Cost estimation of traffic bodily injury damages according to different stages of the victim recovery period

Mercedes Ayuso[*] and Miguel Santolino
RISC –IREA Instituto de Investigación en Economía Aplicada, Regional y Pública
University of Barcelona
Avda. Diagonal, 690
08034 Barcelona
Phone: 93 4021409/934021822 and 93 4024318
Fax: 93 402 1821
(mayuso@ub.edu; msantolino@ub.edu)

Abstract

The World Health Organization (WHO) indicates that automobile bodily injury damages are one of the main health problems in the world and, with the current tendency, it will be the third cause of morbidity in the year 2020. In Europe, the implementation of new legal rules has lead to a decrease in the number of automobile accidents, but the number of victims resulting from them is still very large. The compensation for bodily injury suffered by traffic victims is usually covered by an automobile insurance policy, which must encourage full recovery of injured victims, whenever it is possible. In the context of health economics and insurance economics very little research has been done to analyze the amount of resources that are necessary to cover compensations for different types of traffic victims.

Our objective in this paper is to estimate the capital amount which should be reserved by the insurance company in order to guarantee that future bodily injury compensation payments are covered. We will consider: i) the different stages of bodily injuries recovery process, and ii) the influence of the vehicle type involved in the accident on the victim's severity. Reserves estimations are based on predictions of the severity of bodily injured victims, that are obtained with a heteroscedastic ordered discrete choice model. Using an empirical illustration, we show as the proposed methodology improves the reserve calculations that have traditionally been made by insurance companies.

Keywords: traffic accidents, bodily injury damages, morbidity, reserves, heteroscedastic ordered multiple choice model.
JEL classification: I18, C35, G22.

---

[*] Corresponding author: Mercedes Ayuso, University of Barcelona, Avda. Diagonal 690, 08034 Barcelona; phone: +34 93 402 14 09/ +34 93 402 18 22; Fax: +34 93 402 18 21; e-mail: mayuso@ub.edu

1. Introduction

Despite the fact that the number of traffic accidents is declining in many developed countries, compensation payments to bodily injury (BI) victims are increasing in most of them (because of *judicial* inflation, rising medical expenses, and so on). In Spain, between 2001 and 2005, the compensation for seriously injured victims increased on average by 10% annually (SCOR, 2006). The automobile 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. It 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 BI compensations in order to guarantee that victims involved in a traffic accident will be indemnified for the damage.

In current practice, most motor insurance companies estimate the BI victim compensation payment and, therefore, reserve funds based on their own medical reports. These reports are made by medical experts appointed by the insurance company who examine the victims during the recovery period. Such practice may misestimate the final cost, because there are sometimes significant differences between the severity of injury awarded by the judge and the severity assessed by the company.

Our analysis focuses on predicting the victim's injury severity and on using this prediction to compute the capital funds for which the victim's compensation is guaranteed. In many European markets it is often the severity of injury that matters by insurers. The law limiting insurance claims compensation matching each physical injury, e.g. loss of right/left arm or of sight with the corresponding economic value, is very strict. Besides, mild injury compensations have much less variance than severe injury compensations.

In the paper we estimate the reserve amount to cover BI compensations in function of the predicted severity of injured victims and the empirical cost distribution of the corresponding severity level. This modelling framework allows to calibrate the provision of the victim compensation in response to variations of the expected seriousness, immediately after the new victim information about the recovery status is available (we consider four of the most important phases of the victim's recovery process). 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. Furthermore, since distributional assumptions of the compensation costs are taken into account, the suggested approach can be used by the insurer to predict the upper bound for the reserve amount, with appropriate confidence level.

In the economic literature the focus has mainly been on aggregate BI claims reserving which don't consider the specific characteristics of each victim and accident. The few previous works based on individual information have projected compensation payments according to the victim information available at the accident year (e.g. Norberg, 1993; 1999; Haastrup and Arjas, 1996, Antonio *et al*., 2006; Roholte Larsen, 2007). That means, these techniques did not take into account the variations of the victim information during the recovering period and the effects of these fluctuations on the reserves estimation.

