

Predicción de las reservas individuales para siniestros del automóvil con daños corporales pendientes de liquidación en el contexto de Solvencia II

RESUMEN

Los siniestros del automóvil con daños corporales suelen permanecer sin liquidar durante largos periodos después del accidente. Una adecuada estimación por siniestros comunicados pero no liquidados es por tanto de vital importancia para los aseguradores. Siguiendo las recomendaciones incluidas en el proyecto de Solvencia II, se implementa un método estadístico para la estimación de la reserva. En concreto, se observa que el coste de compensación se distribuye lognormalmente para diferentes niveles de gravedad del daño corporal. La provisión individual del siniestro la estimamos asignando el valor esperado de compensación media según la gravedad predicha de la lesión de la víctima, para la cual también se calcula el límite superior. La gravedad la predecimos mediante un modelo heterocedástico de elección múltiple, porque hallamos evidencias de que la variabilidad en la gravedad latente no es constante para los individuos que viajaban en un turismo. Se demuestra que la metodología propuesta mejora la precisión en la estimación de las reservas en todas las etapas, en comparación con la valoración subjetiva que ha sido tradicionalmente hecha por los peritos de la compañía.

Palabras clave: accidente de automóvil; Solvencia II; siniestros con daños corporales; reserva individual para siniestros pendientes de liquidación.

Clasificación JEL: C35; C53; G22.

2000MSC: G2PO5; G2L12; G2H12.



1. Introduction

Although the number of traffic accidents is declining in many developed countries, in most of them compensation payments to bodily injury (BI) victims are increasing (rising medical expenses, *judicial* inflation and so on). In the Spanish market, from which we have taken the sample used below, the compensation cost for seriously injured victims increased between 2001 and 2005 by an average of 10% annually (SCOR, 2006). Automobile liability insurance is compulsory in Spain. Therefore, BI victims involved in a motor accident have to be compensated by the insurer of the responsible driver. Indeed, the compensation of BI victims represents approximately 60% of the claim costs faced by Spanish motor insurers.

Motor accidents with BI victims involved are usually reported to the insurer shortly after they occur. Nevertheless, claims may remain unsettled for several fiscal years before victims are indemnified. That is why, firstly, the victim must be fully recovered and, subsequently, the compensation amount must be either agreed upon between the parties or set by judicial order. Therefore, insurance companies need accurate methods to calculate the necessary capital funds (reserves) to cover outstanding BI claims liabilities. In particular, they should pay special attention to the provision for Reported but not Settled claims, known as the RBNS reserve.

The actuarial literature has mainly focused on aggregate reserving techniques. Most statistical methods have been developed to compute the reserve for Incurred but not Reported claims (IBNR reserve), and therefore they do not consider the specific characteristics of each victim and accident in the estimation (for a thorough review see England and Verrall, 2002; 2006). Statistical methods based on individual information have projected compensation payments according to the victim information available in the accident year (Norberg, 1993; 1999; Haastrup and Arjas, 1996; Antonio *et al.*, 2006; Roholte Larsen, 2007). Thus, these techniques did not consider any variations in victim information during the claim processing, and the effects of these fluctuations on the reserves estimation.

In current practice, most motor insurance companies calculate compensation liabilities for reported BI claims on a case-by-case basis. Indeed, insurance adjusters assess compensation payments based on the claims information available, especially their own medical reports. In the Solvency II framework the individual evaluation of claims compensation payments is indicated as a permitted technique for reserving purposes

(CEIOPS, 2007). However, the European Committee in charge of the project notes that this valuation technique may be rather subjective. Thus, the Committee recommends applying statistical actuarial methods in order to estimate the RBNS reserve. In fact, Ayuso and Santolino (2007) presented an empirical application where a statistical model underestimated the final cost by 7%, whereas expert assessment was underestimating by 17%.

This paper presents a more general model than Ayuso and Santolino (2007) and discusses details about estimating individual RBNS claim reserves which take into account the compensation cost distribution for different levels of claim severity. The individual provision can be estimated at successive stages during the claim handling process. Our objective is to offer a statistical modelling framework that allows the insurer to calibrate the provision amount for the victim's compensation in response to variations of the expected BI severity of the victim (i.e. immediately after new information about his/her recovery status is available). Furthermore, since distributional assumptions about compensation payments are taken into account, the suggested approach can be used by the company to predict the upper bound reserve amount at the appropriate confidence level. The suggested methodology is applied to the usual stages of claim processing. At each stage, we compare the accuracy of the provision obtained by the proposed methodology with the direct assessment obtained by the insurer, based on internal medical reports.

Several authors have used ordered multiple choice models for the severity of motor accidents (Kockelman and Kweon, 2002; Abdel-Aty, 2003; Lee and Abdel-Aty, 2005; Zajac and Ivan, 2003; Austin and Faigin, 2003; Karlaftis *et al.*, 2003). They all assumed a constant variance in the random term for all individual claims. However, we found evidence that such an assumption seems to be restrictive and may be unrealistic in the case of casualties resulting from accidents involving different types of vehicles. Our model is more general because the severity of the victim's injury is predicted by means of a heteroscedastic ordered multiple choice model (HOMC)¹.

¹ An interesting development of HOMC models was proposed by O'Donnell and Connor (1996), who suggested that the victim's age, the speed, and the time of the accident were predictors of the error variance. More recently, Wang and Kockelman (2005) parameterized the error term variance as a function of vehicle type and vehicle weight. In our case, the error term variance is parameterized according to the victim's vehicle type.

