Modelling the disability severity score in motor insurance claims: an application to the Spanish case♣

By Miguel Santolino[1] and Jean-Philippe Boucher[2]

[1]RFA - IREA. Department of Econometrics, University of Barcelona. Avda. Diagonal, 690. 08034 Barcelona, Spain. Email: msantolino@ub.edu.

[2] Département de Mathématiques, Université du Québec à Montréal 201, avenue du Président-Kennedy, Montréal (Québec), Canada. Email: boucher.jean-philippe@uqam.ca.

Abstract: Bodily injury claims have the greatest impact on the claim costs of motor insurance companies. The disability severity of motor claims is assessed in numerous European countries by means of score systems. In this paper a zero inflated generalized Poisson regression model is implemented to estimate the disability severity score of victims in-volved in motor accidents on Spanish roads. We show that the injury severity estimates may be automatically converted into financial terms by insurers at any point of the claim handling process. As such, the methodology described may be used by motor insurers operating in the Spanish market to monitor the size of bodily injury claims. By using insurance data, various applications are presented in which the score estimate of disability severity is of value to insurers, either for computing the claim compensation or for claim reserve purposes.

Key words: P Motor accident, disability severity, zero-inflated generalized Poisson model, disability scoring scale.

♣ This research was supported by the Spanish Ministry of Science and Innovation and FEDER grants SEJ2005-00741/ECON and ECO2008-01223/ECON. The authors are grateful to Vinzenz Erhardt, for valuable comments regarding the ZIGP regression model, and to the members of the Risk in Finance and Insurance Group at the University of Barcelona, Montserrat Guillén, PhD, Mercedes Ayuso, PhD and Lluís Bermúdez, PhD for their reviews of the manuscript before submission.

1 Introduction

Motor insurance is the most prevalent insurance line in the world and, in Europe, the largest sector in non-life insurance. In 2006, European motor insurance companies generated a total premium income of almost €130bn, though this represented a 1% fall on the 2005 figure. This negative growth is mainly attributable to the high level of competition in the motor insurance industry (CEA, 2007) in which motor insurance companies need to implement adequate claim handling practices as a cost-saving mechanism.

Our aim in this study is to analyze the factors which influence the disability severity sustained by victims of motor vehicle accidents that the insurance company has to compensate. Specifically, the underlying disability severity is modelled here by means of a zero-inflated generalized Poissson (ZIGP) regression model. To the best of our knowledge, a ZIGP model that allows a regression on the overdispersion and zero inflation has not been previously used in the actuarial literature. We show that the estimated injury severity may subsequently be expressed in financial terms and an estimation of the injury claim cost can be provided. Applications in the claim handling process are derived from the cost estimation of injuries, such as the amount required either to reach a compensation agreement or to reserve the claim.

Motor claims involving bodily injury (BI) have different characteristics to material damage claims which means that, in practice, these two types of claim are separately processed by most motor insurance companies. More specifically, BI claims are less frequent, but involve larger compensation payouts, greater variability in the payments and higher litigation rates. As a result, BI claim settlements have the largest impact on insurers' claims expenditure (Bell, 2006; CEA, 2007) and entail a long handling period. This study focuses on BI claims made in Spain where a legislative compensation system is in force for claim settlements. The legislation stipulates the financial compensation to be awarded to victims according to the degree of injuries sustained. To some extent, the financial compensation awarded is automatically fixed according to the severity of the injury. Therefore, the main discrepancies between insurer and plaintiff are concerned with the underlying severity of bodily injuries rather than the sum to be awarded.

In conducting the injury severity evaluation a distinction needs to be drawn between disability and temporary disability. At the claim settlement moment, the victim may have either fully recovered or may present stable injuries. As such, the duration of the period of temporary disability can, in principle, be proved (by the claimant) and verified (by the insurer). By contrast, the disability refers to the physical impairment that the victim will suffer for the rest of his life. In line with the definition provided by the European Committee on Legal Affairs and the Internal Market (EC, 2003), disability can be defined as 'the definitive reduction of physical and/or mental potential which can be identified or explained medically, together with the pain and mental suffering known by the doctor to be a normal concomitant of the sequela plus the everyday consequences which commonly and objectively accompany that sequela', where sequela is any negative after-effect resulting from the accident. In Spain, the evaluation of the disability severity is undertaken using a disability scoring scale. In practice, medical examinations with different severity scores for disability are often submitted as evidence by both parties, the insurer and the plaintiff, to a claim settlement. This means that the severity of the disability is one of the main issues when the compensation is being sought.

In this study the severity score is estimated by means of a ZIGP regression model. The ZIGP model is useful for analysing overdispersed count data with a large amount of zeros and our motor BI claim insurance data show both features. The data present many zeros as many bodily injury victims do not suffer permanent disability and the distribution presents a heavy right tail. Many studies have been undertaken in the field of motor insurance claims dealing with count data that present an excess of zeros and long right tails (see, for instance, Boucher and Denuit, 2008; Boucher *et al.*, 2007; Yip and You, 2005). In these research papers zero-inflated count models are used in modelling the motor claim frequency rather than the claim severity. The severity of motor claims has traditionally been investigated from a financial viewpoint, i.e., modelling the claim cost. Several examples of statistical applications can be found where the event studied is -the logarithmic transformation of- the motor claim payout (for claim reserves, see Antonio *et al.*, 2006; for fraud investigation, Hoyt *et al.*, 2006; Crocker and Tennyson, 2002; Weisberg and Derrig, 1998; for the evaluation of tort reforms, Browne and Puelz, 1996; 1999; Browne and Wells, 1999; and for the analysis of specific effects such as gender and age on the compensation amount awarded, Doerpinghaus *et al.*, 2008).

