Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 3 (2014) 214 – 223

Transportation Research Procedia

www.elsevier.com/locate/procedia

17th Meeting of the EURO Working Group on Transportation, EWGT2014, 2-4 July 2014, Sevilla, Spain

Prediction of road accident severity using the ordered probit model

Rui Garrido[a,*], Ana Bastos[a], Ana de Almeida[b], José Paulo Elvas[c]

[a]*Department of Civil Engineering, University of Coimbra, Coimbra, Portugal*
[b]*Department of Information Science and Technology, ISCTE - IUL University Institute of Lisbon, Lisbon, Portugal*
[c]*Department of Mathematics - Lab. of Geomatic Engineering, Univeristy of Coimbra, Coimbra, Portugal*

Abstract

The ordered probit model is used to examine the contribution of several factors to the injury severity faced by motor-vehicle occupants involved in road accidents. The estimated results suggest that motor-vehicle occupants travelling in light-vehicles, at two-way roads, and on dry road surfaces tend to suffer more severe injuries than those who travel in heavy-vehicles, at one-way roads, and on wet road surfaces. Additionally, the driver's seat is clearly the safest seating position, urban areas seem to originate less serious accidents than rural areas, and women tend to be more likely to suffer serious or fatal injuries than men.

Selection and peer-review under responsibility of the Scientific Committee of EWGT2014

Keywords: Road safety; road accident modelling; injury severity; motor-vehicle occupants; ordered probit model.

1. Introduction

In 2012, almost 28,000 persons died on European Union (EU) roads due to road traffic accidents. Compared to 2011, this represents a decrease of about 9% (EC, 2013). The EU strategic target for the period 2011-2020 is to halve the overall number of road fatalities by 2020 (EC, 2011). From now on, an average annual reduction of about 7% will be required in order to achieve the EU target.

Portugal has been one of the major contributors for the positive results of 2012, by reducing its number of road traffic fatalities by more than 15%. In spite of this remarkable progress, the country still is well above the EU average number of road fatalities. In 2012, within the 27 Member States, there were on average 55 road fatalities per

* Corresponding author.
E-mail address: ruigarrido@hotmail.com

Selection and peer-review under responsibility of the Scientific Committee of EWGT2014
doi:10.1016/j.trpro.2014.10.107

million inhabitants; in Portugal this number grew up to 71, which placed the country on the bottom rung of the ladder (20th position) (EC, 2013).

Fostered by the unsatisfactory position that the country held within the EU, together with the social and economic consequences of road fatalities, the reduction of the number of road traffic fatalities was declared a national challenge. Hence, under the framework of the National Road Safety Strategy (ANSR, 2009), a relevant goal was established: "to place Portugal among the 10 EU countries with a low number of road fatalities". This goal should be accomplished through the reduction of the number of road traffic fatalities "so as to reach 62 deaths per million inhabitants by 2015", the equivalent to a 12.7% decrease compared to 2012.

Data from 2011 highlighted that about 36% of the EU road fatalities took place in urban areas. In Portugal this share was even higher: according to official statistics, the number of road fatalities occurring at urban agglomerations was around 49% in 2011, rising to 50% in 2012 (ANSR, 2012 and 2013).

This study intends to contribute for a better understanding of the factors that affect the occurrence of accidents and, in particular, those that affect its severity. Firstly, a summarized state-of-the-art is presented aiming at identifying the more appropriate methodologies and the most significant explanatory affecting the road accident severity. Then, and resorting to ordered response models, the ordered probit model is applied to a real database, which comprises 4528 records related to road accidents that took place in the municipality of Coimbra, Portugal, between 2007 and 2011.

2. Literature review

An effective road safety management requires a good insight into the factors that are believed to be related to road traffic accidents. Based on this framework, several research studies have been conducted over the years aiming at identifying factors that may influence both the frequency and the severity of road traffic accidents. However, as pointed out by Savolainen et al. (2011), one has to be aware that the factors influencing accident frequency may vary from the ones affecting the severity; hence, it is suggested that their analysis should be performed individually.