In the next section we describe the database used in the empirical analysis, with special attention being paid to the BI compensation cost distribution. Subsequently, the provision for reported BI claims is estimated at each stage of claim processing, and compared with those directly calculated by the insurer. We demonstrate that the proposed methodology can help the insurance company to obtain a more accurate reserve for covering future compensation payments of motor BI victims. The main findings are summarized in the last section.

2. Motor BI claims database

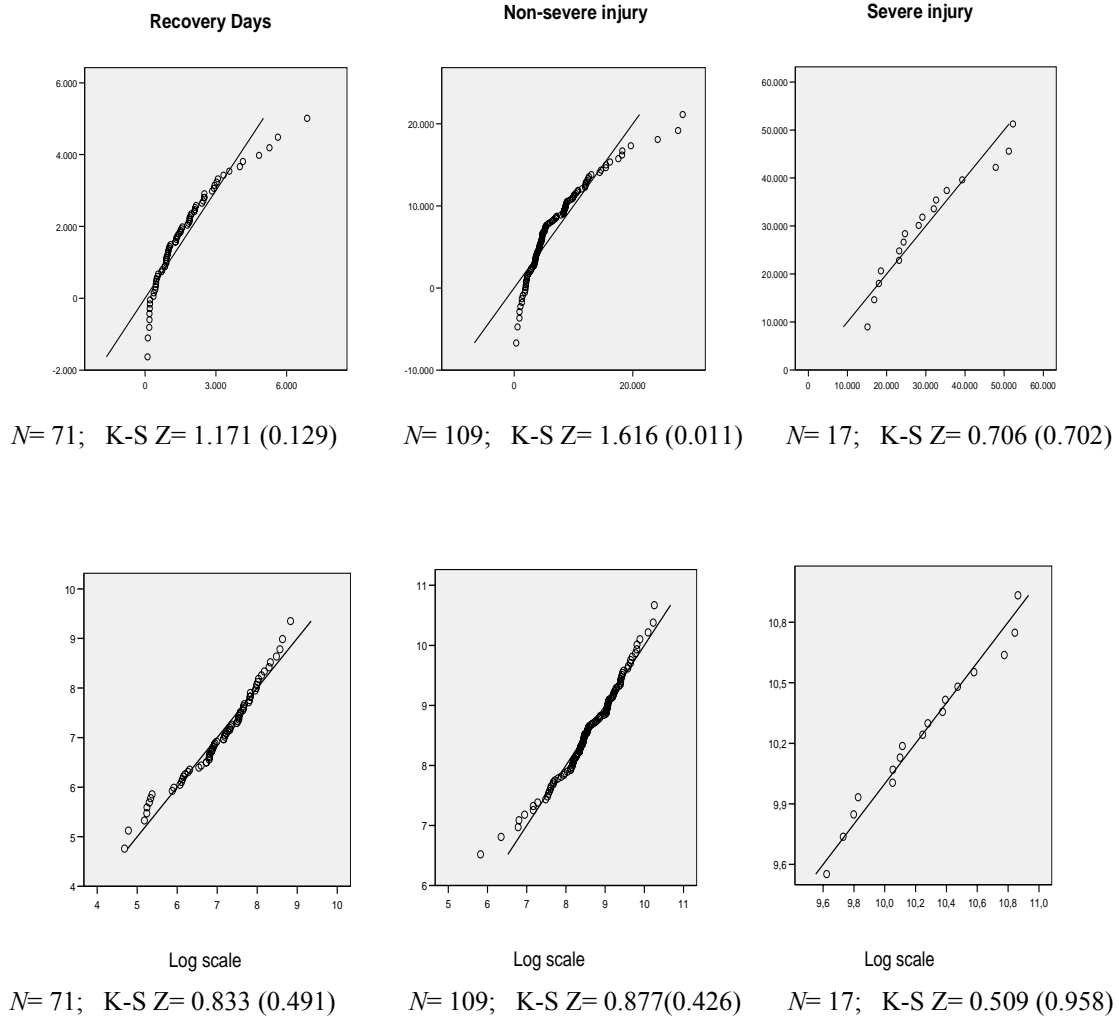
The dataset contains information from 197 non-fatal victims with BI damages involved in traffic accidents. The database was provided by a Spanish insurer who was legally responsible for the compensation payment. The compensation amount for all the victims was established by judicial order for the years 2001 to 2003 because the insurer and the claimant did not reach a prior agreement. Ayuso and Santolino (2007) used this data to model severity with a classical ordered logit model.

The dependent variable of the HOMO regression model is related to the BI severity of the victim. It has three categories: *Recovery Days*, *Non-severe Injury* and *Severe Injury*. These categories are defined according to the severity of injury assigned to sequelae² by the judicial verdict. The category *Recovery Days* represents casualties without sequelae. *Non-severe Injury* means casualties with fewer than 15 points for sequelae, while *Severe Injury* refers to victims with 15 or more points. The HOMO regression model has been estimated by using SPSS v.14.0.

The claim provision for each victim depends on the empirical compensation cost distribution and the severity of the injury. Outstanding BI claims are reserved by allocating the expected mean compensation cost of the forecasted level of BI severity. In Figure 1 the normal Q-Q plot and the Kolmogorov-Smirnov (K-S) test for the compensation cost distribution (in original and logarithmic scale) are presented per severity category.

² Note that sequela is defined as the definitive reduction of a person's physical or mental potential as a result of an accident. The injury severity of a sequela is measured according to a score interval. The definition and score interval of each sequela must agree with those stipulated in the Spanish disability scoring system (LOSSP 30/95). The aggregated score for sequelae ranges from 0 to 100.