An alternative methodology for modelling the size of motor insurance claims is provided by Ayuso and Santolino (2007). The authors deal with the injury severity of victims involved in motor claims. The injuries are categorized by degrees of severity and estimated by means of a sequential ordered logit model at different stages in the claim history. This qualitative approach to motor injury severity, although not well known in the field of insurance, is familiar to other audiences such as medical scientists and public health planners (e.g. Eluru *et al.*, 2008; O'Donell and Connor, 1996; Wang and Kockelman, 2005; Abdel-Aty and Abdelwahab, 2004). In common with Ayuso and Santolino (2007), we also focus on the severity of bodily injuries sustained by motor victims in this study. However, we analyse the observed severity scoring of injuries rather than use a qualitative classification of the claim severity. The advantage of this approach is that modelling the score is more flexible than modelling qualitative levels. In addition, the score estimate may be directly expressed in financial terms since the compensation to be awarded is stipulated by law according to the stated score. By contrast, a summary measure of the compensation cost (as, for instance, the mean compensation) has to be used when severity categories are considered.

In the latter part of this study we present a number of examples in which the severity score estimate of the disability might have practical applications for the insurer. In particular, we show that the information provided by the disability severity estimate can be used as an indicator of the inability to work, to compute the compensation to be awarded or for claim reserving purposes, among others. Although the model results are valid only for Spain, they may be accommodated to other European States that also apply disability scales in the evaluation of the disability in motor insurance claims. Examples of States with medical scales for the assessment of BI claim compensations include Italy, Portugal, Belgium and France. In addition, a project is now underway in the European Union to harmonize disability assessment practices involving the application of a European disability rating scale (EC, 2003).

The rest of the paper is organised as follows: in the second section, we describe the insurance BI claim data set used in the empirical study. In the next section, a revision of the zero-inflated generalized Poisson regression model is provided. In section four, we describe the model regressors and present the main results of this study. Applications of these results to the insurer's claim handling process are described in section five. Finally, in the last section, our main conclusions are summarized.

2 The data

The study consists of a sample of 180 bodily injury claim settlements provided by a Spanish insurance company. Each claim comprises a non-fatal bodily injury victim who has been seeking compensation from the insurance company. Motor liability insurance policy is compulsory in Spain. Therefore, only no-fault victims involved in motor accidents are entitled to compensation for bodily injuries. All claims were settled by Court decision between 2001 and 2003.

Two classes of bodily injury are entitled to financial compensation: disability and temporary disability. The temporary disability can be defined in reference to the period of time from the accident until the victim is fully recovered or presents a stable medical condition. Disability, on the other hand, is any residual impairment suffered by the victim once his medical situation has stabilized. As mentioned above, the degree of disability is evaluated in accordance with the legislative disability scoring scale in force in Spanish[1]. The scale ranges from zero to a hundred points according to the degree of severity. The disability scale describes the possible sequelae resulting from traffic accidents and provides a maximum-minimum score for each. The score awarded for a sequela is an integer inside the established bounds based on its degree of severity. In case where more than one sequela is sustained by the victim, a recursive formula is provided to ensure that the aggregated score for disability sets an integer with an upper bound of 100.

The goal of this study is to analyse and estimate the degree of disability sustained by each injured victim. Therefore, the severity score awarded by Court for the disability is the dependent variable of the regression model, and this is labelled as *score*. The variable *score* is discrete and takes a value of zero when the victims make a full recovery after the period of temporary disability. An examination of the empirical distribution of the variable *score* is provided in Table 1. The empirical distribution seems to show extra of zeros in comparison with the Poisson distribution. Specifically, almost 70 victims only suffered temporary disability as a result of the motor accident (score equal to zero). An overdispersion of data also seems likely. It is broadly known that the variance is equal to the mean in the Poisson distribution. Nevertheless, the empirical distribution shown in Table 1 would have a very heavy right tail to satisfy this condition. The mass of observations concentrates on victims with a score below 15 points, although there are several claims with larger scores. Given these characteristics, the zero-inflated generalized Poisson distribution may provide a better fit to the permanent disability scoring data.

[1] Although the application of the scale is compulsory only for the evaluation of motor victims, it is broadly used by the Courts in the medical assessment of any injured victim.

Table 1. Frequency of the variable score

Scoring awarded by Courts for disability	Number of claims
0	69
1	12
2	15
3	14
4	13
5	8
6	5
7	4
8	3
9	3
10	6
11	3
12	3
13	4
14	3
15	3
16	1
18	1
21	3
22	1
24	1
25	1
26	1
27	1
>=28	2

N=180; *Mean*=4.761; *Variance*=7.203.

The individual characteristics of each BI claim were recorded in 43 covariates, of which five were continuous variables. Twenty seven variables relate to general information about the BI claim. Thus, 13 variables report on the victim's personal attributes (age, sex, marital status and working status) and 14 report characteristics of the accident (e.g. party at fault, position inside the car, etc.). The remaining variables relate to whether follow-up examinations were carried out by the insurance staff during the claim handling or whether independent examiners were responsible for these tasks. One of our objectives was to analyse whether the length of the time that the victim spent recovering from his or her injuries could explain the severity of permanent disability. The temporary disability sustained by the victim was monitored by the number of days required to recover as stated by the forensic scientist [2].

