fraude reviste especial interés en un entorno donde su ejecución es relativamente sencilla. El propio sistema de tramitación de siniestros, el frecuente recurso al "acuerdo mutuo" entre los siniestrados³ y la escasez de medios legales para combatirlo son

³ *Parte de Declaración Amistosa* a nivel europeo o *Sistema No-fault* a nivel norteamericano.

algunas de las principales causas de cara a justificar la elevada presencia de comportamientos fraudulentos entre los asegurados.

Pero, ¿cuáles son las consideraciones teóricas más importantes en cuanto a la caracterización del fraude en el seguro de automóviles?

Realicemos en primer lugar una serie de consideraciones previas.

Si hay un país que sin duda ha de servir de ejemplo en el tratamiento del fraude dentro del marco asegurador y, concretamente, dentro del seguro del automóvil, es EE.UU. En el capítulo anterior ya lo poníamos de manifiesto cuando comentábamos que Clarke (1990) hablaba del mismo como el único país capaz de situarse en la cuarta etapa en el posicionamiento ante el fraude. De esta forma, la presentación de un estudio detallado, como el que se pretende con el presente trabajo, supone la necesidad de considerar algunos de los aspectos más importantes asociados a este mercado asegurador concreto, a las formas más habituales de fraude en el seguro del automóvil estadounidense y a las actuaciones seguidas, tanto a nivel sectorial como institucional, para combatirlo. El objetivo que se persigue es presentar un primer ejemplo de comportamiento para los aseguradores del mercado español.

Lógicamente, siguiendo el mismo criterio, cabe establecer comparaciones con otros países que, sin alcanzar la última etapa considerada, se encuentran mejor posicionados que España en cuanto a la lucha contra el fraude y que forman parte de nuestro propio entorno económico, dentro de la Unión Europea. En cualquier caso, el funcionamiento del seguro del automóvil español, regulado legalmente en consonancia con las Directivas comunitarias⁴, la articulación de la ley 30/1995 de Ordenación y Supervisión de los Seguros Privados, que recoge

⁴ Mediante el Real Decreto Legislativo 1301/1986, de 28 de junio, se produce la adaptación del Texto Refundido de la Ley de Uso y Circulación de Vehículos a Motor al ordenamiento jurídico comunitario (B.O.E. nº 155, de 30 de junio de 1986).

pautas de comportamiento inspiradas en el mercado asegurador europeo⁵ y, el trabajo desarrollado por organismos como el Comité Europeo de Seguros (concretamente por la Comisión de Seguros del Automóvil-Grupo Internacional contra el Fraude), de cara a coordinar la actuación de los diferentes entes que trabajan en los países que conforman el mercado comunitario, son sin duda muestras de un cambio en la actitud del seguro del automóvil español frente al fraude.

Atendiendo a las consideraciones establecidas, y siguiendo el orden de presentación expuesto, comenzaremos el estudio del fraude en el seguro automovilístico español, realizando una introducción al tratamiento y análisis del fraude en el seguro del automóvil estadounidense y europeo.

2.2 El Seguro del Automóvil en EE.UU.: tipología de fraude

En el seguro del automóvil estadounidense, de forma similar a lo que sucede en España, se diferencian básicamente dos coberturas en relación a la responsabilidad civil: la cobertura por daños corporales (*Bodily Injury Liability*)⁶ y la cobertura por daños materiales (*Property Damage Liability*).

Siguiendo a Weisberg y Derrig (1993) podemos definir la cobertura por daños corporales como aquella que *“provee indemnización a la víctima herida en un accidente cuando el conductor asegurado es legalmente responsable”*. La compensación se produce en relación a dos tipos de daños: los *daños especiales* y los *daños generales*. Los primeros hacen referencia fundamentalmente a los gastos médicos y pérdida salarial derivadas del siniestro mientras que los segundos son los asociados a la valoración del daño desde un punto de vista

⁵ La nueva ley ha supuesto la creación de un fichero sectorial de vehículos asegurados denominado F.I.V.A. (*Fichero Informático de Vehículos Asegurados*) que, en coordinación con el Consorcio de Compensación de seguros funciona, por ejemplo, de forma parecida al S.I.T.A. (*Sistema Informático Targhe Assicurate*) italiano. Cabe señalar que la constitución del F.I.V.A. responde a una llamada de los órganos de control comunitarios para hacer frente al elevado número de vehículos que circulan en España sin seguro de responsabilidad civil obligatoria.

⁶ El seguro de responsabilidad por daños corporales en el marco automovilístico estadounidense es el más importante dentro del conjunto de los seguros no-vida (Weisberg y Derrig, 1993).

más subjetivo (daño moral). En cuanto a la cobertura por daños materiales el significado es análogo al que podemos encontrar en España y se relaciona con el resarcimiento de la pérdida ocasionada por los daños al vehículo.

En relación a la legislación existente, la autonomía que impera en los diferentes estados hace que el comportamiento no sea homogéneo, diferenciándose básicamente dos sistemas compensatorios: el tradicional o *Tort System* y el *No-fault System*⁷. Información exhaustiva sobre el funcionamiento de ambos sistemas puede encontrarse en Vaughan y Elliot (1978), en Weisberg y Derrig (1992a, 1992b), en Derrig, Weisberg y Chen (1994) y en Morillas (1992).

Bajo el tradicional *Tort System*, el asegurado que sufre un siniestro siendo suya la culpa es declarado legalmente responsable y ha de compensar al contrario mediante el pago de los daños producidos. Para protegerse de las pérdidas derivadas del suceso, puede contratar la cobertura de responsabilidad civil, bajo la cual la compañía se compromete a defenderle en caso de juicio y a pagar la indemnización correspondiente en función de los límites de la póliza.

A pesar de su vigencia, la existencia de numerosas críticas hacia este sistema, relativas sobre todo a los costes de funcionamiento del mecanismo asegurador y a las dificultades derivadas de litigios largos y costosos (sentencias con indemnizaciones desmesuradas), ha derivado en una mayor importancia del sistema *No-fault*.

Bajo un sistema de compensación *No-fault* ("*no culpa*") la intención de culpar a la parte causante del siniestro se relaja y cada una de las compañías indemniza a su asegurado por los daños derivados del accidente (hasta un límite). A pesar de que algunas propuestas han ido dirigidas a su implantación en los daños al vehículo, su foco principal de acción aparece en el marco de la cobertura por daños corporales.

⁷ Massachusetts fue el primer estado en adoptar un sistema *No-fault* en 1971.

En su sentido más puro, este sistema supondría la abolición total del sistema *Tort System*, dado que cada una de las partes recobraría de su aseguradora la indemnización correspondiente a daños especiales (gastos médicos, ...) y se eliminaría la posibilidad de reclamar por daños generales. Sin embargo, la implantación del sistema *No-fault* no ha ido dirigida a eliminar el sistema tradicional, sino a limitar su campo de acción. El objetivo principal ha sido rebajar el número de litigios (y, por tanto, alcanzar una reducción de costes), en relación a la responsabilidad por daños de cuantía no elevada. De esta forma se limita la posibilidad de demandar por daños morales (daños generales), pero se consigue una rápida compensación por los daños económicos⁸ (daños especiales), independientemente de la culpa.

La existencia de un límite que permite pasar de la compensación base de un sistema *No-fault* a una compensación muy superior (fundamentalmente asociada a daños generales), hace que el fraude cometido en daños corporales acapare una gran atención y sea objeto de un estudio exhaustivo en la literatura existente sobre el tema. La posibilidad de conseguir mayores indemnizaciones ha derivado en una actitud cada vez más frecuente entre los asegurados: incrementar fraudulentamente las consecuencias derivadas del siniestro. Marter y Weisberg (1992) y Weisberg y Derrig (1992b, 1992c) explican cómo los gastos médicos derivados de siniestros por daños corporales se encuentran fuera de control para las aseguradoras. Los asegurados persiguen maximizar las cantidades percibidas por tratamientos, diagnósticos, para alcanzar indemnizaciones mayores a las que les proporciona el sistema *No-fault*.

Teniendo en cuenta lo anterior, las definiciones y estudios que encontramos para el fraude en el seguro automovilístico estadounidense suelen referirse fundamentalmente a la cobertura por daños corporales. La importancia concedida

⁸ Cuando los daños no superan los límites establecidos (monetarios y/o verbales), los demandantes son compensados mediante una indemnización por daños especiales (en Massachussets, por ejemplo, se establece la denominada *Personal Injury Protection -P.I.P.-*, que tiene, además, carácter obligatorio).

a los daños materiales es mucho menor, alegándose la existencia de cantidades defraudadas de magnitud mucho menos relevante.

Weisberg y Derrig (1993) definen el fraude en el seguro del automóvil como el “*intento de engañar al asegurador falsificando algún aspecto pertinente del accidente, de los daños corporales o de los daños generales resultantes del siniestro*”. Bajo esta conceptualización general, pero haciendo especial hincapié en los daños corporales, ambos autores establecen una diferenciación entre lo que han venido a denominar *hard fraud* y *soft fraud*. En el primero de los casos, se hace referencia al fraude *strictus sensu*, con todas sus connotaciones y, que analizaremos a continuación; en el segundo caso, el estudio del fraude supone analizar aquellas situaciones en las que el demandante, en colaboración normalmente con el médico, decide distorsionar al alza los gastos asociados al tratamiento derivado del siniestro para lograr una mayor compensación (Marter y Weisberg, 1991, 1992). De ahí que se denomine, normalmente bajo el nombre de “*build-up*”. Siguiendo a Derrig y Ostaszewski (1994b), este tipo de fraude se define como un intento, por parte del demandante y/o del médico, de sobrevalorar los daños por los que se solicita indemnización.

Respecto al fraude propiamente dicho o *hard fraud*, son dos los tipos básicos de comportamiento que pueden producirse: el fraude planeado (“*planned fraud*”) y el fraude oportunista (“*opportunistic fraud*”⁹).

El fraude planeado aparece en aquellas situaciones en las que la ocurrencia del siniestro no es accidental, sino que es deliberadamente provocada. En este caso, es posible encontrar elevados niveles de sospecha, no sólo en relación al propio accidente, sino también en relación al demandante, al asegurado, al daño producido y al tratamiento médico derivado.

⁹ Para establecer estas consideraciones seguimos las principales conclusiones extraídas por Weisberg y Derrig (1993).

Por el contrario, el fraude oportunista aparece asociado a un siniestro cuya ocurrencia es totalmente legítima, siendo el demandante, los daños corporales y el tratamiento médico los que provocan mayores sospechas. Este tipo de fraude, más frecuente en cuanto a su aparición, tiene lugar sobre todo en aquellos casos en los que el asegurado aprovecha la ocurrencia real del siniestro para actuar en beneficio propio o de un tercero, buscando una compensación que no le corresponde. En definitiva, son casos en que ocurre el siniestro pero no existe daño.

Tomando como referencia lo anterior, podemos decir que la clasificación de siniestros en relación a la existencia o no de fraude, en el marco del seguro automovilístico norteamericano, se concreta en cuatro categorías: siniestro legítimo ("*legitimate*"), siniestro planeado ("*planned fraud*"), siniestro oportunista ("*opportunistic fraud*") y siniestro con aumento deliberado de los gastos asociados al tratamiento médico derivado del mismo ("*build up*").

La categorización presentada ha sido realizada tomando como referencia el estudio centrado en el tratamiento metodológico del fraude de Weisberg y Derrig (1993) y será retomada, posteriormente (en el Capítulo 3), en el análisis de los métodos cuantitativos susceptibles de aplicación a la modelización teórica del comportamiento deshonesto de los asegurados. Sin embargo, desde un punto de vista más general, es posible encontrar una clasificación que recoja las formas de defraudar más típicas y/o los estamentos implicados. Tomando como referencia las principales consideraciones presentadas por el Insurance Fraud Bureau de Massachusetts en el *Annual Report* de 1994, los principales tipos de fraude en el seguro del automóvil estadounidense aparecerían relacionados con las siguientes situaciones:

- a) reclamación por daños corporales inexistentes o por una pérdida de salario ficticia;
- b) aparición de cadenas de fraude, con connivencia de abogados, médicos, etc.;

- c) declaración de un mismo siniestro en varias compañías (aseguramiento múltiple);
- d) falso robo, no sólo del vehículo, sino también de accesorios;
- e) accidente planeado con la utilización de nombres ficticios;
- f) incremento en los daños del vehículo derivados de un siniestro;
- g) incendios provocados de vehículos, y
- h) siniestros en los que el demandante intenta conseguir coberturas para pérdidas no contempladas en la póliza.

Como observaremos a continuación, la similitud con los tipos de fraude cometidos en España es más que evidente. Por lo tanto, las formas de actuación frente al mismo habrán de ser también similares.

2.3 El seguro del automóvil en España y los tipos de fraude más frecuentes dentro del marco europeo

El seguro automovilístico español, al igual que el del resto de Europa (y como hemos visto, el estadounidense), se estructura en torno a dos coberturas básicas: la de los daños corporales y materiales. Desde la entrada en vigor de la Ley de Uso y Circulación de Vehículos de Motor (en diciembre de 1962), pasando por la adaptación al ordenamiento jurídico comunitario (Real Decreto 1301/1986, de 28 de junio) y teniendo en cuenta la implantación en 1988 de la Declaración Amistosa de Accidente y del Convenio de Indemnización Directa Español (C.I.D.E.)¹⁰, la evolución experimentada por el seguro del automóvil español ha sido notable. Sin duda, la culminación de dicha evolución se ha puesto de manifiesto con la publicación del nuevo sistema para la valoración de los daños y perjuicios causados a las personas en accidentes de circulación, en el marco de la ley 30/1995, de 8 de noviembre, de Ordenación y Supervisión de los Seguros Privados.

¹⁰ Dicho convenio busca el "máximo automatismo en la liquidación, tramitación y pago de los siniestros, así como la compensación automática de los expedientes pagados por cuenta de otra entidad". En base al mismo, cada compañía indemnizará a su propio asegurado aunque el responsable sea un tercero. Asimismo, cabe destacar la entrada en vigor en 1990, del convenio A.S.C.I.D.E. (Acuerdo Suplementario del Convenio de Indemnización Directa Español).

Remitiéndonos a Morillas (1992) y a Ayuso (1995) para encontrar una explicación exhaustiva del funcionamiento del seguro automovilístico en España¹¹, nos centraremos, a continuación, en el estudio de la tipología más frecuente de fraude en el mercado asegurador español, siempre dentro del marco europeo en que se integra.

Según el Comité Europeo de Seguros (1996), los tipos de fraude más frecuentes en Europa aparecen asociados, en la mayoría de los casos a las siguientes situaciones:

- a) omisión de información requerida para valorar el riesgo;
- b) no declaración de la existencia de varias pólizas para cubrir la ocurrencia del mismo siniestro;
- c) realización de seguros sobre cosas que no existen o cuyo valor real ha sido exagerado;
- d) provocación deliberada del accidente;
- e) declaración de un siniestro que no ha ocurrido realmente,
- f) intención de conseguir dentro de la indemnización algo que no está incluido en la póliza como cobertura, y
- g) valoración del daño por encima de su precio real.

Pues bien, con la única diferencia de que mientras en España el fraude es perpetrado generalmente de forma individualizada, en el resto de Europa, es más común el fraude cometido por grupos organizados, la tipología existente es similar en todo el marco comunitario. Los tipos de comportamientos fraudulentos más frecuentes, particularizando al seguro automovilístico español, pueden sintetizarse en:

- a) *Falsa declaración de siniestros*. En este contexto las acciones más típicas aparecen relacionadas con el establecimiento de acuerdos

¹¹ Asimismo, en Solana de Ariza (1988) y en U.N.E.S.P.A. (1994f) puede encontrarse un análisis detallado del funcionamiento de los Convenios C.I.D.E.-A.S.C.I.D.E.

ilegítimos entre el asegurado y el contrario, o con la utilización indebida del Parte Europeo de Declaración Amistosa de accidente. También es frecuente la ocultación de datos del siniestro no cubiertos en la póliza, la provocación intencionada de daños por parte del asegurado, así como la inclusión de varios siniestros en un mismo parte, para evitar la penalización o bien para cubrir siniestros anteriores.

b) *Contratación de la póliza después de producirse el siniestro.* En este caso, el asegurado se beneficia de un seguro que no poseía cuando se produjo el accidente (o bien de unas coberturas que en dicho momento no aparecían en la póliza).

c) *Falsa declaración en la póliza.* Utilización de datos falsos, bien referentes al propio asegurado (edad, antigüedad de carnet,...), bien referentes al conductor habitual o al vehículo (averías, antigüedad,...).

d) *Robo.* En este caso, el asegurado simula un intento de robo para cubrir desperfectos anteriores, o bien oculta la aparición de su vehículo para cobrar la indemnización, o simplemente denuncia la sustracción de alguna parte de su automóvil.

e) *Aseguramiento múltiple.* El asegurado contrata diferentes pólizas para cubrir el mismo riesgo en diferentes compañías. El objetivo es cobrar varias indemnizaciones por un siniestro que en la mayoría de los casos provocará.

f) *Fraude de taller.* Recoge aquellos casos en los que, en connivencia o no con el asegurado, el taller falsifica intencionadamente los daños derivados del siniestro para obtener un beneficio.

Una exposición alternativa de las posibles formas de cometer fraude, dentro del mercado español, puede encontrarse en C.E.S. (1992) y en Cobo (1993). En

relación a esta última referencia, cabe destacar la mención de una determinada forma de actividad fraudulenta calificada de forma genérica como *pseudo-fraude* y que responde a “*aquellos supuestos en los que el asegurado pretende prolongar los efectos de la cobertura a unos límites que no corresponden, a diferencia de los verdaderos fraudes, que son los que se encaminan hacia la obtención de un beneficio que bajo ningún concepto está comprendido en la cobertura de la póliza, ya sea en beneficio propio o de un tercero*”. En base a su definición, el *pseudo-fraude* sería en cierto modo equivalente al *soft fraud*, presentado anteriormente para el mercado norteamericano.

Desde otro punto de vista, el análisis de la tipología de fraude puede realizarse tomando como referentes los diferentes estamentos implicados en la ejecución del fraude, a su actuación, así como a los diferentes factores, bien objetivos bien subjetivos, que propician la elevada participación en acciones fraudulentas.

Principalmente, los estamentos con capacidad de fraude en el seguro del automóvil son los asegurados, los peritos, los talleres, los agentes y los grupos organizados. Su actuación no siempre se produce de forma aislada, de modo que a veces se establecen relaciones entre ellos para llevar a cabo acciones de fraude (“*cadena de fraude*”). Lógicamente las formas de defraudar son muy variadas y presentan rasgos diferenciadores según las características del defraudador (en función de si es asegurado, perito,...).

La actuación de las *bandas organizadas*, más o menos profesionales, puede ser considerada como un importante foco de fraude delictivo, a pesar de que en España su implantación no es tan notable como en el resto de Europa. Las acciones de asegurados y de bandas podrían clasificarse como fraude “*desde fuera de la compañía*”, es decir, fraude ejecutado por la parte que, dentro de la celebración del contrato de seguros, cede el riesgo.

Si el estudio lo trasladamos al marco de la compañía de seguros, cabe destacar que las acciones fraudulentas aparecen ligadas sobre todo a la actuación de los

agentes y de los peritos que trabajan para la entidad aseguradora en cuestión. En el caso de los agentes, se deben principalmente a incorrecciones deliberadas en la liquidación de los recibos de prima o bien a actuaciones favorecedoras para su cliente y no para la entidad (retroceso en la fecha de efecto de la póliza para cubrir la ocurrencia de un siniestro). El fraude cometido por los *peritos* se pone en práctica, generalmente, en relación con otros estamentos implicados, como son los talleres o los concesionarios. Una de las acciones fraudulentas más típicas es la que implica la sobrevaloración de la tasación del daño.

La actuación de los talleres, por lo que respecta a daños materiales, y la actuación del personal médico (tratamientos médicos derivados de un siniestro) en relación a los daños corporales, son también importantes focos de fraude.

En España la cobertura más afectada es la de responsabilidad civil por daños materiales (I.C.E.A., 1996) a diferencia de lo que hemos manifestado anteriormente para EE.UU. A una distancia notable se encuentra la cobertura por daños personales (responsabilidad civil por daños corporales y garantía de ocupantes) y la cobertura por daños propios, incendio, rotura de lunas y robo.

Lógicamente, no podíamos acabar este apartado sin hacer referencia a uno de los comportamientos fraudulentos más frecuentes en el mercado asegurador español y europeo en general: la elevada presencia de vehículos que circulan sin ningún tipo de seguro. La preocupación que este punto ha suscitado en los últimos años ha derivado en una mayor concienciación a nivel sectorial (creación de ficheros de vehículos asegurados) e institucional.

2.4 Estadísticas de fraude en el seguro del automóvil

Cuantitativamente hablando, los porcentajes de fraude en el seguro del automóvil son altamente significativos. Como comentábamos al inicio de este capítulo, la evolución experimentada por la siniestralidad (número y coste medio de los

siniestros) en este campo asegurador ha alertado sobre la posible presencia de fraude.

El rápido crecimiento en los costes por siniestro ligados a la cobertura de daños corporales en el seguro del automóvil estadounidense puede atribuirse fundamentalmente a dos hechos: el incremento en el número de reclamaciones por accidente y la presencia de costes médicos cada vez más elevados¹² (Derrig, Weisberg y Chen, 1994). Este comportamiento, que en un principio podría venir explicado por el propio funcionamiento del sistema, se transforma en extraño al observar la tendencia de la siniestralidad en la cobertura por daños materiales, que muestra un claro decrecimiento¹³ (Cummins y Tennyson, 1992; Derrig, Weisberg y Chen, 1994).

Las estimaciones de fraude existentes para el seguro del automóvil en EE.UU hacen referencia, principalmente, a la cobertura de daños corporales. En relación a la misma, y centrandó el estudio en Massachusetts, Weisberg y Derrig (1991, 1992b, 1993) indican que el porcentaje de siniestros con fraude, sin considerar la posible distorsión al alza de los costes médicos derivados del siniestro, se estima aproximadamente en un 10% (un 11.8% en concreto para siniestros ocurridos entre 1985 y 1986 y un 9.1 % para los siniestros de 1989). Sin embargo, el porcentaje de siniestros que son sospechosos (considerando también el incremento provocado en los costes médicos) aumenta de un 34.8% en 1985 y 1986 a un 48.2% en 1989.

Por su parte, Derrig, Weisberg y Chen (1994) estiman que entre el 40 y el 50% de los siniestros relacionados con la cobertura de daños corporales (en Massachusetts) contienen algún grado de sospecha. Además, la proporción de fraudes muestra una tendencia creciente, tal y como evidencia la evolución de la proporción de siniestros con daños corporales.

¹² No obstante, los autores también hacen referencia a la influencia de los sistemas compensatorios existentes.

¹³ Mientras que la frecuencia de siniestros asociados a la cobertura por daños corporales, entre 1989 y 1992, mostraba un incremento medio de aproximadamente un 9% anual, la frecuencia por daños materiales para el mismo periodo, manifestaba un decrecimiento medio del 3% anual (Derrig, Weisberg y Chen, 1994).

En Europa, el coste estimado del fraude se cifra en aproximadamente el 2% de las primas recaudadas en todos los ramos (Comité Europeo de Seguros, 1996). En Alemania, el porcentaje estimado de fraude (Clarke, 1990), para el seguro del automóvil a terceros, es del 11%. Porcentajes similares se mantienen al considerar otros ramos: en Francia, el coste estimado de fraude se cifra en aproximadamente un 5% de los pagos por reclamaciones en no-vida y, en el mercado asegurador británico se sospecha que, del total de reclamaciones que se producen, alrededor de un 10% son fraudulentas. Por su parte, en Italia, el coste estimado de fraude en el ramo de autos se cifra en un 5% de los pagos por reclamaciones (Comité Europeo de Seguros, 1996).

En España, los porcentajes de fraude en relación a la siniestralidad global del seguro del automóvil se estiman en un intervalo que va del 15 al 60%. La gran amplitud del intervalo viene provocada por las diferencias entre las compañías grandes y las pequeñas: mientras que las primeras muestran una mayor homogeneidad en sus estimaciones (entre el 15 y el 20%), las segundas muestran estimaciones más dispares. No obstante hay que señalar que la parte que realmente se detecta es mínima, de forma que únicamente un 1-10% del fraude total llega a descubrirse (C.E.S., 1992). Por su parte, Cobo (1993), estima que, aproximadamente, un 22% de las reclamaciones contienen sospecha de fraude.

Las principales conclusiones extraídas por I.C.E.A. a partir de los Concursos de Detección de Fraudes en el Seguro, celebrados en España desde 1994, reflejan que el ramo de automóviles (frente a otros como el multirriesgo, la responsabilidad civil general o los riesgos personales) es el que presenta un mayor porcentaje de fraudes detectados¹⁴. La cobertura más afectada es la de responsabilidad civil por daños materiales, seguida de la de daños personales (responsabilidad civil por daños corporales y garantía de ocupantes) y otros daños (daños propios, incendio y rotura de lunas). No obstante, los resultados reflejan la gran incidencia del fraude en el coste medio por siniestro con daños

¹⁴ No obstante, los datos presentados habrán de tomarse con la debida precaución, dado que en la aportación de casos de fraude por las entidades participantes puede haber una sobre-representación del seguro de autos.

corporales (en contraposición a la cobertura de daños materiales en la que, a pesar de la elevada frecuencia de fraudes, el coste medio suele ser menor).

En relación al defraudador, las figuras del asegurado y/o del contrario son las más frecuentes. Sin embargo, aunque en porcentajes mucho menos significativos, también se advierte la presencia de fraudes realizados por talleres y por agentes.

Tomando en consideración las modalidades más frecuentes de fraude, destacan las situaciones causadas por el deseo de favorecer a un tercero, la ocultación de casos excluidos en la póliza, la falsificación de la ocurrencia o de las consecuencias del siniestro, la no declaración de determinadas circunstancias personales y la contratación de la póliza con posterioridad a la ocurrencia del accidente. El tipo de vehículo más involucrado es, sin duda, el turismo.

En relación al resultado económico del fraude detectado, calculado en base a las cantidades inicialmente reclamadas, las finalmente pagadas y los costes de investigación, cabe destacar la existencia de elevados porcentajes de ahorro. Asimismo, y considerando únicamente los recursos invertidos en investigación, los rendimientos obtenidos son igualmente muy significativos¹⁵.

Sin embargo, el pequeño porcentaje de casos de fraude que se acaba detectando hace que las entidades continúen pagando elevadas sumas en concepto de indemnizaciones fraudulentas. El campo de investigación abierto es, por lo tanto, muy extenso, sobre todo teniendo en cuenta la necesidad manifiesta del mercado asegurador de disponer de herramientas adecuadas de control y detección del fraude (C.E.S., 1992).

¹⁵ Aproximadamente, por cada peseta invertida en investigación se obtiene un rendimiento de 54 pesetas de ahorro (I.C.E.A., 1997).

2.5 Aspectos prácticos del tratamiento del fraude en el seguro del automóvil

La existencia de fraude es un hecho totalmente asumido por las entidades aseguradoras que, tal y como hemos comentado, han comenzado a desarrollar acciones dirigidas a combatirlo. Según el Comité Europeo de Seguros, las líneas de acción frente al fraude deben concretarse principalmente en tres aspectos: prevención, detección y sanción, siendo necesaria una actuación a nivel interno de cada país, que goce de coordinación supranacional.

Las actividades desarrolladas hasta la actualidad en los diferentes países, tanto del continente norteamericano como del europeo, han sido variadas y, si bien en algunos casos podemos hablar de acciones más marcadas (mayor número de organismos anti-fraude, creación en las entidades de departamentos específicos,...), lo cierto es que, cada vez más, todos los países tienden a homogeneizar su comportamiento frente al fraude en general, y al automovilístico en particular.

Hagamos una breve presentación de algunos de los aspectos prácticos de la gestión del fraude en el seguro del automóvil estadounidense y en los países de nuestro entorno europeo, y pasemos, después, a comentar la situación española.

En EE.UU., la utilización de técnicas avanzadas de investigación de siniestros y el funcionamiento de organismos especializados en el tratamiento del fraude gozan de un elevado nivel de desarrollo¹⁶. La realización de exámenes médicos independientes por parte de la compañía para comprobar la correlación entre el siniestro y los daños producidos, la creación de unidades específicas de investigación (*SIUs*) y el funcionamiento de órganos como el *National Automobile*

¹⁶ Una explicación extensa de las diferentes actividades y organismos que actúan previniendo el fraude en el seguro del automóvil estadounidense puede encontrarse en Clarke (1990) y en Weisberg y Derrig (1993).

*Theft Bureau*¹⁷ o el *Central Index Bureau*¹⁸ sirven como muestra de lo comentado. Sin embargo, hay una institución que ha destacado durante los últimos años por su trabajo en relación a la detección del fraude, sobre la que vamos a hacer un comentario más extenso: el *Insurance Fraud Bureau de Massachusetts (I.F.B.)*¹⁹.

Operativo desde mayo de 1991, el I.F.B. aparece como uno de los principales focos en la lucha contra el fraude dentro del ámbito asegurador, no sólo americano, sino también mundial. Los objetivos de este organismo, para estudiar la presencia de acciones fraudulentas dentro del mercado de seguros, se pueden concretar en tres:

- refinar la definición de fraude,
- estudiar el valor potencial de los "indicadores de fraude", y
- mejorar, en la medida de lo posible, la tramitación de siniestros.

En relación al primer objetivo, el I.F.B. establece una definición que tiene en cuenta el proceso que lleva al asegurado a tomar la decisión de actuar fraudulentamente. Dado que el objetivo principal de nuestro trabajo es realizar una modelización del fraude que tenga en cuenta el esquema de decisión seguido por el asegurado (y/o el tercero) a la hora de comportarse o no fraudulentamente, posponemos para más adelante el análisis detallado de los tipos de actuaciones fraudulentas que el I.F.B. propone. No obstante y, sólo de forma indicativa, señalamos que la clasificación de un siniestro por parte del citado organismo, influenciada por los trabajos de sus investigadores, obedece a alguna de las siguientes categorías:

- Siniestro probablemente legítimo ("Probably legitimate").
- Sospecha únicamente de "build-up" ("Suspected build-up only").

¹⁷ Creado en 1912, gozó de una enorme expansión en los años setenta, representando en la actualidad a más del 95% de los aseguradores de automóviles estadounidenses. Recoge datos relacionados con robos de autos y realiza estimaciones sobre la probabilidad de que algunos de ellos sean fraudulentos.

¹⁸ Organismo que responde a cuestiones planteadas por los aseguradores en relación a siniestros anteriores asociados a una determinada póliza, indicando el seguimiento realizado por el tramitador en cada caso.

¹⁹ Los contactos establecidos con dicho organismo, y concretamente con su vice-presidente, el Dr. Richard Derrig, han influido de forma muy positiva en el desarrollo de este trabajo.

- Sospecha de fraude "oportunistico" ("Suspected opportunistic fraud").
- Sospecha de fraude planeado ("Suspected planned fraud").
- Otros.

El segundo objetivo del I.F.B. se ha centrado en el estudio del valor potencial de los denominados "indicadores de fraude". Aspectos relacionados con las características del accidente, como pueden ser la hora del suceso, el tipo de colisión y los individuos involucrados (incluyendo la presencia de testigos y de policía) son factores que, sin duda, tienen una gran importancia como posibles señales del carácter fraudulento de un siniestro.

La actitud del demandante frente al suceso ha de ser igualmente considerada. Circunstancias como el historial de siniestros que le caracteriza, así como su situación laboral, pueden ser tratados como potenciales indicadores de fraude. Asimismo, las características del daño producido y del tratamiento médico derivado del siniestro generan una serie de variables relacionadas con el perfil de la reclamación (presencia o no de evidencia objetiva de daño, uso de un elevado volumen de visitas al médico,...) que, análogamente, han de ser objeto de estudio. La eventual pérdida de salario que para el demandante se deriva de la ocurrencia del suceso aparece también recogida en una serie de indicadores, a los que se han de añadir aquellos relacionados directamente con las características del conductor asegurado.

En base a todas estas observaciones, el I.F.B. ha creado una amplia serie de indicadores relacionados con el seguro del automóvil. Su utilización conjunta es complicada (sobre todo, por la dificultad de conseguir información para todos ellos), haciéndose necesario el uso de métodos teóricos que permitan discernir qué grupo de indicadores son los que realmente intervienen en la detección de un tipo específico de comportamiento fraudulento y si aportan información estadísticamente significativa.

La creación de un correcto mecanismo que mejore, en la medida de lo posible, la tramitación de siniestros, agilizando la actuación de la entidad aseguradora en la lucha contra el fraude, aparece como el tercer objetivo del I.F.B. El desarrollo de un sistema que permita realizar una investigación eficiente de los siniestros (basado principalmente en los indicadores seleccionados) comportará la optimización de los recursos utilizados. De esta forma, los elevados costes de inversión inicial derivados de la creación del sistema, se verán compensados, con gran probabilidad, por la presencia de importantes beneficios para la entidad.

La actuación del I.F.B. favorece la posibilidad de realizar acciones criminales para aquellos siniestros que hayan sido declarados fraudulentos, con lo que las expectativas de luchar contra el fraude se ven acrecentadas. Es más, él mismo se encarga de llevar a cabo una investigación más profunda y, en su caso, de proceder a realizar los trámites legales, liberando a la compañía de actuar como parte demandante frente a los tribunales.

Analizados, aunque sea de forma muy sintetizada, algunos de los principales rasgos del tratamiento práctico del fraude en EE.UU., ¿cuál es la situación del mismo en Europa?

La decisión de prevenir, detectar y sancionar el fraude reviste grados diversos en los diferentes mercados europeos (Greenouff, 1990)²⁰. En Bélgica, los aseguradores decidieron crear, ya en 1981, una comisión para prevenir el fraude, especialmente en el ramo de autos. Entre sus actividades destaca la puesta en marcha de un sistema de información entre compañías, que permita la difusión de información relativa a siniestros en los que se han detectado acciones fraudulentas. Asimismo, un fichero de vehículos robados permite advertir a las entidades de su aparición. La creación de grupos de inspección para casos especiales, y la importancia concedida a la correcta recogida de datos sobre el conductor habitual son otras de las características de la política anti-fraude desarrollada en este país.

²⁰ Información adicional sobre las medidas adoptadas por los países europeos frente al fraude pueden encontrarse en Clarke (1990), C.E.S. (1992) e I.N.E.S.E. (1996b).

Francia ha sido calificada como el país líder en control de fraude dentro del contexto europeo (Clarke, 1990). Su posicionamiento frente al problema comienza a principios de la década de los setenta, siendo de destacar tanto la actuación sectorial como la individual de cada compañía. A nivel colectivo, destaca la creación de un organismo único para todo el mercado asegurador francés, la Agencia para la Lucha contra el Fraude en el Seguro (A.L.F.A.), así como el asiduo intercambio de información que se produce entre entidades (creación de ficheros para centralizar las informaciones sobre fraudes,...). A nivel individual, el mayor control de riesgos en la suscripción de la póliza, la creación de servicios especiales anti-fraude o el desarrollo de medios técnicos destinados a detectar fraudes mediante el cruce de datos, son algunos de los ejemplos que evidencian los avances conseguidos en este país.

En Austria, cada aseguradora posee un especialista anti-fraude que ayudará, si se considera necesario, en la tramitación de los siniestros. A nivel colectivo, la creación de un departamento especializado ha permitido crear bases de datos con información de reclamaciones fraudulentas.

En 1963, las sociedades de seguros italianas crearon un servicio anti-fraude denominado *Schedario Sinistri*, al que se encuentran adheridas todas las empresas miembros de la A.N.I.A. (Asociación Italiana de Aseguradores). En el mismo se puede encontrar información de todos los siniestros relacionados con la cobertura de responsabilidad civil en autos. Además es posible trabajar con otros ficheros informáticos, como el *Sistema Informatico Targhe Assicurate (S.I.T.A)*, que registra las matrículas de los vehículos y su aseguradora. Es preciso destacar la realización de acciones especiales destinadas a la gestión de siniestros en las zonas donde se detecta elevada presencia de fraude.

En Alemania, el sistema de información *KHI* facilita datos sobre los siniestros sospechosos. No obstante es necesario señalar que las principales acciones se han desarrollado a nivel individual por cada compañía, dada la reticencia de los aseguradores a facilitar información privada (Clarke, 1990).

Suiza es un país en el que destaca la existencia de un sistema centralizado de información (*Zentrale Informations System -Z.I.S.-*) destinado a registrar todos los casos de fraude en seguros. La Asociación Suiza de Aseguradores de Responsabilidad Civil y Automóviles (A.R.C.A.) facilita información sobre los defraudadores y, el tipo de fraude cometido, a todas las entidades miembros. Además, la existencia de un fichero de datos (*RIPOL*) dirigido, entre otras cosas, a la localización de personas sin seguro de responsabilidad civil en autos, es otra muestra importante de los avances existentes en este país en la lucha contra el fraude.

En 1987 se creó en el Reino Unido el *Motor Insurance Anti Fraud and Theft Register (M.I.A.F.T.R.)*. Se trata de un registro antifraude y antirrobo dotado de un fichero informático con datos de siniestros de automóviles y de vehículos robados. El cruce de información permite alertar a las compañías sobre la posible existencia de fraudes.

En Noruega todas las grandes compañías cuentan con investigadores propios de fraude. Asimismo se da una amplia publicidad de los casos detectados para dejar constancia del deseo, por parte de las aseguradoras, de combatir los comportamientos deshonestos. El principal registro incluye siniestros de los ramos no-vida (en funcionamiento desde 1988) y está enfocado, fundamentalmente, a prevenir la existencia del seguro múltiple.

Revisado el comportamiento seguido por los países de nuestro entorno, pasamos a continuación a analizar cuál es la situación actual del mercado de seguros español en relación a las acciones ejecutadas contra el fraude.

En España, a pesar del avance experimentado durante los últimos años, la reacción tanto a nivel individual como sectorial no puede considerarse óptima.

Analicemos detalladamente la situación, dando un doble enfoque al estudio del problema. Por un lado, es necesario determinar por qué las entidades

aseguradoras no han reaccionado decisivamente frente al fraude y, por otro, estudiar por qué los asegurados y demás individuos llevan a término este tipo de acciones.

Las razones por las que las compañías no han adoptado un mayor número de medidas contra el fraude son variadas. Hechos como los gastos de gestión asociados, la necesidad de realizar una gran inversión inicial en la elaboración e incorporación de un mecanismo de detección (diseño de grandes ficheros inter-compañías, creación de departamentos específicos de estudio de fraude,...), así como la alusión a otras razones también de índole monetaria, suelen aparecer como las principales justificaciones de una falta de actuación. No obstante, el gran peso que tiene dentro de la siniestralidad debería suponer un nuevo punto de mira, dado que los costes iniciales y de mantenimiento producidos por la implantación de un mecanismo de detección se verán, con gran probabilidad, sobrepasados por los beneficios inducidos por éste²¹.

Desde el punto de vista del asegurado, los factores que propician las acciones fraudulentas son muy diversos. El funcionamiento de los convenios CIDE/ASCIDE, fundamentalmente en lo referente a la brevedad en el plazo de indemnización; el hecho de que el fraude, no se considere por el colectivo de asegurados como un perjuicio y, la escasez de cooperación entre las compañías son algunos de las razones que inducen al asegurado a actuar en beneficio propio o de un tercero. Igualmente, los agentes, peritos y talleres presentan escasa motivación para modificar su actitud, dadas las relativas facilidades que hasta el momento han encontrado a la hora de actuar fraudulentamente.

²¹ Según los resultados obtenidos del III Concurso de Detección de Fraudes organizado por I.C.E.A. en 1996, las 40 entidades aseguradoras participantes descubrieron un total de 12698 intentos de fraude. La cuantía de fraude detectado ascendió a 7650 millones mientras que las acciones de investigación y prevención supusieron únicamente un gasto del 2% de las reclamaciones (aproximadamente 170 millones de pesetas). En autos (11629 casos) la cuantía de fraude detectado supuso un ahorro de 6500 millones de pesetas (ahorro del 89.03% frente a las reclamaciones iniciales), siendo la inversión media en cada caso investigado de 10480 pts, aproximadamente.

Según el C.E.S. (1992) la lucha contra el fraude debe realizarse a partir de un triple frente de acción coordinado: la actuación realizada por cada entidad en particular (nivel individual) ha de ser trasladada al ámbito del sector asegurador (nivel sectorial), siendo necesaria, asimismo, una estricta colaboración del Estado (nivel oficial).

Respecto a la entidad aseguradora, las propuestas de actuación más importantes se centran, sobre todo, en una mayor implicación en el tema, en un mayor control de la contratación de pólizas y tramitación de siniestros y en una formación continuada de personal especializado en la materia, incluyendo, asimismo, la mejora en la selección de la red pericial y de agentes.

A nivel sectorial, es fundamental fomentar la cooperación entre entidades, siendo también importante una mayor coordinación con otros sectores, como puedan ser los consumidores o la Administración. Asimismo, la incentivación de medidas corporativas (Comisión Técnica de Seguros de Automóviles, Centro Zaragoza,...) y la creación de ficheros sectoriales (como puede ser el Fichero Informatizado de Automóviles -F.I.A.- y el Fichero Informático de Vehículos Asegurados -F.I.V.A.-), pueden constituirse en importantes focos de lucha contra el fraude a nivel español²². El uso de técnicas publicitarias y la denuncia de casos detectados aparecen como otras de las posibles medidas disuasorias que podrían ser utilizadas con mayor asiduidad.

Por último, a nivel oficial, la colaboración de las Fuerzas de Seguridad con las entidades aseguradoras puede agilizar la detección de fraudes, sobre todo en relación a la recuperación de vehículos robados. El desarrollo de mejoras en la

²² Desde U.N.E.S.P.A. se ha reiterado la necesidad de realizar una acción coordinada de control de fraude (I.N.E.S.E., 1996a). La creación de una agencia de prevención de fraude, al estilo de las que funcionan en otros países europeos, es uno de los principales objetivos propuestos por la patronal.

I.T.V.²³ como sistema de control de actuaciones de fraude y la conexión a un fichero centralizado (para las aseguradoras y la policía, por ejemplo) son medidas que pueden ayudar a realizar un control efectivo del problema.

Ahora bien, si la necesidad de actuar individual, sectorial e institucionalmente queda más que justificada, la gran variedad de formas de defraudar, la subjetividad que rodea al problema y la dificultad de realizar un control individualizado exhaustivo de todas las reclamaciones hace necesaria la elaboración de unas “pautas o patrones de comportamiento”, que ayuden al tramitador a realizar un control adecuado de los siniestros. La disponibilidad de modelos, validados estadísticamente, que permitan seleccionar las variables con poder explicativo de la presencia de fraude en el seguro automovilístico español y que hagan posible, asimismo, la cuantificación de la probabilidad de aparición de comportamientos fraudulentos en las carteras de asegurados es un tema pendiente al que damos tratamiento en el presente trabajo.

La aproximación metodológica que realizaremos en los próximos capítulos al estudio del fraude en el seguro del automóvil será, por lo tanto, totalmente novedosa. De hecho, lo es a nivel nacional, pero también a nivel internacional, ya que tendremos en cuenta, en la modelización, el proceso de decisión que lleva al asegurado a actuar fraudulentamente, utilizando, para ello, modelos econométricos adecuados. La disponibilidad de una muestra de expedientes de siniestros facilitada por la principal compañía del mercado asegurador español en autos nos permitirá obtener resultados importantes.