Figure 1. Analysis of normality for the claims compensation cost data clustered by categories of severity, on original scale (*first row*) and logarithmic scale (*second row*)



Note that the null hypothesis of lognormality cannot be rejected for any category of BI severity. In contrast, there is evidence that normality of compensations cannot be accepted for observations classified as *Non-severe Injury*. The same outcome is obtained when the K-S test and the Q-Q plot of the observed compensation cost distribution are carried out for the whole sample. Consequently, we assume that compensation cost data are lognormally distributed. Predictions on the original scale are obtained with the well-known characteristics of the lognormal distribution, i.e. if $\ln(y) \sim N(\mu, \sigma^2)$ then $E[y] = e^{\mu+0,5\sigma^2}$ and $\text{Var}[y] = e^{2\mu+2\sigma^2} (e^{\sigma^2} - 1)$.

Descriptive statistics of the compensation cost data for each category of BI severity are presented in Table 1.

Table 1. Descriptive statistics of the compensation cost variable (in euros)

| | <i>Estimated mean compensation cost (log scale)</i> | <i>Standard deviation (log scale)</i> | <i>Expected mean compensation cost (original scale)</i> | <i>Standard deviation (original scale)</i> |
|--------------------------|-------------------------------------------------------------|-----------------------------------------------|-----------------------------------------------------------------|----------------------------------------------------|
| <i>Recovery Days</i> | 7.110 | 0.953 | 1927.74 | 2345.143 |
| <i>Non-Severe Injury</i> | 8.620 | 0.808 | 7680.44 | 7371.380 |
| <i>Severe Injury</i> | 10.273 | 0.403 | 31388.74 | 13195.383 |
| <i>Total</i> | 8.219 | 1.264 | 8249.01 | 16387.109 |

Regression variables and descriptive statistics for the total sample are presented in Table 2. Explanatory variables refer to attributes of the victim such as age and gender, characteristics of the accident, and medical information collected during the recovery period³. Regarding the accident characteristics, we include as regressors the year that the accident took place, the victim's vehicle type (i.e. if it was a car or another type) and if the casualty was a passenger (not the driver) of the damaged vehicle.

With respect to medical information, we consider the number of sequelae and recovery days⁴ stated by the insurer's medical expert at the first and last examination, and we also incorporate a dichotomous variable which indicates whether the forensic doctor examined the victim and considered that the accident had not caused any sequelae. Under Spanish law the forensic report is compulsory only if the lawsuit follows a penal but not a civil procedure. A control variable (*foren*) was included in the model to prevent civil lawsuits from being treated as missing values in our dataset.

³ Note that here fewer variables than Ayuso and Santolino (2007) have been used. In addition, a new regression including information from the forensic report has been estimated. Variables did not improve the explanatory capacity of the model were excluded.

⁴ A distinction is made between recovery days with and without disability for working.

Table 2. Explanatory variables and descriptive statistics

| | | Mean | Standard Deviation |
|------------------|--------------------------------------------------------------------------------------------|--------|--------------------|
| <i>logcom</i> | Compensation amount (on logarithmic scale) awarded by verdict. | 8.219 | 1.264 |
| <i>year</i> | Accident year (1=1994; 2=1995; ...; 10=2003). | 6.975 | 1.430 |
| <i>year2</i> | Accident year (squared). | 50.680 | 17.151 |
| <i>car</i> | 1 if the victim's vehicle is a car; 0=otherwise (e.g. motorbike, pedestrians). | 0.650 | 0.478 |
| <i>age</i> | Victim's age (1 if age 0 to 9; 2 if 10 to 19; and so forth). | 3.930 | 1.606 |
| <i>gender</i> | 1 if male; 0=otherwise. | 0.497 | 0.501 |
| <i>passen</i> | 1 if the victim is passenger of the insured vehicle; 0=otherwise. | 0.091 | 0.289 |
| <i>seq</i> | Number of sequelae (permanent injuries) expected in first medical report. | 1.092 | 1.340 |
| <i>rdd</i> | Number of recovery days with disability for working expected in first medical report. | 53.563 | 53.971 |
| <i>rdnd</i> | Number of recovery days without disability for working expected in first medical report. | 29.109 | 45.472 |
| <i>same</i> | 1 if last medical report is the same as the first one; 0=otherwise. | 0.316 | 0.467 |
| <i>seq_last</i> | Number of sequelae (permanent injuries) stated in last medical report. | 1.114 | 1.655 |
| <i>varseq</i> | Sequelae number variation across reports (last minus first). | 0.009 | 0.917 |
| <i>rdd_last</i> | Number of recovery days with disability for working stated in last medical report. | 53.131 | 63.027 |
| <i>varrdd</i> | Variation in the number of recovery days unable to work across reports (last minus first). | 2.079 | 37.601 |
| <i>rdnd_last</i> | Number of recovery days without disability for working stated in last medical report. | 37.596 | 59.699 |
| <i>foren</i> | 1 if forensic doctor states the victim has no sequelae; 0 otherwise. | 0.342 | 0.477 |

N=197 (71 victims classified as *Recovery Days*; 109 victims as *Non-severe Injury*; 17 victims as *Severe Injury*).