3 Methodology

Count data with a large zero fraction and a heavy tail are common in a number of applications. The standard Poisson distribution does not fit properly when data present these features. Here the Poisson distribution is extended by overdispersion and zero-inflation parameters, known as the zero-inflation generalized Poisson (ZIGP) distribution. This distribution is particularly flexible for modelling count data with an excess of zeros because it simultaneously allows for the two sources of overdispersion. Specifically, the ZIGP distribution is a mixture of a Bernoulli distribution and a generalized Poisson distribution (Joe and Zhu, 2005; Czado *et al.*, 2007). Joe and Zhu (2005) compared the ZIGP distribution with the zero-inflated negative binomial and concluded that the ZIGP distribution would fit more appropriately when there was a large zero fraction and long tail.

The ZIGP regression model, introduced by Famoye and Singh (2003), assumes that the excess of zeros is generated by a bimodal distribution (Cheung, 2002; Lord *et al.*; 2005; 2007). Here we follow the definition of accident severity provided by Chang and Mannering (1999) in which severity is determined according to the level of injury sustained by the most seriously injured occupant. An interpretation for our data set could be, for instance, that victims with only temporary disability belong to two different groups of accident according to the type of road on which the crash occurred, in an urban area or on an inter-urban road. Motor accidents in urban areas are mild crashes which mainly result in material damages and temporary injuries of a minor severity (e.g., bruising, whiplash). By contrast, inter-urban road crashes are more severe accidents in which the degree of disability sustained by the casualties, including victims with only temporary disability, follows a generalized Poisson process.

[2] In Spain, victims must bring a lawsuit in order to be entitled to compensation payment. When the lawsuit follows a criminal procedure, a forensic doctor examines the recovered victim and files a report.

Let the response variable *score* take the value y_i which is the severity score awarded by the Courts to the motor accident victim i for the disability sustained, $i = 1, \ldots, n$ where n is the number of individuals in the study sample. The probability density function of the zero-inflated generalized Poisson regression model is defined as,

$$\begin{aligned} P(Y = y_i) &= \omega_i + (1 - \omega_i) \exp(-\tfrac{\mu_i}{\varphi_i}) & y_i = 0 \\ &= (1 - \omega_i) \tfrac{\mu_i(\mu_i + (\varphi_i - 1)y_i)^{y_i - 1}}{y_i!} \varphi_i^{-y_i} \exp(-\tfrac{\mu_i + (\varphi_i - 1)y_i}{\varphi_i}) & y_i \neq 0 \end{aligned} \tag{1}$$

where $y_i = 0, 1, 2, \ldots$. The function μ_i satisfies $\mu_i = \exp(x_i'\beta)$ where x_i is the $(p \times 1)$ vector of covariates and β the $(p \times 1)$ vector of unknown parameters. The dispersion function is denoted by φ_i. We should stress that the model definition provided by (1) also allows for underdispersion ($\varphi_i < 1$). Nevertheless, an important constraint in the case of underdispersion is that the lower bound of φ_i will depend on μ_i (for details, see Czado and Min, 2005). Here, we focus on the more frequent case in which the data are overdispersed. In particular, φ_i is modelled as $\varphi_i = 1 + \exp(z_i'\gamma)$ to ensure $\varphi_i > 1$, where z_i is the $(q \times 1)$ vector of covariates and γ the $(q \times 1)$ vector of unknown parameters. Finally, the probability of there being an extra zero $\omega_i \in [0, 1]$ is modelled via the logit function $\omega_i = \exp(\alpha_i'\delta)/(1 + \exp(\alpha_i'\delta))$ where α_i is the $(r \times 1)$ vector of covariates and δ the $(r \times 1)$ vector of unknown parameters. Note that zero-inflation and overdispersion presented in (1) are not constrained to be constant which might be overly restrictive for certain data sets (Bae *et al.*, 2005; Famoye and Singh, 2006; Czado *et al.*, 2007).

The mean and variance (conditional on the covariates) for y_i are given by $E(y_i) = (1 - \omega_i)\mu_i$ and $Var(y_i) = E(y_i)(\varphi_i + \omega_i\mu_i)$, respectively. Model (1) reduces to the zero-inflated Poisson (ZIP) when $\varphi_i = 1$, to the generalized Poisson (GP) when $\omega_i = 0$ and to the Poisson regression when $\varphi_i = 1$ and $\omega_i = 0$. The log-likelihood of the ZIGP$(\mu_i, \varphi_i, \omega_i)$ regression model defined in (1) can be written as,

$$
\begin{aligned}
&\log\left(L\right) = \sum\nolimits_{y_i=0} \log\left[\omega_i + (1-\omega_i)\exp(-\tfrac{\mu_i}{\varphi_i})\right] \\
&+ \sum\nolimits_{y_i>0} \left\{\log\left[\tfrac{(1-\omega_i)\mu_i}{\varphi_i}\right] - \log(y_i!) + (y_i - 1)\log\left[\tfrac{\mu_i+(\varphi_i-1)y_i}{\varphi_i}\right] - \tfrac{\mu_i+(\varphi_i-1)y_i}{\varphi_i}\right\}
\end{aligned} \tag{2}
$$

Parameters are estimated by differentiating (2) and solving simultaneously the likelihood equations. Maximum Likelihood (ML) estimates satisfy the properties of consistency and asymptotic normality although the ZIGP distribution does not belong to the exponential family (Czado and Min, 2005; Czado *et al.*, 2007). Asymptotic Wald statistics are provided to check the significance of the covariate parameters. Goodness of fit and model selection were analysed by the log-likelihood using the Akaike information criterion (see, for instance, Li *et al.*, 2008; Czado *et al.*, 2007; Yip and Yau, 2005). As in previous studies (Li *et al.*, 2008; Lee and Mannering, 2002; Czado *et al.*, 2007), the comparison between the zero inflated generalized Poisson (ZIGP) regression model and the generalized Poisson (GP) regression model is conducted by means of the Vuong's test (Vuong, 1989). For a description of the Vuong's test and its application in testing the hypothesis of zero-inflation, we refer to Lee and Mannering (2002). To test the appropriateness of using the overdispersed model as opposed to the traditional model we also applied the Vuong's test. Additionally, we include the test recently proposed by Clarke for model comparisons (Clarke, 2007). The null hypothesis tested by the Clarke statistic is that the ZIGP and the Poisson (either the GP or the ZIP, as applicable) are equally good regression models against the alternative hypothesis that the ZIGP regression model is better.

