Prediction of individual automobile RBNS claim reserves in the context of Solvency II

By Mercedes Ayuso and Miguel Santolino

RISC–IREA Research Group.
Department of Econometrics, University of Barcelona.
Avda. Diagonal, 690. 08034 Barcelona.
Phone: 93 402 14 09/93 402 18 22 and 93 402 43 18. Fax: 93 402 1821.

Abstract: Automobile bodily injury (BI) claims remain unsettled for a long time after the accident. The estimation of an accurate reserve for Reported But Not Settled (RBNS) claims is therefore vital for insurers. In accordance with the recommendation included in the Solvency II project (CEIOPS, 2007) a statistical model is here implemented for RBNS reserve estimation. Lognormality on empirical compensation cost data is observed for different levels of BI severity. The individual claim provision is estimated by allocating the expected mean compensation for the predicted severity of the victim's injury, for which the upper bound is also computed. The BI severity is predicted by means of a heteroscedastic multiple choice model, because empirical evidence has found that the variability in the latent severity of injured individuals travelling by car is not constant. It is shown that this methodology can improve the accuracy of RBNS reserve estimation at all stages, as compared to the subjective assessment that has traditionally been made by practitioners.

Keywords: Automobile accident, Solvency II, bodily injury claims, individual RBNS reserve.

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). Auto liability insurance is compulsory in Spain. Therefore, bodily injury 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. This is because, 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.

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 compensations 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, and as is shown in this paper, such case-by-case valuation could misestimate the final cost, because there are sometimes significant differences between the final compensation awarded by the judge and that assessed directly by the company staff.

In the actuarial literature the focus has mainly been 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). Statistical methods based on individual information have projected compensation payments according to the victim information available in the accident year (e.g. 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.

This paper presents an empirical application for estimating individual RBNS claim reserves which takes into account the compensation cost distribution for different levels of claim severity. In addition, the individual provision is 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.

In order to estimate the claim reserve at each handling stage, the severity of the victim's injury is predicted by means of a heteroscedastic ordered multiple choice model (HOMC). Several researchers have used ordered multiple choice models in the context 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; Ayuso and Santolino, 2007). Methodologically, these studies 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. An interesting development of HOMC models was proposed by O'Donell 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 bodily injury 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. Estimated parameters from the HOMC model at the successive stages are also presented. 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 bodily injury claims database

The dataset contains information from 197 non-fatal victims with bodily injury 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.

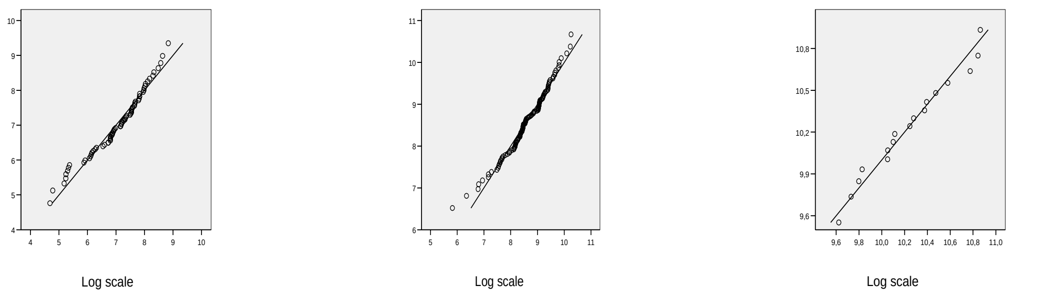
The dependent variable of the HOMC regression model is related to the bodily injury severity of the victim. It has three categories: *Recovery Days*, *Non-severe Injury* and *Severe Injury*. These categories are defined according to the severity of injury assigned to sequelae[1] by the judicial verdict. 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 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.