On the other hand, the severity of the injured victim is predicted by means of a heteroscedastic ordered multiple choice (HOMC) regression model, where the error term variance is parameterized in terms of the vehicle type. Several researchers have used ordered multiple choice models in the context of motor accidents to identify the factors that influence the severity of an injured victim (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, all previous works assume a constant variance in the random term. However, such an assumption seems to be restrictive and may be unrealistic in case of casualties resulting from accidents involving, for instance, different types of vehicles. In accidents with motorbikes injuries may range from very mild ones, when the driver just falls down, to quite severe ones, such as collisions. Since not all circumstances can be measured and incorporated into the model predictor, we can argue that the random error does not necessarily have the same variance for all individual claims. An interesting development of heteroscedastic ordered choice models was offered by O'Donell and Connor (1996). They suggested that the victim's age, the speed, and the time of the accident were predictors of the error variance. More recently, focusing on transportation safety, Wang and Kockelman (2005) parameterized the error term variance as a function of the vehicle type and the vehicle weight.

The remaining part of this paper is structured as follows. In the next section we describe the database used in the empirical analysis, with especial attention to the bodily injury cost distribution. Section 3 is related with the especification of a heteroscedastic ordered logit model to predict the victim bodily injury severity in function of the accident characteristics and the injury status of the casualty during the recovery period. In section 4 the required capital funds to compensate victims, for whom insurers are legally responsible, are estimated and compared with those calculated by the insurer. We demonstrate as the proposed methodology could help insurance companies to obtain a more accurate estimate of the necessary monetary amount to indemnify motor BI victims. Finally, in Section 5, we summarize the main findings.

2. The bodily injury traffic accidents database

The dataset contains information of 197 victims with bodily injury damages involved in traffic accidents. The database was provided by a Spanish insurer. 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 previous agreement.

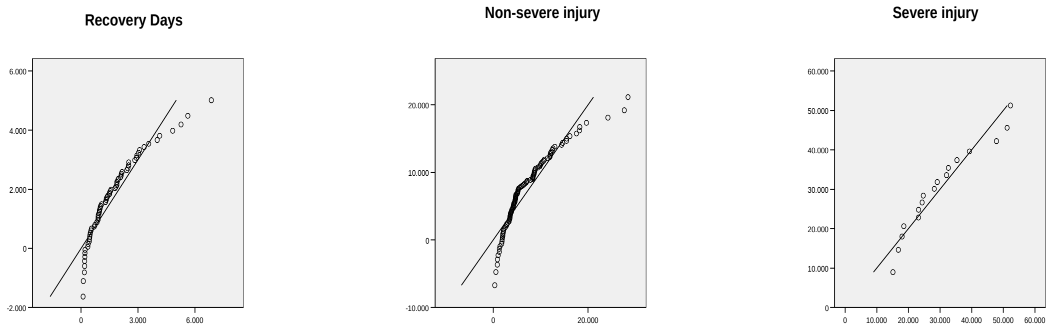
The dependent variable of the HOMC regression model is related with the bodily injury severity of the victim. It has three categories: *Recovery Days*, *Non-severe injury* and *Severe injury*. These categories are defined according to the degree of injury seriousness awarded for sequelae by the judicial verdict. Note that sequela is defined as the definitive reduction of physical or mental potential of a person resulting from an accident. The category *Recovery Days* represents casualties without sequelae. *Non-severe injury* means casualties with less than 15 points for sequelae[†], and finally, *Severe injury* represents victims with 15 or more points.