4 Results

The ZIGP regression model was fitted to our data and adjusted for covariates. The parameter estimation was obtained using maximum likelihood according to the methodology presented in section 3. We conducted the analysis on a variety of regression designs. All regression models were estimated by using the ZIGP package v.2.7 for R available on CRAN (Erhardt, 2008). A sequential backward criterion was followed in the model selection. Only covariates with significant coefficients were included in the models. Design matrices differed between the regression models and, for this reason, the AIC criterion was used for model selection. In Table 2 a detailed description of the explanatory variables selected in the final model is provided and their descriptive statistics are shown.

Table 2. Descriptive statistics of variables

Variable	Description	Mean	Std.Dev.
Dependent variable			
score	Severity scoring awarded by Trial Court for disability.	4.761	7.203
Regressors			
year	Accident year (1=2003, 2=2002, ..., 10=1994).	4.000	1.398
fault	1 if fault of the accident is not clearly attributable according to the internal evaluation of the insurer; 0 otherwise.	0.144	0.352
moto	1 if the injured victim was a motorcyclist; 0 otherwise.	0.239	0.428
ped	1 if the injured victim was a pedestrian or cyclist; 0 otherwise.	0.111	0.315
age	Age of the victim.	33.900	15.758
hrd	Number of days recovering in hospital as reported by the forensic doctor.	2.167	8.155
drd	Number of days recovering out of hospital with a disability preventing victim from working as reported by the forensic doctor.	71.383	97.861
open	Time elapsed between the occurrence of the motor accident and the opening of the claim by the insurer (in days/100).	0.404	1.015
gender	1 if the victim is a man; 0 otherwise.	0.500	0.501

4.1 Parameter estimates

The final model design and the resulting parameter estimates are shown in Table 3. Eight explanatory variables were included in the regression of the mean and one in the overdispersion and zero-inflation regressions, respectively. The Vuong test was estimated comparing the zero-inflated generalized Poisson regression model with the Poisson regression model with equal mean design. The Vuong test results showed that the ZIGP regression model was preferred (v: 5.43; p-value: 0.00). The ZIGP was also preferred in comparison with either the GP regression model (v: 2.87; p-value: 0.00) or the ZIP regression model (v: 3.73; p-value: 0.00) with the same matrix designs. No variations in the model selection were found when the decision was based on the results of the Clarke statistic.

Table 3. Estimates of parameters

	Poisson model		GP model		ZIP model		ZIGP model	
Variable	Coeff	p-value	Coeff	p-value	Coeff.	p-value	Coeff.	p-value
Mean								
Intercept	−0.7378	0.004	−0.7970	0.147	−0.3181	0.259	−1.3289	0.012
year	0.0810	0.003	0.1292	0.020	0.0475	0.098	0.1205	0.013
fault	−0.7575	0.000	−0.6716	0.010	−0.4915	0.000	−0.3960	0.097
moto	1.0374	0.000	1.0540	0.000	0.7823	0.000	1.0308	0.000
ped	0.9092	0.000	1.1255	0.000	0.5303	0.000	1.1058	0.000
age	0.0489	0.000	0.0470	0.036	0.0640	0.000	0.0887	0.000
age^2	−0.0004	0.000	−0.0004	0.087	−0.0009	0.000	−0.0009	0.000
hrd	0.0140	0.000	0.0080	0.313	0.0179	0.000	0.0205	0.002
drd	0.0042	0.000	0.0039	0.000	0.0034	0.000	0.0034	0.000
Over-dispersion								
Intercept	–	–	0.5157	0.002	–	–	−0.0975	0.611
open	–	–	0.3719	0.012	–	–	0.4960	0.003
Zero-inflation								
Intercept	–	–	–	–	−1.1419	0.000	−2.4424	0.000
gender	–	–	–	–	1.0254	0.002	1.8377	0.007
Log-likelihood:	−622.864		−421.746		−461.235		−395.975	
AIC:	1263.728		865.492		944.470		817.951	

N=180; The ZIGP model is the preferred model according to the Vuong statistic when compared with either the Poisson model (v: 5.43; p-value: 0.00), the GP model (v: 2.87; p-value: 0.00) or the ZIP model (v: 3.73; p-value: 0.00). The same model selection decisions are reached with the Clarke test. In all cases, the ZIGP model is preferred at the 1% of significance level.

Odds-ratio and relative risk estimates from the ZIGP regression model are shown in Table 4. The odds-ratio (OR) measure is widely used with logistic regressions. The relative risk (RR) is a standard measure in medical research used to compute the ratio between the risk of an event relative to exposure (for details of both measures, see Simon, 2001). The effects of factors on the severity for disability matched up with the expected direction in most cases. The odds for men were six times greater than those for women victims (OR_{gender}:

6.282), which means that women suffer more severe injuries than men in motor accidents. This result is also supported by previous findings (e.g. Kockelman and Kweon, 2002; Lee and Abdel-Aty, 2005).