3. Estimated reserves for reported but not settled BI claims

In this section we deal with the estimation of claims reserves according to the information about the BI victim available to the insurer at different moments of the claim handling process. These are: i) a first reserve estimation carried out at the time the accident is reported to the insurance company; ii) a second estimation when the company has the initial medical evaluation of BI damages; iii) a third when the victim is fully recovered; and iv) a fourth estimation that is computed when the company has the forensic report. Each outstanding BI claim is reserved by allocating the expected mean compensation cost of the severity level predicted by a heteroscedastic ordered multiple choice model. Finally, the aggregated reserve at each stage of the claim handling process is computed as the sum of individual provisions.

An overview of the heteroscedastic ordered logit model specification and the interpretation of results can be found in Appendix 1.

3.1 Reserve estimation when claims are reported

The initial reserve estimation is based on the information available when claims are reported to the insurer. Thereby, we allocate to each observation the expected mean compensation cost of the severity level predicted by the first HOMC model (Table A.1, Stage 1). A comparison with the total amount which was awarded by judges is presented in Table 3. The first row of Table 3 shows the number of victims according to the BI severity awarded in the judicial verdict. The sum of observed compensations awarded to these victims is reported in second row. The aggregated reserve for outstanding claims is estimated for each category of BI severity (5th row), which is obtained by multiplying the empirical mean compensation cost of each severity category (3rd row and also Table 1) by the predicted frequency of victims derived from the HOMC model (4th row). The sixth and seventh rows of Table 3 present the percentage of observed compensations covered by the estimated reserve and by the upper-bound estimate of the reserve for a 95% confidence level. The same design is followed for the tables in the next stages.

Table 3. Provision derived from the severity level predicted by the HOMC model (victim information available before any medical report)

| | Level of severity | | | Total |
|-------------------------------------------------------------------------|-------------------|-------------------|---------------|------------|
| | Recovery Days | Non-Severe Injury | Severe Injury | |
| Observed frequency (judge) | 71 | 109 | 17 | 197 |
| Total observed amount awarded by judges (euros) | 126693.80 | 799054.14 | 530422.16 | 1456170.10 |
| Expected mean compensation (euros) | 1927.74 | 7680.44 | 31388.74 | 8249.01 |
| Predicted frequency (HOMC model) | 51 | 146 | - | 197 |
| Total provision from the HOMC (euros) | 98314.74 | 1121344.24 | 0 | 1219658.98 |
| Total provision from the HOMC / Total observed amount awarded by judges | 77.60% | 140.33% | 0.00% | 83.76% |
| Confidence limit* of the HOMC / Total observed amount awarded by judges | 99.35% | 158.48% | 0.00% | 93.32% |

* 95% confidence limit.

Note that, at this point, severely injured victims are not correctly predicted by the heteroscedastic ordered logit model. Also, victims without sequelae (classified as *Recovery Days*) are not sufficiently forecasted. Due to these constraints in the prediction of BI severity, the economic resources are concentrated on claims from the second category (*Non-severe Injury*) when the provision is calculated. The overprovision of claims from the intermediate category is not enough to counterbalance the under-provision of claims from the extreme categories. As a result, the total reserve only covers about 84% of the entire compensation

amount of outstanding BI claims. Note that this first estimation of reserves has been carried out with very little information about the victims.

3.2 Reserve estimation after the first medical report

With the first medical report an initial professional assessment of damages is submitted to the insurer. In Table 4 the estimated provision based on the injury severity predicted by the heteroscedastic model (Table A.1, Stage 2) is compared with the provision based on the direct classification of the medical expert. The same criterion of allocating the expected mean cost of the corresponding severity category was applied. Note that the total number of BI victims is now different to that in the previous stage. This is due to the fact that we have taken into account only those victims for whom the first medical report was submitted. As a consequence, the expected mean compensation cost for each level of BI severity, which is directly observed from the sub-sample of BI victims for whom the first medical report was submitted, appears to be slightly different from the one presented for the whole sample (Table 2).

Table 4. Provision derived from the medical expert's classification vs. provision derived from the HOMC model prediction (victim information available after the first medical report)

| | Level of severity | | | | Total |
|----------------------------------------------------------------------------------------|-------------------|---------------|-------------------|---------------|------------|
| | No injury | Recovery Days | Non-Severe Injury | Severe Injury | |
| Observed frequency (judge)* | - | 40 | 67 | 12 | 119 |
| Total observed amount awarded by judges (euros) | - | 66314.27 | 554278.55 | 394015.05 | 1014607.87 |
| Expected mean compensation (euros) | - | 1766.76 | 8465.21 | 33061.09 | 9699.37 |
| Observed frequency (first medical expert classification) | 4 [†] | 42 | 65 | 8 | 119 |
| Predicted frequency (HOMC model) | - | 39 | 70 | 10 | 119 |
| Total provision from medical report/ Total observed amount awarded by judges | - | 111.90% | 99.27% | 67.13% | 87.61% |
| Total provision from the HOMC/ Total observed amount awarded by judges | - | 103.90% | 106.91% | 83.91% | 97.78% |
| Confidence limit ^{††} of the HOMC/ Total observed amount awarded by judges | - | 134.53% | 124.51% | 101.36% | 109.71% |

* Only victims for whom the first medical report was submitted.

[†] Medical expert awarded neither recovery days nor sequelae to the victim.

^{††} 95% confidence limit.

Compensations of severely injured victims were again underprovisioned (Table 4). When the HOMC model was applied, in aggregated terms, the misclassified claims were mainly diverted to the *Non-severe Injury* category. In contrast, following the medical expert's evaluation, the *Recovery Days* claims were primarily overclassified and therefore

overprovisioned. Since the individual provision of a *Non-severe Injury* claim is higher than that of a *Recovery Days* claim, the aggregated provision seems to fit better the proposed methodology. The estimated provision derived from our methodology covered about 98% of the total compensation amount, whereas the provision based on the medical expert's classification covered only 88% of that amount.