²³ Inspección Técnica de Vehículos. En referencia a esta medida cabe señalar la firma en 1994 de un acuerdo entre U.N.E.S.P.A. y A.E.C.A./I.T.V. destinado a hacer frente a dos importantes bolsas de fraude: la elevada presencia de vehículos que circulan sin la cobertura de responsabilidad civil obligatoria y el también elevado número de vehículos que eluden la I.T.V. A partir de dicho acuerdo, las entidades han de exigir la prueba de que se ha pasado la inspección antes de contratar o renovar una póliza, y las I.T.V. comprobar la vigencia del Seguro Obligatorio.

2.6 Categorías de fraude en el seguro del automóvil en España

Teniendo en cuenta la existencia de múltiples formas de defraudar, y atendiendo a la modelización que realizaremos del fraude, presentamos en este apartado un análisis detallado de las diferentes situaciones en las que se puede encontrar el individuo a la hora de decidir entre cometer o no fraude, y que hemos denominado de modo genérico “categorías de fraude”.

Las formas de defraudar en el marco del seguro del automóvil son variadas. Como hemos comentado en apartados anteriores, la acción fraudulenta puede ir asociada tanto al comportamiento del asegurado como a la del contrario, puede hacerse de forma individual o en connivencia con otros miembros del entorno asegurador (peritos, agentes, talleres,...), puede ejecutarse en beneficio propio o de otra persona, puede ser totalmente provocada o generada después de la ocurrencia real del siniestro, etc.

A pesar de que cada una de las formas de defraudar puede incidir de modo diverso en los costes de investigación, la importancia que para nosotros posee el establecimiento de una adecuada categorización del fraude está directamente relacionada con los tipos de modelos que utilizaremos para cuantificar su existencia. Desde este punto de vista, la aplicación de una modelización que tenga en cuenta el proceso de decisión seguido por el asegurado a la hora de actuar o no fraudulentamente, hace necesario analizar cuáles son las diferentes posibilidades que posee en relación a su comportamiento. Siguiendo a Cobo (1993), las formas más frecuentes de fraude en el seguro del automóvil son:

1. Falsa declaración del asegurado para eludir casos excluidos en la póliza.
2. Falsa declaración del asegurado en beneficio propio, sin intervención de un tercero.
3. Falso conductor habitual para eludir los recargos.
4. Ocultación de alcoholemia.

5. Falsa declaración del asegurado para favorecer a un tercero²⁴.
6. Versiones cruzadas para conseguir cobrar ambos implicados.
7. Contratación de la póliza después de ocurrido el accidente.
8. Fraude de taller.

Además, y aunque no han sido recogidas de forma explícita en la tipología presentada, otras formas de defraudar serían:

9. Incremento de los costes médicos derivados del siniestro, gracias al establecimiento de acuerdos entre el demandante y el personal médico.
10. Fraude cometido por peritos y agentes a favor del asegurado o incluso de un tercero.
11. Ausencia de declaración de la recuperación de vehículos previamente indemnizados por robo.

Análogamente, y siguiendo en este caso las modalidades presentadas por I.C.E.A. (1996) como las más frecuentes entre los defraudadores (utilizadas en la elaboración de la "Estadística de Fraudes en Seguros"), cabe destacar:

1. Ocultación de circunstancias personales del asegurado.
2. Falsa información para eludir casos excluidos en la póliza.
3. Versiones cruzadas para cobrar ambos implicados.
4. Contratación de la póliza después de ocurrido el accidente.
5. Falso conductor habitual para eludir los recargos.
6. Falsa declaración para favorecer a un tercero.
7. Falseamiento de origen y/o consecuencias económicas del daño.
8. Simulación /agravación de lesiones.
9. Enfermedad o lesión preexistente.
10. Otras.

²⁴ Es lo que Cobo denomina "aceptación por parte del asegurado condescendiente". El asegurado falsifica la declaración de siniestro con el objetivo de favorecer a un tercero, declarándose para ello culpable ante su compañía aseguradora.

Como se desprende de las clasificaciones anteriores, las posibilidades de actuación fraudulenta al alcance del individuo son variadas. Parece lógico pensar que el asegurado decidirá, en primer lugar, entre si comete fraude o si no lo hace. Si opta por la primera opción tendrá a su alcance diferentes posibilidades, siendo de vital importancia para la compañía conocer cuáles son las circunstancias frecuentemente relacionadas con la aparición de cada tipo de comportamiento.

Ahora bien, trabajar con una categorización tan extensa como la presentada puede ser muy costoso desde el punto de vista econométrico. Resulta adecuado limitar el número de alternativas a conjuntos más reducidos. El examen detallado de las posibilidades enumeradas anteriormente nos ha permitido generar una serie de árboles de decisión, que sintetizan los comportamientos más habituales entre los defraudadores. Los planteamientos pueden ser variados, tal y como veremos a continuación.

- **Fraude *versus* no fraude**

Lógicamente, este es el diseño más sencillo. Se cuantifica, de forma genérica, la probabilidad de que el siniestro sea fraudulento o de que no lo sea, sin tener en cuenta el tipo de fraude. En este caso, el modelo de decisión (Figura 1) se asocia a un árbol con un nivel, lo que simplifica de forma significativa la modelización econométrica ya que el uso de modelos de clasificación sencillos permite obtener resultados utilizando una variable dicotómica que recoja la presencia o la ausencia de fraude.

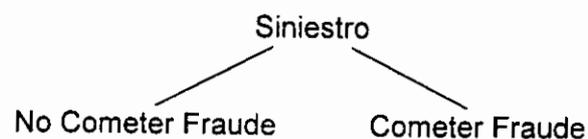


Figura 1

- **Fraude a favor del asegurado *versus* fraude a favor de un tercero**

A pesar de que el siguiente árbol de decisión sería, respecto a lo comentado anteriormente, aquél que recogiese en un segundo nivel todas las posibilidades de defraudar, parece razonable realizar una agrupación de las mismas siguiendo algún criterio preestablecido, con el objetivo de evitar trabajar con una categorización excesiva. De esta manera, en la Figura 2 recogemos el siguiente proceso jerárquico: el asegurado decide, en primer lugar, entre no cometer fraude o hacerlo. Lógicamente, si opta por la primera opción, el siniestro será no fraudulento. Pero si opta por la segunda, su comportamiento podrá venir provocado bien por un deseo de aprovechar la ocurrencia del siniestro en beneficio propio, bien para actuar en favor de un tercero. En este caso, el árbol planteado presenta dos niveles y por lo tanto permite aplicar una modelización econométrica más sofisticada que considere el proceso de forma unificada, como una secuencia jerárquica de decisiones.

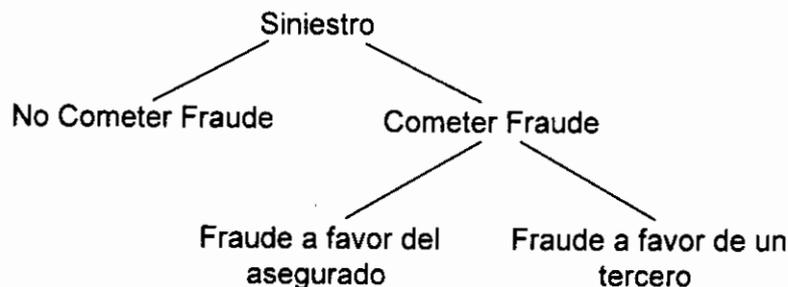


Figura 2

- **Fraude cometido de forma individual *versus* fraude cometido en connivencia**

El planteamiento presentado en este árbol es análogo al anterior. En este caso, si existe fraude, se diferencia entre la posibilidad de que sea cometido de forma individual por el asegurado, de que sea cometido en connivencia con el contrario o de que intervengan terceras personas (talleres, médicos, ...).

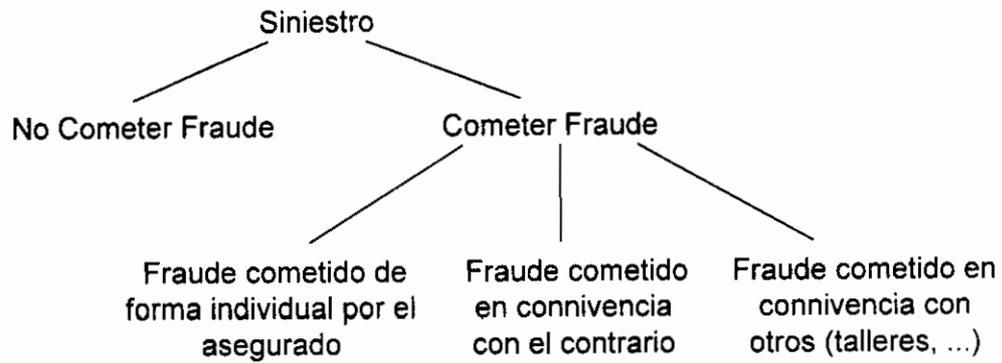


Figura 3

- **Fraude para conseguir una indemnización *versus* fraude para incrementar una indemnización**

Bajo este esquema, en el que de nuevo nos encontramos con un árbol a dos niveles, se diferencian aquellas situaciones en las que el fraude se comete para lograr una indemnización que no corresponde por contrato (al producirse el siniestro bajo condiciones no cubiertas en la póliza) y aquéllas en las que las consecuencias reales del mismo, sobre las que el asegurado posee cobertura, son falsificadas (incrementadas) para lograr una mayor compensación.

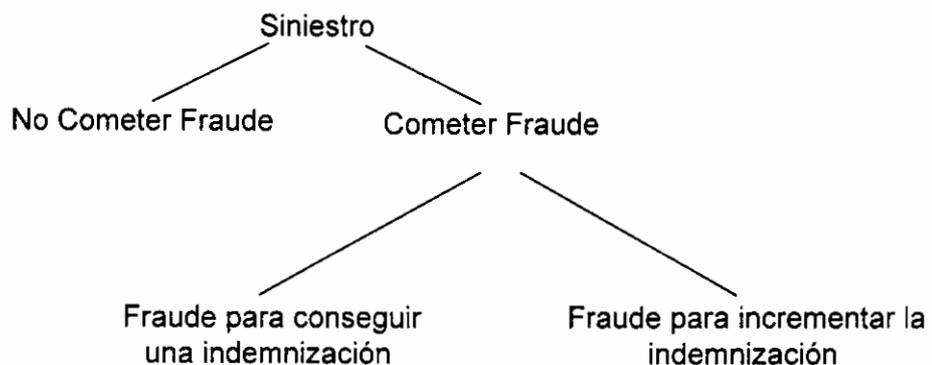


Figura 4

- **Fraude a priori *versus* fraude a posteriori**

Por último, siguiendo este proceso de decisión se intenta diferenciar entre aquellas situaciones en las que existe intencionalidad en la creación del siniestro

(correspondería, aproximadamente a la categoría de *fraude planeado*²⁵) y aquéllas en las que la ocurrencia del siniestro es real y, una vez producido éste, el individuo decide actuar o no fraudulentamente. De ahí que en la definición genérica de la clasificación hablemos de “*a priori*” o “*a posteriori*” para referirnos a la declaración de un siniestro como consecuencia de fraude o a la existencia de fraude tras la ocurrencia real de un accidente, respectivamente.

El árbol de decisión asociado a esta situación es el siguiente:

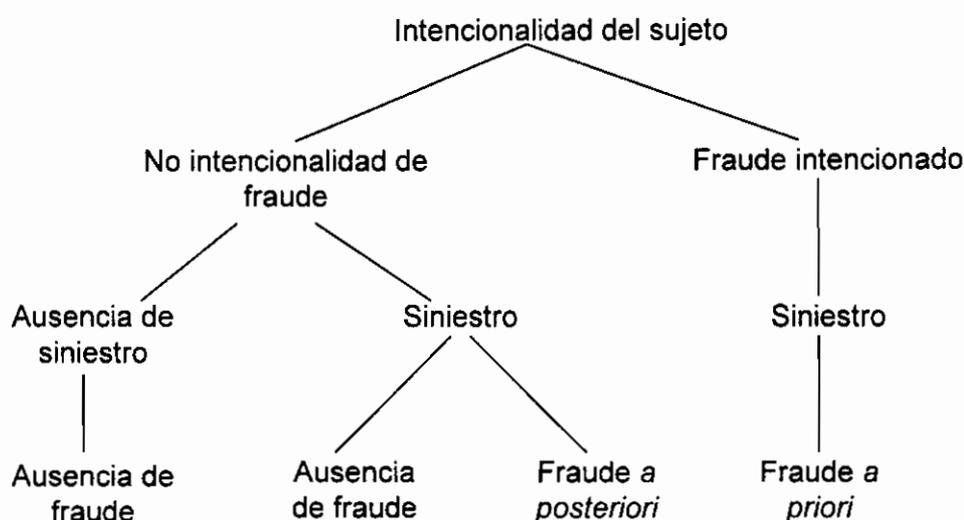


Figura 5

El planteamiento de los árboles de elección presentados y la aplicación de una modelización econométrica adecuada al tratamiento de los mismos ha de permitir obtener avances considerables en la detección y control del fraude. Por un lado, estaremos en condiciones de cuantificar la probabilidad de existencia de fraude, no sólo genéricamente hablando, sino también atendiendo a la tipología existente. Por otro, seremos capaces de analizar qué variables son estadísticamente significativas en la explicación de los diferentes comportamientos fraudulentos.

Sin duda, el alcance de ambos objetivos constituye una de las principales aportaciones del libro que presentamos. En todo momento, la obtención de

²⁵ Planned Fraud.

resultados responderá al proceso de decisión que conduce al individuo a elegir entre alternativas.

Si bien en el presente trabajo nos hemos centrado en modelizar explícitamente las situaciones representadas en las Figuras 2 y 4 (Capítulo 6), en Ayuso (1995) puede encontrarse una aplicación relacionada con la Figura 5, sometida, no obstante, a una simplificación. Asimismo, en Artís, Ayuso y Guillén (1998a) aparecen los resultados de modelizar el árbol de decisión representado en la Figura 1.

Ahora bien, ¿en qué fundamenta el asegurado la elección entre una y otra alternativa? La teoría económica señala que el individuo estructura el proceso de decisión atendiendo a la maximización de la utilidad que espera de cada una de las opciones. El contenido económico de cada uno de los árboles de decisión presentados ha de quedar claramente definido antes de dar paso a la explicación de los modelos utilizados tradicionalmente en la cuantificación del fraude. El uso de una aproximación econométrica que incorpora en su especificación procesos jerárquicos de elección diferenciará claramente el trabajo presentado del realizado por otros autores. Nuestra visión pasa de analizar el fraude desde un punto de vista agregado, a la implementación de modelos para individuos (siniestros). Todos estos aspectos serán ampliamente tratados en el Capítulo siguiente.

3. ENFOQUE ECONÓMICO DEL FRAUDE EN EL SEGURO DEL AUTOMÓVIL Y SU TRATAMIENTO ECONOMETRICO

3.1 Consideraciones Previas

El seguro del automóvil goza de una especial importancia dentro del campo de los Seguros Privados¹, y, como éstos, manifiesta en su comportamiento una elevada correlación con la evolución económica nacional.

El sector asegurador ha ganado un lugar de relevancia en el contexto de la economía española (Ayuso, 1995) y ello ha provocado una mayor preocupación por valorar, en términos estrictamente económicos, la eficiencia de sus principales actuaciones. En el caso del seguro del automóvil esta preocupación se ha visto acrecentada por los resultados técnico-financieros desfavorables que han caracterizado a este ramo asegurador durante los últimos años.

El estudio de la eficiencia productiva en el ámbito de seguros (Weiss, 1991; Picazo, 1995) refleja la necesidad de detallar cuáles son los factores que se consideran fundamentales en la estructuración de la función de costes de la actividad aseguradora. Dichos costes se clasifican fundamentalmente en tres categorías: los gastos de gestión interna (G.G.I.), los gastos de gestión externa (G.G.E.)² y los gastos técnicos.

Los gastos de gestión interna son aquéllos que tienen lugar en el marco de la función administrativa de la entidad, siendo los costes laborales su principal componente. Los gastos de gestión externa están relacionados directamente con la función comercial; en este caso los costes de distribución e intermediación son

¹ Un estudio exhaustivo de la evolución experimentada por el sector asegurador en España, globalmente considerado, y por el seguro del automóvil en particular, puede encontrarse en Ayuso (1995).

² De nuevo, un análisis detallado del comportamiento seguido por los gastos de gestión interna y por los de gestión externa, tanto para el seguro directo en general como para el seguro del automóvil español puede encontrarse en Ayuso (1995).

los más representativos. Por último, los gastos técnicos son los asociados a la reparación de daños y cuantificación de las indemnizaciones derivadas de los siniestros, siendo, por lo tanto, la componente de la función de costes que incorpora un mayor grado de incertidumbre. El estudio de la evolución experimentada por la siniestralidad (en relación al número y coste medio de siniestros) puede ayudar a establecer previsiones acerca de este último tipo de gastos. Sin embargo, la existencia de comportamientos anómalos por parte de los asegurados puede incidir decisivamente en el control de los gastos técnicos por parte de la compañía, influyendo, por lo tanto, en la gestión que ésta ha de hacer de los recursos.

La importancia de lo señalado en el párrafo anterior puede constatarse en la literatura existente sobre el tema. Autores como Cummins y Tennyson (1992), Weisberg y Derrig (1992b, 1993), Bond y Crocker (1997) o Cummins y Weiss (1991, 1993) han reiterado, desde enfoques alternativos, la necesidad de realizar un adecuado control de los costes aseguradores, con el objetivo de evitar un incremento desmesurado en las primas a pagar por los asegurados.

La presencia de acciones fraudulentas incidirá directamente (a través de la función de costes) en la eficiencia de la entidad y lo hará en un doble sentido: a través del pago de indemnizaciones no justificadas y a través del incremento en los gastos asociados a la investigación de siniestros. Cuando el grado de sospecha sobre un siniestro se acentúe, los gastos técnicos sufrirán un crecimiento. Éstos además se verán incrementados cuando, a pesar de la existencia de fraude, la investigación realizada por la compañía no logre detectarlo (a unos mayores costes de investigación se sumará el pago de una indemnización no justificada). Lógicamente, si la situación se repite en un número elevado de casos dentro de la cartera de asegurados, la estructura de costes se puede ver notablemente alterada (mayor peso específico de la componente aleatoria). En última instancia, pueden aparecer ineficiencias productivas (Picazo, 1995).

Teniendo en cuenta los argumentos presentados, ¿no resulta lógico pensar que una parte de la ineficiencia del seguro del automóvil pueda venir provocada por la existencia de factores de riesgo, no controlados de forma decisiva por la entidad, que incidan directamente sobre su función de costes?

El hecho de que la evolución de la siniestralidad en el seguro automovilístico español (Ayuso, 1995) manifieste un crecimiento durante los últimos años en el número medio de siniestros³ (así como un incremento moderado en su coste medio) alerta sobre la posible existencia de comportamientos ajenos al funcionamiento normal del sistema asegurador. La entidad deberá hacer especial hincapié en la política seguida para la gestión de riesgos y tramitación de siniestros. En este contexto, la aplicación de mecanismos de detección y control de fraude está totalmente justificada.

Si el estudio lo realizamos desde el punto de vista del asegurado, la evaluación del impacto económico final derivado de una política de control de fraude, fundamentada en un análisis econométrico, debe tener en cuenta la utilidad esperada por el individuo al actuar fraudulentamente. Dado que la modelización que presentaremos se basará en el comportamiento esperado del individuo, dedicamos la primera parte de este Capítulo (desde el apartado 3.2 hasta el 3.6) a presentar la fundamentación teórica que sustentará nuestro estudio. En el resto de apartados realizaremos una revisión de las principales aplicaciones econométricas que se han realizado, hasta el momento, en el estudio del fraude.

3.2 La Teoría Clásica del Consumidor: ausencia de incertidumbre

El ambiente de incertidumbre en el que se mueve el asegurado a la hora de decidir si actuar o no fraudulentamente supondrá la imposibilidad de aplicar una teoría, como la del consumidor, que parte de la premisa de que cada individuo conoce de antemano los resultados asociados a cada una de sus posibles elecciones. Sin embargo, los conceptos que definen dicha base teórica sí que son, en parte, susceptibles de utilización y nos ayudarán a fundamentar nuestro análisis.

Al igual que en la teoría de la empresa competitiva, donde las funciones de oferta y demanda se obtienen a partir de maximizar el beneficio (previa especificación de un modelo y de las restricciones necesarias), en la teoría del consumidor, será la

³ Las mejoras producidas en la red viaria, las medidas seguidas por el Gobierno, como, por ejemplo, los Planes Renove (en la actualidad "Plan Prever"), las campañas de Seguridad Vial desarrolladas por la Dirección General de Tráfico y las mejoras en la ITV deberían implicar, *a priori*, una reducción en el número medio de siniestros. El hecho de que esto no suceda puede sugerir un comportamiento extraño de los agentes económicos que intervienen (lógicamente, teniendo en cuenta también otras circunstancias: la cada vez mayor densidad circulatoria, la mayor potencia de los vehículos,...).

conducta maximizadora de la utilidad la que prevalecerá a la hora de definir el comportamiento del individuo (Varian, 1992).

Lógicamente, y bajo un enfoque puramente económico, la maximización de la utilidad por parte del consumidor, (en este caso, asegurado), implica la definición previa de lo que en teoría económica se denominan "preferencias del individuo"⁴. Bajo este concepto, los economistas entendemos el conjunto de bienes, (en nuestro caso, situaciones), por los que el consumidor-asegurado muestra un mayor deseo. El *conjunto de consumo* (Varian, 1992) puede ser más o menos amplio, pero cabe suponer que el individuo siempre mostrará un determinado orden de preferencias a la hora de realizar su elección.

La aplicación directa de la teoría del comportamiento, tal y como la conocemos, supondría analizar qué preferencias posee el asegurado a la hora de determinar los tipos de seguros que contrata. Los riesgos para los que desea obtener cobertura son diversos y además se encuentra sujeto a una determinada restricción presupuestaria. Bajo este enfoque, el individuo conoce de antemano las consecuencias derivadas de la compra del producto (elección del bien): la cobertura de uno o varios riesgos que no tienen por qué producirse⁵.

Sin embargo, el planteamiento que ahora presentamos es otro. La situación de partida nos lleva a centrarnos en un individuo que ya está asegurado en una entidad⁶ y cuya decisión versa sobre la elección de una determinada alternativa: comportarse honesta o fraudulentamente. Se enfrenta, por tanto, a una elección entre alternativas discretas. Si elige la primera opción, el contrato de seguro habrá alcanzado su fin: cubrir al asegurado del posible acaecimiento de un riesgo. Tras la ocurrencia del siniestro, la entidad, en función de las coberturas contratadas (asociadas al pago de un determinado montante de prima por parte del asegurado)

⁴ Un amplio desarrollo sobre la clasificación y las propiedades que han de cumplir las "preferencias" manifestadas por el consumidor a la hora de seleccionar conjuntos de bienes, con el objetivo de ser representadas mediante una función de utilidad continua, puede encontrarse en Varian (1992, pp. 113-116) o en González (1993, pp. 74-76), ésta última en relación a la demanda de seguros sanitarios.

⁵ En este caso, la contratación de un determinado tipo de seguro atiende al deseo de eliminar la incertidumbre que rodea la posible materialización de un riesgo. No obstante, es necesario señalar que si el estudio tratase de evaluar la decisión de un individuo a la hora de contratar o no un seguro, teniendo en cuenta la incertidumbre que rodea la ocurrencia o no de un suceso, todo cambiaría, dado que estaríamos frente a una situación en la que se desconocería el resultado final de la elección.

⁶ No obstante, una posible excepción a este planteamiento puede encontrarse en aquellos casos en los que la contratación de la póliza de seguros es posterior a la ocurrencia del siniestro.

responderá de los daños producidos y todo el proceso formará parte de la dinámica propia del sistema. Ahora bien, si el asegurado decide actuar fraudulentamente, el sistema se verá alterado ante un comportamiento no esperado. Lógicamente, las preferencias del individuo entran en juego: elegirá cometer o no fraude en función de la utilidad que espere de cada una de las dos posibles opciones. No obstante, la situación no es la misma que la que caracteriza a la demanda de un bien bajo el enfoque clásico, tal y como veremos a continuación.

Fernández de Castro y Tugores (1988, 1992), en la presentación de la teoría del consumo, establecen básicamente tres hipótesis previas a la modelización. La consideración del individuo⁷ como unidad básica de análisis, la suposición de que el consumidor es racional y el establecimiento de una serie de restricciones fundamentalmente económicas (limitación presupuestaria e integración en un mercado competitivo⁸) son presentados por ambos autores como supuestos de partida. Para ellos, las consecuencias derivadas de cada elección son conocidas *a priori*.

En nuestro tratamiento la unidad de decisión es el asegurado⁹. Éste actúa intentando alcanzar el orden más elevado de sus preferencias y, desde ese punto de vista, puede ser considerado racional¹⁰. Sin embargo, el establecimiento de restricciones previas a la modelización es más complicado. ¿Con qué limitaciones se encuentra el asegurado que quiere actuar fraudulentamente?

Antes de contestar a esta pregunta, recordemos que la perpetración de un fraude puede tomar formas variadas. Como ya sabemos, el asegurado puede producir deliberadamente un siniestro o puede aprovechar, con fines lucrativos, la ocurrencia real del mismo; puede intentar ocultar casos excluidos en la póliza o querer conseguir una indemnización superior a la que le corresponde por contrato; puede cometer fraude en beneficio propio o en beneficio de una tercera persona; puede cometer fraude de forma individual o en connivencia con otras personas, etc.

⁷ En Fernández de Castro y Tugores (1988) se presenta la familia como unidad de decisión básica, con el matiz de que ésta puede estar formada por un único miembro.

⁸ Los precios de los bienes que el consumidor adquiere son fijos, sin que tenga ninguna influencia sobre los mismos.

⁹ No obstante, cuando hablamos de "cadenas de fraude" hacemos referencia al perpetrado por grupos profesionales en los que la unidad de decisión está formada por más de un individuo.

¹⁰ Lógicamente, teniendo en cuenta el contexto en el que nos movemos, pues, estrictamente hablando, la racionalidad llevaría asociada el desarrollo de actuaciones honestas por parte del asegurado.

¿Qué circunstancias rodean cada uno de estos comportamientos? La subjetividad existente en cada situación nos impide establecer restricciones con la misma facilidad con la que fijáramos, por ejemplo, limitaciones presupuestarias en la demanda de un bien.

Obviamente, el individuo que comete una acción fraudulenta es consciente de que una serie de factores pueden actuar en su contra. La existencia inesperada de testigos o la no deseada intervención policial pueden ser algunos de los elementos disuasorios del fraude. Pero, ¿hasta qué punto pueden ser considerados o introducidos como restricción en la modelización de la utilidad esperada por el defraudador? Cada situación en particular, cada forma de defraudar, comportará unas determinadas limitaciones que definirán un ambiente de incertidumbre para el asegurado que ha de decidir entre actuar o no fraudulentamente.

La incertidumbre que rodea el proceso de decisión queda recogida en un hecho fundamental: el asegurado desconoce cuáles van a ser las consecuencias derivadas de su actuación, dado que ignora la probabilidad de que su comportamiento sea detectado por la entidad.

Ante una situación como la planteada, en la que las consecuencias finales de la elección no son conocidas por el individuo, la teoría clásica del consumo pierde poder, siendo la teoría de la utilidad esperada la que aporte el fundamento metodológico necesario para realizar el estudio.

3.3 Fundamentación teórica de la decisión de defraudar: Teoría de la Utilidad Esperada

La decisión de defraudar es, sin duda, una determinación tomada en condiciones de incertidumbre. Como es usual, ante una disyuntiva fraude/no fraude, el consumidor-asegurado también buscará maximizar los resultados de su elección (al igual que haría bajo condiciones de elección cierta o segura). Sin embargo, en este caso, deberá considerar en la representación de sus preferencias la existencia de posibles combinaciones de resultados y las probabilidades de ocurrencia de los mismos.

En un marco de incertidumbre, el conjunto de consumo o espacio de elección estará formado, no por una "cesta de bienes", sino por una serie de distribuciones

de probabilidad (conocidas comúnmente como "loterías") asociadas a cada una de las posibles opciones. Así, se ha de considerar la existencia de factores inciertos¹¹ que, de una u otra forma, inciden en cada uno de los resultados. Dichos factores caen fuera del control del asegurado, pero, de algún modo, son determinantes de la elección final realizada.

Teniendo en cuenta el cumplimiento de los axiomas necesarios para que el orden de preferencias pueda ser representado mediante una función de utilidad continua (Kreps, 1990), la existencia de información imperfecta será tratada, en el marco de la teoría de la utilidad esperada, introduciendo, de alguna forma, la valoración que los individuos hacen de sus decisiones. De este modo, la utilidad esperada por el asegurado será la suma ponderada de la utilidad esperada de cada una de las consecuencias de su comportamiento. Las ponderaciones vendrán dadas por las probabilidades de ocurrencia de cada una de ellas¹².

La *función de utilidad esperada* de una acción (a_i) se define según la expresión,

$$E[U(a_i)] = p_1u(c_{i1}) + p_2u(c_{i2}) + \dots + p_mu(c_{im}) \quad (3.1)$$

donde, $u(c_{ij})$ representa la utilidad asociada a cada una de las consecuencias (c_{ij}) de la acción seguida por el individuo i ; p_j es la probabilidad de cada uno de los estados de incertidumbre s_j , con $j=1,2,\dots,m$, asociados a la elección y, por último, el producto $p_ju(c_{ij})$ es la esperanza matemática de cada uno de los posibles resultados. La aditividad presentada en (3.1) es consecuencia de la independencia entre las utilidades asociadas a cada una de las consecuencias. Asimismo, se supone que la utilidad es lineal en las probabilidades (Fernández de Castro y Tugores, 1992).

Cuando la situación general planteada en la expresión anterior se particulariza al caso de dos posibles consecuencias (derivadas de actuar o no fraudulentamente), mutuamente excluyentes, la utilidad esperada es:

¹¹ *Estados de la naturaleza.*

¹² El desarrollo de la teoría de la utilidad esperada ha dado lugar a enfoques variados. El teorema de Von Neumann y Morgenstern (1944), en relación a la consideración de probabilidades objetivas (fijadas *a priori*) o el desarrollo presentado por Savage (1954), sobre la introducción de probabilidades subjetivas (en base a la valoración propia del individuo), son quizá los principales fundamentos teóricos a la hora de estudiar los conceptos básicos y de interpretación de la utilidad esperada.

$$E[U(a_i)] = pu(c_{i1}) + (1-p)u(c_{i2}) \quad (3.2)$$

Esta expresión se obtiene al considerar la *propiedad de la utilidad esperada* (Varian, 1992), según la cual “la utilidad de una lotería es la utilidad que se espera que reporten sus premios”, de forma que:

$$U(pc_{i1} + (1-p)c_{i2}) = pu(c_{i1}) + (1-p)u(c_{i2}) \quad (3.3)$$

La teoría de la utilidad esperada ha sido aplicada en campos variados dentro del ámbito asegurador. Enfoques como los asociados al análisis de la aversión al riesgo (Pratt, 1964; Borch, 1990) o a la demanda de seguros (Borch, 1990; Mossin, 1992) sirven como ejemplo de ello. Sin embargo, ¿cuál ha sido la aplicabilidad de la teoría de la utilidad esperada en el marco del análisis del fraude en seguros?

Hoyt (1990), a fin de mejorar el conocimiento y las dimensiones del comportamiento fraudulento, y bajo lo que él mismo denomina “*una visión económica del control del fraude en seguros*”, presenta un modelo económico para controlar las actuaciones fraudulentas de los asegurados. Para ello toma como premisa la utilidad esperada por los mismos a la hora de decidir si actuar o no fraudulentamente. El enfoque que presenta es genérico¹³ y no diferencia, en ningún momento, las distintas consecuencias de cada tipo de fraude.

La utilidad que el asegurado defraudador espera obtener de su acción resulta de la adición de dos componentes, llamémoslas *subutilidades*, que presentan signos contrarios: la primera, positiva, se obtendrá directamente de la perpetración de la acción fraudulenta; la segunda, negativa, irá asociada a la posible detección de la misma. Tomando como referencia el comportamiento del consumidor en un ambiente de incertidumbre, y utilizando una medida de la “satisfacción global

¹³ Hoyt (1990) analiza el efecto del fraude en seguros dentro del sistema económico y supone que el tratamiento del fraude en el contexto asegurador ha de tomar en consideración dos perspectivas, la de los asegurados y la de la industria de seguros.

esperada por el individuo¹⁴, este autor plantea una función de utilidad cuyo valor esperado puede escribirse como:

$$E [U(F)] = qU(p) + rU(f) \quad (3.4)$$

donde $E [U(F)]$ es la utilidad esperada del fraude, p la penalización, $U(p)$ su utilidad (negativa), f la cantidad defraudada, $U(f)$ su utilidad, q la probabilidad de que el fraude sea detectado y r la probabilidad de defraudar (de obtener una indemnización asociada a fraude). Nótese que Hoyt (1990) sólo trata la utilidad de la actuación fraudulenta.

A la vista de lo anterior, y con el objetivo de reducir la utilidad esperada del fraude, el autor concluye que cualquier medida de control debería encaminarse a aumentar el primer término de la anterior función (la probabilidad de detección y su penalización) y a reducir el segundo (la probabilidad de defraudar y su utilidad). Para ello, Hoyt señala como necesaria la ejecución de diferentes acciones por parte de las entidades aseguradoras, entre las que destaca la instalación de procedimientos para identificar siniestros fraudulentos, la investigación cuidadosa de los mismos o el establecimiento de "principios generales" para la contratación de las pólizas. La idea que subyace es que, tanto una reducción en la probabilidad de pago de siniestros fraudulentos, como en la cantidad pagada por los mismos, derivará en una disminución de la utilidad esperada por el individuo. Por otro lado, la mayor severidad de las penas incidirá también en una reducción de la utilidad, aunque en este aspecto el autor se muestra más escéptico, dado que la influencia del entorno legal y social, a veces, no es demasiado favorable.

Estudios posteriores, como el realizado por Picard (1996), continúan manteniendo el enfoque económico en el análisis del fraude en seguros. Este autor estudia como incide en el equilibrio de mercado la presencia de tomadores de seguros oportunistas¹⁵ que pueden declarar siniestros fraudulentos.

¹⁴ En realidad, deberíamos hablar de "beneficio final", dado que las consecuencias derivadas se medirán en términos monetarios. Esta aproximación supone analizar el beneficio monetario de cometer fraude sin que sea detectado y la penalización, también monetaria, asociada a la detección de dicho comportamiento. Hoyt realiza una modelización partiendo de la premisa de que existe fraude y por ello, y a diferencia de Picard (1996), no recoge explícitamente la probabilidad de que el asegurado actúe fraudulentamente como situación complementaria a la actuación honesta.

¹⁵ Denominados por el autor "*opportunistic policyholders*", son aquellos asegurados dispuestos a utilizar la póliza de seguros o la ocurrencia del siniestro de forma deshonestamente.

Para Picard la relación entre asegurador y asegurado se modeliza como un juego de información imperfecta, siendo el primero el que ha de decidir entre investigar las causas de un siniestro o no hacerlo. El equilibrio de mercado dependerá fundamentalmente de la capacidad de la entidad para generar sistemas de control (posiblemente descentralizados) que eviten trasladar las consecuencias no deseadas de los comportamientos fraudulentos hacia los asegurados honestos. Basándose al igual que Hoyt (1990) en el concepto de utilidad esperada, considera también que las consecuencias de una determinada acción se miden en términos monetarios. Desarrolla el modelo teórico considerando la "probabilidad de control del fraude" que hace que el asegurado oportunista sea indiferente ante la posibilidad de actuar o no fraudulentamente. Esta probabilidad es el valor \tilde{q} que hace que se verifique la siguiente igualdad¹⁶:

$$U(W-P) = \tilde{q}U(W-P-M(s,P)) + (1-\tilde{q})U(W-P+s) \quad (3.5)$$

siendo W la riqueza inicial del asegurado, P la prima pagada por el seguro, s la cobertura ofrecida por el mismo y $M(s,P)$ la cantidad que el defraudador tendrá que pagar como sanción en caso de detección de su comportamiento y que depende del contrato (s,P) . Nótese que el término de la izquierda corresponde a la utilidad que obtiene el asegurado si no defrauda, en tanto que el de la derecha es la utilidad que le reportará el fraude. Estos dos términos de la expresión son iguales en virtud del valor de \tilde{q} .

A partir de aquí, el asegurado elegirá el valor de t que maximice su utilidad esperada, que vendrá dada por:

$$E[U] = t[qU(W-P-M(s,P)) + (1-q)U(W-P+s)] + (1-t)U(W-P) \quad (3.6)$$

siendo t la probabilidad de que el asegurado cometa fraude y q la probabilidad de que la compañía audite el siniestro¹⁷, de forma que,

¹⁶ La terminología que presentamos no es exactamente la misma que presenta Picard (1996) dado que hemos intentado establecer, en la medida de lo posible, una correspondencia con la utilizada por Hoyt (1990). Picard asume que todo siniestro fraudulento auditado será detectado por el asegurador.

¹⁷ $(1-q) = r$, $M(s,P) = p$, y $s = f$ en relación a la terminología empleada por Hoyt (1990).

$t = 0$, si $q > \tilde{q}$: el asegurado decide no defraudar ante la sospecha fundada de que su comportamiento sería detectado;

$t \in (0,1)$, si $q = \tilde{q}$: el asegurado está indeciso, pues no está claro que el fraude le reporte beneficios, y

$t = 1$, si $q < \tilde{q}$: el asegurado decide defraudar ante el convencimiento de que su actuación no será descubierta por la entidad.

En base a la expresión (3.6), una política adecuada de control de siniestros, que aumente la probabilidad de detectar el fraude y dé publicidad a este hecho, podrá incidir en una reducción de la utilidad esperada por el oportunista, que verá incrementada la posibilidad de tener que pagar una determinada sanción (además de la prima pagada previamente en la contratación del seguro). Ante ello, la probabilidad de que cometa fraude disminuirá, aumentando el último sumando de la igualdad, que refleja el comportamiento del asegurado como honesto.

Sin embargo, la conclusión obtenida está sujeta a una serie de matizaciones. La realización de una auditoría de siniestros dirigida a detectar el fraude por parte de la entidad aseguradora puede comportar la presencia de costes muy elevados, tanto en términos de investigación como de tiempo. Por ello, Picard (1996) introduce la consideración de que las entidades determinarán la probabilidad de controlar o auditar el fraude (q) teniendo en cuenta la maximización del beneficio esperado o la minimización de los costes¹⁸.

3.4 Una nueva aproximación a la aplicación de la Teoría de la Utilidad Esperada en el tratamiento del fraude

La modelización teórica presentada por Hoyt (1990) y por Picard (1996), analizada en el epígrafe anterior, responde a un estudio del fraude considerando el mercado asegurador en general, es decir, sin distinguir coberturas o ramos. Además, como ya comentábamos, no tiene en cuenta la tipología existente de fraude, que puede ser variada en cualquiera de los ramos aseguradores.

¹⁸ El coste esperado para la entidad vendrá definido por la suma de la indemnización y el coste de la investigación del siniestro. A esta cantidad se le habrá de restar aquella pagada por el asegurado en concepto de penalización.

Los diferentes tipos de comportamientos deshonestos existentes en el marco del seguro del automóvil (ver epígrafe 2.6), su consideración en la determinación del beneficio esperado por el asegurado a la hora de actuar o no fraudulentamente, la existencia de costes de investigación diferentes asociados a los mismos y la posibilidad de que la política de control o de detección seguida por la compañía no sea la correcta (clasificación de siniestros fraudulentos cuando no lo son y viceversa) son algunos de los puntos principales que introduciremos en nuestro planteamiento.

A diferencia de Hoyt y Picard, la modelización que presentaremos en el Capítulo 4 implica determinar la utilidad esperada por el individuo considerando características del mismo que pueden incidir en su comportamiento. Así, será fundamental tener en cuenta el hecho de que determinados atributos del asegurado (edad, estado civil, situación laboral, familiaridad con el contrario,...) pueden ser considerados en la cuantificación de la probabilidad de existencia de fraude.

La decisión entre actuar o no fraudulentamente supondrá para el asegurado la necesidad de realizar una determinada elección. Si decide actuar honestamente, su comportamiento estará unívocamente definido. Sin embargo, si decide defraudar, su actuación podrá adoptar formas diversas, tal y como hemos puesto de manifiesto en el epígrafe 2.6, mediante la representación gráfica de diferentes árboles de decisión. Al modelizar la utilidad esperada (que habíamos definido como la suma de la utilidad esperada de defraudar y de la utilidad esperada de no defraudar¹⁹) se tendrán que considerar una serie de parámetros fundamentales, definidos en relación a la utilidad que el asegurado espera obtener de los diferentes tipos de comportamiento fraudulento y, a la probabilidad de que cometa cada uno de ellos.

Retomemos a Picard (1996) y a Hoyt (1990) y veamos cuáles serán los principales objetivos a tener presentes en la modelización.

Para Picard (1996), la utilidad esperada por el asegurado se determina de la forma:

$$E[U] = tE[U(F)] + (1-t)E[U(NF)] \quad (3.7)$$

¹⁹ Multiplicadas, tal y como ya sabemos, por la probabilidad de defraudar y de no defraudar, respectivamente.

donde $E[U(F)]$ y $E[U(NF)]$ recogen la utilidad esperada por el individuo al actuar fraudulentamente o no hacerlo, respectivamente, y t refleja la probabilidad de defraudar²⁰. Ambas utilidades son para el autor función de una parte determinista y una aleatoria. La riqueza de cada individuo, la prima pagada por el seguro, la cobertura contratada y la penalización fijada en base a la póliza, incidirán en la utilidad esperada de la actuación fraudulenta y de la honesta, para cada asegurado de la muestra. Asimismo, la probabilidad de que la entidad audite o no los siniestros influirá, decisivamente, en la utilidad que el individuo espera de su actitud deshonesto, partiendo de la premisa de que siempre que se audita el siniestro se produce un dictamen certero sobre la presencia o no de fraude.

Hoyt (1990) se centra únicamente en la utilidad esperada del fraude, pero la similitud con la formulación presentada por Picard²¹ permite llegar a conclusiones similares.

Sin embargo, ¿cuáles son las variables o circunstancias que influyen en la probabilidad de que el asegurado cometa fraude?, ¿cuál es la probabilidad de que la entidad detecte el comportamiento fraudulento? Como comentábamos en páginas anteriores, es posible que determinadas características del individuo, del siniestro, de la póliza, ..., incidan en la elección finalmente realizada. Asimismo, es igualmente probable que algunas circunstancias influyentes en la decisión final sean omitidas involuntariamente en la modelización. Parece claro, por lo tanto, que la determinación de la utilidad esperada por el individuo está sometida a un cierto margen de error. La aplicación de modelización logística nos permitirá extraer conclusiones para las preguntas formuladas sin alejarnos de la teoría de la utilidad aleatoria.

En la fundamentación teórica de los modelos que presentaremos persistirá al criterio de que el individuo actúa, en la toma de decisiones, bajo la óptica de maximización de la utilidad. Los objetivos perseguidos tras la aplicación de los mismos serán básicamente dos:

²⁰ Esta expresión es equivalente a la número (3.6).