The generalized Poisson part of the model showed that the relative risk of motorcycle victims compared to that of car occupants was higher than one. Specifically, the estimated mean of the severity score for the disability of motorcyclists was 2.8 times higher than that computed for car victims (RR_{moto}: 2.803). Similar results were found for pedestrian victims compared with car victims (RR_{ped}: 3.022). These results are consistent with findings in the existing literature (e.g., Eluru *et al.*, 2008; Chang and Wang, 2006; Majdzadeh *et al.*, 2008). In addition, we found a quadratic linear relationship between the victim's age and the logarithmic of the severity score for disability. The RR of the age factor was higher than one for victims below the age of 44 years. Above this age, the relationship was inverted to a relative risk lower than one. As appointed out by Kima *et al.* (2008), this could be due to the fact that older drivers tend to be more cautious. Another interesting result is that BI claims that remain unsettled for long periods of time are associated with an expected higher severity score for permanent disability (RR_{year}: 1.128).

We should emphasize that the length of time that the victim was temporarily disabled has explanatory capacity over the disability. In particular, the variables that record the number of recovery days in hospital and recovery days out of hospital show risk factors of 1.021 and 1.003, respectively. An interesting result is that the relative risk is lower than one for those victims involved in accidents where the attribution of the fault is unclear. As mentioned above, the Spanish compensation system is based on the liability for the accident. Therefore, the RR_{fault} lower than one means that, on average, the Courts trend to award lower severity scores for disability when the fault is unclear. Finally, we would like to stress that the greater the time period that is allowed to elapse between the occurrence of the crash and the opening of the claim, the greater the overdispersion in the severity score of disability is expected for the BI claim observation (RR_{open}: 1.642).

Table 4. Summary of relative risks and odds ratios from ZIGP regression model (95% CI)

Mean	**Relative risk* (95% CI)**
year	1.128 $(1.026, 1.241)$
fault	0.673 $(0.421, 1.075)$
moto	2.803 $(1.976, 3.978)$
ped	3.022 $(1.920, 4.755)$
$age + age^2$	1.027 $(0.953, 1.107)$
hrd	1.021 $(1.008, 1.034)$
drd	1.003 $(1.002, 1.004)$
Over-dispersion	**Relative Risk$^\pm$ (95% CI)**
open	1.642 $(1.188, 2.269)$
Zero-inflation	**Zero-inflated odds ratio (95% CI)**
gender	6.282 $(1.632, 24.167)$

*For binary regressors, the relative risk (RR) of the k-th factor is computed as $\mathrm{RR}_k = \frac{\hat{\mu}|x_k=1}{\hat{\mu}|x_k=0}$. When the regressor is continuous, $\mathrm{RR}_k = \frac{\hat{\mu}|x_k=\bar{x}+1}{\hat{\mu}|x_k=\bar{x}}$.
$^\pm$ The overdispersion RR is computed as $\mathrm{RR}_k = \frac{\hat{\varphi}-1|x_k=\bar{x}+1}{\hat{\varphi}-1|x_k=\bar{x}}$.

5 Applications

Various applications can be derived from individual estimates of the permanent bodily injury severity recorded in motor accidents claims. First, we show that the severity score estimate may be used to detect victims for whom special follow-up examinations are required. However, the main applications are likely associated with claim cost estimations. In particular, we show that the severity score estimate may be used for assessing compensation for non-pecuniary damages resulting from disability and that this claim cost estimation has implications in various fields of the insurance industry as regards claim settlement or reserve estimation.

5.1 *Claim handling practices*

While only a few motor accident claims involve severely injured victims, they represent the largest cost for insurers. As a result, follow-up examinations are carried out by insurers in the case of bodily injury claims with high severity scores for disability. In such instances, victims registering a high severity score estimate are likely to have suffered injuries that make it impossible for them to work[3]. When victims are unable to work, we can expect large sums of compensation to be awarded as pecuniary damages. Consequently, claims involving an estimated high severity score need to be monitored by the insurers' medical experts on a case-by-case basis so as to determine whether the victim is also incapacitated for work.

A further statistic that needs to be carefully observed by an insurer's staff is the severity score variance, since a particularly high value can negatively affect the representability of the point estimation. For instance, the variance depends positively on the over-dispersion function φ_i which is explained by the time that has elapsed between the occurrence of the accident and the opening of the claim. Therefore, the insurer is able to monitor the effect on the expected variance of a reduction in this time.

5.2 *Individual claim cost*

The bodily injury compensation stipulated by the legal system comprises compensatory awards for non-pecuniary and pecuniary damages. The total sum awarded by the Court to to victim i in the settlement year j is computed as follows:

[3] Obviously, there is no specific severity score which might be considered the threshold for incapacity for work, since this depends on other factors such as the type of injury or the victim's profession. However, it is unlikely that victims with fewer than 20 points on the permanent disability scale will suffer any type of incapacity for work.

$$\pi_{ij} = (y_i * Cpp_{ij} + Nrdh_i * Crdh_j + Nrdi_i * Crdi_j + Nrdwi_i * Crdwi_j) * (1 + cf_{ij}) + CCpI_{ij}$$

where π_{ij} is the financial compensation awarded for bodily injuries. The total compensation for non-pecuniary damages is computed as the sum of the compensation for disability and the compensation for temporary disability. The former is calculated as the total score for disability y_i multiplied by the financial compensation per point Cpp_{ij}, where the value of Cpp_{ij} depends on the settlement year j, the victim's age and the total score recorded for disability y_i. The latter is obtained by multiplying the number of days that the victim spent recovering from his injuries by a financial compensation per day of recovery. The legal system distinguishes between the number of days spent recovering in hospital $Nrdh_i$, the number of days recovering out of hospital while unable to work $Nrdi_i$ and the number of days recovering out of hospital while being able to work[4] $Nrdwi_i$ and establishes a distinct financial award per day for each, $Crdh_j$, $Crdi_j$ and $Crdwi_j$, respectively. The amount stipulated per day of recovery depends only on the year of settlement.