The individual reserve for each victim depends on the empirical BI compensation cost distribution and the estimated category of BI seriousness. Therefore, first we need to test the normality of the observed compensation cost distribution. We test for each category of BI seriousness whether the victim's compensation awarded by judges (*compen*) and the

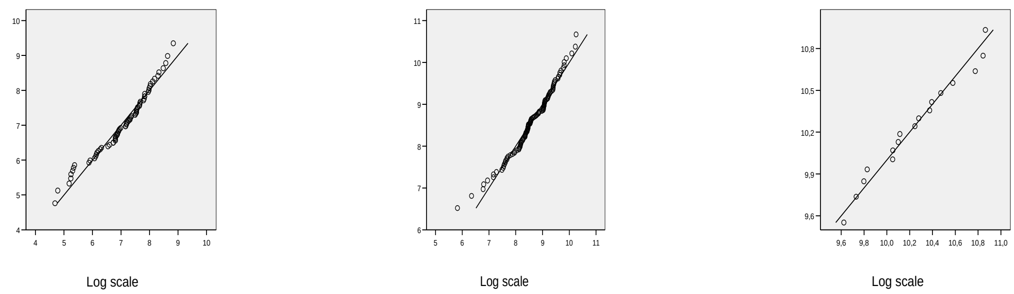
[†] The definition and scoring of sequelae must agree with those stipulated in the Spanish disability scoring system (LOSSP 30/95).

logarithmic transformed scale of this variable (*logcom*) are normally distributed. In Figure 1 the normal Q-Q plot and the Kolmogorov-Smirnov (K-S) test are presented.

Figure 1. Analysis of normality for the BI victims compensation cost data (*compen*) per categories of severity, on original scale (*first row*) and on 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 can not be rejected for any category of BI severity. On the contrary, there are evidences that normality of compensations can not 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. In consequence, we assume that compensation cost data are lognormally distributed. Each victim's compensation is reserved by allocating the expected mean cost of the forecasted level of BI seriousness. Since the compensation cost is lognormally distributed, predictions on the original scale must be obtained with the following expressions (see, e.g., Greene, 1997):

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

Descriptive statistics 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 some descriptive statistics for the total sample are presented in Table 2. Explanatory variables refer to attributes of the victim as the age and the gender, characteristics of the accident or 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. car, motorbike and so on) and if the casualty was a passenger (not the driver) of the damaged vehicle.

Related with the medical information, we consider the valuation of the number of sequelae caused by the accident and the number of recovery days with and without disability for working, at two different moments of the recovery period. Particularly, this information is collected in a first examination when the victim is still recovering (first medical report), and it is again observed when the victim is fully recovered (last medical report). Finally, it is also incorporated a dichotomous variable which indicates if the forensic doctor examined the victim and considered the accident had not caused him/her sequelae.

TABLE 2. Explanatory variables and descriptive statistics

		Mean	Std.Dev.
Logcom	Compensation amount (on logarithmics) 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 if female.	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 valued in first medical report.	1.092	1.340
Rdd	Number of recovery days with disability for working valued in first medical report.	53.563	53.971
Rdnd	Number of recovery days without disability for working valued 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 valued 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 valued 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 valued 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 day*; 109 victims as *Non-severe injury*; 17 victims as *Severe injury*).

3. Predicted victim's bodily injury severity level

In the current section a heteroscedastic ordered logit model is sequentially applied in accordance to the information of the BI victim available to the insurer. Namely, i) a first estimation of the bodily injury severity is carried out just at the moment that the accident is communicated to the insurance company, ii) a second estimation when the company has the initial valuation of the bodily injury damages in the first medical examination (i.e. during the recovering period), iii) a third estimation after the last medical report (i.e. when the victim is fully recovered or with stabilized sequelae), and, finally, iv) a fourth estimation of the BI seriousness is computed after the victim was examined by the forensic doctor.

3.1 The heterocedastic ordered multiple choice model

Ordered multiple choice models are based on a continuous latent variable y^*. This variable may be modelled by means of a linear regression,

$$y_i^* = \boldsymbol{x}_i\boldsymbol{\beta} + \varepsilon_i, \qquad -\infty < y_i^* < +\infty$$

where y_i^* is the true unobserved severity measure for the ith individual claim, i=1,…,N (where N is the sample size), $\boldsymbol{\beta}(K\times1)$ 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. Therefore, we consider that variance may vary across subjects, i.e. the heteroscedastic case. The observed variable y_i is discrete, with J ordered response categories. The functional form introduced by McCullagh and Nelder (1989) that relates the ordered categorical (observed) variable y_i with the latent (unobserved) variable y_i^* is expressed as,