²¹ Para Hoyt, la utilidad esperada del fraude viene dada, como ya sabemos, por la expresión:

$$E[U(F)] = qU(p) + (1-q)U(f)$$

donde la similitud con Picard (1996) se pone de manifiesto en el hecho de que q continua siendo la probabilidad de la compañía detecte el fraude, p es la penalización ($=W-P-M(s,P)$) y f es la cantidad defraudada ($=W-P+s$). No obstante, Hoyt no introduce en su formulación la riqueza inicial del individuo (W) y la prima pagada (P). Por tanto, en realidad, $p = M(s,P)$ y $f = s$.

1. Cuantificar la probabilidad de que el asegurado cometa un determinado tipo de fraude²² y determinar las variables explicativas del mismo.
2. Cuantificar la probabilidad de detectar comportamientos fraudulentos²³, en base a la capacidad predictiva del modelo planteado.

La estimación de las probabilidades de defraudar y de que el fraude sea detectado permitirá cuantificar la utilidad que el individuo espera de su comportamiento. Además, la determinación de las variables explicativas de cada tipo de fraude permitirá enfocar la investigación del siniestro hacia aquel comportamiento para el que el modelo genera una mayor sospecha.

La obtención de las probabilidades estimadas propiciará una aproximación de especial interés para la entidad aseguradora al permitir cuantificar el coste esperado de los siniestros tras la aplicación de un modelo de control de fraude. La combinación adecuada de las probabilidades de cometer o no fraude y de realizar o no una adecuada clasificación de siniestros permitirá, como veremos más adelante, obtener una estimación del coste que para la entidad supondrá la tramitación²⁴ y, en su caso, indemnización del suceso²⁵. No obstante, los resultados que presentaremos, desde un punto de vista aplicado, para este punto, serán simples avances de lo que debería ser un tratamiento más exhaustivo. La no disponibilidad de información completa sobre los costes asociados al siniestro, en la muestra utilizada en la parte empírica del libro (a partir del Capítulo 5), ha impedido realizar un estudio más extenso.

²² Estimación de t dentro de la formulación de Picard (1996). No obstante, la consideración de diferentes posibilidades a la hora de cometer fraude supondrá reformular la expresión presentada por Picard de la forma:

$$E[U] = \sum_{j=1}^m t_j \left[q_j U(W-P-M(s,P)) + (1-q_j) U(W-P+s) \right] + \left(1 - \sum_{j=1}^m t_j \right) U(W-P)$$

donde el subíndice $j=1,2,\dots,m$ recoge los diferentes tipos de fraude (se adecuan a los presentados en el epígrafe 2.6 del presente trabajo).

²³ Estimación de q dentro de la formulación de Picard (1996) o de Hoyt (1990), teniendo en cuenta los comentarios realizados en el pie de página anterior (estimación de q_j).

²⁴ Entendida únicamente bajo la perspectiva de investigación de fraude en el siniestro (es decir, en la determinación del coste esperado sólo tendremos en cuenta aquellos gastos directamente relacionados con el tratamiento de fraude por la entidad).

²⁵ Es posible que al detectarse fraude la entidad no tenga que desembolsar ninguna cantidad en concepto de indemnización. Sin embargo, si aun existiendo fraude, la compañía no lo detecta, la indemnización desembolsada será la que corresponde en función de la póliza contratada.

Ahora bien, antes de dar paso a la explicación de los modelos aplicados en el tratamiento del fraude dentro del presente trabajo, creemos adecuado realizar una revisión de cuáles han sido las principales aplicaciones que, desde un punto de vista econométrico, se ha realizado en relación al tema objeto de estudio. La comparación con las mismas permitirá obtener una visión del libro como un trabajo totalmente diferenciado de los existentes.

3.5 Métodos cuantitativos de detección y control del fraude en el seguro del automóvil

Las aproximaciones realizadas al tratamiento del fraude en el marco del seguro del automóvil han tenido, fundamentalmente, carácter sectorial e institucional, sin que la presencia de una actividad científico-investigadora haya sido intensa.

La caracterización de un suceso como fraudulento debe ir acompañada de la validación estadística de los indicadores de fraude empleados, siendo ésta la principal idea que subyace en los trabajos desarrollados. A pesar de que la literatura existente sobre el tema es escasa, los estudios realizados hasta el momento evidencian la susceptibilidad de aplicación de técnicas econométricas variadas a la modelización de los comportamientos deshonestos. El uso del modelo de regresión múltiple (Weisberg y Derrig, 1993; Derrig y Weisberg, 1998; Cummins y Tennyson, 1996), del análisis cluster y conjuntos borrosos (Derrig y Ostaszewski, 1994a, 1994b), de redes neuronales (Brockett, Xia y Derrig, 1995) y de modelos de elección discreta simples (Ayuso, 1995; Artís, Ayuso y Guillén, 1998a; Belhadji y Dionne, 1997) sirven como ejemplo de lo manifestado. No obstante, los resultados obtenidos se han catalogado en muchos casos de provisionales, debido, sobre todo, a la pequeña dimensión de las bases de datos utilizadas.

- **Aproximación al tratamiento del fraude mediante el modelo de regresión múltiple**

Weisberg y Derrig (1993) presentan como objetivo principal en su estudio el análisis del potencial de la aplicación de métodos cuantitativos para reducir los pagos indebidos de siniestros. En este sentido, la creación de un sistema para identificar siniestros fraudulentos, generado a partir de modelos de indicadores de fraude, ha de permitir obtener resultados dirigidos a mejorar la tramitación de los casos. Los autores seleccionan el modelo de regresión múltiple como técnica

econométrica a utilizar para validar el significado estadístico de los indicadores, manifestando ya *a priori* que el uso de técnicas más sofisticadas podría derivar en la obtención de resultados más satisfactorios.

En estudios anteriores (Weisberg y Derrig, 1991 y 1992b) los resultados obtenidos mostraban un incremento en el número de siniestros con sospecha de fraude. Los trabajos realizados en el seno del *Automobile Insurers Bureau* de Massachussets (*A.I.B.*) habían permitido obtener conclusiones en relación al nivel existente de fraude en el marco de la cobertura por daños corporales. La investigación se había fundamentado en la explotación de dos bases de datos, la primera generada en base a siniestros ocurridos entre 1985 y 1986²⁶ y la segunda en base a siniestros de 1989²⁷.

En 1993, dentro del artículo "Quantitative Methods for Detecting Fraudulent Automobile Bodily Injury Claims", Weisberg y Derrig tratan de validar estadísticamente los resultados extraídos de la explotación de la segunda muestra mencionada. Respecto a la misma, establecen una determinada tipología de fraude que distingue entre las cuatro situaciones posibles a la hora de clasificar un siniestro a las que nos hemos referido en los Capítulos anteriores. De esta forma:

- Si la ocurrencia del siniestro y las consecuencias derivadas del mismo se consideran legítimas, el siniestro es, lógicamente, considerado legal (*"legitimate"*).
- Si el siniestro ha sido totalmente construido o simulado, de forma que el nivel de sospecha es elevado tanto en los datos referidos al accidente, al asegurado y al demandante, como en los referidos al daño y al tratamiento médico derivado, entonces es calificado como fraude planeado (*"planned fraud"*).
- Si la sospecha se centra en torno al demandante, al daño producido o al tratamiento médico, siendo la ocurrencia del accidente real y no

²⁶ Conocida como "*The Baseline Study*", esta base de datos recogía un total de 597 siniestros ocurridos entre el 1 de julio de 1985 y el 30 de junio de 1986. Fue diseñada bajo el propósito de analizar los costes médicos derivados de siniestros con daños corporales, con el objetivo de encontrar casos fraudulentos dirigidos, fundamentalmente, al incremento de dichos costes.

²⁷ En este caso, la muestra estaba constituida por un total de 387 expedientes.

existiendo prácticamente sospecha sobre el conductor asegurado, el siniestro es calificado de oportunista²⁸ (“*opportunistic fraud*”)

- Si el nivel de sospecha es elevado en relación a los costes y tratamiento médico, el siniestro es clasificado como “inflado” (“*build up*”).

La consideración inicial del tipo de fraude existente, en función de la clasificación anterior, y la determinación de los indicadores de fraude que realmente ayuden a validarlo son aspectos esenciales a tener en cuenta en la modelización.

Seleccionada una submuestra no aleatoria de 127 siniestros²⁹, Weisberg y Derrig utilizan un modelo de regresión con el objetivo de modelizar el nivel de sospecha existente (en una escala de 0 a 10), utilizando una doble codificación (una realizada por un grupo de tramitadores de siniestros y la otra por un grupo de investigadores del I.F.B.) que establece una valoración del tipo de fraude³⁰. Adicionalmente, y utilizando también regresión múltiple, modelizan el número de votos a favor de la sospecha de fraude en un siniestro, seleccionando un conjunto de indicadores que pueden incidir en la fundamentación de la misma.

Las variables explicativas introducidas en ambos modelos vienen definidas por los diferentes aspectos y circunstancias de un siniestro. En concreto recogen información relativa al accidente, al demandante, al conductor asegurado, al daño, al tratamiento médico y a la pérdida salarial experimentada³¹. En relación al accidente, las variables consideradas hacen referencia, entre otras cosas, a la presencia o no de informe policial y/o de testigos, a la existencia de un relato del siniestro no plausible o al hecho de que el demandante posea un vehículo antiguo y de escaso valor. Las características del demandante revelan si reside en una

²⁸ Uno de los ejemplos más típicos que identifica a este tipo de fraude es el de un pasajero de autobús que aprovecha la ocurrencia de un siniestro para actuar en beneficio propio, en busca de una indemnización que no le corresponde.

²⁹ De los 387 casos que componían la muestra, sólo en 62 se detectó existencia de fraude. La escasez de datos hizo necesario un enriquecimiento de la muestra, agregando a los 62 siniestros con sospecha de fraude otros 65 siniestros seleccionados aleatoriamente entre los 325 restantes (387-62).

³⁰ La codificación pone de manifiesto diferentes frecuencias de aparición para cada uno de los tipos de fraude. La distinta conceptualización del fraude para los tramitadores de una compañía (buscan hechos que permitan denegar civilmente el siniestro) y para los investigadores del I.F.B. (intentar realizar acciones legales frente al fraude) es la causa principal de dichas diferencias.

³¹ Se dispone de un total de 65 indicadores de fraude relativos a los 127 siniestros. Los modelos que presentan Weisberg y Derrig se basan en la utilización de únicamente 10 indicadores debido a consideraciones prácticas asociadas a la modelización.

ciudad de elevado riesgo, si en las declaraciones que realiza después del siniestro reitera un “exceso de prudencia” y si tiene un historial de accidentes anteriores. Para el conductor asegurado, Weisberg y Derrig únicamente tienen en cuenta la facilidad de cooperación con la entidad, mientras que, en relación al daño producido, el número de indicadores considerado es más elevado (presencia de un daño inusual para el tipo de accidente producido, negativa del demandante a someterse a un examen médico, daños inconsistentes con el informe policial,...). El tratamiento médico derivado del siniestro es igualmente analizado (elevado número de visitas, desmesurado número de prescripciones médicas,...). Por último, los modelos incorporan información relativa a la influencia del siniestro en la capacidad laboral del demandante debido a las repercusiones que puede tener la declaración de larga incapacidad.

La validación de los modelos se realiza a partir del análisis del coeficiente de determinación (R^2) o bondad del ajuste obtenida. No se presenta el poder explicativo de cada una de las variables (en términos de significación individual) ni los resultados de los contrastes de significación global de cada modelo. El modelo que tiene como variable dependiente el índice de sospecha de fraude (de 0 a 10) para los tramitadores presenta una bondad del ajuste del 65%; este coeficiente disminuye al 56% en el modelo con variable dependiente el índice de sospecha de fraude para los investigadores. En el modelo con variable dependiente el número de votos a favor de la presencia de fraude, la bondad del ajuste es del 46%.

Se produce una notable variación en relación a los indicadores de fraude incluidos en la especificación de cada modelo, hecho que según los autores viene provocado por la diferente perspectiva que subyace en la definición de la variable dependiente para tramitadores e investigadores. Ello se hace evidente al considerar la clasificación, por tipos de fraude, establecida por ambos en relación a los 127 siniestros estudiados: mientras que los tramitadores clasifican un 39.4% de los siniestros como legales, un 38.6% como *build-up*, un 15.7% como *opportunistic fraud* y un 6.3% como *planned fraud*, los investigadores del I.F.B. clasifican un 44.9% como legales, un 2.4% como *build-up*, un 40.9% como *opportunistic fraud* y un 8.7% como *planned fraud* (el 3.1% restante lo engloban en la categoría genérica de *otros*). La diferencia más notable, tal y como puede observarse, se encuentra en la clasificación del fraude como *build-up* o como *opportunistic* ya que la dificultad existente en muchos casos para diferenciar

ambos conceptos puede derivar en la concentración de los casos en un único tipo.

El pequeño tamaño de la muestra y el hecho de que no sea aleatoria aparecen como dos de las principales razones para justificar la obtención de unos resultados no satisfactorios en términos de calidad del ajuste. En relación a la segunda razón cabe destacar que el hecho de trabajar con una muestra estratificada implica realizar una ponderación de las observaciones, siendo a veces complicado establecer una medida adecuada para la corrección.

El estudio realizado por Weisberg y Derrig muestra también que cuanto mayor es el número de indicadores considerados más fuerte es la sospecha de fraude y más elevada es la frecuencia de votos a favor de la presencia del mismo. Sin embargo, la consideración de todos ellos en la tramitación de siniestros resultaría ineficiente, fundamentalmente por dos razones. En primer lugar sería demasiado costoso para la compañía recoger de forma habitual toda la información y, en segundo lugar, se obviaría el hecho de que algún indicador (o combinación de indicadores) puede tener especial relevancia en la explicación de un determinado tipo de fraude. Notemos que estas dos conclusiones apoyan la necesidad de aplicar tratamientos econométricos adecuados para realizar una correcta selección de las variables explicativas e identificadoras del fraude.

En 1995 y en 1996 Weisberg y Derrig modelizan la sospecha de fraude, esta vez, utilizando una nueva base de datos, con siniestros ocurridos en 1993.

Sin embargo, estos modelos (de características análogas a los utilizados en estudios anteriores) continúan estando fundamentados en una muestra de siniestros cuyos expedientes ya han sido cerrados por las compañías. La información disponible hace referencia no sólo a las circunstancias fundamentales que rodean la ocurrencia del accidente sino también a la tramitación del mismo por la entidad.

La elaboración de una gran base de datos, (conocida como la "*Detailed Claim Database-D.C.D.*") regulada legalmente, con información de todos los siniestros cerrados en o a partir del 1 de enero de 1994 por las entidades aseguradoras de Massachusetts, supone para los autores mencionados la posibilidad de continuar con sus estudios teniendo a su disposición un elevado número de casos. Ello

supone un importante avance de cara a validar el poder explicativo de los principales indicadores de fraude tradicionalmente utilizados. Las compañías están obligadas a facilitar toda la información relativa al siniestro y, por tanto, las conclusiones obtenidas podrán ser generalizadas al mercado asegurador global. A modo de referencia cabe señalar que en julio de 1997 la *D.C.D.* contenía información para un total de 514583 siniestros de autos.

Sin embargo, el valor potencial de este nuevo sistema de recogida de información parece ligado también a otro aspecto: la posibilidad de analizar la efectividad en coste de un mecanismo automático de control de fraude.

En su artículo más reciente "A.I.B.-P.I.P. Claim Screening Experiment Final Report. Understanding and Improving the Claim Investigation Process" (1998), Derrig y Weisberg presentan los primeros resultados derivados de la creación de un sistema inteligente de detección de fraudes que ayude a mejorar el proceso de investigación de los siniestros.

Si hasta ahora todos sus estudios estaban relacionados con el tratamiento de expedientes de siniestros ya cerrados, el objetivo perseguido bajo lo que los autores denominan el "*Claim Screening Experiment, C.S.E.*" es modelizar la sospecha existente de fraude en diferentes etapas de la "vida" del siniestro. Una vez acaecido el accidente y declarado a la compañía, la información relativa a diferentes aspectos del mismo (fundamentalmente relacionada con el tratamiento médico derivado) puede llegar a la entidad cuando ha transcurrido ya un determinado periodo de tiempo. En base a ello, resulta adecuado generar índices de sospecha que alerten a los tramitadores sobre la conveniencia o no de investigar un siniestro a medida que se va recibiendo información. Normalmente, durante los 30 primeros días se conoce como ha sido el accidente, quién es el demandante y el asegurado. Posteriormente, en los dos meses siguientes, se tienen datos sobre la naturaleza de las lesiones, del abogado del lesionado y de la fase inicial de gestión del siniestro. En el periodo que resta hasta llegar al medio año, se recibe documentación más extensa sobre el tratamiento médico de las lesiones.

Derrig y Weisberg presentan en su estudio una triple clasificación para los siniestros con daños personales (atendiendo a su tramitación). De esta forma diferencian entre:

1. Siniestros con tramitación rápida (“*express*”). Son aquéllos en los que no hay ninguna prueba aparente de sospecha de fraude y las cuantías reclamadas no son elevadas. Por ello, los mismos pueden ser gestionados con rapidez.
2. Siniestros con tramitación no activada (“*duds*”). Para éstos hay una declaración del daño pero no se puede realizar un seguimiento del accidente todavía. Son expedientes que quedan archivados sin que la compañía realice ningún pago por el momento.
3. Siniestros con tramitación prolongada (“*target*”). Son los que despiertan sospecha de fraude y suelen estar caracterizados por elevadas cuantías demandadas (normalmente se incrementan durante el proceso debido, por ejemplo, al alargamiento del tratamiento médico aplicado).

El porcentaje más elevado de siniestros se encuentra en la tercera categoría (60%), según los resultados del C.S.E. La utilización de este sistema de detección puede ayudar a los tramitadores a decidirse sobre la conveniencia o no de investigar un siniestro (diferenciando incluso sobre la necesidad de realizar una investigación ordinaria o especializada).

Los autores modelizan la sospecha de fraude utilizando la técnica de regresión lineal, como ya venían haciendo, pero incorporando ciertas interacciones entre las variables.

Los resultados obtenidos de la aplicación permiten señalar que en determinados momentos de la tramitación estaría justificada una mayor investigación. Además confirman que existe una relación positiva entre el índice de sospecha y la ejecución de investigaciones especiales (*S.I.U.s*,...). Sin embargo, en relación a la efectividad del sistema, las conclusiones ponen de manifiesto un hecho importante. El buen funcionamiento del sistema implica la necesidad de incorporarlo automáticamente en las compañías puesto que los tramitadores se oponen a realizar manualmente la recogida de información requerida para implementar completamente el sistema.

El análisis coste-beneficio del diseño del sistema *C.S.E.* y de la investigación de siniestros derivada de su aplicación aparece como uno de los principales objetivos de estudio a corto plazo. La incorporación de los siniestros experimentados con el *C.S.E.*, que aún permanecen abiertos, al *D.C.D.*, permitirá comparar los costes iniciales y finales de los mismos, considerando además los asociados a la investigación.

El diseño de sistemas inteligentes que ayuden a las entidades a realizar una correcta recogida de datos puede contribuir decisivamente al buen funcionamiento de los sistemas automáticos de control y detección de fraude.

- **Aproximación al tratamiento del fraude mediante análisis cluster**

La aplicación de técnicas multivariantes cluster en el campo asegurador ha ido ganando importancia a lo largo de los últimos años (Lemaire, 1990; Derrig y Ostaszewski, 1994a y 1994b; Cummins y Derrig, 1993, 1997) apareciendo como una metodología susceptible de utilización en la modelización de la incertidumbre.

Conceptuado como un análisis enfocado principalmente a la definición de grupos homogéneos (clasificación de individuos en grupos atendiendo a su similitud en determinadas características), su aplicación en la clasificación de riesgos goza de especial importancia y, desde este punto de vista, ha sido utilizado en diferentes estudios propios del mercado de seguros.

Derrig y Ostaszewski (1994b) seleccionan la técnica cluster para realizar una clasificación de siniestros en términos de sospecha de fraude. Para ello utilizan la misma base de datos formada por 127 siniestros que había sido tratada por Weisberg y Derrig (1993) y que ya ha sido comentada en páginas anteriores. Su objetivo es plantear una solución al problema básico con el que se enfrenta el tramitador de siniestros (o en su caso, el investigador): clasificar o no el siniestro como fraudulento. La técnica utilizada sustituye el usual "cero-uno"³² por una función de medida cuyos valores oscilarán precisamente entre dichos extremos. Si la valoración obtenida tras la aplicación del análisis es 0 ó 1, la asignación a una de las categorías es clara. Si el valor se encuentra a lo largo del intervalo, la técnica generará una determinada clasificación en relación a la sospecha existente de fraude.

³² No fraude-fraude.

Los autores desarrollan un doble estudio. Por un lado, realizan una agrupación de siniestros teniendo en cuenta los niveles existentes de sospecha y, por otro, presentan una clasificación atendiendo a la valoración de fraude realizada, es decir, considerando los diferentes tipos de fraude posibles.

En el primero de los casos, la sospecha de fraude es cuantificada mediante la creación de una escala de cero a diez (ya había sido utilizada en estudios anteriores). Las respuestas son clasificadas en cinco clusters iniciales teniendo en cuenta los niveles de sospecha (según la valoración de los tramitadores de siniestros): ausencia de sospecha (0), sospecha débil (1-3), sospecha moderada (4-6), sospecha fuerte (7-9) y sospecha cierta (10). La agrupación posterior se realiza atendiendo a la información suplementaria aportada por otras dos medidas: el nivel de sospecha de los investigadores³³ (en base a la misma escala que los tramitadores) y el número de “votos” a favor de la presencia de fraude (escalado de cero a tres)³⁴. El algoritmo utilizado provoca una clasificación final de siniestros en la que las diferencias entre los centros de los diferentes clusters pone de manifiesto una divergencia entre el enfoque utilizado por tramitadores e investigadores. No obstante, dicha diferencia se suaviza al considerar sólo dos clusters: el asociado a la “no percepción de fraude” y el que incluye “percepción moderada de fraude” más “percepción fuerte de la existencia de fraude”.

En el segundo caso, la clasificación de siniestros se realiza atendiendo a la tipología de fraude existente. Las categorías consideradas son: “*planned fraud*”, “*opportunistic fraud*”, “*build-up*” y “*no fraud/build-up*”. Cada siniestro es codificado según el nivel de sospecha existente (en una escala de cero a diez) en seis componentes del mismo: el accidente, el daño, el asegurado, el demandante, el tratamiento médico y la pérdida de salario. Los resultados obtenidos muestran una agrupación en cinco clusters³⁵. El centro de cada uno de ellos indica, en base

³³ Cabe recordar que los 127 siniestros fueron codificados paralelamente por tramitadores de siniestros y por investigadores del I.F.B. de Massachusetts.

³⁴ La codificación se realizó por dos tramitadores y dos investigadores (el hecho de que en la escala no aparezca el máximo, es decir, cuatro, se debe a que ninguno de los siniestros fue clasificado como fraudulento por todos los codificadores).

³⁵ La categoría de build-up aparece dividida en dos partes en función del nivel de sospecha existente.

al nivel de sospecha existente para cada una de las seis componentes, la posible presencia de uno u otro tipo de fraude³⁶.

Las conclusiones obtenidas por Derrig y Ostaszewski ponen de manifiesto la susceptibilidad de aplicación del análisis cluster al tratamiento del fraude. En su opinión, su principal conclusión es que hay que otorgar más peso a las componentes asociadas al propio siniestro que a la percepción subjetiva de los tramitadores (y/o de los investigadores).

- **Aproximación al tratamiento del fraude mediante redes neuronales**

Partiendo de la base de datos utilizada por Weisberg y Derrig (1993) y por Derrig y Ostaszewski (1994b), Brockett, Xia y Derrig (1995) presentan una aproximación al estudio del fraude en el seguro del automóvil mediante la aplicación de redes neuronales.

El objetivo es crear un sistema de detección de fraude teniendo en cuenta el vector de características de cada uno de los siniestros. Para ello, se seleccionan los indicadores de fraude (un total de 65 variables) que ya habían sido utilizados por Weisberg y Derrig (1993), y que como sabemos están relacionados con el accidente, el demandante, el asegurado, el daño, el tratamiento médico y la pérdida salarial derivada.

Seleccionando 77 siniestros del total de la muestra (formada por 127 siniestros) para realizar la modelización y dejando 50 siniestros para llevar a cabo predicción ex-post³⁷, la aplicación de redes supone la representación de cada uno de los casos mediante un vector de atributos compuesto, en este caso, por los 65 indicadores de fraude utilizados. El método parte de dos premisas básicas: en primer lugar, modelos de siniestros parecidos tendrán niveles de sospecha similares y, en segundo lugar, cada uno de los indicadores considerados tendrá igual importancia en la explicación de la existencia de fraude.

Así, teniendo en cuenta un input formado por 77 vectores (uno para cada siniestro), el output resultante se centra en la clasificación del suceso en base a

³⁶ A modo de ejemplo, el cluster con centro (0,0,0,0,0,0) recoge siniestros sin sospecha de fraude. En el cluster con centro (7,8,7,8,8,0) aparecen siniestros con alta sospecha de "*planned fraud*" o "*fraude planeado*".

³⁷ Se utiliza una parte de la muestra para validar los resultados obtenidos.

cuatro categorías: “siniestro válido”, “siniestro con débil sospecha de fraude”, “siniestro con sospecha de fraude moderada” y “siniestro con una elevada sospecha de fraude”. El hecho de que los siniestros con vectores de indicadores parecidos (y, por tanto, con pequeñas “distancias” entre ellos) queden más o menos juntos y alejados del resto, dentro del “mapa de representación”, permite crear zonas (“regiones de decisión”) que se identificarán con las cuatro posibilidades (outputs) comentados en relación a la sospecha de fraude.

Introduciendo en el proceso las valoraciones subjetivas de sospecha de fraude de tramitadores e investigadores (ya consideradas en estudios anteriores³⁸), los resultados obtenidos perfeccionan, según el criterio de los autores, la clasificación de siniestros obtenida a partir del uso, únicamente, de una variable dependiente asociada a la sospecha de fraude y de determinados indicadores. No obstante, problemas relacionados con la propia metodología y, en su caso, con los datos, pueden limitar las conclusiones obtenidas del estudio. Así, el hecho de que todos los indicadores sean igualmente ponderados implica no considerar la posible existencia de variables más significativas que otras.

Los resultados del análisis y su comparación con la clasificación de siniestros presentada por los tramitadores y los investigadores, reflejan una elevada calidad en términos de predicción *ex-post*. Por lo que respecta a la predicción *ex-ante*, las conclusiones no son tan favorables. Los autores remiten a una ampliación de la muestra, a una mejora de la información o a una eliminación de los indicadores no significativos como posibles soluciones a la búsqueda de resultados más positivos.

- **Aproximación al tratamiento del fraude mediante el uso de modelos de elección probabilística**

Los modelos de elección discreta ofrecen la posibilidad de cuantificar la probabilidad de aparición de comportamientos fraudulentos en los siniestros cuando la variable dependiente ha sido adecuadamente categorizada para recoger la presencia/ausencia de fraude.

La aplicación de modelos lógit y próbit simples queda patente en la literatura existente sobre el tema. Si bien en Ayuso (1995) y en Artís, Ayuso y Guillén

³⁸ Weisberg y Derrig (1993), Derrig y Ostaszewski (1994b).

(1998a) es posible encontrar una aplicación de modelos lógit sencillos al estudio de la dicotomía planteada, en Belhadji y Dionne (1997) el estudio realizado parte de la aplicación de un modelo próbit.

En el trabajo de investigación "El Fraude en el Seguro del Automóvil" (Ayuso, 1995) se especifica un modelo de regresión logística simple para cuantificar la probabilidad de existencia de fraude. Para ello se utiliza la base de datos facilitada por el Insurance Fraud Bureau de Massachusetts, comentada en páginas anteriores, formada por 127 expedientes de siniestros. Teniendo en cuenta la categorización de fraude presentada por Weisberg y Derrig (1993), el objetivo del estudio se centra en modelizar dos situaciones alternativas:

1. ausencia de fraude *versus* fraude "*a priori*" y,
2. ausencia de fraude *versus* fraude "*a posteriori*".

En el primero de los casos, bajo la denominación de fraude "*a priori*", se recogen todos aquellos siniestros clasificados en la base de datos como sospechosos de fraude planeado (siniestro totalmente construido). En el segundo, bajo el concepto de fraude "*a posteriori*", se consideran todos los casos en los que se sospecha de la existencia de fraude tras la ocurrencia real del accidente (suma de aquéllos en los que existe sospecha de "*opportunistic fraud*" y "*build-up*").

Las variables explicativas utilizadas son seleccionadas del conjunto de indicadores de fraude presentado por Weisberg y Derrig (1993). Están relacionadas con características del accidente y del vehículo (ausencia de atestado policial, ausencia de testigos,...), del demandante (situación laboral), del daño (no existe evidencia objetiva del mismo) y del tratamiento médico derivado del siniestro (elevado número de visitas médicas,...). Los resultados obtenidos tras la modelización son satisfactorios, tanto en términos de significación individual y global de los modelos, como en capacidad predictiva (el porcentaje de aciertos en la clasificación es del 87.9% para el modelo que contiene como variable dependiente la sospecha de existencia de fraude *a priori* y del 80.6% para el que posee como dependiente la existencia de fraude *a posteriori*). No debe olvidarse, sin embargo, el escaso número de casos analizados en la muestra a la hora de extrapolar resultados.

En Artís, Ayuso y Guillén (1998a) puede encontrarse otra aplicación de modelización logística simple, esta vez, con ponderaciones. En este caso, la información contenida en la misma muestra utilizada en el presente libro es usada para cuantificar la probabilidad de existencia de fraude *versus* no fraude (no se distingue, por tanto, la existencia de diferentes tipos de actuación fraudulenta al alcance del asegurado).

Siguiendo un proceso muy similar al desarrollado en el trabajo mencionado en el párrafo anterior, Belhadji y Dionne (1997) utilizan un modelo próbit para estimar la probabilidad de existencia de fraude (detectado o sospechado) frente a la de no fraude. El objetivo planteado por ambos autores es doble. Por un lado, diseñar una herramienta de detección que ayude a los tramitadores de siniestros en el estudio del posible comportamiento fraudulento de los asegurados (implementación automática de los resultados derivados del modelo próbit previa selección de los indicadores más significativos de fraude). Por otro, desarrollar un sistema que evalúe la conveniencia o no de investigar los siniestros con elevada sospecha de fraude teniendo en cuenta el enfoque coste-beneficio.

La representatividad de la muestra utilizada por Belhadji y Dionne queda suficientemente justificada: formada por 2068 siniestros (1937 clasificados como no fraudulentos, 113 con sospecha de fraude y 18 con fraude detectado), ha sido diseñada en base a la información facilitada por 18 de las más grandes compañías que operan en el mercado asegurador automovilístico de Quebec (Canadá). Los expedientes han sido seleccionados aleatoriamente por las entidades entre todos los cerrados durante el periodo que va del 1 de abril de 1994 al 31 de marzo de 1995 (la participación de cada compañía ha sido proporcional a su cuota de mercado).

La selección adecuada de los indicadores de fraude tradicionalmente utilizados goza de especial importancia para los autores mencionados. El elevado número de circunstancias posiblemente relacionadas con la existencia de fraude (definen un total de 50 indicadores³⁹), deriva en la aplicación de un criterio que determine las variables más significativas a la hora de explicar la aparición de comportamientos deshonestos. Aunque plantean la posibilidad de realizar la selección entre indicadores atendiendo al cálculo de las probabilidades

³⁹ Algunos recopilados de la literatura existente y otros elaborados según el criterio de profesionales.

condicionales de fraude para cada uno de ellos, la aplicación de un modelo próbit permite determinar un conjunto de 18 variables como estadísticamente significativas. Éstas aparecen relacionadas, entre otras cosas, con aspectos del accidente y el daño (coste excesivo para un daño menor, existencia de un daño anterior no relacionado con el siniestro y declaración del coche como robado siendo encontrado posteriormente con daños importantes), del demandante y/o del asegurado (conoce el proceso de tramitación y la jerga empleada en seguros y reparaciones, acepta rápidamente su culpa, presenta un elevado número de facturas por daños corporales,...), de su situación financiera, de la rapidez con que pretende llegar a un acuerdo con la entidad, del nerviosismo que presenta durante la investigación, etc.

La selección de criterios probabilísticos alternativos (probabilidad estimada a partir de la cual el siniestro es clasificado como fraudulento) permite establecer, a partir de la muestra utilizada, reglas de actuación para las entidades frente al fraude. Si se elige como criterio de decisión un nivel probabilístico elevado, el número de siniestros a investigar será muy bajo, el nivel de precisión del modelo en relación a los casos que la compañía tiene clasificados como fraudulentos será muy elevado pero la entidad dejará de detectar un elevado número de fraudes. Si el criterio probabilístico es bajo ocurrirá lo contrario: el número de siniestros a investigar será elevado, el modelo clasificará como sospechosos un número de siniestros muy superior al observado en la muestra pero, tras su aplicación, aumentará el número de casos detectados por la compañía.

En relación al deseo de realizar un análisis coste-beneficio de la investigación de siniestros con elevada sospecha de fraude, Belhadji y Dionne muestran un estudio preliminar, ante la insuficiencia de mayor riqueza en los datos. La regla de decisión que plantean es lógica: si el coste de llegar a un acuerdo con el asegurado es más bajo que el coste de la investigación no resulta rentable investigar el siniestro; de otro modo, la investigación será realizada. En su aproximación hacen depender el coste de la misma de cuatro factores: la probabilidad de fraude, la experiencia de los investigadores, la formación de éstos y la existencia de unidades especiales de investigación dentro de la compañía. El tratamiento que realizan es muy sencillo y únicamente les permite extraer como conclusión la conveniencia de investigar el siniestro cuando las ganancias derivadas del proceso cubran, al menos, el coste de realizarlo.

Ambos autores enfatizan la necesidad de crear mecanismos automáticos de detección que indiquen a las entidades la probabilidad de existencia de fraude y que señalen la conveniencia de investigar o no los siniestros.

Frente a una aproximación del fraude que considere la dicotomía asociada a su existencia o no existencia (en muchos casos y, como hemos ido comentando, en relación a la sospecha de la misma), los trabajos realizados en el seno del Departamento de Econometría, Estadística y Economía Española de la Universidad del Barcelona, han estado dirigidos, desde el principio, a la modelización del fraude teniendo en cuenta sus diferentes formas de manifestarse. Desde este punto vista, las técnicas econométricas planteadas como susceptibles de aplicación han estado ligadas a modelos logísticos multinomiales y anidados para los que el libro que presentamos constituye el máximo exponente.

- **Otras aproximaciones al tratamiento del fraude en el seguro del automóvil**

En otro orden de tratamiento, Cummins y Tennyson (1996) presentan una aproximación al fraude considerándolo como una determinada manifestación de azar moral. En su estudio presentan un enfoque sectorial o agregado utilizando variables explicativas relacionadas con el sistema económico en general. Teniendo en cuenta que el asegurado o el demandante pueden realizar acciones después de la ocurrencia de un siniestro que afecten a la distribución de las pérdidas derivadas, es posible hablar de la presencia de "azar moral *ex-post*". Esta situación irá ligada a aquellos casos en los que el demandante tiene más información que el asegurador sobre la situación posterior al accidente.

Teniendo en cuenta que este comportamiento puede afectar tanto al número de siniestros declarados como a las indemnizaciones reclamadas, los autores realizan la modelización de la presencia de fraude (agregadamente a nivel de estado) midiendo su incidencia en la frecuencia de siniestros declarados por daños corporales. Para ello, llevan a cabo una estimación por mínimos cuadrados de un modelo lineal con variable dependiente el ratio entre la frecuencia de siniestros por daños corporales y la frecuencia por daños materiales⁴⁰. Las variables explicativas están relacionadas con características económicas,

⁴⁰ En realidad, la variable dependiente es el logaritmo de dicho ratio. La muestra está formada por observaciones recogidas en 29 estados norteamericanos en 1991 y 1992.

demográficas y legales de los estados. Como indicadores de azar moral aparecen diferentes actitudes o valoraciones de los individuos hacia el fraude (permitir al médico o al abogado extender facturas por servicios no prestados, mentir sobre la indemnización reclamada para recuperar la franquicia, mentir sobre las pérdidas para recuperar las primas pagadas, declarar que se han ocasionado daños corporales en individuos no relacionados con el accidente, participar en cadenas de fraude o aprovechar la existencia del seguro para cubrir un siniestro no relacionado con la póliza).

Adicionalmente, Cummins y Tennyson modelizan la frecuencia de siniestros por daños corporales con mayor probabilidad de estar sujetos a la presencia de azar moral. De esta forma, y considerando trabajos anteriores (Weisberg y Derrig, 1991)⁴¹ utilizan modelos de regresión clásicos para estudiar el comportamiento del porcentaje de siniestros por daños corporales leves (variable dependiente). De nuevo introducen variables demográficas, económicas y legales en la especificación del modelo, teniendo en cuenta, asimismo, variables de actitud relativas a diferentes comportamientos fraudulentos.

A pesar de que en esta segunda modelización la bondad del ajuste (medida por el coeficiente de determinación corregido o ajustado) no presenta valores tan elevados como en la primera, los resultados obtenidos en una y otra ponen de manifiesto dos importantes conclusiones: la existencia de azar moral tiene una incidencia importante en la declaración de siniestros por daños corporales y, además, influye significativamente en la proporción esperada de los mismos asociados a lesiones leves. Todo ello sirve para justificar la influencia del fraude y, de su valoración por parte de la sociedad, en el mercado asegurador.

Una vez analizadas las principales aportaciones realizadas hasta el momento en el estudio del fraude, podemos establecer una serie de consideraciones finales en relación a las mismas.

Los trabajos presentados quedan enmarcados dentro de dos grupos diferenciados: aquéllos de contenido puramente teórico y aquéllos de contenido básicamente aplicado.

⁴¹ Weisberg y Derrig (1991) identifican las lesiones leves (torceduras,...) como las más frecuentemente vinculadas con el fraude en daños corporales.

En relación a los primeros el enfoque es económico y se centra en determinar conceptualmente la utilidad esperada por el individuo al actuar fraudulentamente. En relación a los segundos, el fundamento teórico utilizado aparece poco desarrollado y se centra en presentar resultados derivados de trabajar, en la mayoría de los casos, con muestras de pequeñas dimensiones o de representatividad cuestionable. En la modelización se incluyen variables dotadas, muchas veces, de una gran subjetividad y para las que la compañía no dispone de información de manera inmediata a la ocurrencia del siniestro. La variable dependiente suele ser la sospecha de fraude al no disponerse de muestras representativas con fraude efectivamente detectado.

Todas las aportaciones presentan un enfoque microeconómico a la existencia de comportamientos deshonestos, salvo la de Cummins y Tennyson que enfatiza el estudio del fraude teniendo en cuenta una visión sectorial o agregada.

La modelización que presentaremos a partir del próximo Capítulo pretenderá combinar, de forma adecuada, el marco teórico y el aplicado. El objetivo será estudiar el fraude considerando el comportamiento que sigue el asegurado en la búsqueda de una utilidad máxima. Ello quedará patente en la técnica econométrica seleccionada, para la que se presentará un desarrollo teórico extenso (Capítulo 4). La aplicación empírica se centrará en aplicar los modelos desarrollados a una muestra con información, por primera vez, para fraudes detectados. El tamaño de esta última no es elevado, pero permitirá obtener, tal y como veremos, importantes conclusiones (Capítulos 5, 6 y 7).

4. PRESENTACIÓN DEL MODELO DE ELECCIÓN DE FRAUDE CON MÚLTIPLES ALTERNATIVAS EN EL SEGURO DEL AUTOMÓVIL

4.1 Modelo de elección de fraude: una decisión jerárquica entre alternativas

La aproximación teórica al modelo de elección que recoja el proceso seguido por el asegurado a la hora de actuar o no fraudulentamente puede realizarse atendiendo al diseño de un árbol de decisión. Bajo este planteamiento, la idea principal se centra en el hecho de que, en numerosas ocasiones, la elección entre alternativas no tiene por qué realizarse de forma directa; con frecuencia es posible considerar diferentes etapas en el proceso, que permitirán seleccionar la alternativa final, después de valorar el resto de posibilidades presentes en cada paso. En todo momento, el esquema presentado permitirá describir la elección realizada como el resultado de un proceso de decisión unificado, en el que la selección entre alternativas se llevará a cabo, bien bajo el criterio de maximizar la función de utilidad aleatoria, bien bajo el de minimizar los costes de elección (para nosotros, generalmente, costes de oportunidad en tiempo¹).

La elección entre no cometer fraude o cometerlo (atendiendo a la existencia de diferentes formas de defraudar) puede recogerse mediante el siguiente árbol de decisión:

¹ En Koujianou (1995), se habla, por ejemplo, de la minimización de los costes de búsqueda en la elección de un determinado tipo de vehículo.

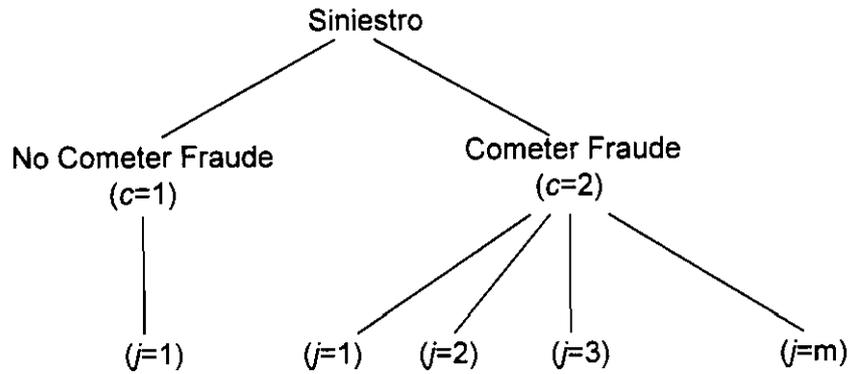


Figura 6

con $c=1,2,\dots,C$, alternativas iniciales y $j=1,2,\dots,m_c$, alternativas finales indexadas en cada opción previa c .

La estimación de un modelo de decisión jerárquico, como el planteado en la Figura 6, puede realizarse atendiendo al criterio de maximización de utilidad aleatoria (McFadden, 1978). Las posibilidades de desarrollar el proceso son fundamentalmente dos.

En primer lugar, el cálculo de las probabilidades asociadas a cada una de las situaciones finales (nivel inferior del árbol) puede realizarse de forma directa, es decir, sin considerar la existencia de niveles intermedios de decisión. Bajo este planteamiento, el uso de un modelo logit multinomial permitirá realizar adecuadamente el proceso, previa adopción de determinadas hipótesis para los términos de error que aparecerán en la definición de la función de utilidad aleatoria.

En segundo lugar, el uso secuencial de modelos logit multinomiales ha de permitir realizar el cálculo de los estimadores teniendo en cuenta la existencia de niveles intermedios. La consideración de la posible existencia de correlaciones entre las alternativas planteadas derivará en la formulación de un nuevo modelo, resultado de la aplicación de regresión logística jerárquica (modelos logit anidados).

Al igual que comentábamos en páginas anteriores, la introducción del criterio de actuación del individuo en base a la maximización de la utilidad implicará considerar la premisa de que, entre varias alternativas, el asegurado siempre elegirá aquella que le reporte mayor utilidad. Según el modelo de utilidad aleatoria (McFadden, 1978), la utilidad que se deriva del comportamiento del individuo puede descomponerse en dos partes, una determinista (que genera la denominada *utilidad estricta*) y una aleatoria:

$$U_{cj} = V_{cj} + e_{cj} \quad \text{con } c=1,2; \quad j=1,2,\dots,m_c. \quad (4.1)$$

La utilidad estricta, V_{cj} , recoge una combinación lineal entre parámetros y variables explicativas propias del individuo y/o de la elección realizada (características del asegurado, del contrario, del siniestro, del vehículo,...). El término aleatorio o de error, e_{cj} , recoge los efectos asociados a variables no consideradas en la parte determinista que pueden influir en la elección, así como las imperfecciones en la percepción de la maximización de utilidad (McFadden, 1978; Maddala, 1983; Greene, 1997).