Once the compensation for non-pecuniary damages has been computed, the compensation for the pecuniary damages resulting from the injury is calculated by multiplying the assessed compensation for non-pecuniary damages by a correction factor[5] cf_{ij}. This correction rate is fixed as a percentage that ranges between zero and 75% as a function of the annual incomes of the victim during the year of settlement. Finally, a complementary compensation for pecuniary damages $CCpI_{ij}$ is awarded in those cases in which the victim has suffered permanent incapacity for work as a result of the accident. The $CCpI_{ij}$ damages are determined on the basis of the degree of incapacity and the year of settlement.

[4] Since 1999 the Spanish legal compensation system has awarded compensation for the days that the victim is still recovering from the accident yet is able to work. This variable was removed from the ZIGP regression model as it did not present a significant coefficient

[5] Since 2000 victims have been entitled to claim the real pecuniary damage sustained during the recovery period whether the fault of the accident is exclusively attributable to the other party (Constitutional Court ruling 181/2000, June 20th). Given the difficulty in attributing the entire fault in a motor accident, most compensation awards for pecuniary damages during the recovery period are, in practice, still computed by applying the correction factors as shown in the standard formula.

5.2.1 *Claim settlement*

Insurers will always seek to reach an amicable agreement with the plaintiff in order to settle BI claims as quickly as possible, since keeping a claim open incurs a financial cost (i.e., interest rate payments, judicial expenses and so forth). In practice, only 1% of motor accident claims are settled in the Courts with most claims being settled before the trial takes place (Lewis, 2006; Derrig and Rempala, 2006). Thus, in this context, the individual claim estimate of the severity score for permanent disability is of great value.

The claim agreement is pursued when the victim is fully recovered. In this way the insurer can establish the period of time it took for the victim's injuries to stabilize. It is not expected for there to be any disagreements as regards the annual income of the victim, his age and the year of settlement. By contrast, differences may appear in the evaluation of the disability and as regards the existence of, and the monetary quantification of, the victim's incapacity for work. As pointed out above, however, very few cases of incapacity for work arise. In addition, the severity score for disability may be used by the insurance company as an indicator of incapacity for work. Thus, the most controversial issue in a compensation agreement is typically determining the severity score y_i. The methodology (1) provides probability estimates of severity scores for disability. Therefore, the point estimate and the upper bound for the expected financial award for non-pecuniary damages that result from the disability can be derived from,

$$\begin{aligned} E[y_i * Cpp_{ij}] &= \sum_{h=1}^{100} \Pr[y_i = h] * y_i * Cpp_{ij|Y_i=h} \\ Var(y_i * Cpp_{ij}) &= E[(y_i * Cpp_{ij})^2] - (E[y_i * Cpp_{ij}])^2 \end{aligned} \tag{3}$$

where 100 is the highest disability severity score for disability, given that the probability of y_i being equal to 100 is negligible so we do not need to sum any more terms. We suggest that this estimate of the expected compensation to be awarded by the Courts is the amount that should be offered by the insurer in the negotiation process for non-pecuniary damages derived from the disability. In addition, the variance estimate of the compensation allows the insurer to estimate the upper-bound of the financial compensation for a given confidence level. The upper-bound estimate is the expected maximum deviation from the point estimate and as such it can be interpreted as the maximum amount of compensation that can be accepted by the insurer in order to avoid legal proceedings.

5.2.2 *Claims reserving*

The disability severity score estimate can also be applied to claim reserving. Traditionally, the actuarial literature has focused on aggregate reserving techniques in computing provisions for motor accident claims (for a review see England and Verrall, 2002; 2006). However, statistical methods based on individual claim information, since their introduction by Taylor and Campbell (2002), have grown in importance in recent years (see, among others, Norberg, 1999; Antonio *et al.*, 2006, Roholte Larsen, 2007, Ayuso and Santolino, 2007; 2008).

In this section we suggest that the claim compensation estimated according to equation (3) can be interpreted by the insurer as the individual claim provision for non-pecuniary damages derived from the disability. As appointed out above, the legal financial compensation per point of disability severity Cpp is positively correlated with the total score recorded for disability. That means, in order to obtain the expected compensation for disability, larger severity scores are multiplied by higher values of Cpp. As such, this would seem to constitute a prudential criterion for claim reserving. Alternatively, $E[Y_i] * Cpp_{ij|Y_i=E[Y_i]}$ could also be used to compute the individual provision[6] of the i-th claim, $i = 1, ..., n$.