$$\Lambda[\gamma_j(x_i, z_i)] = (\mu_j - \boldsymbol{x}_i\boldsymbol{\beta}) / \sigma_i,$$

where the μ's are the model thresholds (unknown parameters), $\gamma_j(\boldsymbol{x}_i, z_i)$ is the cumulative probability that subject i belongs to category j or lower ones, i.e. $\gamma_j(\boldsymbol{x}_i, z_i) = P(y_i \le j | \boldsymbol{x}_i, z_i)$, $\boldsymbol{z}_i(1\times G)$ is the row vector of the G observed regressors which explain the variance of subject i and $\Lambda(\cdot)$ is the link function that relates the cumulative probability with the predictor part. 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). In the homoscedastic case, $\boldsymbol{\tau}=\boldsymbol{0}$ and then σ_i=1. The link function for the ordered logit model is defined as $\log[\gamma_j(\boldsymbol{x}_i, \boldsymbol{z}_i)/1-\gamma_j(\boldsymbol{x}_i, \boldsymbol{z}_i)]$. Therefore the cumulative probability is,

$$\mathrm{P}(y_i \le j \,|\, \boldsymbol{x}_i, \boldsymbol{z}_i) = \frac{e^{(\mu_j - \boldsymbol{x}_i\boldsymbol{\beta})/\sigma_i}}{1 + e^{(\mu_j - \boldsymbol{x}_i\boldsymbol{\beta})/\sigma_i}}, \tag{2}$$

with the discrete probability of the category j estimated as $\mathrm{P}(y_i = j) = \mathrm{P}(y_i \le j) - \mathrm{P}(y_i \le j-1)$, $j=1,\ldots,J$ (J is the number of categories of severity), with $\mu_0 = -\infty$ and $\mu_J = +\infty$. Parameter estimates are usually obtained by maximum likelihood. The Newton-Raphson algorithm has been used in the maximization process.

The marginal effect of a change in a variable depends not only on its own value but also on the other regressors. For this reason, in the estimation of the marginal effects the sample means of the variables are usually used as representative values. Let us suppose that we are interested in estimating the marginal effect of a unit variation in a variable which is a predictor both of the mean and of the variance. In this case, the estimated marginal effect is computed as,

$$\frac{\partial \hat{P}(y=j\,|\,\bar{\boldsymbol{x}},\bar{\boldsymbol{z}})}{\partial x_k, z_g} = \left[f\left(\frac{\hat{\mu}_{j-1} - \bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}}}{\hat{\bar{\sigma}}}\right) - f\left(\frac{\hat{\mu}_j - \bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}}}{\hat{\bar{\sigma}}}\right)\right]\frac{\hat{\beta}_k}{\hat{\bar{\sigma}}} + \left[f\left(\frac{\hat{\mu}_{j-1} - \bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}}}{\hat{\bar{\sigma}}}\right)\frac{\hat{\tau}_g(\hat{\mu}_{j-1} - \bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}})}{\hat{\bar{\sigma}}} - f\left(\frac{\hat{\mu}_j - \bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}}}{\hat{\bar{\sigma}}}\right)\frac{\hat{\tau}_g(\hat{\mu}_j - \bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}})}{\hat{\bar{\sigma}}}\right], \tag{3}$$

where $\bar{x}$ and $\bar{z}$ are the row vectors consisting of the means of each covariate, $\hat{\mu}$'s are the estimated thresholds, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\tau}}$ are the column vectors of the estimated parameters, $\hat{\bar{\sigma}}$ is the estimated mean standard deviation $\hat{\bar{\sigma}} = \exp(\bar{\boldsymbol{z}}\hat{\boldsymbol{\tau}})$, and $f(\cdot)$ is the logistic density function $f(\varepsilon) = e^{\varepsilon}/(1+e^{\varepsilon})^2$. When the variable is a mean predictor but not a variance predictor, then the marginal effect of a unit variation is obtained by computing only the first component of the right-side part of the equation. The previous computation of marginal effects is adequate for continuous variables. When the variable is dichotomous, then the marginal effect is estimated as the difference between the probabilities in the two possible values of the binary variable and representative values for the rest of regressors.