La presencia de variables no observables supone la adopción de una determinada distribución poblacional para las mismas. La estimación de la probabilidad de que se realice una determinada elección c_j vendrá dada por:

$$P_{cj} = \text{Prob}(U_{cj} > U_{c'j'}) \quad \forall c_j \neq c'j'. \quad (4.2)$$

Teniendo en cuenta las consideraciones anteriores en relación a la definición de la función de utilidad y a la existencia de una distribución poblacional para las variables no observadas, la probabilidad anterior puede obtenerse como²:

$$P_{cj} = \int_{e_{cj}=-\infty}^{+\infty} F_{c'j'}(\langle V_{cj} + e_{cj} - V_{c'j'} \rangle) de_{cj}. \quad (4.3)$$

² Ver McFadden (1978).

donde, e_{cj} recoge la componente cj -ésima del vector \bar{e} , formado por los términos aleatorios de cada una de las posibles elecciones $(e_{11}, e_{21}, e_{22}, \dots, e_{2m})$ y F_{cj} es la derivada marginal de la función de distribución acumulada de \bar{e} , $(F(\bar{e}))$, con respecto al argumento (elección realizada).

A modo de ejemplo, la cuantificación de la probabilidad de que el asegurado cometa fraude en beneficio propio (ante tres posibilidades de actuación: no defraudar, defraudar en beneficio propio y defraudar en beneficio de un tercero) supondrá tener en cuenta el siguiente árbol de decisión³:

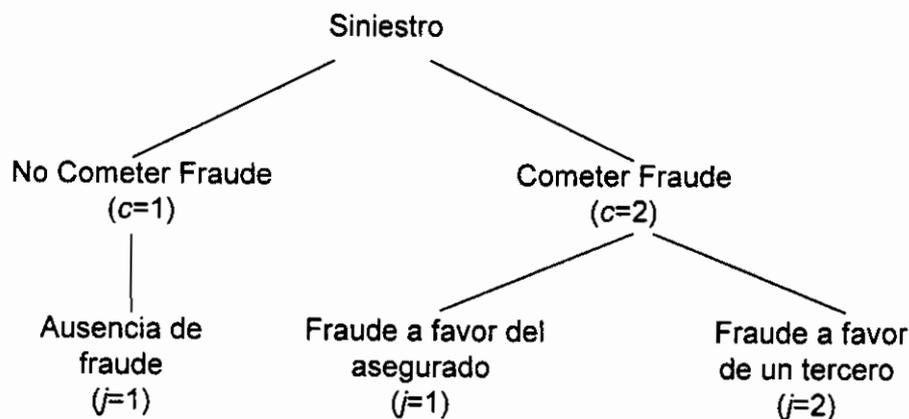


Figura 7

Atendiendo a la definición de una variable dependiente que recoja la elección final observada de la alternativa, de la forma:

$$Y_{21} = f(U_{21}) = \begin{cases} 1, & \text{si } U_{21} = \text{Max}(U_{11}, U_{21}, U_{22}) \\ 0, & \text{de otra forma} \end{cases} \quad (4.4)$$

y teniendo en cuenta que, de forma análoga, se podrían definir variables dependientes asociadas a la elección de no defraudar y de defraudar en beneficio

³ En Artís, Ayuso y Guillén (1998b) puede encontrarse el análisis detallado de este árbol de elección.

de un tercero, la modelización de la probabilidad correspondiente supondría determinar,

$$P_{21} = P(U_{21} > U_{11}, U_{21} > U_{22}). \quad (4.5)$$

Dado que $U_{cj} = V_{cj} + e_{cj}$, de forma que $\bar{e} = (e_{11}, e_{21}, e_{22})$, y suponiendo una determinada función de distribución para el vector de residuos ($F(\bar{e})$), tendremos,

$$\begin{aligned} P_{21} &= P(V_{21} + e_{21} > V_{11} + e_{11}, V_{21} + e_{21} > V_{22} + e_{22}) = \\ &= P(V_{21} + e_{21} - V_{11} > e_{11}, V_{21} + e_{21} - V_{22} > e_{22}) = \\ &= F_{e_{11}, e_{22}}(V_{21} + e_{21} - V_{11}, V_{21} + e_{21} - V_{22}) = \\ &= \int_{e_{21} = -\infty}^{+\infty} \frac{\partial F}{\partial e_{21}}(V_{21} + e_{21} - V_{11}, e_{21}, V_{21} + e_{21} - V_{22}) de_{21}. \end{aligned} \quad (4.6)$$

Dejemos el ejemplo y pasemos a continuación a determinar la expresión de la probabilidad de elegir una determinada alternativa, P_{cj} .

4.2 El Modelo Lógit Multinomial

La generación del modelo lógit multinomial supone una particularización de la expresión general presentada para la probabilidad de elegir una determinada opción (4.3), asumiendo que los residuos, e_{cj} , se distribuyen idéntica e independientemente según una distribución *valor extremo tipo I*⁴ (McFadden, 1978; Maddala, 1983). De esta forma, la probabilidad de elegir una determinada alternativa cj atiende a la expresión,

$$P_{cj} = P(Y_{cj} = 1) = \int_{-\infty}^{+\infty} \prod_{c'j' \neq cj} F(e_{cj} + V_{cj} - V_{c'j'}) \cdot f(e_{cj}) de_{cj}, \quad (4.7)$$

⁴ La función de distribución acumulada para una distribución valor extremo tipo I es $F(e_{cj} < e) = \exp(-\exp^{-e})$ y la función de densidad probabilística es $f(e_{cj}) = \exp(-e_{cj} - \exp^{-e_{cj}})$.

expresión que, una vez realizadas las operaciones oportunas (ver Maddala, 1983), puede escribirse como:

$$P_{cj} = \frac{e^{V_{cj}}}{\sum_{b=1}^C \sum_{J=1}^{m_b} e^{V_{bJ}}}, \quad (4.8)$$

donde b y J recogen el conjunto total de alternativas intermedias (defraudar o no) y finales (tipo de fraude), respectivamente.

Tomemos dicha expresión y adaptémosla a la situación objeto de estudio.

Como ya comentábamos en páginas anteriores, la función de utilidad estricta recoge una combinación lineal entre parámetros y variables, identificativas estas últimas de determinados atributos o características del individuo y/o de la elección.

La generación del modelo logit multinomial, estrictamente hablando, implica introducir únicamente como variables explicativas ciertas características del individuo que pueden incidir sobre su elección final. En este sentido, se contrapone al modelo logit condicional, en el que los atributos considerados son, preferentemente, los asociados a la elección (Greene, 1997).

Bajo esta perspectiva, y realizando la modelización como si de un proceso de decisión único se tratase (estimación de las probabilidades del último nivel del árbol sin considerar la existencia de etapas intermedias), podemos formular la función de utilidad aleatoria de la forma:

$$U_{i(cj)} = V_{i(cj)} + e_{i(cj)} = \beta'_{cj} X_i + e_{i(cj)}, \quad (4.9)$$

donde, cj recoge el conjunto total de alternativas finales entre las que ha de elegir el

individuo⁵, i indica el conjunto de individuos considerados en la muestra, X_i es el vector de características asociadas a los mismos y β_{cj} es el vector de parámetros a estimar para cada una de las posibles elecciones.

Atendiendo a esta notación, el modelo lógit multinomial vendrá especificado por:

$$P_{i(cj)} = P(Y_{i(cj)} = 1 | X_i) = \frac{e^{V_{i(cj)}}}{\sum_{b=1}^C \sum_{J=1}^{m_b} e^{V_{i(bJ)}}} = \frac{e^{\beta'_{cj} X_i}}{\sum_{b=1}^C \sum_{J=1}^{m_b} e^{\beta'_{bJ} X_i}} \quad \forall i=1, \dots, N. \quad (4.10)$$

En esta última expresión N denota el número total de individuos de la muestra disponible.

El árbol decisión presentado en la Figura 6 pone de manifiesto la existencia de un conjunto formado por $m+1$ elecciones finales (podemos, por tanto, reflejar la alternativa seleccionada mediante el subíndice $j(=0, 1, \dots, m)$ ⁶, donde $j=0$ refleja la elección de no defraudar). De esta forma, el sistema de ecuaciones a estimar para determinar las probabilidades asociadas a cada alternativa final, teniendo en cuenta la expresión (4.10), será⁷:

⁵ En el ejemplo presentado en páginas anteriores relacionado con la elección entre no defraudar, cometer fraude en beneficio del asegurado o en beneficio de un tercero, el conjunto de alternativas vendría referenciado por los subíndices "11", "21" y "22", modelizándose la decisión del individuo entre tres opciones finales, es decir, sin considerar la posible existencia de un nivel intermedio (éste queda representado por el primer dígito del subíndice, si bien su presencia queda justificada, únicamente, por el deseo de mantener uniformidad en la notación).

⁶ El subíndice c ha sido obviado a partir de este momento para simplificar la notación.

⁷ El condicionante en X_i ha sido eliminado dado que asumimos que las variables explicativas son deterministas. La expresión obtenida para el lógit multinomial (4.10) podrá escribirse, en base a las consideraciones realizadas, como:

$$P(Y_{ij} = 1) = \frac{e^{\beta'_j X_i}}{\sum_{r=0}^m e^{\beta'_r X_i}} \quad \forall j=0, 1, \dots, m.$$

$$P(Y_{ij}=1) = \frac{e^{\beta_j' X_i}}{1 + \sum_{r=1}^m e^{\beta_r' X_i}} \quad \forall j=1, \dots, m \quad (4.11)$$

$$P(Y_{i0}=1) = \frac{1}{1 + \sum_{r=1}^m e^{\beta_r' X_i}}$$

expresiones obtenidas al imponer la condición de normalización $\beta_0=0$.

De esta forma, como puede observarse, el modelo contiene un vector de parámetros (β_r) asociado a cada alternativa (excepto para la primera, debido a la condición de identificación impuesta).

La estimación de los parámetros se realiza habitualmente por máxima verosimilitud. La aplicación de un método iterativo (normalmente, el método de Newton) permite obtener valores para los estimadores a partir de la log-verosimilitud del modelo.

La interpretación de los coeficientes no es tan inmediata como en el modelo de regresión lineal clásico. En este último, los parámetros estimados miden la variación esperada que se produce en la variable dependiente cuando la variable explicativa correspondiente aumenta en una unidad⁸. En el modelo lógit, la variación esperada que se produce en la probabilidad de elegir una determinada alternativa al aumentar en una unidad la variable explicativa correspondiente no es constante, sino que depende de los valores que tome dicha variable y el resto de las variables explicativas (Maddala, 1983; Greene, 1997). No obstante, las estimaciones obtenidas para los parámetros son susceptibles de interpretación en términos de dirección en la variación de la probabilidad esperada. Asimismo, fijando el valor de todas las variables explicativas y dejando variar en una unidad únicamente una variable, la exponencial del parámetro estimado se interpreta como la variación esperada que se producirá en el cociente de riesgos.

⁸ De esta forma, la variación esperada en la variable dependiente es constante para cualquier incremento unitario en el valor de la variable explicativa, *ceteris paribus*.

La propiedad que cumple este modelo en relación a la “Independencia de Alternativas Irrelevantes - I.I.A.⁹.” (Luce 1959), y que garantiza que el cociente de probabilidades entre la j -ésima y la j' -ésima elección se mantenga constante independientemente del número de alternativas consideradas, simplifica el proceso de estimación del modelo y de predicción (McFadden, 1978; Maddala, 1983).

No obstante, y considerando el cumplimiento de la I.I.A., es posible profundizar en la modelización del esquema o árbol de decisión jerárquico, avanzando un poco más en el uso de los modelos logit multinomiales, con el objetivo de considerar etapas de decisión intermedia.

Atendiendo a este nuevo objetivo, la formulación de la función de utilidad estricta implicará la inclusión de atributos o características propios de cada etapa o nivel de decisión. De este modo, en la función de utilidad (4.9),

$$U_{i(cj)} = V_{i(cj)} + e_{i(cj)},$$

la parte determinista (utilidad estricta) adoptará la siguiente forma:

$$V_{i(cj)} = \beta' X_{i(cj)} + \alpha' Z_{i(c)}, \quad (4.12)$$

siendo $X_{i(cj)}$ un vector de atributos específicos de cada elección final para el individuo i , $Z_{i(c)}$ el vector de variables observadas que varían sólo con la elección intermedia (decisión de defraudar o no defraudar, sin detallar el tipo de fraude que se realiza) y β y α los vectores de parámetros correspondientes. Es posible hablar en términos de estos nuevos vectores de parámetros simplemente considerando que constituyen una concatenación de los parámetros referidos a cada alternativa, redefiniendo convenientemente el vector de explicativas.

⁹ *Independence of Irrelevant Alternatives.*

En base a esta nueva formulación, la estimación de la probabilidad de elegir una determinada alternativa c_j , será ahora el resultado de multiplicar dos probabilidades, una para cada nivel del árbol en el que nos situemos. Así,

$$P_{i(c_j)} = P_{i(j|c)} P_{i(c)}, \quad (4.13)$$

donde:

$P_{i(c_j)}$ es la probabilidad de que el individuo i elija la alternativa (c_j) ,

$P_{i(c)}$ es la probabilidad de que i elija la alternativa intermedia c , y

$P_{i(j|c)}$ es la probabilidad condicionada de elegir la alternativa j una vez el individuo ya se ha decidido por la alternativa c .

Teniendo en cuenta la definición del logit multinomial, la probabilidad de que el individuo i elija una determinada opción final j (atendiendo a la elección previa de c) podrá definirse como:

$$P_{i(j|c)} = \frac{e^{V_i(c_j)}}{\sum_{J=1}^{m_c} e^{V_i(c_J)}} = \frac{e^{\beta' X_i(c_j)}}{\sum_{J=1}^{m_c} e^{\beta' X_i(c_J)}}, \quad (4.14)$$

donde J recoge el conjunto de elecciones posibles en la alternativa intermedia c . Por otro lado, la probabilidad asociada a la elección c atenderá a la expresión:

$$P_{i(c)} = \frac{\sum_{J=1}^{m_c} e^{V_i(c_J)}}{\sum_{b=1}^C \sum_{J'=1}^{m_b} e^{V_i(b_{J'})}} = \frac{e^{\alpha' Z_i(c)} \left[\sum_{J=1}^{m_c} e^{\beta' X_i(c_J)} \right]}{\sum_{b=1}^C \left[e^{\alpha' Z_i(b)} \left[\sum_{J'=1}^{m_b} e^{\beta' X_i(b_{J'})} \right] \right]}, \quad (4.15)$$

donde b recoge el conjunto de alternativas intermedias y J' el de alternativas dentro de cada opción intermedia. Además el subíndice i que indica el individuo especificado varía entre 1 y N .

Teniendo en cuenta la expresión anterior, el valor definido por:

$$I_{i(c)} = \ln \left[\sum_{j=1}^{m_c} e^{\beta' X_{i(cj)}} \right], \quad (4.16)$$

denominado *valor inclusivo*, recogerá la utilidad esperada agregada para un subconjunto de elección o conjunto de alternativas finales asociadas a la intermedia.

Teniendo en cuenta esta especificación y, a partir de las expresiones (4.13), (4.14) y (4.15) podemos reescribir la probabilidad de elegir una determinada alternativa c_j en un modelo logístico anidado, como:

$$P_{i(cj)} = P_{i(j|c)} P_{i(c)} = \frac{e^{\beta' X_{i(cj)}}}{e^{I_{i(c)}}} \frac{e^{\alpha' Z_{i(c)} + I_{i(c)}}}{\sum_{b=1}^C e^{\alpha' Z_{i(b)} + I_{i(b)}}} = \frac{e^{\beta' X_{i(cj)} + \alpha' Z_{i(c)}}}{\sum_{b=1}^C e^{\alpha' Z_{i(b)} + I_{i(b)}}}. \quad (4.17)$$

Continuando con el desarrollo presentado, y siguiendo a McFadden (1978), cabe señalar que la estimación de la probabilidad de elegir una determinada alternativa puede realizarse estimando los parámetros β a partir del modelo condicional (4.14), determinado el valor inclusivo (4.16) y estimando el vector α a partir de la probabilidad marginal definida en (4.15). Esta aproximación no es sino una forma alternativa de modelizar la secuencia jerárquica de decisión, frente a lo que sería la estimación del modelo de forma completa (sin considerar la existencia de alternativas intermedias).

Sin embargo, el desarrollo de los modelos lógit multinomiales, tal y como hemos visto en páginas anteriores, se basa en la adopción de una determinada función de distribución para los términos aleatorios que definen la función de utilidad. Partiendo de la misma, los términos de error aleatorio se suponen idéntica e independientemente distribuidos, sin que se refleje, por lo tanto, la posible existencia de correlaciones entre los mismos. Ello queda de manifiesto en el hecho de que el coeficiente que acompaña al valor inclusivo es igual a la unidad.

La modelización de un proceso jerárquico de decisión, como el que venimos tratando, hace necesario analizar qué ocurre cuándo la hipótesis adoptada para los términos de error es violada, es decir, cuando es posible hablar de existencia de correlación entre las alternativas. Ante esta situación, el proceso econométrico a utilizar está fundamentado en el uso de los modelos lógit anidados, que pasamos a describir a continuación.

4.3 El Modelo Lógit Anidado

La introducción de un coeficiente diferente de la unidad para el valor inclusivo, que aparece en la probabilidad $P_{i(c)}$ de elegir una determinada opción c , dará lugar a un modelo lógit anidado (McFadden, (1978); Maddala (1983)). Este modelo es una generalización del modelo multinomial presentado, de la forma:

$$P_{i(c)} = \frac{e^{\alpha'Z_{i(c)} + (1-\sigma)I_{i(c)}}}{\sum_{b=1}^C e^{\alpha'Z_{i(b)} + (1-\sigma)I_{i(b)}}}, \quad (4.18)$$

donde σ , $0 \leq \sigma \leq 1$, es un coeficiente que en breve interpretaremos.

Siguiendo a Maddala (1983), el modelo lógit multinomial anidado puede derivarse a partir de la adopción, para el vector de residuos $\bar{\epsilon}$, de una distribución *Valor Extremo Generalizado (G.E.V.)*. Su característica principal radica en la posibilidad de estimación del modelo a pesar de la existencia de correlación entre aquellas alternativas que, estando en el mismo nivel de decisión, forman parte de conjuntos de elección diferentes. De esta forma, se relaja la hipótesis de independencia para los errores y, por tanto, no se exige el cumplimiento de la propiedad de Independencia entre Alternativas Irrelevantes (*I.I.A.*).

El modelo lógit anidado puede derivarse, al igual que hemos visto para el lógit multinomial, a partir de la Teoría de maximización de la Utilidad Aleatoria (McFadden, 1978).

Debido a la relación existente entre los modelos lógit anidados y los modelos de valor extremo generalizado¹⁰ (los primeros pueden considerarse un caso particular de los segundos), es posible demostrar como los lógit anidados son también consistentes con la maximización de la utilidad aleatoria. No obstante, dicha consistencia viene marcada por el cumplimiento de unas ciertas condiciones para el coeficiente del valor inclusivo.

Este parámetro recoge una estimación de la similitud entre los términos no observados (errores aleatorios) del nivel inferior del árbol de decisión. Así, el campo de variación para el mismo oscilará entre cero y uno, según exista ausencia de correlación ($\sigma=0$) o correlación máxima ($\sigma=1$). Lógicamente, el grado de independencia ($1-\sigma$) oscilará también entre cero y uno.

El valor inclusivo, tal y como lo hemos definido en páginas anteriores, mide la utilidad agregada esperada asociada a un conjunto de alternativas. Los coeficientes para dichos valores, ($1-\sigma$), se estiman con el resto de coeficientes del modelo y registran la disimilitud entre alternativas que pertenecen a un subconjunto en particular.

Siguiendo a McFadden (1978), la estructura anidada es consistente con la maximización de la utilidad aleatoria si y sólo si los coeficientes de los valores inclusivos se encuentran dentro del intervalo unitario¹¹. Cuando la disimilitud ($1-\sigma$) tiende a uno, σ tiende a cero y, por tanto, la correlación es prácticamente nula (se cumple la hipótesis de *I.I.A.*) y podemos modelizar utilizando un lógit multinomial simple¹². Si el coeficiente tiende a 0, la correlación de los términos de error tiende a uno y, por tanto, los consumidores elegirán aquella alternativa que les proporcione una mayor utilidad estricta (atendiendo a la parte determinista de su función de

¹⁰ Un desarrollo exhaustivo de estos modelos puede encontrarse, asimismo, en McFadden (1978) y Maddala (1983).

¹¹ Ver en McFadden (1978) la demostración de la consistencia de los modelos lógit anidados con la maximización de la Utilidad Aleatoria.

¹² Los términos aleatorios se distribuyen idéntica e independientemente según una distribución *valor extremo tipo I* (McFadden, 1983).

utilidad y, por tanto, en base a sus atributos individuales y/o a los de elección), aunque las alternativas pueden no ser independientes.

La condición establecida en relación al intervalo de variación para $(1-\sigma)$ o para σ , es necesaria y suficiente para que el modelo logit anidado sea consistente con la maximización de la utilidad (McFadden, 1978).

Si el coeficiente del valor inclusivo es mayor que uno ($(1-\sigma) > 1$), se produce un seguimiento en el proceso de decisión a lo largo de los nudos del árbol, de forma que el resultado obtenido no es consistente con la maximización de la utilidad (el anidamiento no es consistente con la maximización de la utilidad).

En base a ello (Koujianou, 1995), cabe destacar que la jerarquía en el toma de decisiones no implica necesariamente que los consumidores (asegurados) elijan secuencialmente. En este caso, el árbol más que plasmar el proceso de selección del individuo, permite reflejar la existencia de posibles correlaciones entre factores no observados. Ello posibilita una modelización econométrica más general.

Así, por ejemplo, la interpretación puede realizarse teniendo en cuenta que el esquema de decisión seguido por el asegurado a la hora de optar por cometer o no fraude puede llevar asociados unos determinados costes de oportunidad (tiempo empleado en la preparación material del siniestro, tiempo invertido en diseñar la forma más verosímil posible del relato a presentar por el propio asegurado y el contrario,...), de forma que, bajo este enfoque, el uso del árbol de decisión se fundamenta más en una forma de eliminar alternativas del conjunto de elección, que en una maximización de la utilidad esperada. Además cabe reseñar (Koujianou, 1995, pág. 914) que la presencia de un valor inclusivo con parámetro negativo no es consistente con la maximización de la utilidad.

El modelo lógit anidado puede ser estimado secuencialmente¹³ (“*estimación por etapas*”). En el primer paso, se realiza la estimación de los parámetros de la log-verosimilitud condicional (nivel inferior). El modelo de elección discreta simple provee estimadores de β . El vector de parámetros estimados y las observaciones muestrales son utilizados para calcular los valores inclusivos. Después, en un segundo paso, se realiza la estimación en relación al nivel superior del árbol de decisión, siendo el valor inclusivo una de las variables exógenas del modelo.

El proceso de estimación secuencial permite obtener estimadores de los parámetros consistentes, pero ineficientes.

Además del método de estimación secuencial, es posible calcular los estimadores de los parámetros a partir del método de Máxima Verosimilitud Completa¹⁴, realizándose la estimación del modelo de forma unificada.

4.4 Optimización del punto de corte en modelos con múltiples alternativas

La determinación del criterio probabilístico óptimo que permita efectuar la clasificación de una observación en una categoría de elección es habitual en las aplicaciones asociadas a modelos lógit sencillos, cuya finalidad sea predictiva.

Cuando la variable dependiente toma únicamente dos valores, recogiendo la ausencia/presencia de una determinada situación, el criterio para asignar un individuo a uno u otro grupo es, por defecto el de una probabilidad estimada superior o no a 0.5. Sin embargo dicho criterio es discutible desde el punto de vista de que, tomando a modo de ejemplo los resultados que obtendríamos al cuantificar la probabilidad de no existencia de fraude ($Y=0$) *versus* fraude ($Y=1$), un siniestro con probabilidad estimada de 0.51 sería clasificado como fraudulento mientras que otro con probabilidad estimada de 0.49 lo sería como no fraudulento. La solución a este problema radica en determinar a partir de qué

¹³ Una aplicación importante puede encontrarse en Koujianou (1995). En este caso, la autora aplica un método secuencial para estimar un modelo anidado con cinco niveles de decisión.

¹⁴ Full Information Maximum Likelihood (FIML).

nivel de probabilidad se debe considerar el siniestro como sospechoso de fraude teniendo en cuenta los resultados obtenidos (análisis del grupo de pertenencia observado y del predicho por el modelo).

Ahora bien, ¿qué ocurre cuando modelizamos teniendo en cuenta la posibilidad de múltiples respuestas de elección? ¿Cómo determinamos el criterio probabilístico que nos permita optimizar la asignación de una observación a una determinada categoría?

La asignación de una observación a una categoría de elección se fundamenta en el criterio de máxima probabilidad. De esta forma, una vez realizado el proceso de estimación, el grupo de pertenencia para la misma (categoría en que queda clasificado cada caso) será aquel para el que se obtiene una mayor probabilidad predicha. La suma de las probabilidades obtenidas ha de ser la unidad, de forma que teniendo en cuenta tres alternativas finales de elección, el modelo será incapaz de decidir cuando las probabilidades predichas sean iguales ($\text{Prob}(Y=0)=1/3$; $\text{Prob}(Y=1)=1/3$; $\text{Prob}(Y=2)=1/3$). Este será el criterio seguido por defecto para delimitar las zonas de clasificación, sin, embargo, no tiene por qué ser el óptimo, tal y como veremos a continuación.

Sean $P(0)$, $P(1)$ y $P(2)$ las probabilidades estimadas de clasificación en el primer, segundo y tercer grupo de elección, respectivamente. Sabiendo que la suma de las mismas es igual a uno, es posible representar gráficamente las zonas de clasificación para cada una de las observaciones de la siguiente forma,

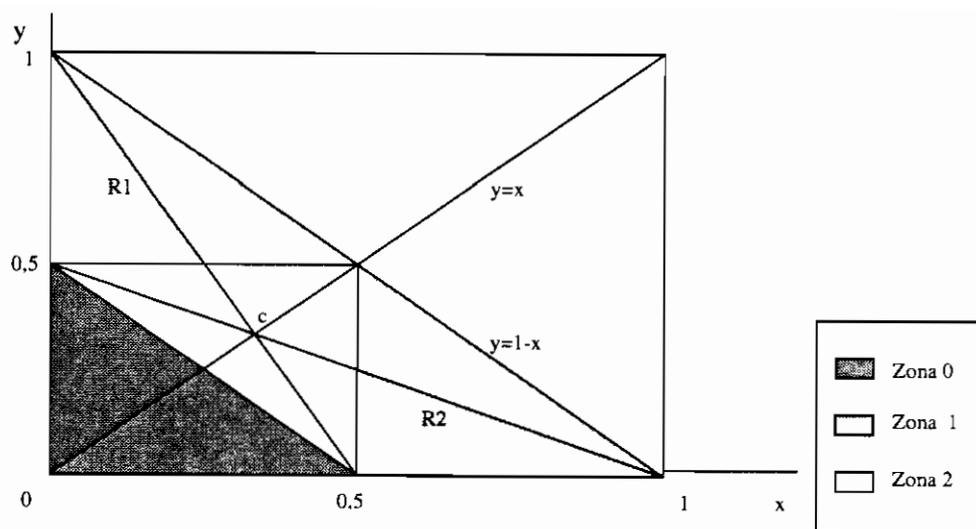


Gráfico 1

Por simplificación notacional realizamos un cambio de variables, siendo $x= P(1)$ e $y= P(2)$. Lógicamente, la utilización de complementarios, $P(0)= 1-(P(1)+P(2))$, permitirá deducir el valor correspondiente a la probabilidad para la primera categoría de elección.

Analicemos, paso a paso, la construcción del gráfico 1.

La hipotenusa del triángulo, representada por la recta $y=1-x$, permite delimitar las zonas de clara aceptación de la alternativa 1 (zona marcada con un 1) y de la alternativa 2 (zona 2), sin más que tener en cuenta, que siempre que la probabilidad estimada para una de ellas sea mayor a 0.5, la observación será clasificada en el grupo correspondiente.

Análogamente, la recta $y=0.5-x$ delimita la zona de aceptación de la alternativa 0 (zona 0), teniendo en cuenta que todos los puntos situados por debajo de la misma tendrían una probabilidad estimada $P(0)>0.5$.

Sin embargo, como se desprende de la representación, queda una zona en la que *a priori* no sabríamos como clasificar la observación, siendo necesario fijar un criterio que nos ayude a incluirla dentro de cualquiera de las tres categorías iniciales, teniendo en cuenta el criterio de máxima probabilidad, es decir, verificando que con el mismo la observación que inicialmente está en zona de incertidumbre ha quedado clasificada en la categoría para la que el modelo otorga una mayor probabilidad estimada.

La representación sobre el triángulo de las rectas,

$$R1 \Rightarrow y = 1 - 2x$$

$$R2 \Rightarrow y = (1-x)/2$$

y la utilización de la diagonal ($y=x$), permite delimitar, teniendo en cuenta el punto de corte de las mismas (c), el criterio general de asignación de las observaciones en la zona 0, en la zona 1 o en la zona 2, respectivamente, siguiendo el siguiente proceso:

$$Si \begin{cases} y < \frac{1-x}{2} \\ y < 1-2x \end{cases}, \quad (4.19)$$

la observación será clasificada dentro de la primera categoría de elección, al ser $P(0)$ la probabilidad estimada máxima.

$$Si \begin{cases} y < x \\ y > 1-2x \end{cases}, \quad (4.20)$$

la observación se clasificará dentro de la segunda categoría, lógicamente, al ser $P(1)$ (es decir, x) la probabilidad máxima y,

$$\text{Si } \begin{cases} y > x \\ y > \frac{1-x}{2} \end{cases} \quad (4.21)$$

se clasificará en la tercera categoría atendiendo a la explicación anterior pero para $P(2)$ (es decir, y).

La demostración de que la probabilidad estimada es máxima para cada una de las situaciones anteriores es sencilla.

Veamos en primer lugar como, si se cumplen las condiciones marcadas en (4.19), $P(0) > P(1)$ y $P(0) > P(2)$.

**Demostración 1.*

Siempre se verifica, por definición del modelo que $P(0) + P(1) + P(2) = 1$

Sea (0) la categoría de asignación si
$$\begin{cases} P(2) < 1 - 2P(1) & (a.1) \\ P(2) < \frac{1 - P(1)}{2} & (b.1) \end{cases}$$

*Aplicando (b.1) $\Rightarrow 2P(2) < 1 - P(1)$
 $1 - P(1) > 2P(2)$*

podemos demostrar $P(0) = 1 - P(1) - P(2) > 2P(2) - P(2) > P(2)$.

Aplicando (a.1) $\Rightarrow -P(2) > 2P(1) - 1$

podemos demostrar $P(0) = 1 - P(1) - P(2) > 1 - P(1) + 2P(1) - 1 > P(1)$.

Haciendo lo propio con $P(1)$, teniendo en cuenta las condiciones que aparecen en la expresión (4.20):

*Demostración 2

Como siempre, se cumple que $P(0)+P(1)+P(2)=1$

Sea (1) la categoría de asignación si $\begin{cases} P(2) < P(1) & (a.2) \\ P(2) > 1-2P(1) & (b.2) \end{cases}$

Aplicando (b.2) $\Rightarrow P(2) > 1-2P(1)$
 $-P(2) < 2P(1)-1$

podemos demostrar $P(1)=1-P(0)-P(2) < 1-P(0)+2P(1)-1;$
 $P(1)-2P(1) < -P(0) \Rightarrow -P(1) < -P(0) \Rightarrow P(1) > P(0),$

y $P(1) > P(2)$ por la propia condición (a.2) .

Por último, considerando las condiciones que aparecen en la expresión (4.21), podemos demostrar para $P(2)$,

*Demostración 3

Análogamente, como en los casos anteriores, se cumple que $P(0)+P(1)+P(2)=1$.

Sea (2) la categoría de asignación si $\begin{cases} P(2) > P(1) & (a.3) \\ P(2) > \frac{1-P(1)}{2} & (b.3) \end{cases}$

Aplicando (b.3) $\Rightarrow 2P(2) > 1-P(1)$
 $-P(1) < 2P(2)-1$

podemos demostrar $P(2)=1-P(1)-P(0) < 1+2P(2)-1-P(0)=2P(2)-P(0)$
 $P(2) < 2P(2)-P(0) \Rightarrow P(2)-2P(2) < -P(0) \Rightarrow -P(2) < -P(0) \Rightarrow P(2) > P(0),$

y $P(2) > P(1)$ por la condición (a.3).

Finalmente, tal como queríamos comprobar, $P(2) > P(0)$ y $P(2) > P(1)$.

Una vez delimitadas las zonas de aceptación, ¿cómo podemos optimizar la clasificación de las observaciones, ampliando, por ejemplo, alguna de las áreas?. Parece lógico pensar que el desplazamiento del punto de corte c (variación del criterio probabilístico) a lo largo de las rectas utilizadas o en otras direcciones permitirá modificar (en el sentido de ampliar o reducir) las zonas de clasificación.

Fijemos la variación de c a lo largo de la recta $y=x$ y determinemos las ecuaciones para las rectas $R1$ y $R2$ teniendo en cuenta que ambas han de pasar por el punto seleccionado. Al establecer este cambio en el punto c estamos actuando de manera simétrica respecto a las zonas 1 y 2. De manera más general se podría proceder a cambiar el criterio para favorecer alguna de las tres elecciones posibles.

La solución al sistema de ecuaciones que pasan por un punto da lugar al siguiente resultado,

$$R1 \Rightarrow \left\{ \frac{x-0}{c} = \frac{y-1}{c-1}; \frac{(c-1)x}{c} = y-1; y = 1 + \frac{(c-1)}{c}x \right.$$

y, (4.22)

$$R2 \Rightarrow \left\{ \frac{x-1}{c-1} = \frac{y-0}{c}; y = \frac{c(x-1)}{c-1} = \frac{c}{c-1}(x-1) \right.$$

De esta forma y teniendo en cuenta las condiciones marcadas anteriormente para delimitar las zonas de clasificación, podemos establecer un criterio para optimizar el punto de corte en un modelo logístico multinomial con tres alternativas, de manera que si,

$$\left\{ \begin{array}{l} y < \frac{c}{c-1}(x-1) \\ y < 1 + \frac{c-1}{c}x \end{array} \right\} \text{ ó } \left\{ \begin{array}{l} P(2) < \frac{c}{c-1}[P(1)-1] \\ P(2) < 1 + \frac{c-1}{c}P(1) \end{array} \right\} \quad (4.23)$$

clasificaremos la observación dentro de la primera categoría (zona 0); si

$$\left\{ \begin{array}{l} y < x \\ y > 1 + \frac{(c-1)}{c}x \end{array} \right\} \text{ ó } \left\{ \begin{array}{l} P(2) < P(1) \\ P(2) > 1 + \frac{c-1}{c}P(1) \end{array} \right\} \quad (4.24)$$

la clasificaremos en la segunda (zona 1) y, si

$$\left\{ \begin{array}{l} y > x \\ y > \frac{c}{c-1}(x-1) \end{array} \right\} \text{ ó } \left\{ \begin{array}{l} P(2) > P(1) \\ P(2) > \frac{c}{c-1}[P(1)-1] \end{array} \right\} \quad (4.25)$$

la clasificaremos en la tercera (zona 2).

La variación de c permitirá, por lo tanto, optimizar los ratios de clasificación.

Terminamos este punto señalando que el criterio establecido por defecto en los programas informáticos econométricos, en el caso de lógit multinomiales, es el de $c=1/3$, baricentro del triángulo utilizado en la demostración. Para los modelos anidados, la aplicación del proceso desarrollado se realizaría teniendo en cuenta las probabilidades finales (calculadas según la expresión 4.17).

El desarrollo presentado en la modelización de elección entre alternativas nos ha de permitir obtener conclusiones en orden al primero de los objetivos fijados al final del epígrafe 3.4: la cuantificación de la probabilidad de que un determinado asegurado cometa fraude¹⁵. Además, se ha establecido un mecanismo de causalidad a través del cual dicha probabilidad depende de un conjunto de factores explicativos.

Con el procedimiento que acabamos de generar, podemos centrarnos ya en el segundo de los objetivos: cuantificar la probabilidad de que el fraude sea detectado por la entidad¹⁶. La generación de una tabla de predicción (en base al uso de

¹⁵ Estimación, por tanto, del parámetro t que aparece en la formulación de Picard (1996).

¹⁶ Estimar, por tanto, el parámetro q de la formulación de Picard (1996).

técnicas de predicción *ex-post*¹⁷) permite determinar el porcentaje de casos que son correctamente clasificados por el modelo (porcentaje de casos fraudulentos y no fraudulentos correctamente clasificados). En base a la relativización al total de observaciones, estaremos en condiciones de obtener una estimación agregada para la probabilidad de detectar el fraude y, por lo tanto, podremos aproximar, dado que conocemos todos los parámetros, la utilidad que el individuo asegurado espera obtener.

No obstante, los resultados obtenidos en relación a la probabilidad de detección de fraude deberán ser objeto de una serie de matizaciones, puesto que vendrán condicionados por el modelo que acabemos seleccionando para cuantificar la probabilidad de existencia de fraude. Sobre ello, realizaremos un análisis detallado en el Capítulo 6.

¹⁷ Esta técnica de predicción consiste en seleccionar, del total de la muestra, un conjunto de observaciones para realizar la estimación, utilizando el resto de casos para validar el modelo.

5. APLICACIÓN EMPÍRICA A LA CUANTIFICACIÓN Y DETECCIÓN DE FRAUDE: DESCRIPCIÓN DE LA MUESTRA

5.1 Comentarios generales

La fundamentación teórica presentada en el Capítulo anterior, en relación al tratamiento del fraude, será materializada, a lo largo de los Capítulos 5 y 6, en lo que constituye la aplicación empírica del trabajo realizado.

Los objetivos presentados en las páginas anteriores en relación a la cuantificación de la probabilidad de presencia de fraude en el seguro del automóvil y de su detección serán sujetos a modelización, en base a las técnicas econométricas comentadas en el Capítulo 4, bajo la aplicación de una muestra, obtenida con la colaboración de la primera entidad aseguradora del mercado asegurador español en automóviles.

La generación de un *“modelo de detección y control de fraude en el seguro automovilístico español”* gozará, por lo tanto, de la particularidad que le confiere el trabajar con datos de una única compañía. No obstante, el funcionamiento más o menos análogo, observado entre entidades, podría permitirnos generalizar los resultados obtenidos, de cara a presentar el comportamiento observado para los principales indicadores de fraude a nivel español.

La disponibilidad de un amplio número de variables en relación a la ocurrencia del siniestro, a las características de la póliza y a determinados atributos de la parte asegurada y de la parte contraria ha permitido realizar un estudio exhaustivo de

cuáles son los principales condicionantes que pueden incrementar la probabilidad de aparición de fraude. Asimismo, la creación de variables adicionales, definidas en base a las anteriores, ha contribuido a detectar situaciones directamente relacionadas con la aparición de comportamientos fraudulentos, tales como, la ocurrencia del siniestro en fin de semana o con anterioridad a la fecha de emisión de la póliza.

La aplicación empírica realizada atiende al siguiente esquema de presentación. A continuación, dentro del Capítulo 5, se detalla, de forma exhaustiva, el contenido de la muestra de siniestros utilizada en la modelización. Además de realizar los comentarios oportunos sobre el tamaño y el carácter aleatorio o no de la misma, se presentan las variables que la componen y se describen estadísticamente. En especial aquéllas que se utilizarán directamente en los modelos presentados (en la mayoría de los casos, las variables seleccionadas no aparecerán directamente en la base de datos sino que habrán sido creadas a partir de la información contenida en la misma).

En el Capítulo 6 se presentan los resultados obtenidos al realizar la modelización bajo la aplicación de modelos lógit multinomiales y anidados. En ambos casos, el objetivo es cuantificar la probabilidad de aparición de fraude y la probabilidad de que la compañía lo detecte. Asimismo, se detallan aquellas variables que resultan estadísticamente significativas de la aparición de comportamientos fraudulentos y se analiza la calidad del ajuste y capacidad predictiva de los modelos presentados.

Por último, en el Capítulo 7 se detallará el resumen de las principales conclusiones extraídas a lo largo del libro, tanto desde un punto de vista teórico como desde un punto de vista aplicado.

5.2 Descripción de la base de datos utilizada en la modelización

5.2.1 Naturaleza de los datos

La muestra utilizada en la estimación de los modelos ha sido obtenida a partir de una selección aleatoria de siniestros de la principal entidad del mercado asegurador español en autos.

La información proporcionada por la misma hace referencia a un total de 1995 expedientes de siniestros, todos ellos acaecidos entre 1993 y 1996, habiendo sido la mitad (998 expedientes) clasificados como legítimos y la otra mitad (997 expedientes) clasificados como fraudulentos por la compañía. Los expedientes fraudulentos son en nuestro caso siniestros donde se ha comprobado la existencia real de fraude (por ejemplo, por propia admisión) y no se reducen a expedientes sospechosos como en otros trabajos. En relación a estos últimos, el tipo de fraude identificado por la entidad atiende a la siguiente clasificación:

- Falsa declaración del asegurado para eludir casos excluidos en la póliza (246 casos).
- Falsa declaración del asegurado para obtener un beneficio sin intervención de un tercero (53 casos).
- Versiones cruzadas para cobrar ambos implicados (9 casos).
- Contratación de la póliza después de ocurrido el accidente (30 casos).
- Ocultación de alcoholemia (16 casos).
- Falso conductor habitual para eludir los recargos (10 casos).
- Fraude del taller (12 casos).
- Falsa declaración del asegurado para favorecer a un tercero (314 casos).
- Identificación de fraude sin disponibilidad de tipo (307 casos).

La información proporcionada posee cobertura nacional, en el sentido de que recoge siniestros ocurridos en prácticamente todas las provincias españolas. Asimismo, hace referencia a una amplia gama de coberturas aunque para su tratamiento éstas han sido englobadas en tres categorías genéricas¹: “a terceros”, “a terceros más complementarios” y “todo riesgo” (con o sin franquicia). Los diferentes tipos de vehículos quedan también suficientemente representados y han sido clasificados en cuatro grupos fundamentales²: “turismos uso particular”, “turismos otros usos”, “motocicletas” y “camiones”.

Los comentarios presentados ponen de manifiesto la disponibilidad de información muestral de elevada calidad y queda suficientemente justificada su aplicación para realizar un correcto tratamiento del fraude.

El análisis de la naturaleza de los datos revela, sin embargo, un hecho significativo: la existencia en la muestra de una sobre-representación para los siniestros fraudulentos. La corrección de este problema se realizará mediante la adecuada ponderación de las observaciones. La disponibilidad de información para los diferentes tipos de fraude nos permitirá modelizar la probabilidad de aparición de éste atendiendo a las diversas formas de comportamiento al alcance del asegurado defraudador. Además, seremos capaces de analizar la significación estadística de determinadas variables en relación a cada uno de los tipos de fraude presentados (normalmente, y con el objetivo de evitar una excesiva categorización de los tipos de fraude, éstos serán agrupados en clases más genéricas, siguiendo la clasificación planteada en los árboles de decisión presentados en el Capítulo 2, epígrafe 2.6).