3.3 Reserve estimation after the last medical report

At this stage the insurer has the last medical report indicating that the victim has fully recovered. Consequently, the sub-sample is composed of victims for whom the insurance company had the first and last medical reports. Thus, as in the previous stage, the expected mean compensations for each severity level have changed (Table 5, third row). The estimated provision according to HOMC predictions (Table A.1, Stage 3) and that directly derived from the medical expert's classification are presented in Table 5. Note that in contrast to the medical expert's classification the number of victims predicted by the HOMC model at each level of BI severity is now closer to the judge's evaluation. It should be emphasized that the estimated reserve is again close to meeting future compensations, with the point and upper-bound estimates covering 95% and 107% of the empirical compensation payments, respectively.

Table 5. Provision derived from the medical expert's classification vs. provision derived from the HOMC model prediction (victim information available after the last medical report)

| | Level of severity | | | | Total |
|----------------------------------------------------------------------------------------|-------------------|---------------|-------------------|---------------|-----------|
| | No injury | Recovery Days | Non-Severe Injury | Severe Injury | |
| Observed frequency (judge)* | - | 40 | 63 | 11 | 114 |
| Total observed amount awarded by judges (euros) | - | 66314.27 | 493812.97 | 364884.22 | 925011.46 |
| Expected mean compensation (euros) | - | 1766.76 | 7980.54 | 33476.80 | 9045.22 |
| Observed frequency (last medical expert classification) | 4 [†] | 50 | 50 | 10 | 114 |
| Predicted frequency (HOMC model) | - | 42 | 63 | 9 | 114 |
| Total provision from medical report/ Total observed amount awarded by judges | - | 133.21% | 80.80% | 91.75% | 88.88% |
| Total provision from the HOMC/ Total observed amount awarded by judges | - | 111.90% | 101.81% | 82.57% | 94.95% |
| Confidence limit ^{††} of the HOMC/ Total observed amount awarded by judges | - | 143.68% | 119.14% | 101.64% | 107.09% |

* Only victims for whom the first and last medical reports were submitted.

[†] Medical expert awarded neither recovery days nor sequelae to the victim.

^{††} 95% confidence limit.

3.4 Reserve estimation after forensic report

Lastly, the reserve was computed when the insurer also had the victim information provided by the forensic report, if one existed. This was the last stage in claim processing before the case was taken to trial. In our analysis the sample size was, at that moment, equal to the sample size used in the previous stage, after the last medical report. Therefore, when estimating the provision the same expected mean compensation costs for the different levels of BI severity were considered (Table 5, third row).

The results are shown in Table 6. According to the classification of severity by the forensic doctor we observed an overprovision for covering the compensations of *Severe Injury* victims. Consequently, the total reserve exceeded the real final compensation amount by more than 26%. In contrast, the proposed methodology provided a more accurate estimation for reserving BI claims. The total provision estimated by means of the HOMC model represented 96% of the total amount.

Table 6. Provision derived from the forensic classification[‡] vs. provision derived from the HOMC model prediction (available victim information after the forensic report)

| | Level of severity | | | Total |
|---------------------------------------------------------------------------------------|-------------------|-------------------|---------------|-----------|
| | Recovery Days | Non-Severe Injury | Severe Injury | |
| Observed frequency (judge)* | 40 | 63 | 11 | 114 |
| Total observed amount awarded by judges (euros) | 66314.27 | 493812.97 | 364884.22 | 925011.46 |
| Expected mean compensation (euros) | 1766.76 | 7980.54 | 33476.80 | 9045.22 |
| Observed frequency (forensic classification) | 40 | 54 | 20 | 114 |
| Predicted frequency (HOMC model) | 40 | 65 | 9 | 114 |
| Total provision from forensic report/ Total observed amount awarded by judges | 106.57% | 87.27% | 183.49% | 126.61% |
| Total provision from the HOMC/ Total observed amount awarded by judges | 106.57%% | 105.05% | 82.57% | 96.57% |
| Confidence limit [†] of the HOMC/ Total observed amount awarded by judges | 137.58% | 122.65% | 101.64% | 108.53% |

[‡] We considered the medical expert's classification (in the last report) for those victims without a forensic report. When the forensic doctor set the sequelae but did not assess them, we considered the mean score of the corresponding interval according to the legislative scale.

* Only victims for whom the first and last medical reports were submitted.

[†] 95% confidence limit.

Summarizing, our methodology presents significant advantages over the provision directly derived from medical reports when it comes to computing the claims reserve⁵. First,

⁵ This happens in all stages, not only in the last stage as illustrated in Ayuso and Santolino (2007).

an initial reserve estimation based on the severity prediction of the victim’s injury may be computed at the time the accident is reported and, therefore, before any medical evaluation is available. For the remaining stages of claims processing, it is shown that our methodology offers a more balanced claims reserve estimation than does the provision derived directly from medical evaluations. In this regard, the reserve based directly on the information collected in internal medical reports covered on average less than 90% of claims payments. However, when the reserve was calculated following the forensic evaluation, claims were more than 26% overprovisioned. In contrast, the reserve estimated by the proposed methodology ranges between 95% and 98% of claims payments for all the stages with medical information.