In Table (5) the total compensation awarded by the insurer to the plaintiffs for non-pecuniary damages for disability is compared with the sum of financial compensations es-

[6] Let us suppose, for instance, that a twenty-year-old victim injured in an accident in 2008 has a score of 10 points on the disability severity scale with a probability of 0.5 or 20 points with the same probability. In this case, the compensation he can expect in the form of non-pecuniary damages derived from the disability is computed as $0.5 \times (10 \times 923.24) + 0.5 \times (20 \times 1233.67)$ which is equal to 16952.9 euros, where 923.24 and 1233.67 are the legally established rates of compensation per point when the total severity scores are 10 and 20 respectively. However, the expected disability severity score for this victim is 15 points which is associated with an established rate per point of 1085.05 euros (16275.75 euros in total, an amount that does not match the expected compensation).

timated by (3). Additionally, the resulting claims provision where $E[Y_i] * Cpp_{ij|Y_i=E[Y_i]}$ is applied to estimate the individual claim provision for the non-pecuaniary damages of the disability is provided. Note that both provision estimates work well. Nevertheless, the claims provision estimated by $\sum_{i=1}^{N} E[Y_i * Cpp_{ij}]$ seems the safest one of the two.

Table 5. Estimated claims provision for non-pecuniary damages resulting from the disability

	Total amount (in euros)	$\frac{\text{Estimated provision}}{\text{Empirical compensation}}$
Empirical compensations	631294.10	-
A) Estimated claims provision (computed as $\sum_{i=1}^{N} E[Y_i * Cpp_{ij}]$).	691922.52	109.60%
B) Estimated claims provision (computed as $\sum_{i=1}^{N} E[Y_i] * Cpp_{ij\vert Y_i=E[Y_i]}$).	609870.82	96.61%

6 Conclusions

The settlement of bodily injury claims represents the largest aggregate claim costs faced by motor insurers. Suitable techniques for estimating the claim compensations during the handling process are widely required by motor insurance companies since the expected size of claim costs will have implications for both the compensation offered in the claim settlement and the required level of reserves, among others.

In the Spanish market we show that this problem can be reduced to an estimate of the disability severity sustained by accident victims, since the remaining factors determining the amount of the claim are known at the time of settlement. The main advantage of modelling injury severity rather than directly modelling the financial compensation is that the former methodology does not depend on economic factors such as the settlement year, the inflation rate or the cost of medical services, among others. This means that any financial effects are withdrawn from the motor accident BI claim assessment allowing insurance companies to monitor the real severity level underlying the claim. As a result, the methodology proposed allows us, for instance, to compare the expected severity of BI claims regardless of the year in which the motor accidents occurred. In the same fashion, it is expected that this methodology remains valid when it is applied to estimate the injury severity of motor claims settled in a period other than the one under review.

References

[1] Abdel-Aty, M.A.; Abdelwahab, H. (2004) "Predicting Injury Severity Levels in Traffic Crashes: A Modeling Comparison" *Journal of Transportation Engineering*, 130(2), 204-210.

[2] Antonio, K., Beirlant, J., Hoedemakers, T., Verlaak, R. (2006) "Lognormal Mixed Models for Reported Claim Reserves" *North American Actuarial Journal*, 10, 30-48.

[3] Ayuso, M., Santolino, M. (2007) "Predicting Automobile Claims Bodily Injury Severity with Sequential Ordered Logit Models" *Insurance: Mathematics and Economics*, 4, 71-83.

[4] Ayuso, M., Santolino, M. (2008) "Prediction of individual automobile reported but not settled claim reserves for bodily injuries in the context of Solvency II" *Journal of Quantitative Methods for Economics and Business Administration*, 6,23-41.

[5] Bae, S., Famoye, F., Wulu, J.T., Bartolucci, A.A., Singh, K.P. (2005) “A Rich family of generalized Poisson regression models with applications” *Mathematics and Computers in Simulation*, 69, 4-11.

[6] Bell, P. (2006) "The Social Construction of Bodily Injury" *The Geneva Papers on Risk and Insurance-Issues and Practice*, 31, 340-356.

[7] Boucher, J.-P., Denuit, M. (2008) "Credibility premiums for the zero-inflated Poisson model and new hunger for bonus interpretation" *Insurance: Mathematics and Economics*, 42, 727-735.

[8] Boucher, J.-P., Denuit, M., Guillén, M. (2007) "Risk Classification for Claim Counts: A Comparative Analyisis of Various Zero-Inflated Mixed Poisson and Hurdle Models" *North American Actuarial Journal*, 11(4), 110-131.

[9] Browne, M.J., Puelz, R. (1996) "Statutory Rules, Attorney Involvement, and Automobile Liability Claims" *Journal of Risk and Insurance*, 63(1), 77-94.

[10] Browne, M.J., Puelz, R. (1999). "The effect of Legal Rules on the Value of Economic and Non-Economic Damages and the Decision to File" *Journal of Risk and Uncertainty*, 18(2), 189-213.

[11] Browne, M.J., Wells, B.P. (1999) "Claims Adjudication in the Personal Automobile Insurance Residual Market" *Journal of Risk and Insurance*, 66(2), 275-290.

[12] CEA (2007) *The European Motor Insurance Market*. CEA Statistics N^{o}32, CEA Insurers of Europe.

[13] Chang, L.Y., Mannering, F.L. (1999) "Analysis of injury severity and vehicle occupancy in truck and non-truck involved accidents" *Accident Analysis and Prevention*, 31 (5), 579–592.

[14] Chang, L.-Y., Wang, H.-W. (2006) "Analysis of traffic injury severity: An application of non-parametric classification tree techniques" *Accident Analysis and Prevention*, 38, 1019–1027.

[15] Cheung, Y.B. (2002) "Zero-inflated models for regression analysis of count data: a study of growth and development" *Statistics in Medicine*, 21,1461–1469.

[16] Clarke, K. A. (2007) "A Simple Distribution-Free Test for Nonnested Model Selection" *Political Analysis*, 15(3), 347-363.