¹ La clasificación inicial permite diferenciar entre un amplio conjunto de coberturas relacionadas con alternativas variadas de contratación de la póliza (normalmente relacionadas con la contratación del seguro obligatorio de responsabilidad civil y de coberturas adicionales).

² La clasificación inicial permitía diferenciar una gran variedad de posibilidades, atendiendo al uso del vehículo (particular o no), al tonelaje de carga para vehículos comerciales, al número de plazas para vehículos de transporte de personas, a la distancia en kilómetros recorrida de forma habitual por vehículos de transporte en general,...

5.2.2 Descripción de variables

La información contenida en la base de datos hace referencia a un total de 88 variables agrupadas, básicamente, en 9 bloques identificadores (atendiendo a la información suministrada):

- **Variables relacionadas directamente con la identificación del siniestro** (número de expediente, año de fabricación del vehículo³, fecha de ocurrencia del siniestro, número de expedientes⁴ que la compañía decide abrir para tramitar las diferentes consecuencias de un siniestro, número de siniestros asociados a la póliza, y culpa del accidente).
- **Variables relacionadas con la póliza** (coberturas contratadas, tipo-uso del vehículo, código del agente de seguros, situación de la póliza en el momento del siniestro, fecha de vencimiento de recibo, fecha de efecto y de emisión de la póliza, número de suplementos⁵ que ha tenido la póliza desde su emisión, existencia y cantidad de franquicia contratada, existencia y cantidad de cobertura por accesorios, clase de responsabilidad civil contratada, forma de pago de los recibos, marca del vehículo asegurado y color del mismo).
- **Variables relacionadas con el siniestro** (hora de ocurrencia del siniestro, provincia de ocurrencia del mismo, presencia o no de testigos, intervención o no de la policía, naturaleza del siniestro⁶, relato del accidente por el asegurado, daños al vehículo asegurado, código del taller del asegurado y fecha de comunicación del siniestro).

³ O de matriculación, si no se conociera la fecha de fabricación del vehículo.

⁴ Asociados a diferentes conceptos, como daños propios, responsabilidad civil, recobros, reclamación de daños y/o lesiones.

⁵ Se entiende por "suplemento" cualquier cambio que afecte a las condiciones de la póliza (coberturas, vehículo, conductor, forma de pago, accesorios,...).

⁶ Ocurrencia del siniestro en zona urbana o en carretera.

- **VARIABLES RELACIONADAS CON EL ASEGURADO-TOMADOR DEL SEGURO** (lugar de residencia⁷, sexo, estado civil, fecha de nacimiento y antigüedad de carnet).
- **VARIABLES RELACIONADAS CON EL PROPIETARIO DEL VEHÍCULO ASEGURADO** (información análoga a la presentada para el tomador del seguro).
- **VARIABLES RELACIONADAS CON EL CONDUCTOR DEL VEHÍCULO ASEGURADO** (información análoga a la presentada para el tomador y para el propietario del vehículo asegurado).
- **VARIABLES RELACIONADAS CON EL CONDUCTOR CONTRARIO** (código de la compañía contraria, marca y color del vehículo contrario, daños ocasionados al mismo, taller de reparación, identificador de coincidencia entre apellidos de la parte asegurada y la contraria, lugar de residencia⁸, sexo, estado civil, fecha de nacimiento y antigüedad de carnet).
- **VARIABLE IDENTIFICADORA DEL TIPO DE FRAUDE EXISTENTE.**
- **VARIABLE QUE MIDE EL COSTE FINAL DEL SINIESTRO** (indicándose, en cada caso, si se trata de pagos o de recobros⁹).

El estudio descriptivo de la información contenida en la muestra disponible pone de manifiesto la presencia de un elevado número de variables de carácter cualitativo que han sido objeto de la codificación oportuna. Las variables de carácter cuantitativo son menos numerosas dentro de la base de datos aunque quizá el propio funcionamiento del sistema asegurador, en términos de la información normalmente recogida, ayude a justificar este hecho. Es muy frecuente la aparición de variables que recogen información sobre fechas (fecha

⁷ Únicamente se dispone del código postal.

⁸ De nuevo, indicado por el código postal.

⁹ La aplicación de los Convenios C.I.D.E. y/o A.S.C.I.D.E. implica que la compañía pague a su asegurado los daños derivados del siniestro y recobre de la compañía contraria un módulo fijo (que puede superar lo pagado).

de ocurrencia del siniestro, fecha de comunicación del mismo,...) siendo su tratamiento una de las partes más elaboradas dentro del estudio que presentamos, dado que ha permitido examinar cada uno de los siniestros dentro de las reglas, que en términos de plazos, rigen en el mercado de seguros español.

Los grupos de edad y de antigüedad de carnet para los conductores asegurados, en el momento del siniestro, quedan representados de forma correcta dentro de la muestra utilizada, observándose intervalos de variación para los mismos suficientemente amplios. De esta forma, la edad del conductor asegurado oscila entre los 16 y los 83 años y la antigüedad de carnet entre los 0 y los 56 años.

En relación a la antigüedad del vehículo, los datos disponibles reflejan la ocurrencia de siniestros que involucran tanto a vehículos nuevos como a vehículos antiguos (la antigüedad del vehículo en el momento del siniestro oscila entre 0 y 37 años). También y, en relación al siniestro, es posible observar la existencia de culpa admitida tanto por parte del asegurado como por parte del contrario.

El intervalo horario y las fechas de ocurrencia del accidente (considerando el día de la semana en que se produjo) permiten modelizar en términos de horas de acaecimiento de los sucesos y de la mayor o menor relación de los mismos con el fin de semana.

El relato del siniestro por parte del asegurado, aunque puede ser breve, es muy informativo. Por ello, es posible realizar un análisis estadístico textual que permita señalar cuáles son las situaciones o circunstancias que tienen un mayor grado de asociación con las declaraciones fraudulentas.

Todas las pólizas se encontraban dadas de alta en el momento del accidente, y las diferentes formas de pago de los recibos (anual, semestral o trimestral) quedan reflejadas en la información facilitada.

En relación al sexo del conductor asegurado, es necesario hablar de una mayor proporción de hombres (1672 hombres frente a 323 mujeres) aunque no cabe esperar una sobre-representación de los mismos puesto que ello ya se espera a nivel poblacional. En relación al estado civil del asegurado, la muestra evidencia casos para las cuatro categorías posibles (casado, soltero, separado y viudo). Para el conductor contrario y, en relación a estas dos variables, se observa una clara ausencia de información.

Atendiendo a las consideraciones que aparecen en la literatura existente (Cobo, 1993; Weisberg y Derrig, 1993; Belhadji y Dionne, 1997) sobre la presencia de determinadas circunstancias (o indicadores) que alertan sobre la existencia de fraude y en base a los resultados que hemos obtenido al realizar un estudio exhaustivo de las características propias de los siniestros con fraude detectado, las variables originales han sido debidamente seleccionadas con el objetivo de crear otras nuevas¹⁰ que permitan modelizar, de forma adecuada, la presencia o no de comportamientos fraudulentos dentro de los expedientes analizados.

Las variables creadas son fundamentalmente dicotómicas y quedan definidas a continuación (omitimos el subíndice *i* que indicaría su referencia a cada individuo de la muestra),

¹⁰ Variables que recogen situaciones comúnmente asociadas a la presencia de fraude.

- **Variables relacionadas con la póliza y el vehículo en el momento del contrato**

DRAMO: variable dicotómica que toma valor 1 si la cobertura contratada es a terceros (con o sin complementarios) y valor 0 si es a todo riesgo (con o sin franquicia).

DUSO: variable dicotómica que toma valor 1 si el vehículo asegurado es un turismo de uso particular y valor 0 en caso contrario (si se encuentra dentro de las categorías de “turismo otros usos”, “motocicletas” o “camiones”).

SFRANQUI: variable dicotómica que toma valor 1 si existe franquicia en la póliza; toma valor 0 en caso contrario.

ACCESORI: variable dicotómica que toma valor 1 si en la póliza existe cobertura de accesorios; toma valor 0 en caso contrario.

DPAGO1: variable dicotómica que toma valor 1 si el pago de los recibos es semestral; toma valor 0 si es anual o trimestral.

DPAGO2: variable dicotómica que toma valor 1 si el pago de los recibos es trimestral; toma valor 0 si es anual o semestral.

- **Variables relacionadas con el asegurado en el momento del siniestro**

HISTSTR: número de siniestros asociados a la póliza anteriores al estudiado.

EDAD: edad del conductor asegurado en el momento del siniestro.

CARNET: años de antigüedad de carnet del conductor asegurado en el momento del siniestro.

DSEX: variable dicotómica que toma valor 1 si el conductor asegurado es hombre.

DECC: variable dicotómica que toma valor 1 si el conductor asegurado está casado; toma valor 0 en caso contrario (soltero, separado o viudo).

- **Variables basadas en la relación con el conductor contrario**

DESTAF: variable dicotómica que toma valor 1 si existe coincidencia de apellidos entre la parte asegurada en la entidad (asegurado/ propietario/ conductor) y la parte contraria; vale 0 en caso contrario.

DDOMICIL: variable dicotómica que toma valor 1 si existe proximidad de domicilios¹¹ entre el conductor asegurado y el contrario; vale 0 en caso contrario.

- **Variables relacionadas con el siniestro y el vehículo en el momento del accidente**

DCULPA: variable dicotómica que toma valor 1 si la culpa es del asegurado y valor 0 si la culpa es del contrario.

ANTIGUO: antigüedad (en años) del vehículo en el momento del siniestro.

¹¹ La ausencia de información en relación a domicilios exactos (calle,...) ha supuesto la necesidad de analizar dicha proximidad atendiendo únicamente a los códigos postales de residencia declarados por el conductor asegurado y por el contrario, con todas las matizaciones que ello puede llevar asociado.

DCOMOCUR: variable dicotómica que toma valor 1 si la comunicación del siniestro se produce con posterioridad al séptimo día de su ocurrencia¹²; toma valor 0 en caso contrario.

ALTAHORA: variable dicotómica que toma valor 1 si el siniestro ha ocurrido a altas horas (entre las 11 de la noche y las 5 de la madrugada); vale 0 en caso contrario.

DZONA1: variable dicotómica que toma valor 1 si el siniestro ha ocurrido en zonas de siniestralidad alta¹³; toma valor 0 si ha ocurrido en zonas de siniestralidad media o baja.

DZONA3: variable dicotómica que toma valor 1 si el siniestro ha ocurrido en zonas de siniestralidad baja; toma valor 0 si ha ocurrido en zonas de siniestralidad alta o media.

CARRET: variable dicotómica que toma valor 1 si el siniestro ha ocurrido en carretera; vale 0 si el siniestro ha ocurrido en zona urbana.

SABDOM: variable dicotómica que toma valor 1 si el siniestro ha ocurrido en sábado o domingo; vale 0 en caso contrario.

DEFOCEMI: variable dicotómica que toma valor 1 si la fecha de ocurrencia del siniestro es anterior o igual a la fecha de emisión de la póliza y

¹² La ley de Contrato del Seguro en su Art. 16 señala: "el tomador del seguro o el asegurado o el beneficiario deberán comunicar al asegurador el acaecimiento del siniestro dentro del plazo máximo de siete días de haberlo conocido, ...".

¹³ Atendiendo a la clasificación de zonas de siniestralidad en España presentada por U.N.E.S.P.A. (1993a). Teniendo en cuenta, entre otras cosas, factores climáticos y de infraestructuras, se diferencia entre:

- Zona 1 (elevada siniestralidad): Galicia, Asturias, Cantabria, País Vasco y Navarra.
- Zona 2 (siniestralidad media): Madrid y Cataluña.
- Zona 3 (siniestralidad baja): Resto de Comunidades, Ceuta y Melilla.

posterior o igual a la fecha de efecto; toma valor 0 en caso contrario.

STEST: variable dicotómica que toma valor 1 si hay testigos del siniestro; vale 0 en caso contrario.

DRELATO: variable dicotómica que toma valor 1 si en la declaración del asegurado aparecen alguno de los relatos relacionados con: estacionamiento, aparcamiento, marcha atrás, adelantamiento y/o cruce; toma valor 0 en caso contrario.

SAUTDAD: variable dicotómica que toma valor 1 si hay intervención de policía en el siniestro; toma valor 0 en caso contrario.

5.3 Estadísticos descriptivos básicos

El análisis descriptivo de los datos refleja la existencia de variables en las que no se observan grandes diferencias entre las submuestras de fraude y no fraude. De este modo, por ejemplo, la media de la variable edad para los dos grupos considerados no muestra diferencias significativas (37 y 39 años). Lo mismo podría señalarse para las variables que recogen los años de antigüedad de carnet del conductor asegurado o la antigüedad del vehículo en el momento del siniestro. Las frecuencias relativas para variables como la existencia de testigos o la forma de pago semestral de la póliza reflejan un comportamiento análogo. De esta forma y, aunque deberíamos ejecutar algún tipo de contraste (por ejemplo, de medias o proporciones) que nos permitiese ratificar lo anterior, parece lógico pensar que el estudio de la influencia que las variables presentadas tienen sobre la elección de defraudar se ha de realizar considerándolas de forma conjunta. La aplicación de técnicas univariantes que recogiesen el efecto de las variables una a una, impediría obtener resultados significativos para aquellos indicadores con comportamientos similares en ambas submuestras. Por lo tanto, la modelización

a aplicar ha de estar ligada a técnicas multivariantes que recojan el efecto conjunto entre las variables consideradas.

El análisis descriptivo revela como determinadas situaciones, asociadas, por ejemplo, a la ocurrencia del siniestro en fecha anterior o igual a la fecha de emisión de la póliza y posterior o igual a la fecha de efecto (variable *DEFOCEMI*), la comunicación del siniestro fuera del plazo establecido en la Ley de Contrato del Seguro (variable *DCOMOCUR*) y/o la ocurrencia del accidente a altas horas de la noche/madrugada (variable *ALTAHORA*) son más frecuentes en los casos en los que se ha detectado fraude. El mismo comportamiento se observa para la variable que recoge la aparición en la declaración del asegurado de relatos típicos como estacionamiento, marcha atrás,... (variable *DRELATO*). La coincidencia de apellidos entre la parte asegurada y la contraria (variable *DESTAF*) y la aceptación de culpa por parte del asegurado (variable *DCULPA*) presentan el mismo comportamiento.

A diferencia de las anteriores la intervención de policía en el siniestro (variable *SAUTDAD*) muestra una mayor frecuencia relativa para los casos de ausencia de fraude.

No queremos relegar a un segundo plano el tratamiento textual de la variable que explica cómo ocurrió el siniestro. Aunque es una información que aporta el propio asegurado y por tanto su fiabilidad puede quedar cuestionada, tiene una característica esencial. Es una de las pocas magnitudes sobre las que el asegurado puede ejercer algún tipo de manipulación. Muchas variables son observables directamente por la compañía o son más fácilmente comprobables o, simplemente, no permiten la flexibilidad que aporta una descripción verbal o escrita de lo ocurrido.

Para abordar el análisis de las respuestas textuales que aparecen en los expedientes hemos realizado una aproximación estadística básica al estudio de

este tipo de datos, si bien somos conscientes de que un tratamiento textual más avanzado tendría mucho interés. Con él sería posible buscar asociaciones entre palabras e incluso poder plasmar distancias entre conceptos y proximidades de ciertos relatos a la presencia y ausencia de fraude. Nuestro análisis, si bien preliminar, se ha reducido a tener en cuenta la frecuencia de aparición de ciertos términos. Para ello, se ha considerado la raíz semántica de la palabra y el contexto de su aparición. Este tratamiento es ciertamente laborioso, y ha permitido identificar ciertos relatos que se asocian con mayor frecuencia a la presencia de fraude. Esta aproximación es muy novedosa puesto que los estudios empíricos conocidos hasta la fecha no realizan ningún análisis de este tipo. Sin embargo, sí incluyen variables que intentan recoger el mismo tipo de información.

Es usual en otros estudios valorar la actitud del asegurado al realizar la reclamación y suele ser normal encontrar variables que indican si el mismo coopera con la entidad, es excesivamente explícito e incluye muchos detalles en su declaración o incluso, muestra mucha habilidad en el uso de la terminología utilizada en la tramitación, dando a entender que conoce a la perfección el funcionamiento del mercado asegurador (y, por tanto, es más fácil que sepa como actuar deshonestamente). La principal crítica que han recibido estas variables es su elevado grado de subjetividad puesto que dependen de la valoración que realiza el tramitador. En nuestro caso, no hemos podido disponer de ninguna valoración respecto a actitudes y comportamientos del asegurado al efectuar la declaración (excepto la dilación entre la ocurrencia del siniestro y su comunicación). Sin embargo, creemos que la utilización de los relatos del asegurado en la propia declaración aporta elementos informativos sobre su forma de proceder.

Igualmente se ha incluido el indicador de coincidencia entre los apellidos de las partes implicadas como variable potencialmente explicativa de circunstancias sospechosas.

6. APLICACIÓN EMPÍRICA: MODELIZACIÓN LOGÍSTICA MULTINOMIAL Y ANIDADA

6.1 Introducción

Las consideraciones que hemos ido estableciendo a lo largo del libro en relación al estudio y tratamiento del fraude en el seguro del automóvil, son materializadas a lo largo de este Capítulo en la determinación de un conjunto de modelos que permitan cuantificar su probabilidad de existencia.

Como hemos establecido, el asegurado realiza una elección entre alternativas discretas, persiguiendo, en cualquiera de los casos, maximizar su utilidad esperada. Sin embargo, el tratamiento econométrico susceptible de aplicación es variado y está sujeto a la disponibilidad de información adecuada.

En la generación de un modelo de detección y control del comportamiento fraudulento es necesario tener en cuenta un hecho importante: poblacionalmente hablando, los niveles existentes de fraude son de difícil cuantificación. Las divergencias presentadas en el Capítulo 2 (apartado 2.4) en relación a los porcentajes estimados de fraude (para diferentes países y ramos) y la poca claridad con que éstos son presentados por los diferentes autores (a veces es difícil diferenciar entre si los porcentajes son sobre fraude detectado o sobre sospecha de fraude) dificulta la obtención de valores poblacionales fiables.

Atendiendo a lo anterior, la determinación del tamaño muestral óptimo que asegure la representatividad de la población puede estar sujeta a matizaciones, más aún, cuando nada garantiza que en muchos de los siniestros que las

compañías tienen calificados como de no fraudulentos puede existir fraude subyacente¹. No obstante, la generación de una muestra, como la analizada en el Capítulo anterior, en la que se dispone de información suficiente para expedientes de siniestros sin fraude detectado y para expedientes fraudulentos nos ha permitido desarrollar una aproximación innovadora al tratamiento del mismo. El criterio de selección muestral seguido en el diseño y recogida de información y, la introducción de ponderaciones que adecuen, en la medida de lo posible, el comportamiento muestral al poblacional serán factores importantes a tener en cuenta en la estimación de los modelos.

La modelización del fraude por parte del asegurado admite, desde la óptica de los modelos de elección², una doble aproximación:

- especificación de modelos de regresión logística simple,
- especificación de modelos lógit multinomiales y anidados.

La primera y más sencilla supone definir como variable dependiente del modelo la dicotomía ausencia/presencia de fraude. El estudio se realiza en sentido genérico (sin diferenciar entre tipologías de fraude) y los resultados, en términos de significación estadística de las variables, permiten determinar qué circunstancias tienen un mayor poder explicativo en la aparición de actitudes deshonestas.

El uso de modelos lógit multinomiales y anidados supone no sólo trabajar con una metodología más avanzada sino también, obtener conclusiones en relación a las diferentes formas de comportamiento que puede presentar el asegurado defraudador. Los tipos de fraude son variados (como veíamos en el Capítulo anterior, la variable SUBFRAU presenta hasta nueve valores diferentes asociados

¹ Sólo los siniestros que han sido investigados exhaustivamente por la entidad en los que se ha reconocido la acusación o ha existido condena responden a una clasificación correcta en términos de fraude. Para los que se clasifican como no fraude cabría dudar sobre si la investigación ha sido suficiente.

² Modelos de respuesta cualitativa o de elección individual

a comportamientos fraudulentos distintos³ -fraude a favor del asegurado, a favor del contrario, del taller,...-) y la determinación de las variables estadísticamente significativas en la explicación de cada uno de ellos puede ser decisiva a la hora de establecer patrones de investigación por parte de la compañía. En definitiva se trata de “matizar” y de proponer la metodología correspondiente para distinguir formas de fraude (aspecto, el de la existencia de diversos tipos de comportamientos deshonestos, sobre el que existe un amplio consenso entre los especialistas). Además, como veremos, la capacidad predictiva del modelo (fundamentada en el porcentaje de casos correctamente clasificados por el mismo⁴) ha sido mejorada gracias al diseño, dentro del presente trabajo (Capítulo 4), de un método que permita seleccionar el punto óptimo de corte en modelos multinomiales.

El hecho de disponer únicamente de información relativa al individuo (asegurado/contrario), a la póliza y al siniestro y no de la elección (características propias de cada uno de los posibles comportamientos) supondrá la aplicación a priori de modelos *lógit* multinomiales propiamente dichos, en contraposición a los denominados *lógit* condicionales⁵ (en éstos se tienen en cuenta, además, atributos específicos de cada una de las alternativas sobre las que versa el proceso de decisión). La conversión de determinadas variables en atributos específicos de una elección concreta permitirá la aplicación de modelización logística anidada.

En este Capítulo nos centraremos en el análisis de resultados obtenidos tras la aplicación de un modelo *lógit* multinomial, contrastándolos con los derivados al aplicar modelos jerárquicos o anidados. Las principales conclusiones derivadas de la aplicación de un *lógit* simple pueden encontrarse en Artís, Ayuso y Guillén (1998a).

³ No obstante, cabría hablar únicamente de ocho clases de fraude recogidas en la muestra dado que la última categoría recoge casos fraudulentos sin identificación de tipo.

⁴ Porcentaje de casos no fraudulentos y fraudulentos (teniendo en cuenta la tipología existente) correctamente clasificados como tales.

⁵ *Modelo lógit condicional de McFadden* (Maddala, 1983).

6.2 Criterio de selección muestral. Determinación de las ponderaciones a introducir en los modelos

La información muestral que hemos utilizado en la estimación de los modelos logísticos, descrita ampliamente en el Capítulo anterior, proviene de una selección de siniestros declarados aleatoria estratificada con estratificación endógena ("muestreo basado en la elección"). Desde este punto de vista, su diseño no atiende a los criterios propios de muestreo aleatorio simple en sentido estricto, sino que viene fundamentado en técnicas de recogida de información basadas en la variable dependiente del modelo⁶. De esta forma, la probabilidad de que un individuo esté en la muestra depende de la alternativa que haya elegido.

El muestreo basado en la elección es útil cuando la realización de una recogida de información completamente aleatoria o basada en alguna de las variables explicativas (o todas) incluidas en el modelo, sólo consigue una escasa representatividad, ya que deriva en la obtención de un número insuficiente de individuos eligiendo una alternativa particular. En la literatura existente (Maddala, 1983; Amemiya, 1985; Pudney, 1989) se hace referencia a esta técnica de muestreo aludiendo al diseño de una muestra relacionada con decisiones de transporte. A pesar de que una de las ventajas manifestada por los autores es la económica (recoger información en las paradas de autobús o de metro es menos costoso que hacer entrevistas en las casas), el hecho de que gracias a la misma es posible alcanzar una cuota adecuada de representación de todas las alternativas la hace más atractiva.

⁶ Amemiya (1985) diferencia básicamente entre tres tipos de técnicas de muestreo: *muestreo aleatorio*, *muestreo exógeno* (realizado en base a las variables independientes o explicativas) y *muestreo endógeno* (según la variable dependiente). Esta última técnica se conoce, dentro de los modelos de respuesta cualitativa, como "muestreo basado en la elección (*choice-based sampling*)".

A pesar de la dificultad comentada en la introducción de este Capítulo en relación a trabajar con porcentajes poblacionales de fraude fiables, parece lógico pensar que el número de expedientes de siniestros sin fraude sea notablemente superior al de fraudes. Las compañías, hasta el momento, no han realizado políticas agresivas de detección de fraude y, por ello, la obtención de información en base a muestreo aleatorio hubiera derivado, probablemente, en una escasa presencia de expedientes de este tipo. Es por ello que la recogida de datos se realizó mediante una estratificación endógena: se seleccionaron aleatoriamente entre los expedientes sin fraude (o al menos sin fraude detectado) un total de 1000 casos y de los expedientes clasificados como fraudulentos y, también aleatoriamente, un total de 1000. La selección aleatoria de estos últimos ha supuesto una notable variación en el número de casos clasificados en los diferentes tipos de fraude. Finalmente se rechazaron 2 expedientes sin fraude y 3 con fraude, debido a la incongruencia en la información que suministraban.

Ahora bien, ¿qué implicación tiene el hecho de trabajar con una muestra formada por un 50% de siniestros de cada clase?. Si retomamos lo mencionado al inicio del párrafo anterior, dando como válido el hecho de que poblacionalmente cabe esperar que el número de no-fraudes sea muy superior al de fraudes, debemos hacer algún tipo de corrección con el objetivo de ajustar los resultados obtenidos tras la modelización al comportamiento de la población. La introducción de pesos o ponderaciones dentro del modelo permitirá corregir la infra-representación de los no fraudes y la sobre-representación de los fraudes.

¿Cómo determinaremos los pesos a introducir en el modelo?.

La muestra ha sido diseñada deliberadamente para contener un 50% de casos sin fraude y un 50% de casos fraudulentos mientras que las proporciones en la población (tomando como referencia las cifras presentadas por Cobo, 1993) son del 78% y 22%, respectivamente. De esta forma, los casos de no fraude están infra-representados por un factor de 0.64 ($=0.5/0.78$) mientras que los de fraude

aparecen sobre-representados por un factor de 2.27 ($=0.5/0.22$). Para obtener una representación adecuada en la población será necesario, por tanto, ponderar las observaciones de no fraude por un factor de 1.56 ($=0.78/0.50$) y las de fraude por 0.44 ($=0.22/0.50$).

Lógicamente, la consideración de diferentes tipos de fraude supondrá modificar adecuadamente los factores de ponderación. Si bien los pesos anteriores serán los utilizados en la especificación y estimación del modelo lógit simple (modelizando directamente la probabilidad de presencia de fraude versus no fraude), la agrupación de las diferentes clases de comportamientos fraudulentos en categorías más genéricas supondrá trabajar con diferentes tamaños muestrales y por lo tanto con diferentes múltiplos del factor de ponderación (construidos, no obstante, bajo el criterio presentado en el párrafo anterior).

La aplicación de modelos lógit multinomiales se ha realizado teniendo en cuenta el diseño previo de dos esquemas de elección que se presentan en las Figuras 8 y 9. En el primero de ellos, el asegurado ha de elegir entre tres posibles alternativas: no cometer fraude; cometer fraude a favor de sí mismo o hacerlo a favor del contrario.



Figura 8

En el segundo, y bajo el objetivo de recoger de algún modo la posible intencionalidad en la creación del siniestro, las alternativas modelizadas son: no

cometer fraude; cometer fraude para conseguir una indemnización o hacerlo para incrementar la indemnización que corresponde en función de la póliza contratada.



Figura 9

Lógicamente, el objetivo perseguido es modelizar la probabilidad de que el individuo cometa fraude atendiendo, no a todas y cada una de las formas posibles que tiene de hacerlo, sino a criterios generales fundamentados en el propio sistema asegurador.

El análisis de la información disponible para los tipos de fraude pone de manifiesto, tal y como hemos analizado en el Capítulo anterior, la existencia de nueve formas diferentes de clasificar los expedientes en los que la compañía ha detectado existencia de comportamiento fraudulento. Modelizar por separado los posibles comportamientos detectados hubiera derivado en un modelo con un elevado número de parámetros, más aún cuando los pocos casos englobados en algunas de las categorías hubiera impedido extrapolar poblacionalmente algunos de los resultados obtenidos. El objetivo perseguido se centra en justificar determinados patrones de comportamiento, analizando qué circunstancias los favorecen. Englobar categorías en otras más genéricas no supone ningún impedimento, sobre todo, cuando a veces resulta difícil comprender cuál es la razón que lleva al tramitador a englobar el siniestro dentro de un tipo u otro de fraude.

¿Cuál ha sido el criterio seguido a la hora de diseñar las alternativas finales?.

El seguro del automóvil español se fundamenta en dos aspectos de gran importancia a la hora de intentar dar un enfoque al estudio del fraude. Por un lado, las coberturas contratadas, si bien pueden estar relacionadas con la protección de daños sobre el propio vehículo, siempre tienen en cuenta la reparación de daños a un tercero⁷. Por otro, el hecho de que la mayoría de entidades españolas “estén en convenio⁸” facilita la posible colaboración entre las partes (asegurado y contrario). La eliminación de procesos judiciales y la rapidez con que se han de pagar las indemnizaciones son, sin duda, factores que favorecen los comportamientos deshonestos o al menos dificultan su detección.

En base a lo anterior parece lógico realizar una agrupación de tipos de fraude como la presentada en el primer esquema (Figura 8). La clasificación presentada en el segundo (Figura 9) pretende diferenciar, básicamente, entre aquellos casos en los que se persiguen evitar exclusiones de cobertura (el siniestro se produce bajo determinadas circunstancias que quedan fuera de las condiciones de la póliza) y aquéllos en los que el objetivo es aumentar la indemnización correspondiente (el siniestro está cubierto pero el asegurado quiere obtener un beneficio no justificado). En cierto modo, esta clasificación guardará analogías con la utilizada por Weisberg y Derrig (1993) y que diferencia entre fraude *a priori* (intencionalidad en la creación del accidente) y fraude *a posteriori* (uso indebido de la ocurrencia real del siniestro). No obstante, el hecho de que utilizando la información disponible no podamos diferenciar con exactitud en qué casos el siniestro ha sido creado premeditadamente o en qué casos se ha producido bajo situaciones excluidas impide establecer comparaciones con los resultados obtenidos por los autores mencionados.

⁷ Siempre teniendo en cuenta la calificación que de “tercero” aparece en el condicionado de la póliza. Por ejemplo, es posible que en el marco de la cobertura de responsabilidad civil de suscripción voluntaria no tengan la consideración de terceros los familiares hasta tercer grado de consanguinidad. Una explicación análoga puede aplicarse, cuando el asegurado sea una persona jurídica, para sus representantes o los familiares de los mismos.

⁸ Aplican convenios (C.I.D.E./A.S.C.I.D.E).

La agrupación de tipos se podría haber realizado bajo la premisa de otras consideraciones. Por ejemplo, sería posible analizar si el fraude ha sido cometido de forma individual o en connivencia. Aún más, si la connivencia se ha realizado con la parte contraria u otros estamentos (talleres, médicos, abogados,...). Sin embargo, para este último caso, no disponemos de información completa en la muestra utilizada.

La categorización del fraude en dos grupos según su pretensión de beneficiar al asegurado o beneficiar a un tercero se ha realizado del siguiente modo. Aquellos casos en los que la compañía ha detectado falsas declaraciones del asegurado para eludir casos excluidos en la póliza (246 expedientes) o para obtener un beneficio sin intervención de un tercero (53 expedientes) han sido utilizados como submuestra de la categoría fraude a favor del asegurado. Aquellos en los que queda de manifiesto, de forma expresa, que el asegurado ha realizado una declaración falsa para favorecer a un tercero (314 casos) determinan la submuestra de fraude a favor de un tercero. El tamaño muestral utilizado en la estimación del primer modelo es, por tanto, de 1611 observaciones (teniendo en cuenta, lógicamente, los 998 casos en los que no se ha detectado comportamiento fraudulento) perdiéndose un total de 384 observaciones.

Para realizar la agrupación de casos teniendo en cuenta si el objetivo del fraude es conseguir una indemnización o incrementar la correspondiente según la póliza contratada hemos seguido el siguiente criterio. El conjunto de expedientes clasificados por la entidad como fraudulentos por poseer falsa declaración del asegurado para eludir casos excluidos en la póliza (246 casos), por haberse contratado la póliza después de haber ocurrido el accidente (30 casos), por detectarse ocultación de alcoholemia (16 casos) o porque se haya declarado un falso conductor habitual para eludir recargos (10 casos) han sido agrupados bajo la categoría genérica de fraude para conseguir una indemnización. Los expedientes en los que se han detectado comportamientos del tipo de falsa declaración del asegurado para obtener un beneficio sin intervención de un

tercero (53 casos), presentación de versiones cruzadas para cobrar el asegurado y el contrario (9 casos) o declaración falsa del asegurado para favorecer a un tercero (314 casos) han determinado la clase de fraude para incrementar la indemnización. Finalmente, la muestra utilizada para realizar la estimación de este segundo modelo está constituida por un total de 1676 casos (319 menos que en la inicial).

Una vez diseñadas las muestras a utilizar en cada uno de los casos, las ponderaciones introducidas en cada uno de los modelos para corregir la sobre-representación de expedientes con fraude han sido las siguientes. En el primero, las observaciones de no fraude han sido ponderadas por un factor de 1.26 ($=0.78/0.62$), las de fraude a favor del asegurado por 0.58 ($=0.11/0.19$) y las de fraude a favor de un tercero también por 0.58 ($=0.11/0.19$). Para los casos fraudulentos y dado que poblacionalmente aceptamos que en un 22% de los casos existe sospecha⁹ de fraude (Cobo, 1993), adoptamos la hipótesis de que un 11% serán de un tipo y el 11% restante del otro. Esta hipótesis está fundamentada en una ausencia de información sobre el comportamiento de ambas clases en la población aunque, en este caso, la gran similitud entre el número de observaciones obtenidas a partir de la muestra para cada una de ellas nos aporta elementos para justificarla.

Para el segundo modelo, las ponderaciones son muy similares. Los casos no fraudulentos han sido ponderados por un factor de 1.3 ($=0.78/0.60$); los casos en los que el motivo del fraude es conseguir una indemnización por 0.61 ($=0.11/0.18$) y aquéllos en los que se persigue incrementar la cantidad a percibir por la ocurrencia del siniestro por 0.5 ($=0.11/0.22$).

⁹ Lógicamente, este porcentaje puede estar sujeto a matizaciones dado que puede ser que exista sospecha de fraude y sin embargo no sea posible demostrar su existencia (Cobo, 1993). No obstante, teniendo en cuenta que a nivel de mercado y a nivel científico es comúnmente aceptada la existencia de un elevado porcentaje de fraudes no detectados (a modo de ejemplo Derrig, Weisberg y Chen (1994) estiman que el porcentaje de siniestros con sospecha de fraude estaría entre el 40 y el 50%) y dado que el porcentaje presentado surge como resultado de un proceso de muestreo creemos justificada su aplicación.

Una vez se ha puesto de manifiesto cuál ha sido el criterio seguido en el diseño de la muestra y se ha justificado el uso de ponderaciones en los modelos, pasamos a continuación a presentar el método de estimación utilizado y los principales resultados obtenidos.

6.3 Estimación logística multinomial de la probabilidad de existencia de fraude a favor del asegurado o a favor de un tercero (MODELO 1)

Presentamos a continuación el análisis detallado de los principales resultados obtenidos para el primer modelo de elección (Figura 8).

6.3.1 Análisis de la significación individual y global de modelo

La especificación del modelo que nos permita cuantificar la probabilidad de que el asegurado no cometa fraude, de que lo haga a favor de sí mismo o para beneficiar a un tercero (Figura 8) se ha realizado teniendo en cuenta una amplio conjunto de variables explicativas, todas ellas presentadas en el Capítulo anterior, y directamente relacionadas con aspectos del propio sistema asegurador (póliza, siniestro,...). El hecho de no disponer en la muestra de atributos específicos para cada una de las elecciones ha sugerido la aplicación de modelos lógit multinomiales. Así, el árbol que presentábamos para este mismo esquema de decisión en el Capítulo 2, en el que teníamos en cuenta una etapa intermedia de decisión,

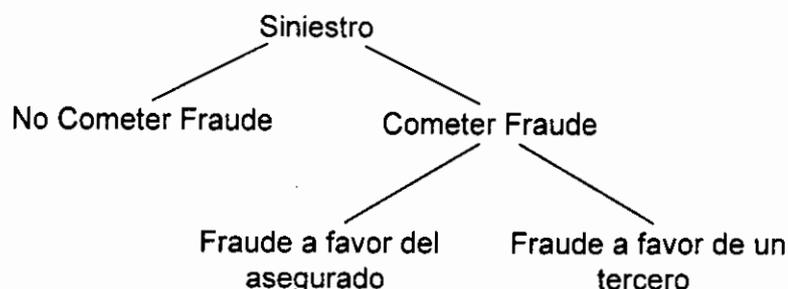


Figura 10

ha sido simplificado, teniéndose en cuenta únicamente las alternativas finales (Figura 8).

Al no disponer de ninguna variable que tome valores diferentes para las dos categorías de fraude presentadas parece lógico pensar en una modelización de este tipo, siendo necesario, no obstante, demostrar que ambas alternativas son independientes desde un punto de vista econométrico (en realidad el esquema de decisión subyacente es el presentado en la Figura 10, y por lo tanto podría suceder que los términos de error de las funciones de utilidad para las alternativas estuvieran correlacionados). El test de Hausman-McFadden nos permitirá, más adelante, demostrar el cumplimiento de la hipótesis de Independencia de Alternativas Irrelevantes.

La variable dependiente es el tipo de siniestro que hemos codificado por: $Y=0$, si el siniestro no es fraudulento; $Y=1$, si el asegurado ha defraudado en beneficio propio e, $Y=2$, si existe fraude para beneficiar a un tercero. Las observaciones de dicha variable han sido ponderadas durante el proceso de estimación para tener en cuenta el diseño muestral. Consideramos la categoría de los siniestros no fraudulentos como la de referencia.

En la especificación del modelo se han introducido diecinueve variables explicativas (incluido el término independiente). Las variables, definidas en el capítulo anterior no se refieren a las elecciones finales, es decir, no cambian con la alternativa elegida. Todas ellas están relacionadas con aspectos de la póliza, del automóvil y del siniestro, siendo la inmediatez su principal característica (esta conclusión podría matizarse para la variable que recoge la culpa del siniestro). La compañía dispone de la información que las caracteriza en un breve periodo de tiempo, lo que sin duda constituye un factor importante a considerar en la agilización de la investigación de siniestros.

Los parámetros estimados aparecen en la página siguiente:

Tabla 1. Resultados de la estimación de un modelo lógit multinomial
[MODELO 1]

variable	coeficiente	t-test	P-valor
Fraude para beneficio del asegurado:			
CONSTANTE	-3.403	-5.357	0.000
DRAMO	1.067	2.516	0.012
DUSO	-0.720	-2.664	0.008
HISTSTR	0.196	4.132	0.000
EDAD	-0.028	-1.987	0.047
CARNET	0.003	0.171	0.864
DESTAF	1.504	4.793	0.000
DCULPA	0.229	1.104	0.270
ANTIGUO	0.029	1.421	0.155
DCOMOCUR	1.324	6.590	0.000
ALTAHORA	1.366	5.199	0.000
CARRET	0.607	1.855	0.064
SABDOM	0.298	1.500	0.134
DEFOCEMI	0.729	0.696	0.486
STEST	-9.862	-15.206	0.000
DZONA1	0.751	2.286	0.022
DZONA3	1.036	4.763	0.000
DRELATO	0.623	3.154	0.002
SAUTDAD	-1.675	-3.600	0.000
Fraude a favor de un tercero:			
CONSTANTE	-4.010	-6.297	0.000
DRAMO	0.801	2.135	0.033
DUSO	-0.384	-1.325	0.185
HISTSTR	0.171	2.981	0.003
EDAD	-0.021	-1.486	0.137
CARNET	0.001	0.022	0.982
DESTAF	0.747	1.785	0.074
DCULPA	2.670	12.760	0.000
ANTIGUO	0.006	0.267	0.790
DCOMOCUR	1.335	6.511	0.000
ALTAHORA	1.084	3.618	0.000
CARRET	0.208	0.542	0.588
SABDOM	0.274	1.286	0.198
DEFOCEMI	0.163	0.166	0.868
STEST	2.315	2.160	0.031
DZONA1	-0.525	-1.277	0.202
DZONA3	0.491	2.347	0.019
DRELATO	0.486	2.466	0.014
SAUTDAD	-1.245	-2.834	0.004
número de observaciones	1611	Chi-cuadrado	573.399
ln función verosimilitud	-808.878	grados de libertad	36
ln verosimilitud restringida	-1095.577	nivel signif.	0.000

Los primeros diecinueve parámetros definen el vector de coeficientes estimados para la probabilidad de que el asegurado cometa fraude en beneficio propio. El segundo grupo de parámetros estimados se refiere a la probabilidad de que exista fraude en beneficio de un tercero. Ambas especificaciones usan las mismas variables, presentándose, para todas ellas, los estadísticos de significación individual y los valores de la probabilidad de la cola.

El test de la razón de verosimilitud rechaza la hipótesis de que todos los coeficientes (excepto las constantes) en las dos ecuaciones estimadas sean iguales a cero; el logaritmo neperiano de la verosimilitud decrece de -808.9 a -1095.6, mientras que se ganan 36 grados de libertad al imponer las restricciones de igualdad. En nuestro caso, el estadístico (también conocido como estadístico Chi-cuadrado), toma un valor igual a 573.4 y su comparación con el valor en tablas para una distribución χ^2 con 36 grados de libertad permite aceptar la hipótesis alternativa de significación global del modelo al 1% de significación.

Los valores obtenidos muestran hechos importantes en relación a la probabilidad estimada de fraude. Como se desprende de la Tabla 1, la significación estadística de las variables es notablemente distinta al tener en cuenta ambos tipos de fraude. Mientras que en el primer grupo encontramos catorce coeficientes significativos (trece al 5% y uno al 10%) en el segundo sólo once lo son (diez al 5% y uno al 10%). El hecho de que haya variables que no tengan poder para explicar la existencia de un determinado tipo de fraude y lo tengan para explicar el otro es, sin duda, un punto importante a tener en cuenta en la investigación de siniestros y la evidencia empírica a nuestro entender más deseable.

En la interpretación de los resultados, para ambos vectores de parámetros y por separado, debe recordarse que la categoría de referencia de la variable dependiente es la de los siniestros no fraudulentos.

Para aquellas variables creadas con el objetivo de recoger situaciones claramente sospechosas de fraude, el signo de la estimación máximo verosímil de los parámetros asociados es el esperado. Para el resto de variables, y sin ser capaces en algunos casos de extraer conclusiones en cuanto a signos esperados (el propio mecanismo asegurador dificulta, en algunos casos, obtener conclusiones sobre el comportamiento que se espera para algunos regresores en relación al fraude), los resultados son claramente significativos.

Comencemos por los términos independientes del modelo. La aparición de un coeficiente estadísticamente significativo (al 1%) y con signo negativo en el vector de parámetros asociado a la categoría de fraude a favor del asegurado permite señalar que, si el resto de variables introducidas en el modelo son nulas, es inferior la probabilidad de aparición de este tipo de comportamiento en relación al no fraude. Una conclusión análoga se extrae a partir del coeficiente para el término independiente de la categoría de fraude a favor del tercero.