[1] 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*)

N= 71; K-S Z= 1.171 (0.129) *N*= 109; K-S Z= 1.616 (0.011) *N*= 17; K-S Z= 0.706 (0.702)



N= 71; K-S Z= 0.833 (0.491) N= 109; K-S Z= 0.877(0.426) N= 17; K-S Z= 0.509 (0.958)

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 expressions (see Greene, 1997):

$$\begin{aligned}&\text{if } \ln(y): \ N(\mu,\sigma^2) \quad \text{then}\\ &E[y]=e^{\mu+0,5\sigma^2}\\ &\mathrm{Var}[y]=e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right).\end{aligned} \tag{1}$$

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)

	Estimated mean compensation cost (log scale)	*Standard deviation* (log scale)	*Expected mean compensation cost* (original scale)	*Standard deviation* (original scale)
Recovery Days	7.110	0.953	1927.74	2345.143
Non-Severe Injury	8.620	0.808	7680.44	7371.380
Severe Injury	10.273	0.403	31388.74	13195.383
Total	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 period. 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 examination of the victim made by medical experts appointed by the insurance company at two different points during the recovery period. In particular, we consider the expert valuation of a) the expected number of sequelae and recovery days[2] due to the accident at the first examination, i.e. when the victim is still recovering (first medical report), and b) both these variables when the victim is fully recovered (last medical report). Finally, we also incorporate a dichotomous variable which indicates whether the forensic doctor examined the victim and considered that the accident hadn't 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.

[2] A distinction is made between recovery days with and without disability for working.

TABLE 2. Explanatory variables and descriptive statistics

		Mean	SD
Logcom	Compensation amount (on logarithmic scale) awarded by verdict.	8.219	1.264
year	Accident year (1=1994; 2=1995; ...; 10=2003).	6.975	1.430
year2	Accident year (squared).	50.680	17.151
car	1 if the victim's vehicle is a car; 0=otherwise (e.g. motorbike, pedestrians).	0.650	0.478
age	Victim's age (1 if age 0 to 9; 2 if 10 to 19; and so forth).	3.930	1.606
gender	1 if male; 0=otherwise.	0.497	0.501
passen	1 if the victim is passenger of the insured vehicle; 0=otherwise.	0.091	0.289
seq	Number of sequelae (permanent injuries) expected in first medical report.	1.092	1.340
rdd	Number of recovery days with disability for working expected in first medical report.	53.563	53.971
rdnd	Number of recovery days without disability for working expected in first medical report.	29.109	45.472
same	1 if last medical report is the same as the first one; 0=otherwise.	0.316	0.467
seq_last	Number of sequelae (permanent injuries) stated in last medical report.	1.114	1.655
varseq	Sequelae number variation across reports (last minus first).	0.009	0.917
rdd_last	Number of recovery days with disability for working stated in last medical report.	53.131	63.027
varrdd	Variation in the number of recovery days unable to work across reports (last minus first).	2.079	37.601
rdnd_last	Number of recovery days without disability for working stated in last medical report.	37.596	59.699
foren	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 bodily injury 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.

Parameter estimates for variables used at each stage in the prediction of the individual BI severity level are shown in Table 3. At the bottom of Table 3 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 didn't examine the victim we considered the severity classification made in the last internal medical report. An overview of the heteroscedastic ordered logit model specification and the interpretation of results can be found in Appendix 1. The significance of the scale parameter *car* at three of the four analyzed stages suggests that the variance of the error term varies with vehicle type. 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. Therefore, heterocedastic variance specification is accepted.