Since the marginal effect depends on the value of all explanatory variables, the interpretation of coefficients of the ordered choice model is not direct. In the homoscedastic case, the direction of the probability variation for extreme categories may be inferred from the sign of the coefficient. Moreover, the marginal effect over the probabilities of extreme categories is always opposite. This result is an important constraint of the homoscedastic ordered logit model. Several authors believe that certain factors may influence the probability of two oppositional extreme situations in the same direction. One possible example is the use of airbags. It is believed that the use of airbags reduces the probability of both any injury and fatal or severe injury. However, airbags tend to cause minor abrasions. Therefore their use increases the probability of a mild injury (Ulfarsson and Mannering, 2004).

The limitation of opposite effect on the two extreme categories may be reduced by the inclusion of scale parameters $\boldsymbol{\tau}$ that model the error variance. Let us suppose that the variables of the two sets of regressors (mean and variance) are different, $x_k \neq z_g \quad \forall k \in K, g \in G,$ and we want to analyze the effect of a unitary variation in the regressor z_g. The direction of the probability variation for the extreme categories depends on the coefficient sign of $\hat{\tau}_g$ and the sign of the expression $(\hat{\mu}_j - \boldsymbol{x}_i\hat{\boldsymbol{\beta}})$ for j=1,…J, i=1,…N. If we assume that, for example, the response variable has three categories (J=3), the regressor coefficient $\hat{\tau}_g$ is negative, and for a given individual i, $(\hat{\mu}_1 - \boldsymbol{x}_i\hat{\boldsymbol{\beta}}) < 0$ and $(\hat{\mu}_2 - \boldsymbol{x}_i\hat{\boldsymbol{\beta}}) > 0,$ then if the regressor value rises Δz_{ig}, the probability of the extreme categories decreases. Obviously, in this case the probability of the intermediate category (y_i=2) increases. When the expression $(\hat{\mu}_j - \boldsymbol{x}_i\hat{\boldsymbol{\beta}})$ takes the same sign for all j, then a unit variation in the regressor value z_{ig} affects inversely the probability of the extreme categories, so as in the homoscedastic case.

3.2 The estimation results

Four models are estimated in this section. The probability of different BI severity levels for each victim is estimated at the time when the claim is reported to the insurance company, after the internal first medical report (carried out by the doctor of the company), after the internal last medical report (carried out also by the doctor of the company), and finally, when the insurer has the forensic report. Information to complete in the subsequent medical examinations is not much different. Therefore, in the last two models we did not include the information from the first medical report, but we incorporated the variations across reports.

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 avoid that the civil lawsuits were treated as missing values in our dataset. Besides, in the case that forensic doctor examines the victim, he/she may assess the victim's sequelae in order to assist the adjudicator but he/she is not forced to do it. Forensic doctor only must describe the victim's sequelae in concordance with the Spanish disability rating scale. In our modelling framework, the variable *foren* takes value 1 if the forensic doctor checked the victim and decided that the casualty did not have sequelae, and 0 if the forensic doctor awarded sequelae to the victim or there was no forensic report (civil lawsuit).

The parameter estimates for each stage of new information about the BI victim are shown in Table 3. At the bottom of the table is presented the percentage of BI victims for which the model correctly predicted the final BI severity, as it has been finally established in the judicial verdict. 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 information of medical reports and then any percentage is included.

TABLE 3. Estimation of the parameters (sequential heteroscedastic ordered logit model)

	Stage I (Model when the accident is communicated to the insurer)		Stage II (Model after the first medical report)		Stage III (Model after the 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)	
BI damages correctly predicted by the model (%)	63.452%		72.269%		78.070%		91.228%	
BI damages correctly classified by medical reports (%)	-		62.185%		61.403%		83.333%†	

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

As shown in Table 3, the chi-square statistic is significant in all the phases. For a given stage of the estimation, it was computed as the difference between minus two times the log-likelihood for the model in the previous phase and that for the current model. The statistical significance then means that the incoming information at each stage has explanatory capacity in regards to the severity of a victim's injury.