[17] Crocker, K.J., Tennyson, S. (2002) "Insurance Fraud and Optimal Claims Settlement Strategies" *The Journal of Law and Economics*, 45, 469-507.

[18] Czado, C., Min, A. (2005) "Consistency and asymptotic normality of the maximum likelihood estimator in a zero-inflated generalized Poisson regression" *Sonderforschungsbereich 386*, Paper 423. (http://epub.ub.uni-muenchen.de/)

[19] Czado, C., Erhardt, V., Min, A., Wagner, S. (2007) "Zero-inflated generalized Poisson models with regression effects on the mean, dispersion and zero-inflation level applied to patent outsourcing rates" *Statistical Modelling*, 7(2), 125-153.

[20] Derrig, R.A., Rempala, G.A. (2006) "A statistical analysis of the effect of anchoring in the negotiation process of automobile bodily injury liability insurance claims" In: CAS predictive modelling seminar, 4 October 2006, Boston.

[21] Doerpinghaus, H.I., Schmit, J.T., Yeh, J.J.-H. (2008) "Age and gender effects on auto liability insurance payouts", *Journal of Risk and Insurance*, 75(3), 527-550.

[22] EC (2003) *Draft report with recommendations to the Commission on a European disability rating scale*. Committee on Legal Affairs and the Internal Market 2003/2130(INI).

[23] Eluru, N., Bhat, C.R., Hensher, D.A. (2008) "A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes" *Accident Analysis and Prevention*, 40, 1033–1054.

[24] England, P., Verrall, R. (2002) "Stochastic Claims Reserving in General Insurance" *British Actuarial Journal*, 8, 443-544.

[25] England, P., Verrall, R. (2002) "Predictive distributions of outstanding liabilities in general insurance" *Annals of Actuarial Science*, 1(2), 221-270.

[26] Erhardt, V. (2008) *The ZIGP Package*, R manual.

[27] Famoye, F., Singh, K. (2003) “On inflated generalized Poisson regression models” *Advances and Applications in Statistics*, 145-158.

[28] Famoye, F., Singh, K. (2006) “Zero-inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data” *Journal of Data Science*, 4, 117-130.

[29] Hoyt, R., Mustard, D., Powell, L. (2006). “The Effectiveness of State Legislation in Mitigating Moral Hazard: Evidence from Automobile Insurance” *Journal of Law and Economics*, 49(2), 427-450.

[30] Joe, H., Zhu, R. (2005) “Generalized Poisson Distribution: the Property of Mixture of Poisson and Comparison with Negative Binomial Distribution” *Biometrical Journal*, 47 (2), 219–229.

[31] Kima, J.-K., Ulfarsson, G.F., Shankar, V.N., Kim, S. (2008) "Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis" *Accident Analysis and Prevention*, 40, 1695–1702.

[32] Kockelman, K., Kweon, Y. (2002) “Driver Injury severity: an application of ordered probit models” *Accident Analysis and Prevention*, 34(3), 313-321.

[33] Lee, C., Abdel-Aty, M. (2005) “Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida” *Accident Analysis and Prevention*, 37(4), 775-86.

[34] Lee, J., Mannering, F. (2002) “Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis” *Accident Analysis and Prevention*, 34, 149–161.

[35] Lewis, R. (2006) "How important are insurers in compensating claims for personal injury in the U.K.?" *The Geneva Papers on Risk and Insurance-Issues and Practice*, 31(2), 323-339.

[36] Li, Z., Knight, S., Cook, L.J., Hyde, L.K., Holubkov, R., Olson, L.M.(2008) “Modeling motor vehicle crashes for street racers using zero-inflated models” *Accident Analysis and Prevention*, 40, 835-839.

[37] Lord, D., Washington, S., Ivan, J.N. (2005) "Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory" *Accident Analysis and Prevention*, 37, 35–46.

[38] Lord, D., Washington, S., Ivan, J.N. (2007) "Further notes on the application of zero-inflated models in highway safety" *Accident Analysis and Prevention*, 39, 53–57.

[39] Majdzadeh, R., Khalagi, K., Naraghi, K., Motevalian, A., Eshraghian, M.R. (2008) "Determinants of traffic injuries in drivers and motorcyclists involved in an accident" *Accident Analysis and Prevention*, 40, 17–23.

[40] Norberg, R. (1999) "Prediction of Outstanding Liabilities II: Model Variations and Extensions" *ASTIN Bulletin*, 29, 5-25.

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

[42] Simon, S.D. (2001) "Understanding the Odds Ratio and the Relative Risk" *Journal of Andrology*, 22 (4), 533-536.

[43] Roholte Larsen, C. (2007) "An Individual Claims Reserving Model" *ASTIN Bulletin*, 37, 113-132.

[44] Taylor, G., Campbell, M., (2002) "Statistical case estimation" *Research paper number 104*, Centre for Actuarial Studies, The University of Melbourne.

[45] Vuong, Q. (1989) "Likelihood ratio tests for model selection and non-nested hypotheses" *Econometrica*, 57, 307-334.

[46] Yip, K., Yau., K. (2005) "On Modeling Claim Frequency Data in General Insurance with Extra Zeros" *Insurance: Mathematics and Economics*, 36, 153–63.

[47] Wang, X., Kockelman, K. (2005) "Occupant Injury Severity using a Heteroscedastic Ordered Logit Model: Distinguishing the Effects of Vehicle Weight and Type" *Transportation Research Record: Journal of Transportation Research*, 1908, 195-204.

[48] Weisberg, H.I., Derrig, R.A. (1998) "Quantitative methods for detecting fraudulent automobile bodily injury claims" *Risques,* 35, (July–September), 75–99.