TABLE 3. 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**
year	0.795	0.129	1.237	0.118	1.429	0.152	2.291	0.330
year2	-0.082	0.066*	-0.104	0.111	-0.124	0.129	-0.179	0.357
car	-1.462	0.000***	-0.608	0.159	-0.770	0.112	-1.074	0.443
age	0.142	0.101	0.194	0.082*	0.245	0.049**	0.590	0.086*
gender	-0.895	0.003***	-0.877	0.016**	-1.162	0.008***	-2.369	0.072*
passen	0.472	0.319	0.678	0.134	0.643	0.206	-0.275	0.878
seq	-	-	0.701	0.002***	-	-	-	-
rdd	-	-	0.015	0.001***	-	-	-	-
rdnd	-	-	0.008	0.078*	-	-	-	-
same	-	-	-	-	-0.823	0.065*	-2.686	0.170
seq_last	-	-	-	-	0.676	0.006***	1.825	0.039**
varseq	-	-	-	-	-0.686	0.031**	-2.172	0.049**
rdd_last	-	-	-	-	0.014	0.005***	0.021	0.110
varrdd	-	-	-	-	-0.014	0.018**	-0.036	0.082*
rdnd_last	-	-	-	-	0.007	0.059*	0.017	0.046**
foren	-	-	-	-	-	-	-10.159	0.005***
car (scale)	-0.165	0.377	-0.649	0.028**	-0.582	0.055*	0.695	0.074*
	N= 197; pseudo-R^2= 0.189; χ^2= 33.844(0.000)		*N*= 119; pseudo-R^2= 0.611; χ^2= 56.046(0.000)		*N*= 114; pseudo-R^2= 0.647; X^2= 6.847(0.077)		*N*=114; pseudo-R^2= 0.861; χ^2= 57.511(0.000)	
Claims correctly predicted by the model (%)	63.452%		72.269%		78.070%		91.228%	
Claims correctly classified by medical reports (%)	-		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.

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 3, Stage 1). A comparison with the total compensation awarded by the judge is presented in Table 4. The first row of Table 4 shows the number of victims according to the BI severity awarded in the judicial verdict. The aggregated reserve for outstanding claims is estimated for each category of BI severity (4th row), which is obtained by multiplying the empirical mean compensation cost of each severity category (2nd row and also Table 1) by the predicted frequency of victims derived from the HOMC model (3rd row). The fifth and six rows of Table 4 present the percentage of empirical 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 4. Provision derived from the severity level predicted by the HOMC model (Victim information available before any medical report)

	Level of severity			
	Recovery Days	**Non-Severe Injury**	**Severe Injury**	**Total**
Observed frequency (judge)	71	109	17	197
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 amount awarded by the judge	77.60%	140.33%	0.00%	83.76%
Confidence limit* of the HOMC / Total amount awarded by the judge	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.

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 5 the estimated provision based on the injury severity predicted by the heteroscedastic model (Table 3, 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 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 5. 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				
	No injury	**Recovery Days**	**Non-Severe Injury**	**Severe Injury**	**Total**
Observed frequency (judge)*	-	40	67	12	119
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 amount awarded by the judge	-	111.90%	99.27%	67.13%	87.61%
Total provision from the HOMC/ Total amount awarded by the judge	-	103.90%	106.91%	83.91%	97.78%
Confidence limit†† of the HOMC/ Total amount awarded by the judge	-	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 5). 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.

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 6, second row). The estimated provision according to HOMC predictions (Table 3, Stage 3) and that directly derived from the medical expert's classification are presented in Table 6. 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 6. 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				
	No injury	Recovery Days	Non-Severe Injury	Severe Injury	Total
Observed frequency (judge)*	-	40	63	11	114
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 amount awarded by the judge	-	133.21%	80.80%	91.75%	88.88%
Total provision from the HOMC/ Total amount awarded by the judge	-	111.90%	101.81%	82.57%	94.95%
Confidence limit†† of the HOMC/ Total amount awarded by the judge	-	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.

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 6, second row).

The results are shown in Table 7. 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 7. Provision derived from the forensic classification‡ *vs.* provision derived from the HOMC model prediction (Available victim information after the forensic report)

	Level of severity			
	Recovery Days	Non-Severe Injury	Severe Injury	Total
Observed frequency (judge)*	40	63	11	114
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 amount awarded by the judge	106.57%	87.27%	183.49%	126.61%
Total provision from the HOMC/ Total amount awarded by the judge	106.57%%	105.05%	82.57%	96.57%
Confidence limit† of the HOMC/ Total amount awarded by the judge	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 reserve. First, 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 8. 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 8. 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).

4. Conclusions

The time period between the occurrence of a motor accident and the point at which victims are compensated for bodily injury (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 at successive stages of the claim 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

This work has received support from the Spanish Ministry of Education and Science and FEDER grant SEJ2005-00741/ECON.