Para las variables relacionadas con la póliza y el vehículo en el momento del contrato, cabe destacar como la existencia en la póliza de cobertura a terceros está positivamente relacionada con la aparición de los dos tipos de fraude considerados, siendo el coeficiente estimado para la variable DRAMO estadísticamente significativo (al 5%) en los dos vectores de parámetros. Cuando ocurre el siniestro, el asegurado que posee cobertura de daños propios no necesita cometer fraude para conseguir reparar, por ejemplo, los daños de su vehículo, mientras que con la cobertura a terceros las inquietudes del individuo por conseguir una indemnización se ven incrementadas.

El tipo de vehículo que posee el asegurado tiene diferente poder explicativo en los dos tipos de fraude considerados. Así, mientras que el coeficiente que acompaña a la variable DUSO es estadísticamente significativo (al 1%) en la categoría de fraude a favor del asegurado, deja de serlo en la categoría de fraude a favor del tercero. El signo negativo obtenido para el mismo permite señalar que

la posesión de un turismo de uso particular disminuye tanto la probabilidad de que el asegurado cometa fraude en beneficio propio como para un tercero.

En relación a la variable que recoge el número de siniestros previos declarados en la compañía (dentro del conjunto de variables relacionadas con el asegurado en el momento del siniestro), los resultados son claramente significativos. El valor del coeficiente estimado, positivo y estadísticamente significativo al 1% en ambos vectores de parámetros, indica que los asegurados con siniestros previos tienden a cometer fraude. La explicación a este hecho puede ser variada. Quizá el asegurado, ante la no detección por parte de la compañía de fraudes anteriores, decida reincidir creando nuevos siniestros. Quizá esté en la peor clase de bonus-malus y por lo tanto piense que su situación, aunque tenga un nuevo siniestro, no puede ser peor.

La edad del conductor asegurado sólo es significativa en la explicación del fraude cometido en beneficio propio. La aparición de un coeficiente con signo negativo permite señalar que la probabilidad de fraude para beneficio propio disminuye a medida que aumenta la edad del individuo. Los años de antigüedad de carnet del conductor asegurado ha resultado ser una variable no significativa en la explicación del fraude.

La coincidencia de apellidos entre la parte asegurada y la contraria (variable basada en la relación con el conductor contrario) está también directamente asociada con la probabilidad de aparición de ambos tipos de fraude. No obstante el hecho de que el coeficiente goce de mayor significación estadística en el primer vector de parámetros (para el segundo sólo podemos aceptar la significación al 10%), nos lleva a pensar en una mayor influencia de la variable en la aparición de fraude a favor del asegurado. Una justificación para este resultado puede encontrarse, por ejemplo, en aquellos casos en los que el asegurado, que posee cobertura a terceros y ha sufrido un accidente sin involucración de un contrario,

conviene con un familiar que sea éste quien acepte la culpa del siniestro para así reparar sus daños.

Para el resto de variables, relacionadas con el siniestro y el vehículo en el momento del accidente, los resultados obtenidos son también relevantes.

La culpa del siniestro es una variable estadísticamente significativa en la explicación de fraude a favor del contrario y, sin embargo, no goza de significación en la categoría de fraude a favor del asegurado. El signo positivo que acompaña al coeficiente estimado en el segundo vector de parámetros permite señalar que si el asegurado admite la culpa del siniestro aumenta la probabilidad de que defraude a favor de un tercero. Si el asegurado posee cobertura a terceros (situación más frecuente) y admite la culpa del siniestro parece lógico pensar que sea para favorecer al contrario ya que no tiene cubiertos sus daños. No obstante, cabe suponer que este último compensará de una u otra forma la pérdida que experimenta el asegurado al entrar en la escala de malus (o descender en la de bonus). Lógicamente, si el asegurado posee cobertura de daños propios puede admitir su culpa para que, a la vez que repara sus daños, beneficie a un contrario que posee únicamente cobertura a terceros.

La antigüedad del vehículo en el momento del siniestro no tiene poder explicativo en la aparición de ninguno de los dos tipos de fraude considerados. Así lo indica el coeficiente estimado para dicha variable, que aún presentando signo positivo en ambos vectores de parámetros no reviste significación estadística. No obstante, el hecho de que el nivel de significación al que aceptaríamos la hipótesis alternativa (coeficiente individualmente significativo) sea notablemente inferior para el coeficiente que aparece en el primer vector, nos permite justificar lo que sería en principio un comportamiento esperado: de tener influencia, la antigüedad del vehículo del asegurado aparecerá relacionada con la existencia de fraude en beneficio propio.

El hecho de que el siniestro sea comunicado a la compañía con posterioridad a la primera semana desde su ocurrencia y el que haya ocurrido a altas horas de la noche son variables con coeficientes estadísticamente significativos y relacionados positivamente con la probabilidad de aparición de ambos tipos de fraude.

El primero de los anteriores coeficientes (relativo a la dilación en la comunicación del accidente a la entidad) puede interpretarse atendiendo al hecho de que, cuando no existe fraude, cabe esperar que el asegurado comunique el siniestro a su compañía con la mayor brevedad posible. La preparación de un fraude puede suponer una demora en la comunicación del accidente (indecisión sobre la posibilidad de comportarse fraudulentamente, diseño del plan a seguir con el contrario,...).

Cuando el siniestro ocurre a altas horas de la noche puede resultar más fácil cometer fraude. La ausencia de testigos y la posibilidad de que en el siniestro intervengan asegurados jóvenes (más propensos, como hemos visto, a cometer fraude) son algunos de los factores que pueden explicar el comportamiento observado para esta variable.

La ocurrencia del accidente en una zona de alta siniestralidad respecto a otras zonas (siniestralidad media o baja) incrementa la probabilidad de fraude en beneficio propio mientras que reduce la probabilidad de fraude a favor de un tercero. No obstante, para este último caso carece de significación estadística. Por otro lado, el acaecimiento del siniestro en zonas de baja siniestralidad (respecto a las de siniestralidad alta o media) presenta coeficientes positivos y estadísticamente significativos en ambos vectores de parámetros.

Los resultados obtenidos para la segunda variable considerada que nos llevan a señalar un incremento de la probabilidad de fraude en las zonas con siniestralidad

baja, pueden sugerir que la definición de zonas de siniestralidad utilizada quizá no sea la adecuada para discriminar comportamientos diferenciales.

La ocurrencia del siniestro en carretera aumenta la probabilidad de fraude a favor del asegurado. El coeficiente que acompaña a la variable CARRET, positivo, es significativo (no obstante, a un 10%) en la primera categoría de fraude, mientras que carece de significación estadística en la categoría de fraude para un tercero.

El hecho de que el siniestro suceda en fin de semana no tiene poder explicativo en la aparición de los dos tipos de fraude. Esta conclusión se observa también para el hecho de que el siniestro ocurra entre la fecha de efecto y la fecha de emisión de la póliza. La obtención de coeficientes no significativos para ambas variables justifica tales afirmaciones. No obstante la modelización logística anidada nos ha ayudado a confirmar la relación de la segunda variable con la aparición de fraude a favor del asegurado, partiendo de la suposición de que, desde un punto de vista empírico, el retroceso en la fecha de efecto puede venir provocado muchas veces por el deseo de beneficiar al asegurado (por parte de personal de la compañía y de los agentes).

Algunos parámetros estimados sugieren que la influencia de las variables asociadas es opuesta en las dos ecuaciones. Así, la presencia de testigos disminuye la probabilidad de fraude en beneficio propio, mientras que está positivamente relacionada con la probabilidad de defraudar a favor del contrario. En este último caso cabe pensar en la existencia de testigos falsos que colaboran en la ejecución del fraude.

Por último, la presencia en la declaración del asegurado de los relatos sospechosos en el estudio de los siniestros fraudulentos¹⁰ aumenta la probabilidad de fraude, tanto a favor del asegurado como del tercero. La

¹⁰ Recordemos que éstos estaban relacionados con descripciones que incluían referencias a aparcamientos, adelantamientos, marcha atrás,...

intervención de policía en el siniestro disminuye la probabilidad para ambos tipos, resultado lógico si tenemos en cuenta que la presencia de autoridad y la elaboración de un atestado reduce notablemente el espacio para cometer fraude. Ambas variables revisten significación estadística en la explicación de los dos tipos de comportamiento.

La interpretación de los coeficientes, tal y como comentamos en el Capítulo 4, no es inmediata. El hecho de que la variación que se espera se produzca en la probabilidad de elegir una determinada alternativa al aumentar en una unidad una determinada variable explicativa no sea constante, sino que dependa de los valores que tome dicha variable y el resto de las explicativas incluidas en el modelo, hace que los coeficientes sólo sean directamente interpretables en términos de dirección de variación de la probabilidad.

Una solución se encuentra en el cálculo de los efectos marginales de los regresores en las probabilidades. En nuestro caso el cómputo de los mismos ha derivado en la obtención de valores muy bajos. No obstante, el hecho que la mayoría de las variables explicativas consideradas sean dicotómicas hace que su interpretación sea complicada, teniendo en cuenta que las derivadas parciales de las probabilidades con respecto al vector de características para un determinado individuo son evaluadas en el vector de medias de las variables explicativas consideradas. Es por ello que obviamos su presentación.

6.3.2 Independencia de Alternativas Irrelevantes

Como vimos en el Capítulo 4, desde el punto de vista de la estimación del modelo ha de ocurrir que el cociente entre las probabilidades de elegir dos alternativas cualquiera sea independiente de los atributos o de la disponibilidad de una tercera alternativa de elección. Este hecho, comúnmente denominado Independencia de Alternativas Irrelevantes, viene fundamentado en la hipótesis de independencia para los términos de error en el modelo original (independencia

entre los términos de error de las funciones de utilidad planteadas¹¹), y puede contrastarse mediante la aplicación del test de Hausman-McFadden (1984).

Para nuestro caso, los resultados del test son 0.7 y 1.0, respectivamente, cuando eliminamos cada tipo de fraude. El contraste con una distribución Chi-cuadrado, en este caso con 19 grados de libertad, permite señalar que no podemos rechazar la hipótesis nula de Independencia de Alternativas Irrelevantes, incluso al 1% de significación y por lo tanto, los cocientes planteados, $P_{i(1)}/P_{i(0)}$ y $P_{i(2)}/P_{i(0)}$ permanecen constantes cuando la tercera y la segunda alternativa no son tenidas en cuenta, respectivamente.

En la demostración de la independencia de alternativas ha sido necesario realizar una corrección de los pesos introducidos en el modelo, teniendo en cuenta, en cada caso, el nuevo tamaño muestral y el hecho de trabajar con una elección menos.

6.3.3 Análisis de la calidad del ajuste

En los modelos de elección probabilística es frecuente utilizar como medida de calidad del ajuste el porcentaje de casos correctamente clasificados por el modelo, teniendo en cuenta las alternativas planteadas. Tras la aplicación del logit multinomial, el cómputo de las probabilidades predichas por el modelo para cada una de las observaciones supondrá una clasificación del siniestro en aquella categoría para la que se obtiene una mayor probabilidad¹². La comparación, para cada observación, entre la clasificación real u observada y la predicha queda sintetizada en la siguiente tabla,

¹¹ Ver Capítulo 4.

¹² Lógicamente, la suma de las tres probabilidades predichas por el modelo para cada una de las alternativas ha de sumar la unidad.

Tabla 2. Frecuencias de clasificación
(utilizando el criterio de máxima probabilidad)

	Elección predicha			Total
	Legítimo	Beneficio propio	Beneficio tercero	
Elección Observada				
Legítimo	961	12	25	998
Beneficio propio	222	37	40	299
Beneficio tercero	177	9	128	314
Total	1360	58	193	1611

Los resultados obtenidos permiten señalar que 1126 siniestros han sido clasificados en el grupo correcto. De esta forma el porcentaje de clasificación correcta es del 70%. Sin embargo, y a pesar de que el porcentaje de siniestros sin fraude detectado correctamente clasificados como tales es elevado (96.3%), no podemos decir lo mismo para los siniestros con fraude. Así, tan sólo un 12.4 % y un 40.8% de los siniestros han sido correctamente clasificados en las categorías de fraude a favor del asegurado y fraude a favor de un tercero, respectivamente.

En relación a los siniestros clasificados de forma errónea, los resultados no son buenos. Las frecuencias obtenidas reflejan una elevada proporción de casos predichos por el modelo como no fraudulentos. De esta forma 222 expedientes con fraude observado a favor del asegurado y 177 con fraude observado a favor de un tercero son clasificados por el modelo como no fraudulentos. Lógicamente y, desde este punto de vista, la capacidad predictiva del modelo planteado es baja, dado que tras su aplicación, la entidad dejaría de detectar un 74.2% de siniestros de la primera tipología de fraude y un 56.4% de la segunda. El porcentaje de siniestros sin fraude detectado clasificados por el modelo como fraudulentos, y el de siniestros fraudulentos de un tipo clasificados en la categoría alternativa de fraude no son elevados. Además, y para el último caso, las consecuencias de una clasificación incorrecta no son tan preocupantes, pues en

ambas se detecta fraude (lógicamente, ésto se podría matizar al tener en cuenta los costes derivados de investigar uno u otro tipo de fraude).

La obtención de unos resultados como los presentados para los siniestros fraudulentos (de uno u otro tipo) puede venir provocada por el hecho de que la muestra utilizada, en relación a los mismos, no sea suficientemente representativa. El hecho de que en la alternativa de no fraude dispongamos de un número de observaciones notablemente superior al de las alternativas de fraude puede ser una de las razones para que el modelo cometa un mayor error en la clasificación de los siniestros fraudulentos. Sin embargo, los resultados de clasificación presentados pueden mejorar notablemente al optimizar el criterio probabilístico utilizado para clasificar un siniestro dentro de una determinada categoría.

Como veíamos en el Capítulo 4 (la optimización del punto de corte en un modelo lógit multinomial constituye, sin duda, una de las principales aplicaciones de este libro y su fundamentación teórica ha sido presentada en el apartado 4.4) el punto de corte o probabilidad utilizada por defecto en la delimitación de las zonas de clasificación de los siniestros en cada una de las tres categorías presentadas es de $1/3$. La selección del baricentro presentado en el Gráfico 1 (ver Capítulo 4), delimita áreas de aceptación para cada una de las alternativas con idénticas dimensiones. Sin embargo, tras la aplicación del modelo, los resultados obtenidos en términos de calidad del ajuste pueden sugerir la conveniencia de modificar dicho criterio probabilístico, fundamentalmente, al tener en cuenta la elevada presencia de siniestros de una determinada categoría clasificados dentro de otra.

La optimización del punto de corte, destinada fundamentalmente a reducir el porcentaje probabilístico que permite calificar al siniestro como fraudulento (ampliación de las zonas de aceptación de ambos tipos de fraude) ha permitido obtener la siguiente clasificación de siniestros, estando el punto de corte fijado en el vector de probabilidades (0.15, 0.15, 0.15).

Tabla 3. Frecuencias de clasificación
(utilizando el criterio $c=0.15$)

	Elección predicha			Total
	Legítimo	Beneficio propio	Beneficio tercero	
Elección Observada				
Legítimo	753	111	134	998
Beneficio propio	104	129	66	299
Beneficio tercero	48	34	232	314
Total	905	274	432	1611

El porcentaje de siniestros correctamente clasificados dentro de sus categorías respectivas es ahora del 69.1%. Un total de 753 siniestros sin fraude detectado han sido clasificados de forma adecuada por el modelo, de forma que el porcentaje de aciertos para la primera alternativa es del 75.5%. En relación a las alternativas de fraude, para los siniestros en los que se observa fraude a favor del asegurado, el porcentaje de aciertos es del 43.1% mientras que para aquellos en los que se observa fraude a favor de un tercero dicho porcentaje es superior, del 73.9%.

La comparación con los resultados que aparecen en la Tabla 2 muestran, sin duda, una notable mejora. Mientras que el porcentaje total de siniestros clasificados de forma correcta prácticamente no sufre variación, el incremento de aciertos para los casos de fraude es una señal de la mayor capacidad predictiva del modelo en relación a este tipo de comportamiento.

¿Qué ocurre con los siniestros clasificados de forma errónea?

Como veíamos en páginas anteriores, el error cometido al clasificar como no fraudulentos siniestros en los que existe fraude detectado es especialmente preocupante para la compañía ya que, en esos casos, al no realizarse la investigación pertinente, resulta difícil detectar comportamientos deshonestos.

Los porcentajes que para este error de clasificación observábamos en la Tabla 2 son ahora muy inferiores, pasando a ser del 34.8% para el primer tipo de fraude y del 15.3 para el segundo, advirtiéndose por lo tanto una mejora importante. Un 22.1% de los siniestros con fraude a favor del asegurado son clasificados de forma errónea como fraudes a favor del tercero mientras que un 10.8% de los siniestros con fraude a favor del contrario son clasificados como fraudes a favor del asegurado. No obstante, el hecho de que la compañía continúe detectando fraude (lógicamente esta conclusión podría matizarse si tuviéramos en cuenta la existencia de costes de investigación muy diferentes para cada tipo de comportamiento fraudulento) hace que estos resultados no sean *a priori* tan preocupantes. Por último, un 24.5% de siniestros sin fraude detectado han sido clasificados como fraudulentos por el modelo (111 expedientes sin fraude son clasificados dentro de la primera categoría de fraude y 134 dentro de la segunda) siendo el aumento innecesario en los costes de investigación la principal desventaja asociada a este error de clasificación.

Los resultados presentados en la Tabla 3 permiten señalar que el modelo lógit multinomial, tras la optimización del punto de corte, proporciona una técnica de predicción aceptable, fundamentalmente en lo que a siniestros fraudulentos se refiere. Además, y dado que para los siniestros sin fraude el porcentaje de aciertos es también muy elevado podemos ratificarnos en la afirmación de que la aplicación del Modelo 1 ayudará a la compañía a realizar una correcta clasificación de siniestros, teniendo en cuenta la tipología de fraude presentada.

No obstante, para la entidad aseguradora podría resultar interesante el planteamiento de un escenario adicional. Así, si asumimos que los costes de oportunidad en tiempo y los costes monetarios asociados a la investigación de siniestros no son elevados (la compañía está dispuesta a aceptar un mayor error en la clasificación de siniestros no fraudulentos como fraudulentos), podemos mejorar aún más la capacidad del modelo en términos de predicción de fraude.

De esta forma, la selección del punto de corte (0.1, 0.1, 0.1) permite obtener los resultados que aparecen a continuación,

Tabla 4. Frecuencias de clasificación
(utilizando el criterio $c= 0.10$)

	Elección predicha			Total
	Legítimo	Beneficio propio	Beneficio tercero	
Elección Observada				
Legítimo	642	203	153	998
Beneficio propio	63	167	69	299
Beneficio tercero	25	49	240	314
Total	730	419	462	1611

observándose unos porcentajes de clasificación correcta para los siniestros con fraude a favor del asegurado y del tercero, del 56% y 76%, respectivamente. El número de siniestros fraudulentos clasificados erróneamente en la categoría de no fraude es muy bajo (63 casos para el primer tipo de fraude y 25 para el segundo). El porcentaje global de casos correctamente clasificados es ahora del 65.1% demostrándose, de nuevo, una elevada capacidad predictiva del modelo.

Los resultados obtenidos de la optimización del punto de corte evidencian la importancia del proceso dentro de la modelización logística multinomial. Así, hemos podido demostrar como, al optimizar el criterio probabilístico utilizado en la definición de las zonas de aceptación de cada una de las alternativas, el porcentaje de fraudes detectados ha aumentado de forma muy importante. Con todo ello, la compañía puede encontrar en el Modelo 1 una buena herramienta, no sólo para analizar qué variables tienen mayor poder explicativo en la aparición de fraude a favor del asegurado y de fraude a favor del tercero, sino también para saber cuál sería el margen de error cometido en la detección de este tipo de fraudes dentro de su cartera. La comparación entre el coste de investigación de aquellos siniestros que el modelo clasifica como fraudulentos y los beneficios finalmente derivados de la detección de ambos tipos de fraudes permitirá a la

entidad extraer las primeras conclusiones en relación a las ventajas o, en su caso desventajas, de instalar un mecanismo de control.

Una vez realizada la validación del Modelo 1, determinada la probabilidad estimada de existencia de cada tipo de comportamiento y cuantificada la probabilidad de clasificación correcta e incorrecta de cada tipo de siniestro, hemos dado respuesta al primer bloque de objetivos que presentábamos en el Capítulo 3. Ahora bien, los resultados obtenidos pueden utilizarse, tal y como comentábamos en el mencionado Capítulo, en la consecución de dos objetivos finales: por un lado, estimar la utilidad esperada por el individuo al elegir cada una de las alternativas y , por otro, cuantificar el coste por siniestro esperado por la compañía a partir de la aplicación de un modelo, como el presentado, de control de fraude. El planteamiento de escenarios nos ayudará, en los siguientes apartados, a analizar las principales conclusiones para ambos, teniendo siempre en cuenta las alternativas de elección presentadas y el tipo de modelo que hemos utilizado.

6.3.4 Estimación de la utilidad esperada por el individuo al elegir entre las alternativas

Como veíamos en el Capítulo 3, la estimación de la utilidad que el asegurado espera de su comportamiento en términos de fraude puede realizarse teniendo en cuenta una serie de parámetros básicos. Estableciendo una analogía con la formulación presentada por Hoyt (1990) y Picard (1996), la obtención de la información necesaria para estimar la utilidad esperada por el individuo supondrá tener en cuenta básicamente dos conceptos. Por un lado, habremos de cuantificar la probabilidad de que el asegurado elija una determinada alternativa y, dado que en su decisión intervendrá la información que posee en relación a la mayor o menor probabilidad de que la compañía audite el siniestro, necesitaremos estimar la probabilidad de que se realice auditoría o investigación de los casos.

Al realizar la estimación del Modelo 1 hemos conseguido cuantificar la probabilidad de que el asegurado decida no defraudar, de que cometa fraude en beneficio propio o de que cometa fraude en beneficio de un tercero. De esta forma disponemos de información relativa al primero de los conceptos mencionados y la podremos utilizar directamente en la estimación de la utilidad esperada.

La probabilidad de que la compañía audite el siniestro estará determinada por los resultados obtenidos tras la aplicación del modelo en relación a la existencia de fraude: sólo si la probabilidad estimada supera el criterio probabilístico establecido como punto de corte se realizará la investigación del siniestro en cuestión. Tras la aplicación del mecanismo de control, la entidad será capaz de detectar un elevado número de siniestros fraudulentos de uno y otro tipo (ver Tabla 3), pero cometerá error en la clasificación de determinados expedientes. Cuando existiendo fraude la compañía no realiza ningún tipo de investigación, la utilidad esperada por el individuo que actúa fraudulentamente se ve incrementada. Para el resto de siniestros erróneamente clasificados la auditoría alcanzará los objetivos deseados, pues tras la investigación supondremos que se detectará el tipo real de fraude existente o se confirmará la no aparición de comportamiento fraudulento.

La existencia de diferentes alternativas quedará reflejada en la especificación de la función de utilidad. Si el objetivo fuera analizar la utilidad que el individuo espera al actuar fraudulentamente frente a un comportamiento honesto bastaría con agrupar los resultados obtenidos para los dos tipos de fraude considerados. De esta forma, la probabilidad de cometer fraude vendría determinada por la suma de la probabilidad asociada al primer tipo de comportamiento fraudulento más la probabilidad asociada al segundo. Asimismo, la probabilidad de que la compañía clasificase un siniestro fraudulento en cualquiera de los tipos de fraude (probabilidad de detección de fraude) vendría determinada por el cociente entre el número de siniestros fraudulentos correctamente clasificados en sus categorías

respectivas más aquellos clasificados incorrectamente en las categorías alternativas y el número total de siniestros fraudulentos observados. Una aplicación de este escenario puede encontrarse en Artís, Ayuso y Guillén (1998b).

Sin embargo, cuando el objetivo es determinar la utilidad esperada por el individuo teniendo en cuenta la elección entre tres alternativas (sin realizar, por lo tanto, una agrupación previa entre las dos categorías de fraude) se ha de tener en cuenta un amplio abanico de probabilidades. Éstas estarán relacionadas no sólo con la probabilidad de que el individuo elija una opción determinada sino también, con la probabilidad de que la compañía detecte o no su comportamiento.

¿Cómo podemos determinar la utilidad esperada por el asegurado al cometer fraude en beneficio propio?

A partir de los resultados presentados en la Tabla 3¹³ podemos determinar q_1 , definida como la probabilidad de que la compañía detecte la existencia de fraude a favor del asegurado y que estimamos es igual a 0.431. La probabilidad de que la compañía no detecte este tipo de comportamiento, $(1-q_1)$, será por lo tanto de 0.569. Sin embargo, este último porcentaje ha de ser objeto de corrección al tener en cuenta que, en determinadas situaciones, el individuo también será penalizado dado que el siniestro será clasificado erróneamente dentro de la categoría de fraude a favor del tercero ($q_1' = 0.221$). La probabilidad de detección será, por tanto, $q_1 + q_1' = 0.652$, siendo $(1-0.652)$ la probabilidad de que el fraude considerado no sea detectado. Suponemos que clasificarlo en una categoría errónea de fraude implica que al final se identificará el comportamiento deshonesto debido a que se iniciará una investigación. Es importante destacar que hemos estimado la probabilidad de detección sin tener en cuenta que ésta puede ser diferente para los distintos individuos, ya que no sabemos si los errores de clasificación son mayores o menores en función de ciertas características del expediente.

¹³ De forma análoga podrían utilizarse los resultados presentados en la Tabla 4.

Dada W (riqueza inicial del individuo), P (prima pagada por la cobertura) y s (indemnización derivada del siniestro fraudulento¹⁴) y, suponiendo que la penalización para el asegurado ante la realización de cualquier tipo de fraude es siempre la misma y supone la expulsión de la compañía¹⁵ (cuantificada, por lo tanto, en base a la nueva prima que ha de pagar en otra compañía para gozar de un seguro que es obligatorio, P'), nuestro modelo proporciona una estimación de la utilidad que el asegurado espera al cometer fraude a favor de sí mismo,

$$E[U(1)] = [0.652U(W-P-P') + 0.348U(W-P+s)].$$

En relación a la utilidad esperada por el asegurado que decide cometer fraude en beneficio del tercero, el proceso de cálculo es análogo al anterior. La probabilidad de que la compañía detecte dicho comportamiento es ahora 0.739 (q_2), y a la misma habremos de adicionar la probabilidad de que la compañía detecte, de forma errónea, fraude a favor del asegurado ($q_2' = 0.108$) pues en tal caso supondremos que también penalizará al individuo porque lo detectará. La probabilidad de que el fraude no sea detectado será por lo tanto de 0.153:

$$E[U(2)] = [0.847U(W-P-P') + 0.153U(W-P+s)].$$

Si el objetivo es cuantificar la utilidad que el individuo espera de su comportamiento teniendo en cuenta, de forma conjunta, la existencia de tres posibilidades de elección, será necesario adicionar el resultado proporcionado por el modelo en términos de utilidad esperada para cada una de las situaciones, teniendo en cuenta la probabilidad estimada de que el individuo elija cada una de las alternativas.

¹⁴ Dicha indemnización puede ser simplemente la que corresponde por contrato (en caso de la creación de un siniestro o de la falsedad de datos para conseguir cobertura en situaciones excluidas en la póliza) o la misma pero incrementada (la ocurrencia legítima del siniestro es utilizada para conseguir una indemnización mayor a la que corresponde en base al tipo de seguro contratado).

¹⁵ Asumimos que el castigo impuesto no varía de un individuo a otro. Las compañías tienden a aplicar una regla común en orden a proteger los derechos de los demandantes honestos. Normalmente es la expulsión del asegurado de la entidad aunque podría ocurrir que la penalización impuesta consistiera únicamente en la devolución de la parte de indemnización no justificada.

Para cada individuo y dadas sus características, el modelo genera estimaciones puntuales de t_0 (probabilidad estimada de que el individuo se comporte honestamente), de t_1 (probabilidad estimada de que el asegurado cometa fraude en beneficio propio) y de t_2 (probabilidad estimada de que cometa fraude a favor del tercero). Teniendo en cuenta los resultados anteriormente presentados en relación a la utilidad esperada para cada uno de los comportamientos fraudulentos, la utilidad esperada del comportamiento del asegurado vendrá dada por la expresión (el subíndice i relativo al individuo no ha sido considerado, para simplificar la notación):

$$E[U] = t_1[0.652U(W-P-P') + 0.348U(W-P+s)] + t_2[0.847U(W-P-P') + 0.153U(W-P+s)] + t_0 U(W-P+s_0).$$

siendo s_0 la indemnización a percibir por el asegurado derivada de un comportamiento honesto.

La sustitución, en la expresión anterior, de las probabilidades estimadas por el modelo de elegir cada una de las alternativas (t_j , con $j=0,1,2$), y de los valores fijados para W , P , P' , s y s_0 , permitirá cuantificar el resultado esperado por el asegurado tras su actuación.

6.3.5 Cuantificación del coste esperado por siniestro tras la aplicación de un modelo de control del fraude

Las probabilidades estimadas de elección de cada una de las alternativas y de clasificación de los siniestros pueden utilizarse, asimismo, en la estimación del coste que la compañía espera del siniestro, tras la aplicación de un mecanismo de control de fraude.

Una posible solución a la estimación del coste esperado tras la materialización del riesgo, teniendo en cuenta la existencia de fraude, vendría por agrupar los

resultados obtenidos (probabilidades de elección y de clasificación de siniestros) tras la aplicación del lógit multinomial, para las dos categorías de fraude consideradas. De esta forma acabaríamos cuantificando el coste esperado final teniendo en cuenta el coste esperado para los siniestros no fraudulentos y para los fraudulentos, estos últimos, genéricamente considerados.

Sin embargo, la estimación del coste esperado teniendo en cuenta la existencia de tres alternativas finales de elección constituye un proceso más elaborado, tal y como veremos a continuación.

Retomemos las definiciones presentadas en el apartado anterior para t_0 , t_1 , t_2 , q_1 , q_1' , q_2 y q_2' . Sea q_0 la probabilidad de que la compañía clasifique correctamente un siniestro legítimo y q_0' la probabilidad de que lo clasifique como fraude a favor del asegurado. Para todas estas probabilidades disponemos de alguna estimación mediante los resultados que proporciona el modelo.

Supongamos que el coste que tiene que asumir la compañía resulta de la adición de dos componentes: la indemnización finalmente pagada más los costes asociados a la investigación del siniestro. De esta forma, y para cada una de las situaciones, el coste esperado del siniestro queda recogido en la Tabla 5.

Tabla 5. Coste esperado del siniestro

	Clasificado como		
	Legítimo	Beneficio propio	Beneficio tercero
Tipo de siniestro			
Legítimo	s_0	c_1+s_0	c_1+s_0
Beneficio propio	s	c_2	c_2
Beneficio tercero	s	c_2	c_2

Donde s es el coste del accidente que debe ser pagado de acuerdo con la póliza contratada (s_0 , es el coste real del siniestro), c_1 es el coste de investigar un

siniestro sospechoso que no es fraudulento (legítimo) y, c_2 es la cantidad gastada en identificar un fraude.

En relación a la cantidad pagada en concepto de indemnización y, de cara a simplificar la interpretación de la Tabla 5, hemos supuesto que es la misma para los dos tipos de fraude (s). Las situaciones, sin embargo, pueden ser diversas. La aplicación de los convenios C.I.D.E.-A.S.C.I.D.E., implica que en algunos casos el coste del siniestro sea únicamente la cantidad pagada por módulo. La diferencia puede ser importante. Así, por ejemplo, si el asegurado, que posee cobertura a terceros, acepta la culpa del siniestro para beneficiar al contrario, el coste del siniestro para la compañía en concepto de indemnización será únicamente de 90000 pts. Si es el contrario el que acepta la culpa para beneficiar al asegurado, el coste para la compañía será el asociado a la reparación de los daños (previo recobro del módulo de la compañía contraria). Generalizando, la consideración del tipo de cobertura existente en la póliza, la identificación de la culpa del siniestro, la aplicación o no de convenios,..., puede derivar en indemnizaciones de diferente cuantía para la entidad.

Sin embargo y, dado que nuestro objetivo es presentar sólo una aproximación al tratamiento de los costes¹⁶, hemos creído conveniente simplificar las hipótesis y diferenciar únicamente entre el coste del siniestro sin fraude y del siniestro con fraude. Aunque podrían aparecer situaciones en las que la existencia de comportamiento fraudulento implicase un menor coste para la entidad (en el ejemplo planteado en primer lugar en el párrafo anterior la entidad únicamente ha de pagar el módulo), la aparición de fraude suele representar, en la mayoría de los casos, un coste para la compañía superior al esperado ($s \geq s_0$).

¹⁶ La no disponibilidad, en la muestra original, de información completa en relación a los costes asociados a los siniestros con fraude detectado (diferencia entre la cuantía del siniestro y la cantidad finalmente pagada, si es que se paga alguna cantidad) ha impedido realizar un estudio más exhaustivo de este tema.

En relación a los costes de investigación, la simplificación es también evidente. Así, hemos supuesto que el coste de identificar como legítimo un siniestro con sospecha de fraude a favor del asegurado o con sospecha de fraude a favor de un tercero, es el mismo y toma el valor c_1 . Siguiendo el mismo criterio, el coste de detectar cualquiera de los tipos de fraude es siempre c_2 . Tomando en consideración ambos costes, podríamos esperar una relación del tipo $c_1 \leq c_2$. Posiblemente, la entidad invertirá mayores cantidades monetarias y su coste de oportunidad en tiempo será superior cuando persiga identificar la existencia de fraude en el siniestro, a diferencia de aquellos casos en los que, durante el proceso de investigación, no tiene suficientes pruebas para sospechar de la existencia de comportamiento fraudulento.

Teniendo en cuenta las consideraciones anteriores, el coste total esperado del siniestro, $E(T)$, puede escribirse como:

$$E(T) = t_1[(q_1 + q_1')(c_2 - s) + s] + t_2[(q_2 + q_2')(c_2 - s) + s] + t_0[q_0 s_0 + (1 - q_0)(s_0 + c_1)].$$

Nuestro modelo provee un método para estimar la probabilidad de elección de cada tipo de comportamiento (t_1 , t_2 y t_0) dadas las características del siniestro y/o del individuo. Como los parámetros de coste (s_0 , s , c_1 y c_2) son fijos y conocidos y las probabilidades de la compañía clasifique correcta (q_0 , q_1 y q_2) y erróneamente (q_0' , q_1' y q_2') a los siniestros en su categoría correspondiente pueden inferirse de los resultados del Modelo, es posible estimar el coste esperado del siniestro.

Retomemos, para acabar este apartado, los resultados que aparecen en la Tabla 3 y planteemos un escenario sencillo.

La probabilidad de clasificar un siniestro legítimo como tal es 0.755, siendo 0.245 la probabilidad de clasificarlo incorrectamente como fraudulento. Como ya sabemos, para los siniestros con fraude a favor del asegurado el porcentaje de aciertos es 0.431, mientras que la probabilidad de clasificarlos como fraudes a

favor del tercero es 0.221. Además la probabilidad de clasificar de forma correcta un siniestro con fraude a favor del tercero es 0.739. La probabilidad de clasificarlo como fraude a favor del asegurado es 0.108.

Sea s_0 el coste real del siniestro y fijemos que $s=1.5s_0$ es la indemnización pagada por la entidad atendiendo a la declaración presentada por el asegurado. Si asumimos que $c_1= 0.10s_0$ y, $c_2= 0.15s_0$, el coste total esperado estimado del siniestro es igual a:

$$E(T)= [t_1(0.62s_0) + t_2(0.36s_0) + t_0(1.02s_0)].$$

Si asumimos que la probabilidad estimada de que un asegurado cometa fraude a favor de sí mismo (t_1) es 0.44 y la de que cometa fraude a favor del tercero (t_2) es 0.30, entonces, en término medio, el coste esperado por la compañía se cifra, aproximadamente, en un 65% de la cantidad monetaria que ésta debería pagar en concepto de indemnización (s_0), en caso de que el siniestro fuese legítimo.

Es decir de no aplicar ningún control del fraude, la compañía pagaría $1.5s_0$ para este siniestro, mientras que de aplicar el sistema se espera que pague $0.65s_0$. En realidad ello supone un ahorro esperado de $(1.5-0.65)s_0$, es decir, $0.85s_0$.

En base a los resultados la instalación de un mecanismo de detección de fraude quedaría justificada.

6.4 Estimación logística multinomial de la probabilidad de existencia de fraude para conseguir una indemnización o para incrementar una indemnización (MODELO 2)

Una vez realizado un estudio exhaustivo en relación al primer modelo planteado, pasamos a continuación a analizar los resultados obtenidos al modelizar el segundo árbol de decisión considerado. Tendremos en cuenta, por tanto, una clasificación alternativa para los tipos de fraude que diferencia entre fraude para conseguir una indemnización y fraude para incrementar una indemnización.

6.4.1 Análisis de la significación individual y global del modelo

La cuantificación de la probabilidad de que un asegurado no cometa fraude, de que lo haga para conseguir una indemnización que no le corresponde por contrato o para incrementar la que le correspondía en función del mismo, ha sido realizada, de nuevo, mediante la especificación de un modelo lógit multinomial (Figura 9). La carencia de atributos específicos para cada una de las alternativas finales ha impedido realizar la modelización teniendo en cuenta el esquema de decisión presentado en el Capítulo 2, mediante la aplicación de un lógit condicional.

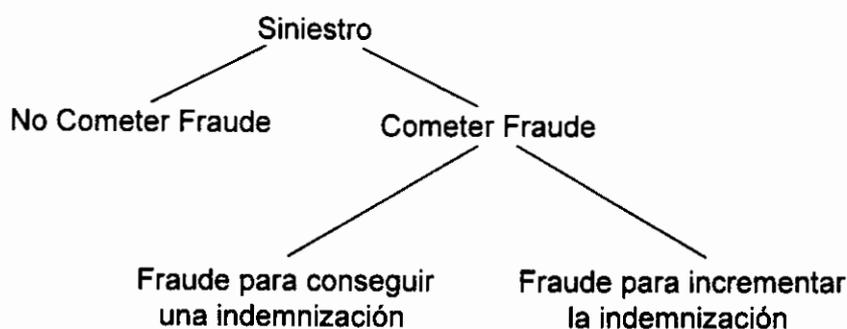


Figura 11

La muestra utilizada en la estimación del modelo está formada, tal y como hemos comentado en el apartado 6.2, por 1676 observaciones. La variable dependiente continúa siendo el tipo de siniestro, lógicamente teniendo en cuenta las categorías de fraude ahora analizadas. Así, hemos utilizado la codificación siguiente: $Y=0$, si el siniestro no es fraudulento (categoría de referencia); $Y=1$, si el fraude ha sido cometido para conseguir una indemnización e $Y=2$, si la ocurrencia del siniestro ha sido utilizada para incrementar la indemnización que correspondería por contrato. Las observaciones han sido ponderadas con la finalidad de reflejar el proceso de obtención de la muestra.

Las variables introducidas en la especificación del modelo son, con algunas variaciones, las mismas que hemos considerado en el Modelo 1. De nuevo utilizamos información que puede ser obtenida por la entidad en un breve periodo de tiempo y que recoge los principales aspectos relacionados con el seguro contratado, las partes intervinientes y el siniestro.

Los estimaciones obtenidas para los coeficientes del modelo, los correspondientes estadísticos de significación individual y los valores de la probabilidad de la cola aparecen en la página siguiente:

Tabla 6. Resultados de la estimación de un modelo lógit multinomial
[MODELO 2]

variable	coeficiente	t-test	P-valor
Fraude para conseguir una indemnización:			
CONSTANTE	-3.119	-5.028	0.000
DRAMO	0.426	0.959	0.338
DUSO	-0.658	-2.476	0.013
SFRANQUI	0.015	0.021	0.983
ACCESORI	-0.150	-0.417	0.676
HISTSTR	0.188	4.080	0.000
EDAD	-0.022	-2.985	0.003
DESTAF	1.379	4.459	0.000
DCULPA	0.537	2.761	0.006
ANTIGUO	0.032	1.608	0.108
DCOMOCUR	1.240	6.326	0.000
ALTAHORA	1.489	5.866	0.000
CARRET	0.662	2.169	0.030
SABDOM	0.344	1.733	0.083
DEFOCEMI	2.734	4.805	0.000
STEST	-9.946	-15.261	0.000
DZONA1	0.968	3.009	0.003
DZONA3	1.195	5.371	0.000
DRELATO	0.427	2.288	0.022
SAUTDAD	-1.784	-3.968	0.000
Fraude para incrementar una indemnización:			
CONSTANTE	-3.603	-5.744	0.000
DRAMO	0.886	2.110	0.035
DUSO	-0.403	-1.450	0.147
SFRANQUI	0.100	0.142	0.887
ACCESORI	-0.429	-1.137	0.255
HISTSTR	0.170	3.109	0.002
EDAD	-0.021	-2.710	0.007
DESTAF	0.908	2.333	0.020
DCULPA	2.156	11.332	0.000
ANTIGUO	0.006	0.280	0.779
DCOMOCUR	1.286	6.596	0.000
ALTAHORA	1.146	4.152	0.000
CARRET	0.244	0.673	0.501
SABDOM	0.270	1.339	0.180
DEFOCEMI	0.461	0.519	0.604
STEST	1.646	1.498	0.134
DZONA1	-0.422	-1.126	0.260
DZONA3	0.452	2.277	0.023
DRELATO	0.444	2.322	0.020
SAUTDAD	-1.211	-3.032	0.002
número de observaciones	1676	Chi-cuadrado	535.637
In función verosimilitud	-880.948	grados de libertad	38
In verosimilitud restringida	-1148.767	nivel signif.	0.000

El primer vector de parámetros es el asociado a la alternativa de fraude para conseguir una indemnización y el segundo, se refiere la alternativa de fraude para incrementar los costes. Como se observa, en ambas especificaciones se han utilizado las mismas variables.

El modelo es globalmente significativo. El test de la razón de verosimilitud es igual a 535.6, siendo el logaritmo neperiano de la verosimilitud del modelo general (con todas las variables explicativas) igual a -880.9 y el correspondiente a la verosimilitud del modelo restringido igual a -1148.8. La comparación del estadístico con el valor en tablas para una χ^2 con 38 grados de libertad permite rechazar al 1% de significación la hipótesis nula de que todos los coeficientes (excepto las constantes) son iguales a cero en las dos ecuaciones estimadas.

Los resultados obtenidos para los coeficientes estimados permiten destacar la existencia de variables con efectos claramente diferenciados en la explicación de ambos tipos de fraude. Teniendo en cuenta que la categoría de referencia de la variable dependiente es la de los siniestros no fraudulentos y que por tanto la interpretación de los resultados, para los dos vectores de parámetros por separado, debe hacerse en términos de la misma, pasamos a analizar la influencia de las variables consideradas en los comportamientos modelizados.

Los coeficientes que acompañan a las constantes del modelo son estadísticamente significativas (al 1%) en ambos conjuntos de parámetros. Su signo negativo permite interpretar una reducción en la probabilidad de aparición del primer y segundo tipo de fraude, cuando suponemos que el resto de variables incluidas en el modelo son nulas.

Para las variables relacionadas con la póliza y el vehículo en el momento del contrato, los resultados obtenidos evidencian una notable diferencia en su influencia sobre los dos tipos de fraude.

La existencia de cobertura a terceros está directamente relacionada con la probabilidad de aparición de ambos tipos de fraude, pero sólo goza de poder explicativo cuando el fraude ha sido cometido para incrementar una indemnización. La obtención de este resultado puede estar fundamentada en la propia definición de la cobertura, dado que cuando hablamos de conseguir una indemnización parece lógico pensar que sea a favor del asegurado y con la contratación de esta garantía quedan excluidos los daños sufridos por el mismo (o por su vehículo).