The variable *gender* has a significant coefficient in all the phases and its negative sign indicates minor severity injuries for male victims. The variable *age* behaves in a similar way, but in this case the positive sign shows a higher severity for older people. In relation to the information from medical reports, both the number of recovery days (on disability and not on 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 initial medical report provides information relevant to the explanation of the injury severity, even when the company already has the final report or the forensic examination results. Ayuso and Santolino (2007) suggested that the first medical report information was still relevant in subsequent stages because the initial medical examination pursued a different goal inside the company that the last medical evaluation. Concerning the last phase of the estimation, let us emphasize that the percentage of cases accurately estimated by the model increased notably when the forensic report information was included. This relationship between the forensic report and accuracy of estimations indicates a strong influence of the forensic evaluation on the level of severity awarded by the judge.

We would like to point out that the scale parameter is statistically significant in three of the four phases. Since there is only one scale parameter in the model, this result shows that the heteroscedastic variance specification is accepted. The significance of the scale parameter *car* suggests that the variance for the BI severity varies with the vehicle type. Individuals travelling by car at the moment of the accident exhibit different variability on the latent injury severity than those travelling by motorbikes or pedestrians.

In Annex 1 we can find an example of computing marginal effects for a unit variation in the variable *car*, with the same sign effect in the extreme categories (i.e. only recovery days and severely injured), justifying the use of an heteroscedastic ordered logit model.

4. Estimated capital funds requirements to cover BI casualties

In this section we implement the heteroscedastic multiple choice model results in the computation of the necessary capital funds (reserves) to guarantee the BI victim compensation. In order to calculate these reserves, we must estimate the expected individual compensation of each BI victim. Namely, the estimation of the reserve is calculated as the sum of individual estimates of the victims' compensations. Each indemnity (individual provision) is firstly estimated when the claim is reported, and it is later revised in the successive phases of the victim's recovering, according to variations in the predicted category of the casualty severity. A comparison with the observed BI compensation amount –awarded by the judge- and with the provision directly derived from the insurer's medical examination is given to evaluate the accuracy of the estimated reserves obtained by the proposed methodology.

The first BI cost estimation: when the claim is communicated to the insurer

In the first phase, we allocate to each observation the expected mean compensation of the victim corresponding to the severity level predicted by the heteroscedastic ordered logit model. A comparison with the compensation awarded by the judge is presented in Table 4. In the first row of this table it is shown the observed number of victims in function of the BI severity awarded in the judicial verdict. Per each category of BI severity, the estimated reserve for victims' compensations (4th row) is obtained multiplying the empirical mean compensation (2nd row) by the predicted frequency of victims according the HOMC model (3rd row). The percentage of the empirical compensations covered by the estimated reserve is presented in row 5th. The upper-bound estimate of the reserve for a 95% confidence level appears in the last row. The same design is followed for tables presented in next stages.

TABLE 4. Estimated reserves from the severity level predicted by the HOMC model (Victim information when the accident is communicated to the insurer)

	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. Because of these constraints in the prediction of the BI seriousness, when the provision is calculated, the economic resources are concentrated on the second category (*Non-severe injury*). The overprovision of the intermediate category is not enough to counterbalance the under-provision of the extreme categories. As a result, the total reserve only covers about 84% of the entire compensation amount of the BI victims. Note that this first estimation of reserves has been carried out with very little information about the victims.

The second stage: reserves after the first insurer's medical report

With the first medical report, an initial professional assessment of the damages is submitted to the insurer. In Table 5 the estimated provision based on the injury severity predicted by the heteroscedastic model is compared to the provision based on the direct classification of the medical expert. The same criterion of allocating the expected mean cost of the corresponding category of severity was applied.