Before concluding, an example of the individual provision of BI claims is presented. Let us suppose that the insurer wants to estimate the capital required to meet the compensation liabilities of four traffic victims with BI damages. Moreover, the available accident information is not the same for each victim because they refer to different stages of claim processing. Let us suppose that victim A is at stage I (before any medical report), victim B is at stage II (after the first medical report), victim C is at stage III (after the last medical report), and finally, victim D is at the last stage (after the forensic report). The results for this example are reported in Table 7. The allocated individual provision is the expected mean compensation for the corresponding predicted severity of the injured victim, and thus it depends on both the severity level and the claim information stage.

Table 7. An example of provisions for four outstanding BI claims according to the HOMC model prediction (in brackets, the insurer classification of BI severity and the obtained provision)

| Victim | Information stage | Observed Cost (euros) | Predicted severity* | Individual provision | Confidence limit (95%) of the claim provision |
|--------|-------------------|-----------------------|---------------------|---------------------------------|-----------------------------------------------|
| A | 1 | 19661 | NSI (-) | 7680.44 (8249.01 [†]) | 19805.67 (35189.50) |
| B | 2 | 553 | NSI (RD) | 8465.21 (1766.76) | 20127.42 (5018.66) |
| C | 3 | 968 | RD (RD) | 1766.76 (1766.76) | 5018.66 (5018.66) |
| D | 4 | 3370 | NSI (NSI) | 7980.54 (7980.54) | 18762.41 (18762.41) |
| Total | | 24552 | | 25892.95 (19763.07) | 46137.76 (49143.15) |

* RD: Recovery Days; NSI: Non-Severe Injury.

[†] The estimated mean compensation of all victims is allocated.

When analyzing victim A, we observed a final compensation amount of 19661 euros. When there was no medical information available (stage I) the HOMC model predicted this victim as *Non-severe Injury*, and 7680.44 euros were allocated to reserves. At that moment, if the HOMC prediction was not available, the estimated mean compensation of all victims would be 8249.01 euros. Victim B is at stage II, i.e. right after the insurer received the first internal medical evaluation of BI damages. According to the HOMC model prediction the victim was classified as *Non-severe Injury*, and so 8465.21 euros were allocated. However, since the medical expert considered that the victim did not have sequelae (only *Recovery Days*), the insurer reserved 1766.76 euros. The individual provision for the remaining victims, C and D, was computed in the same fashion.

According to our results, the estimated reserve of these BI claims based on the insurance staff's evaluation was not enough to cover their compensation payments. In contrast, when the reserve was computed by means of the proposed methodology the estimated provision represented 105.5% of the final compensation payments. On the other hand, when the limit of the provision with a confidence level of 95% was computed, we observed that the upper limit of the total provision derived from the medical expert's classification was larger than the one from the model. This was due to the large sample variance of the provision allocated to victim A, which is estimated from the entire compensation cost distribution (Table 1), and proves that the statistical model that takes into account non constant variance performs better than an approach based on expert assessment rather than statistical analysis.

4. Conclusions

The time period from the occurrence of a motor accident to the point at which victims are compensated for BI damages is long. As a consequence, the insurer is faced with calculating reserves for Reported but not Settled claims. With the aim of promoting objective techniques, insurers are encouraged by the Solvency II project to implement statistical actuarial methods for reserve estimation (CEIOPS, 2007). Insurance companies traditionally assess the compensation cost for a known BI claim (not yet paid) according to their own medical reports. Subsequently, they compute the total RBNS reserve as the sum of individual provisions. Unfortunately, there are often substantial differences between the claim compensation assessed by insurance staff and the amount finally awarded by the judicial verdict.

In this paper a statistical reserving methodology for outstanding BI claims based on individual data has been presented. Empirical compensation cost data grouped by severity levels of victims' injuries are shown to follow a lognormal distribution function. Each claim is provisioned by allocating the expected mean compensation cost of the predicted BI severity. The upper limit of the reserve with an appropriate confidence level is also estimated. The RBNS reserve is computed at the main stages of the claim handling process. It is shown that the proposed methodology is able to estimate the RBNS reserve for claims with different levels of available information. As compared to reserves based on internal medical reports, our methodology performs better at all stages.

Another feature of this paper is that we apply a heteroscedastic ordered multiple choice model to predict the severity of victims' injuries, showing that individuals travelling by car present different variability in latent severity. This qualitative modelling approach allows us to monitor the probability transition of expected severity of each claim at successive stages of the handling process without making additional assumptions regarding price variations, such as the evolution of the inflation rate, the cost of medical services or wages. In addition, and due to the independence of economic factors, the methodology could be applied to estimate the reserve for BI claims settled in a period other than the period under review, without any substantial changes in the explanatory variables' behaviour being expected. In this regard, only assumptions concerning the evolution of the mean compensation cost per severity level would be required.

Acknowledgements

The authors are grateful for the valuable suggestions from research affiliates of Risk in Finance and Insurance at the University of Barcelona. We also thank the two anonymous referees for the general and specific comments. This work has received support from the Spanish Ministry of Science and Innovation and FEDER grants SEJ2005-00741/ECON and ECO2008-01223/ECON.

References

- ABDEL-ATY, M. (2003) Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models. *Journal of Safety Research*, 34, 597–603.
- ANTONIO, K., BEIRLANT, J., HOEDEMAKERS, T. & VERLAAK, R. (2006) Lognormal Mixed Models for Reported Claim Reserves. *North American Actuarial Journal*, 10, 30–48.
- AUSTIN, R. & FAIGIN, B. (2003) Effect of Vehicle and Crash Factors on Older Occupants. *Journal of Safety Research*, 34, 441–452.