Con la variable que recoge el tipo de vehículo que posee el asegurado, el efecto observado es el contrario. La aparición de un coeficiente con signo negativo en ambos vectores de parámetros indica una disminución en la probabilidad de aparición de fraude cuando el vehículo asegurado es un turismo de uso particular. Sin embargo, únicamente se observa significación estadística (al 5%) para el coeficiente asociado a la categoría de fraude destinado a conseguir una indemnización.

Para los coeficientes asociados a las variables SFRANQUI y ACCESORI, relacionadas con la existencia de franquicia en la póliza y la contratación de la cobertura de accesorios, respectivamente, los resultados ponen de manifiesto una ausencia de significación estadística en ambos vectores de parámetros. No obstante y, a pesar de carecer de poder explicativo, la aparición de un signo positivo para el primer coeficiente permitiría interpretar que la existencia de franquicia aumenta la probabilidad de aparición de ambos tipos de fraude mientras que el signo negativo para el segundo supondría una interpretación en sentido contrario. Cuando existe franquicia el asegurado sabe que ha de asumir una parte del coste del accidente y ésto le puede llevar, por ejemplo, a declarar en un mismo parte de siniestro daños anteriores del vehículo con el objetivo de hacer frente a una única franquicia. Cuando el asegurado posee cobertura de accesorios disminuye la necesidad de cometer fraudes para recuperar, por

ejemplo, un radio-cassette robado o para intentar cambiar uno antiguo por otro más nuevo.

El historial de siniestros del asegurado tiene poder explicativo en la aparición de los dos tipos de fraude. El coeficiente observado, positivo y estadísticamente significativo (al 1%) en ambos vectores podría interpretarse de forma análoga a como lo hacíamos en el Modelo 1. El mayor conocimiento de la forma de actuar de la compañía, la no detección de fraudes anteriores o el hecho de que su clasificación en la escala de bonus-malus no pueda ser peor de lo que es, son algunas de las razones que ayudan a interpretar el valor obtenido para el coeficiente.

La edad del asegurado aparece multiplicada por un coeficiente con signo negativo y estadísticamente significativo (al 1%) en ambas categorías de fraude. El resultado obtenido permite señalar que la probabilidad de que el asegurado cometa cualquiera de los dos tipos de fraude disminuye con la edad.

La coincidencia de apellidos entre la parte asegurada y la contraria aumenta la probabilidad de aparición de fraude. El coeficiente que acompaña a esta variable es positivo y estadísticamente significativo en las dos alternativas. La consideración, dentro de la primera categoría de fraude, de aquellos casos en los que el asegurado declara en falso para eludir casos excluidos en la póliza (posibilidad de crear un siniestro con un familiar) permite interpretar el resultado para el coeficiente que aparece en el primer vector de parámetros. Para el coeficiente que aparece en el segundo vector, bastaría pensar, por ejemplo, en un acuerdo entre familiares que se declaran mutuamente culpables en sus respectivas compañías.

La aceptación de culpa por parte del asegurado está directamente relacionada con un aumento de la probabilidad de aparición de ambos tipos de fraude siendo estadísticamente significativa en la explicación de ambos comportamientos. La

interpretación de este resultado puede resultar complicada dada la gran variedad de situaciones que pueden plantearse y, en todo caso, estará sujeta a matizaciones. No obstante, el planteamiento de escenarios concretos puede ayudarnos en la explicación del valor obtenido para el coeficiente. Así, por ejemplo, si la fecha de efecto de la póliza ha sido retrocedida con el objetivo de cubrir un siniestro, parece más lógico pensar en la culpa del asegurado que en la del contrario. De la misma forma, la aceptación de culpa por parte de un asegurado que posee cobertura de daños propios permitirá cubrir los daños del contrario y a la vez reparar los propios, existiendo la posibilidad de que se beneficien ambos implicados.

El coeficiente para la variable que recoge la antigüedad en años del vehículo del asegurado no es estadísticamente significativo en ninguna de las dos alternativas de fraude, aunque para la categoría de fraude destinado a conseguir una indemnización prácticamente podríamos aceptar su significación al 10%. La aparición de un signo positivo para el parámetro estimado nos permite señalar que la probabilidad de existencia de fraude (sobre todo del primer tipo) aumenta cuanto más antiguo es el vehículo. La necesidad de reparar las cada vez más frecuentes averías puede ser una de las principales razones para que el asegurado prepare el siniestro.

Al igual que ocurría en el Modelo 1, las variables que recogen el hecho de que el siniestro sea comunicado a la compañía con posterioridad a la primera semana desde su ocurrencia o el que haya ocurrido a altas horas de la noche presentan coeficientes estadísticamente significativos y relacionados positivamente con la probabilidad de aparición de ambos tipos de fraudes. La interpretación que dábamos entonces para los mismos puede utilizarse ahora, en términos de que cuando el siniestro es legal cabe esperar brevedad en su comunicación a la compañía y de la mayor facilidad que puede existir para cometer fraudes en siniestros ocurridos a altas horas de la noche.

La ocurrencia del siniestro en carretera está directamente relacionado con la probabilidad de existencia de fraude dirigido a conseguir una indemnización. La obtención de un coeficiente positivo y significativo (al 5%) en el primer vector de parámetros permite realizar la afirmación anterior, sin que podamos decir lo mismo para el coeficiente asociado al segundo tipo de fraude. En este último caso, el coeficiente carece de significación estadística y por tanto la variable analizada no puede considerarse relevante en la explicación de fraude para aumentar la indemnización.

Un comportamiento análogo se observa para la variable que recoge la ocurrencia del siniestro en fin de semana y para la variable que indica la ocurrencia del siniestro entre la fecha de efecto y de emisión de la póliza. Para la primera, la aparición de un parámetro positivo y estadísticamente significativo (al 10%) en el primer vector permite señalar que cuando el siniestro ocurre en sábado o domingo aumenta la probabilidad de que el asegurado defraude para conseguir una indemnización sin que la variable goce de significación en la explicación del segundo tipo de comportamiento fraudulento. Para la segunda, la conclusión es la misma, con la salvedad de que el coeficiente es significativo al 1%. Cabe destacar, en relación a esta variable, como los resultados obtenidos permiten validar los comentarios que para la misma hacíamos en el Modelo 1. De esta forma cabe esperar que el retroceso en la fecha de efecto esté directamente relacionado con la existencia de fraude para conseguir una indemnización e, implícitamente podríamos pensar en la existencia de fraude para beneficiar al asegurado.

Para las variables que recogen la presencia de testigos y la ocurrencia del siniestro en una zona de elevada siniestralidad, los coeficientes estimados muestran la existencia de efectos contrarios. Así, la existencia de testigos disminuye la probabilidad de que el asegurado cometa el primer tipo de fraude mientras que incrementa la probabilidad de existencia del segundo tipo, aunque en este segundo caso el coeficiente no resulta significativo. La interpretación para

el primer resultado parece plausible. En relación al segundo podría pensarse en situaciones extremas en las que por ejemplo se diera la existencia de testigos falsos. Para la variable que recoge la ocurrencia del siniestro en una zona de elevada siniestralidad, el coeficiente es positivo y significativo en la primera categoría de fraude mientras que, para la segunda, presenta signo contrario y carece de significación estadística.

La acaecimiento del siniestro en una zona de baja siniestralidad presenta coeficientes positivos y estadísticamente significativos en ambos vectores de parámetros. Nótese de nuevo que tal vez la clasificación de zonas de siniestralidad utilizada no es la adecuada para discriminar entre tipos de fraude.

Por último, los coeficientes asociados a la existencia en la declaración relatada del asegurado de cierta terminología y a la intervención de policía en el siniestro muestran un comportamiento análogo en ambas categorías de fraude, siendo en todos los casos estadísticamente significativos. Para la primera variable, la aparición de valores positivos permite señalar una relación directa con la probabilidad de aparición de ambos tipos de fraude. Para la segunda, la relación es inversa (disminución de la probabilidad de aparición de fraude cuando la policía interviene en el siniestro).

Al igual que veíamos para el Modelo 1, los coeficientes obtenidos sólo son interpretables en términos de dirección de variación de la probabilidad. Desde este punto de vista, su análisis únicamente nos ha permitido señalar si las variables a las que acompañan tiene una relación positiva o negativa con la probabilidad de elegir cada una de las alternativas de fraude (tomando como categoría de referencia la categoría de no fraude). Cuando la relación es positiva (coeficiente con signo positivo), un incremento en el valor de la variable (o la presencia de un valor igual a 1 si es dicotómica) va asociado a un aumento de la probabilidad correspondiente; cuando es negativa, la variación de la probabilidad se produce en sentido inverso. Los resultados obtenidos tras efectuar el cálculo

de los efectos marginales de los regresores en las probabilidades, como solución a la no interpretación directa de los estimadores, reflejan valores muy bajos. Al igual que ocurría en el Modelo 1, el hecho de que numerosas de las variables consideradas sean dicotómicas dificulta su interpretación (recordemos que los efectos marginales son computados teniendo en cuenta el vector de medias de las variables explicativas y dicha medida estadística carece de sentido cuando la variable es categórica).

6.4.2 Independencia de Alternativas Irrelevantes

La aplicación del test de Hausmann y McFadden (1984) permite demostrar la independencia entre las alternativas consideradas en el modelo. Al eliminar la alternativa de fraude para incrementar la indemnización, el test es igual a 0.004 y su comparación con el valor en tablas para una χ^2 con 20 grados de libertad supone no poder rechazar la hipótesis nula de independencia de alternativas irrelevantes, al 1% de significación. Una conclusión análoga ha sido obtenida al eliminar la alternativa de fraude para conseguir una indemnización, siendo el valor del test en este caso igual a 0.005.

Cabe señalar, que al igual que ocurría en la demostración de Independencia de Alternativas en el Modelo 1, las ponderaciones introducidas han sido debidamente corregidas para tener en cuenta los tamaños muestrales y el número de elecciones finales consideradas.

Al no poder rechazar la hipótesis nula de Independencia de Alternativas Irrelevantes, el cociente entre la probabilidad de que el asegurado elija la primera alternativa de fraude y la probabilidad de que no defraude ($P_{i(1)}/P_{i(0)}$) permanece constante cuando no consideramos el segundo tipo de fraude. Lo mismo podemos señalar para el cociente entre la probabilidad de que el asegurado elija el segundo tipo de comportamiento fraudulento y la probabilidad de que no

cometa fraude ($P_{i(2)}/P_{i(0)}$), esta vez, al no considerar la alternativa de fraude para conseguir una indemnización.

6.4.3 Análisis de la calidad del ajuste

Para cada observación de la muestra, la comparación entre los valores observados y predichos para la variable dependiente queda recogida en la siguiente tabla de clasificación.

Tabla 7. Frecuencias de clasificación
(utilizando el criterio de máxima probabilidad)

	Elección predicha			Total
	Legítimo	Conseguir una indemnización	Incrementar una indemnización	
Elección Observada				
Legítimo	964	12	22	998
Conseguir una indemnización	213	48	41	302
Incrementar una indemnización	243	15	118	376
Total	1420	75	181	1676

Los datos anteriores reflejan que un total de 1130 expedientes han sido clasificados por el modelo dentro la categoría correspondiente siendo el porcentaje de aciertos del 67.4%. Sin embargo, mientras que el porcentaje de siniestros sin fraude detectado clasificados de forma correcta es elevado, de un 96.6%, el error de clasificación observado para los siniestros fraudulentos es notable. Tan sólo un 15.9% de los siniestros con fraude destinado a conseguir una indemnización son clasificados correctamente. En relación a los siniestros con fraude destinado a incrementar la indemnización derivada de la ocurrencia del siniestro el porcentaje, si bien mayor, continua siendo preocupante, tomando un valor del 31.4%.

La situación es parecida a la que encontrábamos en el análisis de la Tabla 2, asociada al Modelo 1. De nuevo el problema no es sólo el escaso número de fraudes detectados, sino también el elevado porcentaje de siniestros fraudulentos clasificados por el modelo como legítimos. Así, el hecho de que 213 y 243 expedientes, asociados a la primera y la segunda categoría de fraude, respectivamente, sean clasificados como no fraudulentos induce a pensar en la escasa calidad predictiva del modelo desde la óptica de fraude (tras su aplicación la compañía dejaría de detectar el 70.5% de los siniestros que presentan el primer tipo de comportamiento fraudulento considerado y el 64.6% de los siniestros que presentan el segundo).

La solución al problema anterior, producido tal vez porque la muestra utilizada no sea suficientemente representativa, se encuentra en la optimización del punto de corte o criterio probabilístico utilizado en la delimitación de las zonas de aceptación de cada una de las alternativas. Como resultado de dicho proceso, la selección del punto definido por el vector (0.15, 0.15, 0.15) deriva en los resultados presentados en la Tabla 8.

Tabla 8. Frecuencias de clasificación
(utilizando el criterio $c= 0.15$)

	Elección predicha			Total
	Legítimo	Conseguir una indemnización	Incrementar una indemnización	
Elección Observada				
Legítimo	776	101	121	998
Conseguir una indemnización	106	121	75	302
Incrementar una indemnización	89	58	229	376
Total	971	280	425	1676

El modelo predice correctamente el comportamiento observado de 1126 siniestros, siendo, por tanto, el porcentaje de aciertos del 67.2%. Un total de 776 siniestros sin fraude detectado son clasificados de forma correcta (porcentaje de

aciertos del 77.8%), y, en relación a los siniestros fraudulentos, la capacidad de predicción puede considerarse elevada. Así, un 40.1% de los siniestros con fraude destinado a conseguir una indemnización y un 60.9% de los siniestros destinados a incrementarla son clasificados correctamente dentro de sus categorías respectivas. El error cometido al clasificar siniestros fraudulentos como no fraudulentos es notablemente inferior al que aparecía en la Tabla 7. De esta forma, el porcentaje de siniestros que conteniendo el primer tipo de fraude son clasificados por el modelo como no fraudulentos es ahora del 35.1%; en relación al segundo tipo de fraude el porcentaje de siniestros incorrectamente clasificados como no fraudulentos es ahora del 23.7%.

La clasificación obtenida para los siniestros fraudulentos puede mejorarse al fijar el punto de corte en el vector de probabilidades (0.1, 0.1, 0.1). En este caso los resultados son los que se muestran en la Tabla 9.

Tabla 9. Frecuencias de clasificación
(utilizando el criterio $c=0.10$)

	Elección predicha			Total
	Legítimo	Conseguir una indemnización	Incrementar una indemnización	
Elección Observada				
Legítimo	656	180	162	998
Conseguir una indemnización	69	154	79	302
Incrementar una indemnización	47	83	246	376
Total	772	417	487	1676

Los porcentajes de siniestros fraudulentos correctamente clasificados son del 51% y del 65.4%, respectivamente, para cada tipo de fraude. Los porcentajes de siniestros fraudulentos incorrectamente clasificados como no fraudulentos son ahora notablemente inferiores, pasando a ser del 22.8% y del 12.5%, respectivamente.

La selección de uno u otro criterio probabilístico dependerá, fundamentalmente, del margen de error que la compañía esté dispuesta a aceptar en relación a la clasificación errónea para los siniestros no fraudulentos. Si los costes de investigación no son elevados tal vez se decante por los resultados presentados en la segunda tabla, dado que, en ese caso, los beneficios obtenidos de una mayor detección de fraudes compensarán los costes de investigar un elevado número de siniestros en los que finalmente no existirá comportamiento fraudulento. En cualquiera de los casos, los resultados presentados en las Tablas 8 y 9 evidencian una elevada capacidad predictiva del Modelo 2 de forma que la aplicación del mismo para cuantificar la probabilidad de que el asegurado elija cada una de las alternativas planteadas queda justificada.

Una vez realizada la validación del modelo, estimada la probabilidad de existencia de cada tipo de fraude y determinada la existencia de una elevada bondad del ajuste, disponemos de la información necesaria para estimar la utilidad esperada por el individuo al elegir una alternativa concreta y para estimar el coste esperado del siniestro tras la aplicación por la compañía de una política de detección de fraude mediante el uso del Modelo 2. Ambos aspectos serán tratados en los apartados siguientes.

6.4.4 Estimación de la utilidad esperada por el individuo al elegir entre las alternativas

Al igual que veíamos para el Modelo 1, la estimación de la probabilidad de que el individuo elija cualquiera de las tres alternativas analizadas y, la determinación de la probabilidad de que la compañía clasifique correctamente cada uno de los siniestros dentro de la categoría correspondiente, permite ajustar la utilidad que el asegurado espera de su comportamiento.

Dadas W (riqueza inicial del individuo), P (prima pagada en la entidad), s_0 (coste real del accidente), s (indemnización incrementada tras la actuación fraudulenta)

y, P' (prima pagada en otra entidad como consecuencia de la expulsión de la compañía inicial, en concepto de penalización tras la detección del fraude), la estimación de la utilidad esperada por el asegurado al cometer fraude para conseguir una indemnización es:

$$E[U(1)] = [0.649U(W-P-P') + 0.351U(W-P+s_0)].$$

Así, la probabilidad de que la compañía detecte la existencia de fraude es 0.649 (la probabilidad de que identifique fraude para conseguir una indemnización, q_1 , es 0.401 y la de que detecte fraude para incrementar la indemnización, q_1' , es 0.248). La probabilidad de que no se detecte fraude es, por tanto, de 0.351. Las estimaciones utilizadas han sido extraídas de los resultados presentados en la Tabla 8.

Siguiendo un proceso análogo al del apartado 6.3.4, la utilidad esperada por el individuo que comete fraude para incrementar la indemnización derivada de un siniestro, cuya ocurrencia es real y está cubierto por póliza, es:

$$E[U(2)] = [0.763U(W-P-P') + 0.237U(W-P+s)].$$

En este caso, la probabilidad de que la compañía detecte fraude es 0.763 (la probabilidad de que la entidad identifique fraude para incrementar la indemnización, q_2 , es 0.609 y la de que detecte erróneamente fraude para conseguir una indemnización, q_2' , es 0.154). Por tanto, la probabilidad de no detectar fraude es de 0.237.

La cuantificación de la utilidad esperada por el asegurado considerando, conjuntamente, las tres alternativas de elección será, por tanto:

$$E[U] = t_1[0.649U(W-P-P') + 0.351U(W-P+s_0)] + t_2[0.763U(W-P-P') + 0.237U(W-P+s)] + t_0U(W-P+s_0).$$

Las probabilidades t_0 , t_1 y t_2 corresponderán a las estimadas por el modelo de que el asegurado se comporte honestamente, de que cometa el primer tipo de fraude o el segundo, respectivamente.

6.4.5 Cuantificación del coste esperado por siniestro tras la aplicación de un modelo de control del fraude

Consideremos las definiciones presentadas en el apartado anterior para t_0 , t_1 , t_2 , q_1 , q_1' , q_2 y q_2' . Sea q_0 la probabilidad de que la compañía clasifique correctamente un siniestro legítimo y q_0' la probabilidad de que lo clasifique como fraude para conseguir una indemnización. La estimación del modelo nos proporciona la información necesaria para evaluar todas estas probabilidades. Sea el coste del siniestro para la compañía el resultado de sumar la indemnización finalmente pagada y las cantidades monetarias invertidas en investigación. De esta forma, el coste esperado del siniestro es el que se muestra en la Tabla 10, según los diferentes casos.

Tabla 10. Coste esperado del siniestro

Tipo de siniestro	Clasificado como		
	Legítimo	Fraude para conseguir una indemnización	Fraude para incrementar la indemnización
Legítimo	s_0	c_1+s_0	c_1+s_0
Fraude para conseguir una indemnización	s_0	c_2	c_2
Fraude para incrementar la indemnización	s	c_2	c_2

En este caso, s_0 y s han sido definidos previamente y, c_1 y c_2 recogen el coste de investigar un siniestro sospechoso para el que no se detecta fraude (y que es, por tanto, legítimo) y el asociado a la identificación de fraude, respectivamente.

De nuevo y, con el objetivo de simplificar la interpretación y la nomenclatura, suponemos que los costes de investigación que deben sufragarse para la identificación como legítimo, de un siniestro en el que existe sospecha de fraude para conseguir una indemnización o, de un siniestro con sospecha de fraude para incrementar una indemnización, son iguales (c_1). Análogamente, la confirmación de la existencia de fraude lleva unos gastos asociados iguales, en cualquiera de los casos, a c_2 . Teniendo en cuenta las mismas consideraciones que veíamos para el Modelo 1, cabe esperar que $c_1 \leq c_2$.

El coste total esperado del siniestro, $E(T)$, puede escribirse ahora como:

$$E(T) = t_1[(q_1 + q_1')(c_2 - s_0) + s_0] + t_2[(q_2 + q_2')(c_2 - s) + s] + t_0[q_0 s_0 + (1 - q_0)(s_0 + c_1)].$$

Dado que todos los parámetros que aparecen en la expresión anterior son conocidos o, pueden estimarse a partir del modelo, es posible presentar un ajuste del coste esperado del siniestro.

Terminemos este apartado, y el Capítulo en sí, presentando un pequeño ejemplo en relación a la estimación del coste que la entidad espera del accidente (utilizaremos los resultados que aparecen en la Tabla 8).

La probabilidad de clasificar correctamente un siniestro legítimo es ahora 0.778. El porcentaje de error en relación a la clasificación de los mismos es, por tanto, de 0.222. Como ya sabemos, para los siniestros con fraude dirigido a conseguir una indemnización, la probabilidad de acierto es 0.401, mientras que la probabilidad de que sean clasificados dentro de la tipología alternativa de fraude es 0.248. La probabilidad de clasificar como tal un siniestro con fraude dirigido a incrementar una indemnización es 0.609, siendo la probabilidad de clasificarlo como fraude dirigido a conseguir una indemnización de 0.154.

Sea s_0 el coste real del siniestro y $s=1.5s_0$ la indemnización pagada por la entidad atendiendo a la declaración presentada por el asegurado tras la ocurrencia real del siniestro (bajo condiciones cubiertas en la póliza). Si asumimos que $c_1= 0.10s_0$ y, $c_2= 0.15s_0$, el coste total esperado estimado del siniestro será igual a:

$$E(T)= [t_1(0.45s_0) + t_2(0.47s_0) + t_0(1.02 s_0)].$$

Supongamos que la probabilidad estimada de fraude para conseguir una indemnización (t_1) es 0.44 y la de fraude para incrementarla (t_2) es 0.30. El coste esperado del siniestro, tras la aplicación del modelo de detección es, aproximadamente, un 40% más pequeño que el que se derivaría de la no detección del fraude (pago de s_0). Ello supondría para este individuo un ahorro esperado para la entidad de $(1.00-0.60)s_0$, es decir $0.40s_0$.

El escenario que se acaba de presentar, de forma análoga al que presentábamos en el apartado 6.3.5, es extremo. Si un asegurado tiene una probabilidad de no defraudar del 26%, lógicamente, el sistema que utiliza el Modelo 2 debe conducir a la conclusión de que, en términos de costes esperados, implementar dicho modelo como herramienta de clasificación y por lo tanto, de detección es rentable. Sin embargo, en casos más dudosos el beneficio esperado podría no ser tan rotundo, más aún, teniendo en cuenta el hecho de que aunque una compañía clasifique un siniestro como fraudulento y lo audite no implica necesariamente que descubra y logre desenmascarar el fraude.

6.5 Modelización logística anidada.

Los resultados obtenidos para el Modelo 1 y el Modelo 2, en cuanto a la significación estadística de los coeficientes que acompañan a las variables explicativas consideradas, han sido ratificados tras la aplicación de modelos jerárquicos o anidados.

¿Cuál es la fundamentación teórica de estos modelos?

Cada individuo, a la hora de realizar una determinada elección, hace intervenir una serie de factores que, sin duda, condicionan la alternativa final seleccionada. La existencia de restricciones económicas, familiares, sociológicas,..., determinan su comportamiento, de forma que la toma de decisiones que realiza suele ir acompañada de una abstracción sobre las posibilidades existentes.

Pensar en aquellas situaciones en las que se elige, implica reflexionar, en la mayoría de los casos, mediante un proceso de decisión por etapas.

A la hora de comprar un vehículo (Koujianou, 1995) el individuo estructura mentalmente sus posibilidades. Primero decide entre si quiere comprar o no. Si decide comprar, puede pensar en un vehículo nuevo o en un vehículo de segunda mano. Quizá si opta por la segunda alternativa sus posibilidades se vean reducidas, pero si opta por la primera un amplio conjunto aparece a su alcance. A la decisión sobre la clase de vehículo que quiere (turismo, motocicleta,...) le seguirá la decisión sobre las características que busca en el mismo. Las preferencias sobre el lugar de fabricación, sobre el tamaño o sobre la cilindrada son algunos de los factores que permitirán continuar con el proceso de elección, de forma sucesiva, hasta llegar a la alternativa final seleccionada. La elección del modo de viajar es otro de los ejemplos utilizado en la literatura (Hensher, 1986; Greene, 1995) para analizar procesos de decisión por etapas.

Ahora bien, ¿qué implicaciones tiene la toma de decisiones de forma sucesiva en el diseño de un modelo que intente explicar los factores que la determinan?. ¿Cuál es la diferencia con lo que sería la modelización en base a alternativas finales, sin tener en cuenta etapas intermedias?. En definitiva, ¿en qué se diferencia la modelización anidada de la presentada en los apartados anteriores del presente Capítulo?.

Anidar significa tener en cuenta que las alternativas de elección pueden no ser independientes. En los apartados 6.3 y 6.4, se ha supuesto un enfoque que implica considerar una única etapa de decisión. El asegurado decide de forma directa entre no defraudar, defraudar en favor de sí mismo o defraudar para beneficiar a otra persona y, en la modelización de este comportamiento, únicamente se tienen en cuenta características relacionadas con diferentes aspectos que, en ninguno de los casos, son propios de una alternativa específica de elección. La información contenida en la muestra disponible no refleja, para ninguna de las observaciones, características propias de un determinado tipo de fraude. No disponemos de atributos de elección y, por lo tanto, únicamente podemos determinar qué factores o circunstancias generales del siniestro pueden tener mayor impacto en la explicación de comportamientos, desde la óptica de fraude.

Los modelos presentados en los capítulos anteriores suponen la existencia de una elección múltiple y los resultados obtenidos evidencian que determinadas variables pueden tener una influencia diferente en la probabilidad de elegir cada una de las alternativas posibles. La existencia de niveles de significación diferentes o la presencia de coeficientes con signos contrarios para una misma variable en categorías alternativas permite obtener importantes conclusiones relacionadas, por ejemplo, con la información que la entidad debe utilizar para detectar ciertos comportamientos. Igualmente, puede orientar sobre la forma de encaminar la investigación sobre el fraude.

La modelización del proceso de decisión por etapas deriva en un proceso econométrico más complejo, siendo necesario transformar las variables utilizadas en atributos específicos de una determinada alternativa de elección. En nuestro caso y, una vez realizado el pertinente rediseño de la base de datos, los resultados obtenidos en la nueva estimación de los Modelos 1 y 2 ratifican los ya obtenidos en la modelización multinomial con lo que obviamos su presentación específica (confirman además la importancia que la antigüedad del vehículo

puede tener en la aparición de fraude a favor del asegurado). En el siguiente Capítulo del libro (de Conclusiones) pueden encontrarse las principales consideraciones a establecer en relación a esta modelización alternativa.

Los procesos de decisión analizados atienden a los esquemas que aparecen en las Figuras 12 y 13, ya presentadas en páginas anteriores del presente libro.

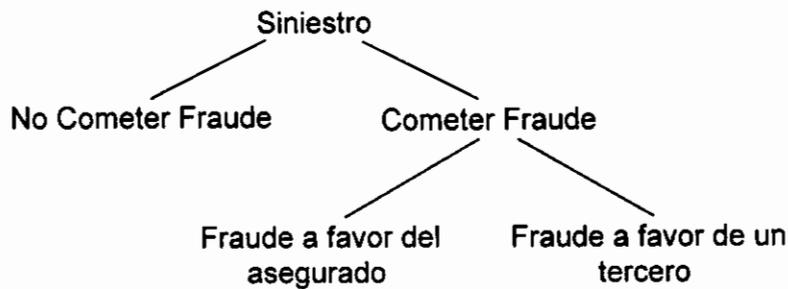


Figura 12

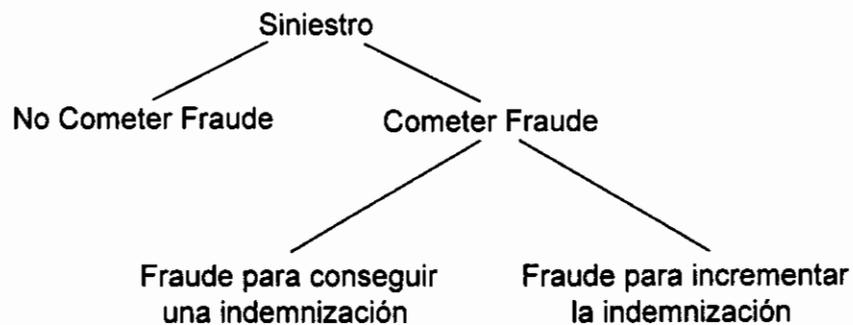


Figura 13

La complejidad que se genera desde el punto de vista de la modelización se ve compensada por la obtención de unos resultados muy atractivos de cara a generar herramientas muy sofisticadas de detección de fraude. Los coeficientes estadísticamente significativos en el nivel intermedio del árbol de decisión indicarán que variables deben ser analizadas para diferenciar entre comportamientos fraudulentos y no fraudulentos. Los coeficientes estadísticamente significativos en el nivel final de dicho árbol indicarán qué variables se han de analizar para detectar cada tipo de fraude y, por tanto,

determinarán cómo debe la entidad orientar la investigación de los siniestros. Se remite a las Conclusiones la presentación de las variables estadísticamente significativas en cada nivel de decisión.

7. CONCLUSIONES

La existencia de fraude en el seguro de automóviles provoca un incremento en el número y coste medio de los siniestros que ha despertado el interés de las entidades. La solución tradicional utilizada por las compañías ante la presencia de costes superiores a los esperados ha sido el incremento de las primas aplicadas. Sin embargo, este mecanismo no ha podido evitar los resultados técnicos negativos que han caracterizado al ramo de automóviles en los últimos años.

El objetivo perseguido en el presente libro es modelizar el comportamiento deshonesto en las declaraciones de siniestros del seguro del automóvil. Para ello, se ha utilizado información de una muestra de siniestros del mercado asegurador español y se ha aplicado la metodología econométrica para especificar, estimar y validar los modelos.

La mayor preparación del personal asegurador, la cooperación entre compañías y la colaboración con las fuerzas de seguridad, entre otras posibilidades, son esenciales. Sin embargo, no son totalmente efectivas si no van acompañadas de mecanismos de control suficientemente validados desde un punto de vista estadístico, hecho que ha dado lugar a una involucración de la comunidad científica en el tratamiento del fraude.

Los resultados obtenidos constituyen un modelo explicativo a la vez que un mecanismo de detección. En el tema estudiado a lo largo de este trabajo confluyen dos intereses: 1) realizar un análisis microeconómico del fraude y 2) elaborar un procedimiento que permita agilizar la tramitación y la investigación de los siniestros.

Este segundo aspecto permite desarrollar una herramienta de decisión que puede ayudar a las entidades aseguradoras a controlar el fraude. Más aún, el libro profundiza en la identificación de los factores que influyen en los diversos comportamientos deshonestos. Por lo tanto, los modelos no sólo ayudan a decidir si se debe investigar o no un siniestro sospechoso, sino a concretar de qué forma debe investigarse. Los modelos estudiados permiten concluir que algunas características se asocian a un tipo de fraude y no a otro y, de este modo, aportan evidencias de cómo encaminar el seguimiento de la reclamación.

A pesar de que en otros países (fundamentalmente, EE.UU, Canadá y Francia) es posible encontrar grupos de trabajo o instituciones directa o indirectamente asociadas al estudio del fraude, hasta el momento no podíamos decir lo mismo a nivel nacional. La investigación presentada es, por tanto, totalmente novedosa en el marco del seguro automovilístico español, aunque las diferencias metodológicas en relación al tratamiento econométrico realizado por otros autores permiten ampliar dicho calificativo a nivel internacional.

Las principales conclusiones obtenidas a lo largo del trabajo se detallan a continuación:

1.- La importancia concedida en la última década a la detección de comportamientos fraudulentos en el mercado asegurador en general queda patente a nivel nacional e internacional. La prevención del fraude en la selección de riesgos y en la liquidación de siniestros es uno de los objetivos reflejados en la nueva Ley de Ordenación y Supervisión de los Seguros Privados en España, en vigor desde 1995. Además, como resultado de la integración comunitaria, todos los países miembros están homogeneizando sus organismos sectoriales de lucha contra el fraude. El seguro del automóvil es el más afectado y, dado que suele llevar asociados resultados técnicos negativos, es el que ha despertado un mayor interés por el tratamiento del problema, aunque se acepta la existencia de fraude en la práctica totalidad de los ramos aseguradores. A pesar de todo ello, no existen en Europa trabajos científicos que traten el fenómeno del fraude en el

seguro del automóvil desde un punto de vista cuantitativo. La única excepción son las investigaciones de Pierre Picard, que mediante un enfoque de teoría económica demuestran la necesidad de crear organismos de control de fraude. Otras aportaciones no van más allá de la mera sugerencia de porcentajes de fraude y de una constante, pero pasiva, preocupación por el problema.

2.- Las estadísticas existentes sobre el fraude en el seguro automovilístico español muestran divergencias notables entre el porcentaje de casos en los que se sospecha una posible actuación deshonestas y el porcentaje de casos en los que realmente llega a confirmarse dicho comportamiento. Mientras que para los primeros, las cifras revelan valores superiores en algunos casos al 20% de los siniestros declarados, para los segundos, la cifra es notablemente inferior y tiende a situarse entre el 1 y el 10% (C.E.S., 1992). Estos resultados son completamente concordantes con los observados por investigadores de EE.UU. y Canadá. Así, cualquier sistema dirigido a identificar las actuaciones deshonestas sospechosas que realmente responden a comportamientos fraudulentos será de gran valía para las compañías y la sociedad en general.

3.- La existencia de tipos de comportamientos muy variados dificulta la obtención de una definición genérica de fraude. Resulta patente que, en cualquiera de los casos, la actuación deshonestas por parte del asegurado supone la alteración del principio de buena fe, que sustenta el contrato de seguro. Así pues, la consideración de las diferentes posibilidades que el individuo tiene a su alcance a la hora de actuar fraudulentamente puede ayudar de modo notable a la compañía para generar un mecanismo de control.

La tipología existente de fraude en el marco del seguro automovilístico español se concentra fundamentalmente en ocho categorías. Sin embargo, en el presente trabajo se ha adoptado una categorización más genérica en base a la agrupación de algunas de las clases anteriores. El objetivo es ceñirse a las características específicas del funcionamiento del sistema de seguros de automóviles en España. La clasificación se ha realizado en un doble sentido. Por un lado, se ha

considerado el fraude realizado por el asegurado en beneficio propio, *versus* el perpetrado en beneficio de un tercero; por otro, el fraude destinado a conseguir una indemnización que no corresponde por contrato, *versus* el dirigido a incrementar la indemnización derivada de la ocurrencia de un siniestro sujeto a cobertura.

La determinación de variables significativas en la explicación de las probabilidades de aparición de cada una de las categorías puede ayudar a que la compañía optimice su política de control de fraude. La sospecha sobre el tipo de fraude existente inducirá a analizar prioritariamente el comportamiento de aquellas variables para las que se ha demostrado poder explicativo sobre el mismo.

4.- La dificultad para obtener muestras con información completa sobre las variables que pueden ser utilizadas en la modelización del fraude es una de las principales razones que justifican la escasez de tratamientos metodológicos exhaustivos. Weisberg y Derrig (1993) utilizan una muestra no aleatoria formada por 127 siniestros (aproximadamente en un 60% de los mismos se sospecha la existencia de fraude). Belhadji y Dionne (1997) usan una muestra aleatoria formada por 2.068 siniestros (un 93.77% de los casos no son fraudulentos; en un 5.46% existe sospecha de fraude y tan sólo en un 0.87% se detecta existencia de fraude).

En nuestro caso, la muestra utilizada contiene información para un total de 1995 expedientes de siniestros, de los cuáles el 50% (997 casos) contienen fraude detectado. El 50% restante corresponde a siniestros clasificados por la compañía como no fraudulentos. Para estos últimos resulta más correcto hablar en términos de "ausencia de fraude detectado", dado que la entidad los ha clasificado como legítimos, pero es posible que, aunque haya realizado alguna investigación, no haya logrado verificarlo. La existencia en el diseño de la muestra de estratificación endógena (con el objetivo de obtener una muestra representativa para los siniestros fraudulentos, la selección de expedientes no se ha realizado

de forma estrictamente aleatoria) ha estado presente en la modelización mediante la incorporación de ponderaciones en los procesos de estimación.

Los datos, referidos a un amplio conjunto de variables, han sido depurados con el objetivo de recoger información precisa para un conjunto de indicadores, presentados en la literatura como señales de fraude. Las variables literales (por ejemplo, la que recoge la descripción del relato del siniestro) han sido objeto de tratamiento textual. De esta forma se ha conseguido averiguar, entre otras cosas, cuáles son las circunstancias más propicias a la ocurrencia de siniestros fraudulentos. Mediante un identificador de coincidencia de apellidos entre asegurado y contrario se ha pretendido detectar relaciones familiares entre las partes que intervienen en el accidente.

El estudio de las fechas relacionadas con diferentes aspectos del siniestro y de la póliza ha permitido realizar tratamientos exhaustivos asociados, entre otras cosas, a la determinación del plazo que transcurre entre la ocurrencia del siniestro y la comunicación del mismo a la entidad, o a la detección de casos en los que la ocurrencia del siniestro se produce entre la fecha de efecto y la fecha de emisión de la póliza.

Para numerosos atributos, como los relativos a la intervención policial en el siniestro o a la participación de testigos, se han utilizado variables *dummies* que han permitido realizar una recodificación en términos de la presencia o ausencia de una determinada situación. Otras variables, como la edad del individuo y su antigüedad de carnet en el momento del accidente, o el historial de siniestros que posee en la entidad, han sido directamente introducidas en las ecuaciones de regresión.

5.- Aunque las razones que pueden llevar al asegurado a cometer fraude pueden ser variadas (deseo de recuperar la prima pagada, intención de beneficiar a un amigo, conseguir la cobertura para situaciones excluidas de la póliza,...), su actuación está fundamentada principalmente en dos hechos concretos: la mayor

o menor probabilidad con la que cree que será detectado y las circunstancias en las que se produce el siniestro (que incluyen el caso de que el accidente sea provocado por el asegurado).

La utilidad final esperada por el individuo, definida por Picard (1996) como el resultado de adicionar la utilidad esperada al actuar de forma honesta y la esperada al actuar fraudulentamente (ambas utilidades ponderadas por la probabilidad de elegir una u otra alternativa, respectivamente), debería recoger la existencia de diferentes clases de fraude. Para cada una de ellas el asegurado tendrá en cuenta dos situaciones posibles. Si su comportamiento es detectado por la compañía sufrirá una penalización (en el caso español normalmente será la expulsión de la compañía o el pago únicamente de la indemnización que corresponde por contrato). Si no lo es, él y/o el contrario percibirán una cantidad monetaria en concepto de reparación de daños, no justificada.

El proceso necesario para estimar la utilidad esperada por el asegurado a la hora de decidir si comportarse honestamente o cometer un determinado tipo de fraude supondrá la cuantificación de dos conceptos básicos. Por un lado, se determinará la probabilidad de que el asegurado elija cada una de las alternativas; por otro, se cuantificará la probabilidad de que, tras la aplicación del modelo, se detecte la existencia de actuación ilícita. El alcance de ambos objetivos ha de estar presente en la selección de la técnica econométrica utilizada en la modelización.

6.- El trabajo desarrollado está dirigido a la aplicación de métodos cuantitativos en el diseño de un modelo de control de fraude, teniendo en cuenta la tipología existente para el mismo. El objetivo principal se ha centrado en considerar, dentro de la modelización, el proceso de decisión seguido por el asegurado a la hora de elegir una determinada alternativa. Este proceso puede ser unietápico o polietápico. En el segundo caso, el individuo decide, en primer lugar, cometer fraude o bien no hacerlo y, si opta por la primera alternativa, selecciona la forma más conveniente de realizarlo. La representación gráfica de árboles de decisión ayuda a comprender el proceso seguido, teniendo en cuenta, por tanto, una

elección secuencial o por etapas. La aplicación de técnicas de regresión logística (modelos lógit multinomiales y anidados) ha permitido cuantificar la probabilidad de elegir cada una de las alternativas, determinando, asimismo, las variables estadísticamente significativas en la explicación de cada comportamiento.

Los resultados obtenidos permiten interpretar el árbol de decisión atendiendo a un doble criterio: el asegurado realiza una elección por etapas bajo el deseo de maximizar la utilidad esperada de su comportamiento (conclusión obtenida tras la aplicación de los modelos lógit multinomiales), o bien lo hace para estructurar mentalmente las diferentes opciones que posee con el objetivo de minimizar los costes -en tiempo, en preparación del siniestro,...- asociados a la elección (conclusión derivada tras aplicar modelos lógit anidados).

No podemos decidir si es mejor utilizar un modelo multinomial o un modelo anidado. Ambos proporcionan conclusiones complementarias. El primero constituye una herramienta de clasificación fácil de implementar pero, a diferencia del anidado, no orienta suficientemente sobre la forma de conducir la investigación. Sirve para identificar las variables relevantes o aquellas características en que deberían fijarse los tramitadores para luego enviar el expediente a una unidad especial de investigación. Creemos que algunas variables importantes no aparecen como significativas en este modelo porque no son capaces de discriminar entre las categorías de la variable dependiente (incluyen los siniestros honestos y los distintos tipos de fraude). En cambio, el modelo anidado, al tener en cuenta probabilidades condicionadas, puede tener una utilidad muy diferente. En este caso, el modelo discrimina entre las situaciones sospechosas y las que no lo son. Una vez que identifica el comportamiento deshonesto es capaz de encaminar la investigación hacia el tipo de fraude más probable. Esto se consigue determinando qué variables son explicativas para cada una de las alternativas de fraude. Los resultados permiten concluir que determinadas variables que no son significativas en el modelo multinomial, sí lo son en este caso. Una vez se sospecha de la existencia de

fraude, ciertas circunstancias resultan explicativas del tipo de proceder más probable.

La implementación de un modelo anidado, aunque más compleja, permite estudiar el comportamiento de las probabilidades condicionadas, aspecto que no había sido tratado hasta este momento.

Los modelos planteados permiten, asimismo, determinar la capacidad de predicción de los comportamientos fraudulentos. Los resultados obtenidos para el porcentaje de siniestros correctamente clasificados en sus categorías respectivas son interpretados como la capacidad que la compañía tiene de detectar, en caso de que exista, un determinado tipo de fraude. La optimización del criterio probabilístico utilizado en la delimitación de las zonas de clasificación de los siniestros dentro de una determinada categoría ha permitido mejorar notablemente la calidad predictiva de los modelos utilizados. La demostración del proceso a seguir para optimizar el punto de corte en el marco de la modelización logística multinomial y anidada constituye, sin duda, una de las principales aportaciones, en lo que a la vertiente teórica se refiere.