Note that the total number of BI victims is now different than in the previous phase. This is due to the fact that we have taken into account the BI victims for which only the first medical report was made. As a consequence, the expected mean compensation for each level of BI seriousness, which is directly observed from the subsample of BI victims for whom the first medical report was made, appeared to be slightly different from the one presented for the whole sample (Table 2).

TABLE 5. Estimated reserves from the insurer's medical expert classification *vs.* those 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 that have the first medical report.
† Medical expert awarded neither recovery days nor sequelae to the victim.
†† 95% Confidence limit.

Severely injured victims were again underprovisioned. When the HOMC model was applied, in aggregated terms, the misclassified victims were mainly diverted to the *Non-severe injury* category. On the contrary, following the inusrer's medical expert evaluation, the *Recovery days* victims were primarily overclassified and therefore overprovisioned. Since the individual provision of a *Non-severe injury* victim is higher than that of a *Recovery days*, the aggregated provision seems to fit better with the proposed methodology. The estimated provision in this last case covered about 98% of the total compensation amount, whereas the provision based on the medical expert's classification from the insurance company covered only 88% of that amount.

The third stage: reserve after the last insurer's medical report

At this stage the insurer has the last medical report indicating that the victim is fully recovered (with or without sequelae). Consequently, the sample is composed of victims for which the insurance company had the first and the last medical report. As in the previous phases, the estimated provision according to the victim severity level predicted by the HOMC model is compared to the provision directly derived from the insurer's classification (Table 6). Note that, in contrast to this last one, the number of victims predicted by the HOMC model at each level of BI severity was now closer to the judge's evaluation. Let us emphasize that the estimated reserve is again proper to meet BI victims' compensations, covering the point and the upper-bound estimates the 95% and 107% of the empirical compensations, respectively.

TABLE 6. Estimated reserves from the insurer's medical expert classification *vs.* those 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 that have the first and the last medical report.
† Medical expert awarded neither recovery days nor sequelae to the victim.
†† 95% Confidence limit.

The last stage before the trial: reserves after forensic report (if it exists)

Lastly, the reserve was estimated when the insurer also had the victim's BI damages information provided by the forensic report, if one existed. This phase is the last one before the case is taken to trial. In our analysis, at that moment, the sample size was equal to the sample size we had in the previous phase, after the last medical report. For this reason, in the estimation of the provision, the same expected mean compensations for the distinct levels of BI severity were considered.

The results are shown in Table 7. Following the classification of severity made by the forensic doctor, we observed an overprovision for covering the compensation of *Severe injury* victims. Consequently, the total reserve exceeded the real final compensation amount by more than 26%. On the contrary, the proposed methodology provided a more accurate estimation for reserving BI victims' compensations. The total provision estimated by means of the HOMC model represented in this case a 96% of the total amount.

TABLE 7. Estimated reserves from the forensic classification‡ *vs.* those derived from the HOMC model prediction (Available claim 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 classification (in last report) for those claims without 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 that have the first and the last medical report.
† 95% Confidence limit.

The advantage of case by case estimation methods is that they use the (available) specific information of the BI victim in order to compute the necessary capital funds to cover his/her compensation amount. However, as it has been appointed by the European Committee in charge of the Solvency II regulation (CEIOPS, 2007), these methods can be directly related with the claims settlement staff of the insurance company, and thus can be rather subjective valuation methods. In the paper it is shown that statistical methods based on individual data overcome this limitation keeping the advantage of using the specific victim information in the estimation.

5. Conclusions

When a traffic accident occurs the insurer must face the calculation of the necessary capital funds (reserves) to meet BI damages compensations. Insurance companies traditionally assess compensations of BI victims according to their own medical reports. Subsequently, they compute the total reserve as the sum of the estimated individual provisions. Unfortunately, there are often substantial differences between the severity level assessed by the medical reports and the level awarded by the judicial verdict.

In this paper we apply a heteroscedastic ordered multiple choice model to estimate the individual reserves for a sample of BI traffic victims. Firstly, we analyze the variables that are explaining different BI severity levels and demonstrate that individuals travelling by car at the moment of accident (in contrast to those on motorbikes or pedestrians) present different variability on the latent injury severity level. Thus, the homoscedastic assumptions should be relaxed. The results show that if the victim was either travelling by motorbike or was a pedestrian, both the probability of being severely injured and of not suffering any sequelae increased.