APPENDIX 1

The heteroscedastic ordered logit model

The ordered logit model is based on a continuous unobserved variable y^* that is modelled by means of a linear regression. The observed variable y_i is discrete, with J ordered response categories. The cumulative probability that subject i belongs to category j or lower ones is,

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

where the μ's are the model thresholds (with $\mu_0=-\infty$ and $\mu_J=+\infty$), $\boldsymbol{\beta}(K\times1)$ is the column vector of K unknown parameters, and $\boldsymbol{x}_i(1\times K)$ 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 $(\hat{\mu}_j - \boldsymbol{x}_i\hat{\boldsymbol{\beta}})$ is the predictor of the expected mean value, and σ_i is the standard deviation. Usually, σ_i is parameterized as $\exp(\boldsymbol{z}_i\boldsymbol{\tau})$ to ensure its positivity, and $\boldsymbol{z}_i\hat{\boldsymbol{\tau}}$ is the variance predictor, with $\hat{\boldsymbol{\tau}}$ $(G\times1)$ the column vector of G unknown scale parameters (O'Donell 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

Table 3 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 3). 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 is statistically significant in three of the four stages, and therefore the heteroscedastic specification is accepted. The significance of the parameter *car* indicates that the variance depends on vehicle type.

References

Abdel-Aty, M., 2003, Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models, *Journal of Safety Research*, 34(5): 597-603.

Antonio, K., J. Beirlant, T. Hoedemarkers, and R. Verlaak, 2006, Lognormal Mixed Models for Reported Claim Reserves, *North American Actuarial Journal*, 10(1): 30-48.

Austin, R., and B. Faigin, 2003, Effect of Vehicle and Crash Factors on Older Occupants, *Journal of Safety Research,* 34(4): 441-452.

Ayuso, M., and M. Santolino, 2007, Predicting Automobile Claims Bodily Injury Severity with Sequential Ordered Logit Models, *Insurance: Mathematics and Economics,* 41(1): 71-83.

CEIOPS, 2007, *QIS 4 Technical Specifications,* CEIOPS DOC-23/07.

England, P., and R. Verall, 2002, Stochastic Claims Reserving in General Insurance, *British Actuarial Journal,* 8(3): 443-544.

Greene, W.H., 1997, *Econometric Analysis*, Third edition, New York: Prentice Hall International.

Haastrup, S., and E. Arjas, 1996, Claims Reserving in Continuous Time: a Non-parametric Bayesian Approach, *ASTIN Bulletin,* 26(2): 139-164.

Karlaftis, M.G., I. Kotzampassakis, and G. Kanellaidis, 2003, An Empirical Investigation of European Drivers' Self-Assessment, *Journal of Safety Research,* 34(2): 207-213.

Kockelman, K., and Y. Kweon, 2002, Driver Injury Severity: An Application of Ordered Probit Models, *Accident Analysis & Prevention*, 34(3): 313-321.

Lee, C., and M. Abdel-Aty, 2005, Comprehensive Analysis of Vehicle-Pedestrian Crashes at Intersections in Florida, *Accident Analysis & Prevention,* 37(4): 75-786.

Norberg, R., 1993, Prediction of Outstanding Liabilities in Non-Life Insurance, *ASTIN Bulletin*, 23(1): 95-115.

Norberg, R., 1999, Prediction of Outstanding Claims II: Model Variations and Extensions. *ASTIN Bulletin*, 29(1): 5-25.

O'Donell, C.J., and D.H. Connor, 1996, Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice, *Accident Analysis & Prevention*, 28(6): 739-756.

Roholte Larsen, C., 2007, An Individual Claims Reserving Model, *ASTIN Bulletin*, 37(1): 113-132.

SCOR, 2006, *Nivel y Evolución del Coste Medio Daño Corporal Grave por Accidentes de Circulación Ocurridos en España*, SCOR Global P&C.

Wang, X., and K. Kockelman, 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., and J. Ivan, 2003, Factors Influencing Injury Severity of Motor Vehicle-Crossing Pedestrian Crashes in Rural Connecticut, *Accident Analysis & Prevention,* 35(3): 369-379.