En la presentación final se ha optado por dar mayor relevancia a la capacidad explicativa de los modelos estimados. Es por ello que no se ha dividido la muestra con la finalidad de poder validar los resultados con datos no empleados en la estimación. A ello se ha sumado la intención de mantener el tamaño muestral que conduce a la estimación de los parámetros, lo mayor posible.

En relación a las variables explicativas cabe destacar un hecho importante: para todas ellas la compañía dispone de información de forma inmediata, de manera que la investigación sobre la existencia o no de fraude puede realizarse sin una demora excesiva.

A diferencia de los estudios realizados por otros autores, los modelos presentados no incluyen indicadores subjetivos o aquéllos para los que la

información se recaba durante la tramitación del siniestro pero no en el momento inicial de su apertura.

7.- Las diferencias existentes con las aportaciones realizadas por otros autores se hacen patentes en el tratamiento metodológico aplicado y en los indicadores de fraude utilizados, además de que aquí consideramos un conjunto de alternativas de elección más amplio que el asociado a la dicotomía fraude/no fraude.

7.1.- El enfoque planteado es distinto al utilizado por Picard (1996), que realiza un análisis microeconómico teórico del comportamiento deshonesto de los asegurados y sugiere la creación de un organismo (agencia) común a todas las aseguradoras, que se encargue total o parcialmente de los gastos derivados de la investigación de siniestros, para fomentar que las entidades auditen el fraude sin alejarse del criterio económico de minimización de costes.

Nuestro trabajo está más en la línea de los estudios realizados por Weisberg y Derrig (1993, 1998) y por Belhadji y Dionne (1997), ya que presentan análisis cuantitativos para identificar los indicadores de fraude.

7.2.- Weisberg y Derrig (1993, 1998) tienen en cuenta la existencia de categorías genéricas de fraude. Se consideran cuatro posibilidades a la hora de clasificar el siniestro: legítimo, con sospecha de fraude planeado -siniestro construido o simulado-, con sospecha de fraude oportunista -ocurrencia del accidente real pero sin daños aparentes- y con sospecha de gastos médicos inflados.

Sin embargo, mientras que en este libro la variable dependiente es cualitativa y viene definida por el tipo de comportamiento observado para los individuos de la muestra, Weisberg y Derrig especifican como variable dependiente la "sospecha" de existencia de una determinada actuación puntuada de 0 a 10. En su modelización hacen intervenir una serie de indicadores relacionados con diferentes aspectos del siniestro, de las partes que intervienen y de los daños

derivados (otorgan especial importancia a los costes médicos). Mediante la aplicación de técnicas de regresión múltiple, determinan cuáles resultan estadísticamente significativos en la explicación de la dependiente. Las diferencias entre nuestro trabajo y el de dichos autores quedan patentes no sólo en la metodología aplicada, sino también en el objetivo final perseguido. Para estos últimos, el resultado final del estudio es la determinación de las características básicas que permiten fundamentar la existencia de sospecha de un tipo de fraude (los patrones de comportamiento definidos son susceptibles de implementación automática). En este libro, el análisis de la significación estadística de las variables introducidas en el modelo goza de una especial importancia, pero forma parte de un proceso más amplio, dirigido a cuantificar la probabilidad de existencia de cada tipo de fraude y a aumentar la probabilidad de que la compañía logre detectarlos lo antes posible.

7.3.- En relación al trabajo realizado por Belhadji y Dionne (1997), las principales similitudes se encuentran en los objetivos perseguidos tras la modelización. La cuantificación de la probabilidad de existencia o sospecha de fraude (previa selección de los indicadores a introducir en la especificación del modelo y análisis estadístico de los mismos) y la determinación del criterio probabilístico óptimo a utilizar por la compañía para clasificar correctamente los siniestros en su categoría respectiva (fraude/no fraude) son los principales resultados extraídos del análisis realizado por ambos autores. Presentan, también, una aproximación al tratamiento del fraude teniendo en cuenta los costes de investigación (análisis coste-beneficio asociado a la realización de un estudio exhaustivo del siniestro). Sin embargo las diferencias se hacen evidentes al considerar el tratamiento metodológico aplicado. Así, y respecto a nuestra aportación, Belhadji y Dionne no tienen en cuenta la existencia de diferentes alternativas al alcance del individuo que quiere actuar fraudulentamente. En su estudio, todos los resultados aparecen asociados a la dicotomía fraude/no fraude, sin que puedan interpretarse teniendo en cuenta un proceso de decisión por etapas. La técnica econométrica utilizada es, de esta forma, más simple: frente a la aplicación de modelos lógit multinomiales y anidados, ellos proponen el uso de

un modelo próbit. Entendemos que esta limitación les viene provocada por el tipo de información estadística de que disponen. Al tratarse de una muestra aleatoria de siniestros, el porcentaje de casos fraudulentos es excesivamente bajo.

La definición de patrones de comportamiento por parte de estos autores queda limitada al caso general de presencia-ausencia de fraude y no es posible determinar la existencia de atributos directamente relacionados con formas alternativas de fraude. Ello es especialmente peligroso para aquellas variables cuya influencia en la probabilidad de cometer un cierto tipo de fraude difiere de la de otro tipo, sobre todo si el parámetro asociado cambia de signo.

Tanto Weisberg y Derrig como Belhadji y Dionne utilizan indicadores de fraude caracterizados, en algunos casos, por una elevada subjetividad (por ejemplo, la valoración que los tramitadores hacen de la conducta del asegurado) y, en otros, por una tardanza en su disposición (número de días en que el asegurado ha recibido tratamiento médico,...). Ésta es también una de las grandes diferencias con nuestro trabajo, dado, que como hemos manifestado anteriormente, las variables utilizadas en la modelización poseen una elevada objetividad y para todas ellas la compañía dispone de información de forma inmediata. No obstante, la comparación con los indicadores utilizados a nivel internacional nos ha alertado sobre la existencia de ciertas variables, cuya información puede ayudar notablemente en el estudio del fraude, que no son recogidas actualmente en los partes de siniestros de las entidades españolas (por ejemplo, la situación socioeconómica del demandante o indicadores indirectos de la misma).

Respecto al análisis de la significación individual de las variables no se han encontrado diferencias alarmantes respecto a las conclusiones de otros autores, o que no puedan ser explicadas por los diferentes sistemas aseguradores.

8.- Pasamos a continuación a destacar las principales conclusiones de los resultados obtenidos en la modelización, tanto multinomial como anidada.

Las variables explicativas que tienen un efecto significativo en todos los modelos son la coincidencia de apellidos entre las partes implicadas, la intervención policial, los antecedentes de siniestralidad que el asegurado posee en la compañía y la descripción textual de lo ocurrido. Podríamos calificar a estas variables como las más relevantes o las que se ha de considerar siempre en la investigación del fraude. Esta conclusión coincide con la de otros investigadores que insisten en la conveniencia de que las autoridades emitan informes de siniestros con más frecuencia. El resto de variables corresponden a circunstancias propias del suceso y, por lo tanto, poco se puede hacer institucional o corporativamente para cambiarlas.

Cuanto mayor es el historial de siniestros que el asegurado posee en la compañía, mayor es la probabilidad de que cometa cualquier tipo de fraude (el conocimiento de la forma en que la entidad tramita los siniestros -rapidez, peritajes poco exhaustivos,...- y la no detección de posibles fraudes anteriores, pueden ser algunas de las razones que ayudan a interpretar el resultado obtenido). La coincidencia de apellidos entre la parte asegurada y la parte contraria es, sin duda, uno claro indicador de cooperación que induce a sospechar de la existencia de fraude. La presencia en la declaración del asegurado de descripciones alusivas al estacionamiento, adelantamiento, cruce y marcha atrás aparece relacionada con un aumento en la propensión de que exista cualquiera de los tipos de fraude comentados.

A modo de síntesis presentamos otras variables que muestran cierto poder explicativo en los modelos multinomiales y, por lo tanto, las podemos considerar interesantes para diferenciar tipos de comportamientos.

- La ocurrencia del siniestro a altas horas de la madrugada y la demora en la comunicación del siniestro son variables directamente relacionadas con la aparición de fraude. En relación a la existencia de testigos del accidente se observa una influencia contrapuesta en los tipos de fraude analizados: produce una disminución en la probabilidad de aparición de fraude para conseguir una

indemnización y de fraude cometido en beneficio propio, e incrementa la probabilidad de aparición de fraude cometido en beneficio de un tercero (existencia de testigos falsos). Ello constituye una evidencia de que las variables pueden tener influencia distinta en la probabilidad de aparición de cada tipo de fraude. Este resultado es muy interesante puesto que otros autores concluyen que la influencia de esta variable en la presencia de fraude no es significativa, pese a que éste no sea el resultado esperado. En nuestro caso se confirma su diferente influencia en los distintos tipos de comportamiento, que se anula si se consideran éstos agregadamente.

- Cuando el vehículo asegurado es un turismo de uso particular disminuye la probabilidad de aparición de fraude (la variable es estadísticamente significativa en la explicación de fraude en beneficio propio y de fraude para conseguir una indemnización). La propensión a defraudar es mayor, por tanto, en aquellos casos en los que la utilización del turismo es diferente o cuando los vehículos pertenecen a otras categorías como motocicletas o camiones.
- A medida que aumenta la edad del individuo disminuye la probabilidad de que cometa fraude, siendo más frecuente la aparición de comportamientos fraudulentos en los individuos jóvenes (con edades, básicamente, entre 25 y 34 años, en el momento del siniestro).
- La antigüedad de carnet del conductor asegurado, considerada en la modelización multinomial de fraude a favor del asegurado y a favor del tercero, no goza de significación estadística. La misma conclusión se obtiene para la existencia de franquicia en la póliza y para la contratación de cobertura de accesorios, introducidas al modelizar la probabilidad de aparición de fraude para conseguir una indemnización o para incrementarla.

Respecto a la modelización del tipo de fraude en las ramas inferiores de un esquema anidado, como hemos señalado anteriormente, las variables

significativas indican cómo dirigir la investigación del siniestro, es decir, como identificar cada tipo de fraude.

- Cuanto más antiguo sea el vehículo del asegurado mayor es la probabilidad de que actúe fraudulentamente en beneficio propio y no en beneficio de otro (por ejemplo, reparando averías cada vez más frecuentes) aunque se ha detectado la existencia de un punto de inflexión (aproximadamente, en el séptimo año de antigüedad) a partir del cual dicha probabilidad disminuye, como consecuencia, probablemente, del bajo valor venal del automóvil.
- Cuando se detecta que el siniestro ha ocurrido entre la fecha de efecto y la fecha de emisión de la póliza (por este orden) aumenta la probabilidad de que el asegurado haya cometido fraude para conseguir una indemnización y no para incrementarla.
- Cuando el asegurado tiene contratada únicamente cobertura a “terceros” y cuando admite la culpa del siniestro, aumenta la probabilidad de que cometa fraude en beneficio de un tercero frente a que lo haga en el suyo propio, como es de esperar.

Todos los modelos presentados en los Capítulos anteriores son globalmente significativos y los estadísticos de ajuste son satisfactorios.

9.- Las líneas de mejora propuestas para un futuro, en relación al proceso de recogida de datos, se centran básicamente en dos aspectos: 1) el análisis de la existencia de comportamientos fraudulentos considerando categorías de fraude alternativas a las presentadas en el libro y, 2) el diseño de una base de datos más amplia que nos permita trabajar con un mayor número de casos para cada una de las alternativas consideradas.

9.1.- En relación a la primera línea, creemos que sería positivo, no sólo a nivel investigador, sino también a nivel de entidades, el diseño de “patrones de

tipos de fraude”, adicionales a los presentados en el libro, que ayudase a los tramitadores a clasificar los siniestros con fraude detectado en la categoría conveniente. En la actualidad ya se trabaja con categorías de fraude preestablecidas, pero la existencia de un importante número de casos en los que la compañía no identifica el tipo de fraude nos ha alertado de la dificultad asociada al proceso de clasificación. El tramitador puede tener problemas a la hora de identificar de forma explícita la clase de fraude existente en el siniestro. En este sentido podría pensarse en alternativas adicionales de clasificación a las presentadas en este trabajo, entre las que queremos destacar la que nos llevaría a diferenciar (siguiendo la pauta establecida por el Insurance Fraud Bureau de Massachusetts) entre aquellos siniestros en los que existe fraude “*a priori*” (el asegurado fabrica deliberadamente el siniestro) y aquellos en los que existe fraude “*a posteriori*” (la ocurrencia del siniestro es legal, pero el asegurado y/o el contrario la utilizan para obtener un beneficio no justificado).

9.2.- El reducido número de observaciones para cada una de las submuestras utilizadas (teniendo en cuenta cada una de las categorías de fraude planteadas) ha dificultado la realización de un estudio de indicadores más exhaustivo dirigido, fundamentalmente, a la inclusión de interacciones entre variables. La disponibilidad de un mayor número de casos para cada tipo ayudaría a refinar los patrones de comportamiento definidos, teniendo en cuenta la posibilidad de recoger en el modelo, mediante la creación adecuada de variables, algunas de las normas que regulan el funcionamiento del seguro del automóvil (considerando, por ejemplo, de forma conjunta la culpa del siniestro y el tipo de cobertura,...). Las interacciones entre variables utilizadas en la estimación han estado dirigidas a la definición de atributos de elección específicos para cada una de las alternativas consideradas.

10.- Desde el punto de vista metodológico, las líneas de mejora irían encaminadas al tratamiento de un aspecto fundamental: el estudio longitudinal de la vida de los siniestros. Ello requeriría trabajar con muestras de expedientes desde su apertura hasta su liquidación. De esta forma, se lograría detectar qué

indicadores inciden en cada etapa de la tramitación y, por tanto, qué momento es óptimo para realizar los controles de fraude.

11.- Para terminar, las líneas de mejora propuestas, desde el punto de vista de la eficiencia asociada a la implantación de un mecanismo de control de fraude por parte de una entidad, son fundamentalmente dos: 1) la realización de un análisis exhaustivo de la influencia que el fraude tiene en el coste del siniestro para la entidad y 2) la determinación de la rentabilidad asociada a la instalación de un sistema automático de control de fraude teniendo en cuenta la relación coste-beneficio asociada al mismo.

8. BIBLIOGRAFÍA

- A.I.B. (1997) 1997 Annual Report. *Automobile Insurers Bureau of Massachusetts*. Boston.
- Addis, K.K. (1989) "The Fraud Squad". *Security Management*, September, vol. 33, nº 9, 53-57.
- Allingham, M.G. y A. Sandmo (1972) "Income Tax Evasión: A Theoretical Analysis". *Journal of Public Economics*, vol. 1, nº 3/4, 323-338.
- Amemiya, T. (1985) *Advanced Econometrics*. Ed. Basil Blackwell Ltd., Oxford, U.K.
- Anderson, K. (1990) "The Phoney-Claims Drain". *Insurance Review*, vol. 51, nº 8, 20-23.
- Artís, M., M. Ayuso y M. Guillén (1997) "Modelling Different Types of Automobile Insurance Fraud Behaviour in the Spanish Market". Ponencia presentada en el primer Congreso Internacional en *Insurance: Mathematics and Economics*, celebrado en Amsterdam del 25 al 27 de agosto de 1997.
- Artís, M., M. Ayuso y M. Guillén (1998a) "Aplicación de un Modelo Lógit a la Cuantificación de la Probabilidad de Fraude en el Seguro del Automóvil". *Working Paper*. Departamento de Econometría, Estadística y Economía Española. Universidad de Barcelona.
- Artís, M., M. Ayuso y M. Guillén (1998b) "Modelling Different Types of Automobile Insurance Fraud Behaviour in the Spanish Market". Aceptado para su publicación en la revista *Insurance: Mathematics and Economics*, Elsevier Science Publishers, North-Holland. Número de páginas: 31.
- Ayuso, M. (1995) *El Fraude en el Seguro del Automóvil*. Trabajo de Investigación realizado dentro del Programa de Doctorado *Economía y Territorio: Análisis Cuantitativo*. Departamento de Econometría, Estadística y Economía Española.
- Ayuso, M. (1998) *Modelos Económicos para la Detección del Fraude en el Seguro del Automóvil*. Tesis Doctoral. Departamento de Econometría, Estadística y Economía Española. Universidad de Barcelona.
- Ayuso, M. y M. Guillén (1995) "Modelos de Elección del Fraude en el Seguro de Automóviles". *Actas de la IX Reunión de la Asociación Científica Europea de Economía Aplicada ASEPELT-España* celebrada en Santiago de Compostela, 22 y 23 de junio de 1995, vol. II (*Economía Sectorial*), 383-394.
- Ayuso, M. y M. Guillén (1996) "Modelos de Detección de Fraude en el Seguro del Automóvil". Aceptado para su publicación en *Cuadernos Actuariales*. Col·legi d'Actuaris de Catalunya. Número de páginas: 18.

- Ayuso, M., Guillén, M. y M. Artis (1998) "Técnicas Cuantitativas para la Detección del Fraude en el Seguro del Automóvil". Aceptado para su publicación en los *Anales del Instituto de Actuarios Españoles*. Número de páginas: 18.
- Belhadji, E.B. y G. Dionne (1997) "Development of an Expert System for the Automatic Detection of Automobile Insurance Fraud". *Working Paper 97-06*. École des Hautes Études Commerciales. Université de Montréal.
- Bernouilli, D. (1954) "Exposition of a New Theory of the Measurement of Risk". *Econometrica*, vol. 22, nº 1, 23-36.
- Betés de Toro, A. (1996) "La Estafa en el Seguro". *Actualidad Aseguradora*, nº 2, 22 de enero, 9.
- Bollinger, C. R. y M. H. David (1997) "Modeling Discrete Choice With Response Error: Food Stamp Participation". *Journal of the American Statistical Association*, vol. 92, nº 439, 827-835.
- Bond, E.W. y K.J. Crocker (1997) "Hardball and the Soft Touch: The Economics of Optimal Insurance Contracts with Costly State Verification and Endogenous Monitoring Costs". *Journal of Public Economics*, vol. 63, nº 2, 239-264.
- Borch, K. (1990) *Economics of Insurance*. Advanced Textbooks in Economics, vol. 29, Ed. North-Holland, Amsterdam.
- Brockett, P.L., X. Xia y R. A. Derrig (1995) "Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud". Ponencia presentada en *American Risk and Insurance Association Annual Meeting*, Seattle.
- Brotman, B.A. y P. Fox (1988) "The Impact of Economic Conditions on the Incidence of Arson: Comment". *Journal of Risk and Insurance*, vol. 55, nº 4, 751-755.
- Brundin, I. y F. Salanié (1997) "Fraud in the Insurance Industry: An Organizational Approach", preliminary, April 1997. *mimeo*.
- Bujold, L., G. Dionne y R. Gagné (1997) "Assurance Valeur à Neuf et Vols d'Automobiles: Une Étude Statistique". *Cahier de recherche 97-01*, février 1997. École des Hautes Études Commerciales. Université de Montréal.
- Caron, L. y G. Dionne (1996) "Insurance Fraud Estimation: More Evidence from the Quebec Automobile Insurance Industry". *Working Paper 96-02*, december 1996. École des Hautes Études Commerciales. Université de Montréal.
- C.E.S. (1992) *El Fraude en el Seguro de Automóviles*. Centro de Estudios del Seguro. Madrid.
- Cercós, J. y R. Mayol (1990) Seguro de Automóviles. *Tratado General de Seguros, Teoría y Práctica de los Seguros Privados*, tomo III, vol. 3. Ed. Consejo General de los Colegios de Agentes y Corredores de Seguros de España, 145-301.
- Chao, J. (1994) "Nuevo Frente contra el Fraude". *Mercado Previsor*, nº 195, 30 de octubre, 12-13.

- Clarke, M. (1990) "The Control of Insurance Fraud. A comparative view". *The British Journal of Criminology*, vol. 30, nº 1, 1-23.
- Cloninger, D.O. (1990) "Arson and Abandonment: A Restatement". *The Journal of Risk and Insurance*, vol. 57, nº 3, 540-545.
- Cobo, P. (1993) *Manual de Investigación de Siniestros y Lucha contra el Fraude en el Seguro de Automóviles*. Ed. Mapfre, Madrid.
- Comité Europeo de Seguros (1996) *Le Guide de l'anti-fraude à l'assurance en Europe*. CEA Info, Hors - Série nº 4, Mai 1996.
- Correlation Research, Inc. (1989) *Marb Baseline Study of Bodily Injury Liability Claims*.
- Correlation Research, Inc. (1990) *AIB Study of 1989 Bodily Injury Liability Claims*.
- Correlation Research, Inc. (1991) *AIB Study of 1989 PIP Claims*.
- Correlation Research, Inc. (1993) *AIB Study of Suspected Fraud. Supplementary Coding Form for 1989 BI and PIP Claims*.
- Crocker, K.J. y J. Morgan (1998) "Is Honesty the Best Policy? Curtailing Insurance Fraud through to Optimal Incentive Contracts". *Journal of Public Economics*, vol. 106, nº 2, 355-375.
- Cummins, J.D. y R.A. Derrig (1993) "Fuzzy Trends in Property-Liability Insurance Claim Costs". *Journal of Risk and Insurance*, vol. 60, nº 3, 429-465.
- Cummins, J.D. y R.A. Derrig (1997) "Fuzzy Financial Pricing of Property-Liability Insurance". *North American Actuarial Journal*, vol. 1, nº 4, 21-44.
- Cummins, J.D. y S. Tennyson (1992) "Controlling Automobile Insurance Costs". *Journal of Economic Perspectives*, vol. 6, nº 2, 95-115.
- Cummins, J.D. y S. Tennyson (1996) "Moral Hazard in Insurance Claiming: Evidence from Automobile Insurance". *Journal of Risk and Uncertainty*, vol. 12, nº 1, 29-50.
- Cummins, J.D. y M.A. Weiss (1991) "The Effects of No-Fault on Automobile Insurance Loss Costs". *Geneva Papers on Risk and Insurance*, vol. 16, nº 1, 20-38.
- Cummins, J.D. y M.A. Weiss (1993) "Measuring Cost Efficiency in the Property-Liability Insurance Industry". *Journal of Banking and Finance*, vol. 17, nº 3, 463-482.
- Derrig, R.A. y L.K. Krauss (1994) "First Steps to Fight Workers' Compensation Fraud". *Journal of Insurance Regulation*, vol. 12, nº 3, 390-415.
- Derrig, R.A. y K.M. Ostaszewski (1994a) Fuzzy Sets. *Cas Ratemaking Seminar*. Atlanta, Georgia, March 10-11, 1994.
- Derrig, R.A. y K.M. Ostaszewski (1994b), "Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification", *4th International Conference on Insurance Solvency and Finance*. Philadelphia, PA. April 25-27, 1994.
- Derrig, R.A. y H.I. Weisberg (1995) "Identification and Investigation of Suspicious Claims. A Report on the AIB Study of 1993 Personal Injury Protection Claims". *AIB Filing on Fraudulent Claims Payment*, DOI Docket R95-12, Boston.

- Derrig, R.A. y H.I. Weisberg (1996) "A Report on the AIB Study of 1993 Personal Injury Protection and Bodily Injury Liability Claims. Coping with the Influx of Suspicious Strain and Sprain Claims". *AIB Filing on Cost Containment and Fraudulent Claims Payment*, DOI Docket R96-36, Boston.
- Derrig, R.A. y H.I. Weisberg (1997) "AIB PIP Claim Screening Experiment Interim Report. Understanding and Improving the Claim Investigation Process". *AIB Filing on Fraudulent Claims Payment*, DOI Docket R97-37, Boston.
- Derrig, R.A. y H.I. Weisberg (1998) "AIB PIP Claim Screening Experiment Final Report. Understanding and Improving the Claim Investigation Process". *AIB Filing on Fraudulent Claims Payment*, DOI Docket R98-41, Boston.
- Derrig, R.A., H.I. Weisberg, y X. Chen (1994) "Behavioral Factors and Lotteries Under No-Fault with Monetary Threshold: A Study of Massachusetts Automobile Claims". *Journal of Risk and Insurance*, vol. 61, nº 2, 245-275.
- Dionne, G. y R. Gagné (1997) The Non-Optimality of Deductible Contrats Against Fraudulent Claims: An Empirical Evidence in Automobile Insurance. *Working Paper 97-05*, june 1997. École des Hautes Études Commerciales. Université de Montréal.
- Elguero, J. M. (1996) "Código Penal: ¿Cómo Afecta al Fraude en el Seguro?. *Actualidad Aseguradora*, nº 2, 22 de enero, 36-37.
- Fagart, M.C. y P. Picard (1997) "Optimal Insurance Under Random Auditing". Preliminary and incomplete, September 1997. *mimeo*.
- Fernández de Castro, J. y Tugores, J. (1988) *Fundamentos de Microeconomía*. Ed. McGraw-Hill, Madrid.
- Fernández de Castro, J. y Tugores, J. (1992) *Fundamentos de Microeconomía*. Segunda Edición. Ed. McGraw-Hill, Madrid.
- García, J.L. (1989) "Fraud: The Hidden Element in Health Care Costs". *Compensation and Benefits Management*, vol. 6, nº 1, 49-51.
- Garrigues, J. (1974) *Curso de Derecho Mercantil*, vol. 2, 6ª edición revisada con la colaboración de F. Sánchez Calero, Ed. Aguirre, Madrid.
- González, Y. (1993) *La Demanda de Seguro Sanitario: Una Elección Bajo Incertidumbre*. Tesis Doctoral. Departamento de Econometría, Estadística y Economía Española de la Universidad de Barcelona.
- Greene, W.H. (1993) *Econometric Analysis*. Second Edition. Ed. Macmillan Press. New York.
- Greene, W.H. (1995) *LIMDEP, Version 7.0, User's Manual*. Ed. Econometric Software, Inc. New York.
- Greene, W.H. (1997) *Econometric Analysis*. Third Edition. Ed. Prentice-Hall International, Inc. New York.

- Greenouff, M. (1990) "El Fraude en el Seguro del Automóvil. Una comparación Internacional". Información Comité Europeo de Seguros, C.E.A. IV Jornadas Comunitarias del Seguro del Automóvil. *Previsión y Seguros*, nº 8, noviembre-diciembre, 109-120.
- Hausman, J. (1975) "An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models", *Econometrica*, vol. 43, nº 4, 727-738.
- Hausman, J. y D. McFadden (1984) "Specification Tests for the Multinomial Logit Model", *Econometrica*, vol. 52, nº 5, 1219-1240.
- Hensher, D. (1986) "Simultaneous Estimation of Hierarchical Logit Mode Choice Models" *Working Paper*, 24, MacQuarie University, School of Economic and Financial Studies.
- Hershberger, R.A. y R.K. Miller (1978) "The Impact of Economic Conditions on the Incidence of Arson: A Reply", *Journal of Risk and Insurance*, vol. 45, nº 2, 275-290.
- Hershberger, R.A. y R.K. Miller (1988) "The Impact of Economic Conditions on the Incidence of Arson: A Reply", *Journal of Risk and Insurance*, vol. 55, nº 4, 755-757.
- Hevia, E. y J.J. Lafuente (1992) *Como Luchar contra el Fraude en la Empresa*. Ed. Instituto de Auditores Internos de España, Madrid.
- Hoyt, R.E. (1990) "The Effect of Insurance Fraud on the Economic System". *Journal of Insurance Regulation*, vol. 8, nº 3, 304-315.
- I.C.E.A. (1995) "Manual de Lucha contra el Fraude en el Seguro de Automóviles". *Informe nº 607*, abril 1995.
- I.C.E.A. (1996) "Estadística de Fraudes en Seguros, 1995". *Informe nº 642*, mayo 1996.
- I.C.E.A. (1997) "Fraude: Una Lucha Permanente". *Actualidad Aseguradora*, nº12, 7 de abril, 60-62.
- I.C.E.A. (1998) "El Fraude al Seguro Español en 1997 (Cuánto, Cómo, Dónde)". *Informe nº 733*, julio 1998.
- Institute of Internal Auditors, Inc. (1985) "Deterrence, Detection, Investigation and Reporting of Fraud", *Statement on Internal Auditing Standards (SIAS)*, nº 3, 1-20.
- I.F.B. (1994) *1994 Annual Report*. Ed. Insurance Fraud Bureau of Massachusetts, Boston.
- I.F.B. (1994) *Insurance Fraud Research Register*. Ed. Insurance Fraud Bureau of Massachusetts, Boston.
- I.F.B. (1995) *1995 Annual Report*. Ed. Insurance Fraud Bureau of Massachusetts, Boston.
- I.F.B. (1995) *Insurance Fraud Research Register*. Ed. Insurance Fraud Bureau of Massachusetts, Boston.
- I.F.B. (1996) *1996 Annual Report*. Ed. Insurance Fraud Bureau of Massachusetts, Boston.
- I.F.B. (1996) *Insurance Fraud Research Register*. Ed. Insurance Fraud Bureau of Massachusetts, Boston.
- I.F.B. (1997) *1997 Annual Report*. Ed. Insurance Fraud Bureau of Massachusetts, Boston.

- I.F.B. (1998) "Focus Fraud". *Insurance Fraud Bureau of Massachusetts Publications*, vol. 5, nº 1.
- I.N.E.S.E. (1996a) "Organización del Sector en la Lucha contra el Fraude". Departamento de Investigación. *Actualidad Aseguradora*, nº 2, 22 de enero, 2-4.
- I.N.E.S.E. (1996b) "El Fraude en el Ramo de Autos en Europa". *Actualidad Aseguradora*, nº 2, 22 de Enero, 5-8.
- I.N.E.S.E. (1996c) "Fraude al Seguro: Alarma Roja". *Actualidad Aseguradora*, nº 24, 1 de julio, 2-5.
- I.N.E.S.E. (1996d) "¡Ojo al Fraude!". *Actualidad Aseguradora*, nº 24, 1 de julio, 6-8.
- Johnston, J. (1992) *Métodos de Econometría*. Ed. Vicens-Vives, Barcelona.
- Kleinbaum, D. G. (1992) *Logistic Regression: a self-learning text*. Ed. Springer, New York.
- Kreps, D.M. (1990) *A Course in Microeconomic Theory*. Ed, Harvester Wheatsheaf, New York.
- Koujianou, P. (1995) "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry". *Econometrica*, vol. 63, nº 4, 891-951.
- Lara, I. (1996) "La solución se llama FIVA". *Actualidad Aseguradora*, nº 2, 22 de enero, 34-35.
- Lemaire, J. (1985) *Automobile Insurance*. Ed. Kluwer-Nijhoff Publishing, U.S.A.
- Lemaire, J. (1990) Fuzzy Insurance. *Astin Bulletin*, vol. 20, nº 1, 33-55.
- Lemaire, J. (1995) *Bonus-Malus Systems in Automobile Insurance*. Ed. Kluwer Academic Publishers, U.S.A.
- Ley 122/1962, de 24 de diciembre, de Uso y Circulación de Vehículos de Motor (Boletín-Gaceta nº 310 de 27 de diciembre de 1962).
- Ley 50/1980, de 8 de octubre, de Contrato de Seguro (B.O.E. nº 250 de 11 de octubre de 1980).
- Ley 30/1995, de 8 de noviembre, de Ordenación y Supervisión de los Seguros Privados (B.O.E. nº 268 de 9 de noviembre de 1995).
- Ley 10/1995, de 23 de noviembre, Código penal (B.O.E. nº 281 de 24 de noviembre de 1995).
- Luce, R.D. (1959) *Individual Choice Behaviour*. Ed. Wiley, New York.
- Macho, I y D. Pérez (1994) *Introducción a la Economía de la Información*, 1ª edición, Ed. Ariel, Madrid.
- Maddala, G.S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs. Ed. Cambridge University Press, Cambridge.
- Manski, C. y S. Lerman (1977) "The Estimation of Choice Probabilities from Choice-Based Samples". *Econometrica*, vol. 45, 1977-1988.
- Manski, C y D. McFadden (eds.) (1981) *Structural Analysis of Discrete Data with Econometric Applications*. Ed. MIT Press, Cambridge, Massachusetts.

- Marter, S.S. y H.I. Weisberg (1992) "Medical Costs and Automobile Insurance: A Report on Bodily Injury Liability Claims in Massachusetts". *Journal of Insurance Regulation*, vol. 9, nº 3, 381-422.
- Marter, S.S. y H.I. Weisberg (1992) "Medical Expenses and The Massachusetts Automobile Tort Reform Law: First Review of 1989 Bodily Injury Liability Claims". *Journal of Insurance Regulation*, vol. 10, nº 4, 462-514.
- Martin Boyer, M. (1997) Insurance Taxation and Insurance Fraud. *Working Paper 97-10*, September 1997. École des Hautes Études Commerciales. Université de Montréal.
- Martín Pliego, F.J. (1994) *Introducción a la Estadística Económica y Empresarial*. Ed. AC, Madrid.
- McFadden, D. (1978) "Modelling the Choice of Residential Location". En *Spacial Interaction Theory and Planning Models (Studies in Regional Science and Urban Economics*, vol. 3). Ed. A. Karlquist et al., North-Holland, Amsterdam, 75-96.
- McFadden, D. (1981) "Econometric Models of Probabilistic Choice". En Manski, C. y D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications*, Ed. MIT Press, Cambridge, Massachusetts. 198-272.
- McFadden, D. (1983) "Econometric Analysis Qualitative Response Models". En Griliches, Z. y M. Intrilligator (eds.) *Handbook of Econometrics*, vol. 2, cap. 24, Ed. North-Holland, Amsterdam, 1395-1457.
- Morillas Jarillo, M.J. (1992). *El Seguro del Automóvil: El Aseguramiento Obligatorio de la Responsabilidad Civil Automovilística*. Ed. José M^a Bosch, Barcelona.
- Mossin, J. (1992) "Aspects of Rational Insurance Purchasing". En Dionne, G. y S. Hamigton (eds.) *Foundations of Insurance Economics (Readings in Economics and Finance)*. Ed. Kluwer Academic Publishers, Boston, 119-133.
- Münchener Rück (1988) *El Fraude en el Seguro*. Ed. Münchener Rückversicherungs-Gesellschaft. München.
- Münchener Rück (1994) *El Fraude en el Seguro de Personas*. Ed. Münchener Rückversicherungs-Gesellschaft, München.
- Norusis, M.J. (1990) *SPSS Advanced Statistics User's Guide*. SPSS Inc, U.S.A.
- Novalés, A. (1993) *Econometría*. Segunda Edición. Ed. McGraw-Hill, Madrid.
- Paul, M. (1995) "El Fraude en el Seguro de Autos enciende la Alarma en Europa". *La Tribuna de Automoción*, 15 de noviembre, 8-9.
- Picard, P. (1996) "Auditing Claims in the Insurance Market with Fraud: The Credibility Issue". *Journal of Public Economics*, vol. 63, nº 1, 27-56.
- Picazo, A. (1995) "La Eficiencia en los Seguros". *Revista de Economía Aplicada, sección de Notas*, vol. 3, nº 8, 197-215.
- Pratt, J. W. (1964) "Risk Aversion in the Small and in the Large". *Econometrica*, vol. 32, nº 1-2, 122-136.

- Pudney, S. (1989) *Modelling Individual Choice. The Econometrics of Corners, Kinks and Holes*. Ed. Basil Blackwell Ltd, New York.
- Real Decreto Legislativo 1301/1986 de 28 de junio 1986, (Ministerio de Economía y Hacienda). Automóviles-Comunidad Económica Europea. Adaptación de la Ley de Uso y Circulación de Vehículos de Motor al Derecho de las Comunidades Europeas (B.O.E. nº 155 de 30 de junio de 1986).
- Savage, L. (1954) *The Foundations of Statistics*. Ed. John Wiley and sons., New York.
- Solana de Ariza, G. (1988) *La Declaración Amistosa de Accidente de Automóvil y el Convenio de Indemnización Directa Español (C.I.D.E.)*. Ed. Einsa, Madrid.
- Taricco, A. (1988) "Medical Consultant Key to Dealing Effectively with Insurance Fraud and Abuse", *Physician Executive*, September/October, 22-25.
- Texto refundido de la Ley de Uso y Circulación de Vehículos de Motor, aprobado por Decreto 632/1968 de 21 de marzo (B.O.E. nº 85 de 8 de abril de 1968).
- Tennyson, S. (1997) "Economic Institutions and Individual Ethics: A Study of Consumer Attitudes toward Insurance Fraud". *Journal of Economic Behavior and Organization*, vol. 32, 247-265.
- U.N.E.S.P.A. (1990) *Estadística de Seguros Privados. Análisis Económico Actuarial del Sector Asegurador Español, Ejercicio 1989*, Ed. Aseguradora, Madrid.
- U.N.E.S.P.A. (1991a) *Estadística de Seguros Privados. Información Estadística del Seguro Privado, 1981-90*, vol. 2. Ed. Aseguradora, Madrid.
- U.N.E.S.P.A. (1991b) *Informe Actuarial sobre Determinación de las Primas de Riesgo. Recargo de Seguridad y Recargo para Bonificaciones por No Siniestralidad (Sistema Bonus) en el Seguro de Responsabilidad Civil de Automóviles. Agrupación Nacional de Seguros de Automóviles*, Ed. Aseguradora, Madrid.
- U.N.E.S.P.A. (1992) *Estadística de Seguros Privados. Información Estadística del Seguro Privado, 1982-91.*, Ed. Aseguradora, Madrid.
- U.N.E.S.P.A. (1993a) *Informe Actuarial sobre Determinación de las Primas de Riesgo. Recargo de Seguridad y Recargo para Bonificaciones por No Siniestralidad (Sistema Bonus) en el Seguro de Responsabilidad Civil de Automóviles. Comisión Técnica de Seguros de Automóviles*, Ed. Aseguradora, Madrid.
- U.N.E.S.P.A. (1993b) *Estadística de Daños Propios, Incendio, Robo y Rotura de Lunas del Seguro del Automóvil. Datos 1991. Comisión Técnica de Seguros de Automóviles*, Ed. Aseguradora, Madrid.
- U.N.E.S.P.A. (1993c) *Estadística de Seguros. Cifras de Avance e Informe Económico, 1992*, Ed. Aseguradora, Madrid.
- U.N.E.S.P.A. (1993d) *Estadística de Seguros Privados. Información Estadística del Seguro Privado, 1983-92.*, Ed. Aseguradora, Madrid.

- U.N.E.S.P.A. (1993e) *Estadística del Seguro de Responsabilidad Civil del Automóvil. Datos 1991. Comisión Técnica de Seguros de Automóviles. Ed. Aseguradora, Madrid.*
- U.N.E.S.P.A. (1993f) *Informe Actuarial sobre Determinación de las Primas de Riesgo. Recargo de Seguridad y Recargo para Bonificaciones por No Siniestralidad (Sistema Bonus) en las Modalidades de Daños Propios, Robo y Lunas del Seguro de Automóviles 1993. Comisión Técnica de Seguros de Automóviles, Ed. Aseguradora, Madrid.*
- U.N.E.S.P.A. (1994a) *Estadística de Daños Propios, Rotura de Lunas e Incendio del Seguro del Automóvil. Datos 1992. Comisión Técnica de Seguros de Automóviles, Ed. Aseguradora., Madrid.*
- U.N.E.S.P.A. (1994b) *Estadística de Seguros. Cifras de Avance e Informe Económico, 1993, Ed. Aseguradora, Madrid.*
- U.N.E.S.P.A. (1994c) *Estadística de Seguros Privados, 1984-93, Ed. Aseguradora, Madrid.*
- U.N.E.S.P.A. (1994d) *Estadística del Seguro de Responsabilidad Civil del Automóvil. Datos 1992. Comisión Técnica de Seguros de Automóviles, Ed. Aseguradora, Madrid.*
- U.N.E.S.P.A. (1994e) *Informe Actuarial sobre Determinación de las Primas de Riesgo. Recargo de Seguridad y Recargo para Bonificaciones por No Siniestralidad (Sistema Bonus) en el Seguro de Responsabilidad Civil de Automóviles. Comisión Técnica de Seguros de Automóviles, Ed. Aseguradora, Madrid.*
- U.N.E.S.P.A. (1994f) *Manual de Criterios. Comisión de Vigilancia y Arbitraje. C.I.D.E. y A.S.C.I.D.E. Comisión Técnica de Seguros de Automóviles, Ed. Aseguradora, Madrid.*
- Varian, H. (1992) *Análisis Microeconómico. Tercera Edición. Ed. Antoni Bosch editor, Barcelona.*
- Vaughan, E.J. y C.M. Elliot (1978) *Fundamentals of Risk and Insurance. Second Edition. Ed. Wiley/Hamilton, New York.*
- Von Neumann, J. y O. Morgenstern (1944) *Theory of Games and Economic Behaviour. Ed. Princeton University Press.*
- Weiss, M.A. (1991) "Efficiency in the Property-Liability Insurance Industry". *Journal of Risk and Insurance*, vol. 58, nº 3, 452-479.
- Weisberg, H.I. y R.A. Derrig (1991) "Fraud and Automobile Insurance: A Report on the Baseline Study of Bodily Injury Liability Claims in Massachusetts". *Journal of Insurance Regulation*, vol. 9, nº 4, 497-541.
- Weisberg, H.I. y R.A. Derrig (1992a) "Massachusetts Automobile Bodily Injury Tort Reform". *Journal of Insurance Regulation*, vol. 10, nº 3, 384-440.
- Weisberg, H.I. y R.A. Derrig (1992b) "Compensation for Automobile Injuries Under No-Fault with a Monetary Threshold: A Study of Massachusetts Claims". *AIB Filing for 1993 Private Passenger Automobile Rates, DOI Docket G92-21, 28-94.*

- Weisberg, H.I. y R.A. Derrig (1992c) "The System Misfired". *Best's Review*, December, vol. 93, nº 8, 37-40 y 87.
- Weisberg, H.I. y R.A. Derrig (1992d) "Pricing Auto No-Fault and Bodily Injury Liability Coverages Using Micro-Data and Statistical Models". *Casualty Actuarial Society*, 1993 Ratemaking Call Paper Program.
- Weisberg, H.I. y R.A. Derrig (1993) "Quantitative Methods for Detecting Fraudulent Automobile Bodily Injury Claims". *AIB Cost Containment/Fraud Filing*, 49-82.
- Wilmot, R.T.D. (1991) "World Insurance Forum: London". *Best's Review*, (Property/Casualty), vol. 92, nº 7, 97-98.

El trabajo realizado en el presente libro pretende dar respuesta a una necesidad ampliamente reiterada desde el sector asegurador: avanzar en la prevención, detección y control del fraude en el seguro del automóvil. El tratamiento adecuado de toda la información que rodea a la póliza y al siniestro derivará en el diseño de herramientas suficientemente preparadas para agilizar el trabajo de los tramitadores y para estructurar una adecuada política de investigación de casos. La aplicación de métodos estadísticos dirigidos a cuantificar la probabilidad de existencia de comportamientos fraudulentos en los expedientes de siniestros automovilísticos ha sido la principal línea de trabajo desarrollada por Mercedes Ayuso en el seno del equipo de investigación **Modelos Econométricos del Riesgo**. Operativo, desde la década de los ochenta, dentro del Departamento de Econometría, Estadística y Economía Española de la Universidad de Barcelona, este grupo de trabajo ha dirigido sus esfuerzos a la elaboración de estudios directamente aplicables dentro del mundo financiero y asegurador (campo crediticio, seguro de vida, sistemas de tarificación,...), para los que el libro *El Fraude en el Seguro del Automóvil: Cómo Detectarlo*, puede servir de referencia.

MAP 322 AYU FRA
27004



© FUNDACIÓN MAPFRE

Prohibida la reproducción total o parcial de esta obra sin el permiso escrito del autor o de FUNDACIÓN MAPFRE