Once the predicted severity is available, the bodily injury damages for the victim are provisioned by allocating the expected mean cost of the estimated severity level. The model is sequentially applied at different stages, according information about the victim recovery process (including information from the forensic report, if it exists). In this paper we demonstrate as reserves estimated with the heterocedastic ordered logit model are more accurate to those finally awarded by the judge, in comparison to those based on the insurance company medical reports. Due to the estimation of the moments of the cost distribution, this methodology also allows us to compute the limit of the capital funds to cover bodily injury damages with an appropriate confidence level.

Acknowledgements

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

Annex 1

We estimated the marginal effects for a unit variation in the variable *car*, as an example of a predictor with the same sign effect on extreme categories. This variable was included into the model as a linear predictor of the mean (x_{car}) and of the variance (z_{car}). Since *car* is a binary variable, the marginal effect was estimated as the difference between probabilities in the two possible values, and taking the simple mean values for the remaining regressors. For instance, when the marginal effect was estimated in this way at stage II, for a victim travelling by car (compared to one travelling by motorbike or a pedestrian), the probability of *Recovery Days* decreased by 1.5%, and the probability of *Severe injury* by 6.5%. These results were counterbalanced by an increase in the probability of *Non-severe injury* (8%). This result can now be illustrated by an example.

Let us suppose that we are interested in estimating the probabilities of BI severity for a 20-year-old victim, male and driver of a vehicle involved in an accident that took place in year 2000. The first medical report is available and the medical expert decided that the victim had 4 sequelae, and he would need 40 days (20 days with disability for working and 20 days without disability) to fully recover. For a victim travelling by car, the probability of the severity leading to *Recovery Days* is 3.35%, the probability of *Non-severe injury* is 95.30% and of *Severe injury* is 1.35%. However, for a victim travelling by motorbike or for a pedestrian, the probability of *Recovery Days* is 8.59%, the probability of *Non-severe injury* 75.11% and *Severe injury* 16.3%. Note that both extreme probabilities rise if the victim does not travel by car. This seems to make sense. Although almost all motorbike or pedestrian accidents result in personal injury, these may be either very mild injuries (motorbike driver falls down, pedestrian is hit by a vehicle in the city) or very severe injuries (collision of a motorbike with other vehicles, or a pedestrian hit on a major road).

If we use the same explanatory variables but assume that the model is homoscedastic, the results are different. In that case, for a victim travelling by car, the probability of the levels of severity is 8.47% (*Recovery Days*), 86.51% (*Non-severe injury*) and 5.02% (*Severe injury*), and for a victim travelling by motorbike or a pedestrian the probability is 3.60% (*Recovery Days*), 84.81% (*Non-severe injury*) and 11.59% (*Severe injury*). Note that the marginal effect over the extreme categories is now opposite, increasing the probability of *Severe injury*, but decreasing the probability of *Recovery Days.* This result would contradict the data on BI claim compensations provided by the Spanish automobile insurers association. They show that the percentage of motorbike BI claims with low compensations is higher than the percentage of car BI claims with low compensations (UNESPA, 2006). Therefore, there is evidence that certain factors may affect extreme probabilities in same direction, and we see that the heteroscedastic model overcomes this constraint of the classical ordered logit

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.

Greene, W.H., 1997, *Econometric Analysis*, Third edition, 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.

McCullagh, P., and J.A. Nelder, 1989, *Generalized Linear Models,* Second edition, London: Chapman & Hall.

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.

Ulfarsson, G.F., and F.L. Mannering, 2004, Differences in Male and Female Injury Severities in Sport-Utility Vehicle, Minivan, Pickup and Passenger Car Accidents, *Accident Analysis & Prevention,* 36(2): 135-147.

UNESPA, 2006, *Memoria Social del Seguro Español 2005*, UNESPA Asociación Empresarial del Seguro.

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.