- AYUSO, M. & SANTOLINO, M. (2007) Predicting Automobile Claims Bodily Injury Severity with Sequential Ordered Logit Models. *Insurance: Mathematics and Economics*, 4, 71–83.
- CEIOPS (2007) QIS 4 Technical Specifications. Committee of European Insurance and Occupational Pensions Supervisors CEIOPS DOC-23/07.
- ENGLAND, P. & VERALL, R. (2002) Stochastic Claims Reserving in General Insurance. *British Actuarial Journal*, 8, 443–544.
- ENGLAND, P. & VERALL, R. (2006) Predictive distributions of outstanding liabilities in general insurance. *Annals of Actuarial Science*, 1, 221–270.
- HAASTRUP, S. & ARJAS, E. (1996) Claims Reserving in Continuous Time: a Non-parametric Bayesian Approach. *ASTIN Bulletin*, 26, 139–164.
- KARLAFTIS, M.G., KOTZAMPASSAKIS, I. & KANELLAIDIS, G. (2003) An Empirical Investigation of European Drivers' Self-Assessment. *Journal of Safety Research*, 34, 207–213.
- KOCKELMAN, K. & KWEON, Y. (2002) Driver Injury Severity: An Application of Ordered Probit Models. *Accident Analysis & Prevention*, 34, 313–321.
- LEE, C. & ABDEL-ATY, M. (2005) Comprehensive Analysis of Vehicle-Pedestrian Crashes at Intersections in Florida. *Accident Analysis & Prevention*, 37, 775–786.
- NORBERG, R. (1993) Prediction of Outstanding Liabilities in Non-Life Insurance. *ASTIN Bulletin*, 23, 95–115.
- NORBERG, R. (1999) Prediction of Outstanding Claims II: Model Variations and Extensions. *ASTIN Bulletin*, 29, 5–25.
- O'DONELL, C.J. & CONNOR, D.H. (1996) Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice. *Accident Analysis & Prevention*, 28, 739–756.
- ROHOLTE LARSEN, C. (2007) An Individual Claims Reserving Model. *ASTIN Bulletin*, 37, 113–132.
- SCOR (2006) Nivel y Evolución del Coste Medio Daño Corporal Grave por Accidentes de Circulación Ocurridos en España. SCOR Global P&C (Spanish document).
- WANG, X. & KOCKELMAN, K. (2005) Use of Heteroscedastic Ordered Logit Model to Study Severity of Occupant Injury: Distinguishing the Effects of Vehicle Weight and Type. *Transportation Research Record*, 1908, 195–204.
- ZAJAC, S. & IVAN, J. (2003) Factors Influencing Injury Severity of Motor Vehicle-Crossing Pedestrian Crashes in Rural Connecticut. *Accident Analysis & Prevention*, 35, 369–379.

APPENDIX 1

The heteroscedastic ordered logit model

The heteroscedastic ordered logit model is based on a continuous unobserved variable y^* , which collects the bodily injury severity, that is modelled by means of a linear regression. The observed variable y_i is discrete, with J ordered response categories. In our case, it consists of three categories: *Recovery Days*, *Non-severe Injury* and *Severe Injury*. The cumulative probability that subject i belongs to category j or lower ones is,

$$P(y_i \leq j | \mathbf{x}_i, \mathbf{z}_i) = \frac{e^{(\mu_j - \mathbf{x}_i \boldsymbol{\beta}) / \sigma_i}}{1 + e^{(\mu_j - \mathbf{x}_i \boldsymbol{\beta}) / \sigma_i}}, \quad j=1, \dots, J, \quad i=1, \dots, N,$$

where the μ_j 's are the model thresholds (with $\mu_0 = -\infty$ and $\mu_J = +\infty$), $\boldsymbol{\beta} (K \times 1)$ is the column vector of K unknown parameters, and $\mathbf{x}_i (1 \times K)$ is the row vector of K observed regressors. We assume that the residual term ε_i follows a normal distribution with zero expected value and σ_i^2 variance. Note that $(\hat{\mu}_j - \mathbf{x}_i \hat{\boldsymbol{\beta}})$ is the predictor of the expected mean value, and σ_i is the standard deviation. Usually, σ_i is parameterized as $\exp(\mathbf{z}_i \boldsymbol{\tau})$ to ensure its positivity, and $\mathbf{z}_i \hat{\boldsymbol{\tau}}$ is the variance predictor, with $\hat{\boldsymbol{\tau}} (G \times 1)$ the column vector of G unknown scale parameters (O'Donnell and Connor, 1996; Wang and Kockelman, 2005). The constraint of the homoscedastic ordered logit model related to opposite marginal effects on the two extreme categories is reduced by the inclusion of scale parameters $\boldsymbol{\tau}$. Parameter estimates are usually obtained by maximum likelihood, applying any algorithm (e.g. Newton-Raphson) in the maximization process.

The estimation results of the heteroscedastic ordered logit model

Parameter estimates for variables used at each stage in the prediction of the individual BI severity level are shown in Table A.1. At the bottom of Table A.1 we list the percentage of BI victims for which the model correctly predicted the final BI severity. In order to make comparisons the percentage of victims for which the severity was accurately classified in medical reports is also indicated. Note that in the first stage there is not yet any information from medical reports and so no percentage is included. In the last stage, if the forensic doctor did not examine the victim we considered the severity classification made in the last internal medical report.

Table A.1. Estimation of parameters at the successive stages (heteroscedastic ordered logit model)

| | Stage I (Model before any medical report) | | Stage II (Model after first medical report) | | Stage III (Model after last medical report) | | Stage IV (Model after forensic report) | |
|-----------------------------------------------------------|------------------------------------------------------------------------------------|----------|------------------------------------------------------------------------------------|----------|-----------------------------------------------------------------------------------|----------|-----------------------------------------------------------------------------------|----------|
| | Coefficient | P-value | Coefficient | P-value | Coefficient | P-value | Coefficient | P-value |
| μ_1 | -0.041 | 0.979 | 4.167 | 0.097* | 4.114 | 0.178 | 3.744 | 0.592 |
| μ_2 | 2.996 | 0.059* | 8.168 | 0.004*** | 8.462 | 0.011** | 17.331 | 0.049** |
| <i>year</i> | 0.795 | 0.129 | 1.237 | 0.118 | 1.429 | 0.152 | 2.291 | 0.330 |
| <i>year2</i> | -0.082 | 0.066* | -0.104 | 0.111 | -0.124 | 0.129 | -0.179 | 0.357 |
| <i>car</i> | -1.462 | 0.000*** | -0.608 | 0.159 | -0.770 | 0.112 | -1.074 | 0.443 |
| <i>age</i> | 0.142 | 0.101 | 0.194 | 0.082* | 0.245 | 0.049** | 0.590 | 0.086* |
| <i>gender</i> | -0.895 | 0.003*** | -0.877 | 0.016** | -1.162 | 0.008*** | -2.369 | 0.072* |
| <i>passen</i> | 0.472 | 0.319 | 0.678 | 0.134 | 0.643 | 0.206 | -0.275 | 0.878 |
| <i>seq</i> | - | - | 0.701 | 0.002*** | - | - | - | - |
| <i>rdd</i> | - | - | 0.015 | 0.001*** | - | - | - | - |
| <i>rdnd</i> | - | - | 0.008 | 0.078* | - | - | - | - |
| <i>same</i> | - | - | - | - | -0.823 | 0.065* | -2.686 | 0.170 |
| <i>seq_last</i> | - | - | - | - | 0.676 | 0.006*** | 1.825 | 0.039** |
| <i>varseq</i> | - | - | - | - | -0.686 | 0.031** | -2.172 | 0.049** |
| <i>rdd_last</i> | - | - | - | - | 0.014 | 0.005*** | 0.021 | 0.110 |
| <i>varrdd</i> | - | - | - | - | -0.014 | 0.018** | -0.036 | 0.082* |
| <i>rdnd_last</i> | - | - | - | - | 0.007 | 0.059* | 0.017 | 0.046** |
| <i>foren</i> | - | - | - | - | - | - | -10.159 | 0.005*** |
| <i>car (scale)</i> | -0.165 | 0.377 | -0.649 | 0.028** | -0.582 | 0.055* | 0.695 | 0.074* |
| | <i>N</i> = 197; pseudo- <i>R</i> ² = 0.189; χ^2 = 33.844(0.000) | | <i>N</i> = 119; pseudo- <i>R</i> ² = 0.611; χ^2 = 56.046(0.000) | | <i>N</i> = 114; pseudo- <i>R</i> ² = 0.647; χ^2 = 6.847(0.077) | | <i>N</i> =114; pseudo- <i>R</i> ² = 0.861; χ^2 = 57.511(0.000) | |
| <i>Claims correctly predicted by the model (%)</i> | 63.452% | | 72.269% | | 78.070% | | 91.228% | |
| <i>Claims correctly classified by medical reports (%)</i> | - | | 62.185% | | 61.403% | | 83.333% [†] | |

*** indicates 1% significance level; ** indicates 5% significance level; * indicates 10% significance level.

[†]We consider the medical expert's classification (in the last report) for those claims without a forensic report. When the forensic doctor sets the sequelae but he/she does not assess them, we consider the mean score of the corresponding interval according to the legislative scale.

Table A.1 shows that the chi-square statistic is significant at all stages. For a given stage of the estimation the statistic was computed as the difference between minus two times the log-likelihood for the model with the variables included in the previous stage and that for the current model. The statistic estimation takes into account the sample size at each stage. The statistical significance thus means that the incoming information at each stage has explanatory power with regard to the severity of a victim's injury.

The variable *gender* has a significant coefficient at all stages, and the variable *age* behaves in a similar way. The negative sign of the *gender* parameter indicates that men are less likely to suffer serious injuries than are women. On the other hand, the *age* parameter is positive and thus older victims have a greater probability of having serious injuries resulting from the accident. In relation to the information from medical reports, both the number of recovery days (regarding disability and no disability for working) and the number of sequelae considered by the insurer's medical expert are positively related to the severity of a victim's injury. Notice that the variables *varseq* and *varrdd* are significant in the last two stages. These variables register variations in the expert evaluation across medical reports. Therefore, the parameters' significance illustrates that the initial medical report provides information that is relevant to the explanation of injury severity, even when the company already has the final report or the forensic examination results. Concerning the last phase of the estimation, it should be emphasized that the percentage of cases accurately estimated by the model increased notably when the forensic information was included (Table A.1). This relationship between the forensic report and the accuracy of estimations indicates a strong influence of the forensic evaluation on the level of severity awarded by the judge.

Finally, we would like to point out that the scale parameter *car* is statistically significant in three of the four analyzed stages and, therefore, the heteroscedastic variance specification is accepted. Individuals travelling by car at the moment of the accident exhibit different variability in latent injury severity compared to those travelling by motorbike or pedestrians.