In the area of accident severity research, continuous efforts have been conducted in order to investigate the relationship between the level of severity (dependent variable) and a set of explanatory variables, which usually include: driver attributes (e.g., age and gender), vehicle features (e.g., body type, vehicle age and number of vehicles involved in the accident), road characteristics (e.g., number of lanes, road surface conditions, intersection control and types of road), and accident characteristics (e.g., accident's main cause). Occasionally, the influence of other variables on accident severity like speed limit, day of the week, time of the day, average traffic characteristics (AADT), weather and traffic conditions have also been scrutinized (Delen et al., 2006; Manner and Ziegler, 2013; Torrão, 2013).

From previous studies, it is worth mentioning a recent study carried out by Christoforou et al. (2010), mainly because it offers a comprehensive literature review on the subject. By crossing the results from 28 studies, the authors found that the factors which have consistently shown to be connected to an increased severity were: (1) aging; (2) driving while intoxicated; (3) head-on collisions; (4) crashes with heavy-vehicles and motorcycles; (5) poor lighting conditions; (6) vertical and horizontal curvature; (7) rural *versus* urban areas; and (8) speeding. In contrast, the use of helmet or seat belt was associated to a high decrease of the severity outcome. Other factors like gender, intersection type, road surface conditions, seating position, weather and average traffic characteristics led to conflicting findings – in some studies these factors were related to increased severity, while in other studies the opposite was concluded.

In summary, the accident severity investigations attempted to examine the influence of driver attributes, vehicle features, accident characteristics, and road, weather, and traffic conditions on the severity outcome. A wide variety of methodological approaches have been used to fulfil this propose. Recently, Savolainen et al. (2011) conducted a research, which intended to assess the characteristics of road accidents severity data and the methodological approaches most commonly used for the analysis of such data. The authors highlighted that the "appropriate methodological approach can often depend heavily on the available dataset, including the number of observations, quantity and quality of explanatory variables, and other data-specific characteristics". Still, they found that the majority of the modelling approaches were framed in the discrete response models which include: binary response models (e.g., binary probit and binary logit), ordered discrete response models (e.g. bivariate ordered probit and

generalized ordered logit), unordered multinomial discrete response models (e.g. multinomial logit, Markov switching multinomial logit, nested logit, and mixed logit); other less common approaches have also been used and include: artificial neural networks, and some data mining techniques such as the classification and regression tree (CART) analysis. A complete review of this literature is provided pointing out the strengths and the weaknesses of each one of the approaches typically employed for examining accident severity data.

Despite the methodological evolution registered, which improved the statistical validity of findings, it is worth noting that statistical regression models such as the logistic regression model still are broadly employed in the identification of contributing factors to accident severity (e.g., Bédard et al., 2002; Konnen et al., 2011). For example, Bédard et al. (2002) applied a multivariate logistic regression to assess the independent contribution of driver, accident, and vehicle characteristics to fatal injuries sustained by drivers. They found that factors such as increased driver age, female gender, blood alcohol concentration greater than 0.30, non-belted drivers, driver side impacts, travelling speeds greater than 70 mph, and older vehicles were associated to higher odds of fatal outcomes; on the opposite, a 25 cm increase in vehicle size (wheelbase) was translated into a reasonable reduction of the fatality risk.

CART is a non-parametric model that has been pointed out by some authors (e.g., Das et al., 2009; Kuhnert et al., 2000) as a valid alternative to the frequently used logistic regression models in the analysis of road injury severity. Das et al. (2009) applied Random Forests, which are ensembles of individual trees grown by CART algorithm, to identify traffic and roadway design, as well as driver and vehicle information significantly related to fatal and severe injuries on urban arterials for different accident types. This methodology identified alcohol/drug use, posted speed limits, failure to use safety equipment by all drivers, presence of a driver/passenger in the vulnerable group (more than 55 years or less than 3 years), slow moving vehicles (cycles/moped), and lower ADT as contributing factors leading to an increase of severe outcomes.

Gray et al. (2008) used ordered probit models to identify specific accident characteristics that increase the likelihood of accident severity involving young male car drivers in Great Britain. Slight, serious, or fatal injuries were the possible severity outcomes considered. The authors concluded that driving in the darkness, during early morning, between Friday and Sunday, on the main roads, during overtaking manoeuvres, and on the single carriageway of speed limit 60 mph were the most significant characteristics related to serious and fatal injuries. On the other hand, characteristics like driving in daylight, between Monday and Thursday, on roads with a speed limit of 30 mph or less, at a roundabout, waiting to move, and when an animal is on the carriageway were associated to light injuries.

The ordered probit modelling methodology was also used by Abdel-Aty (2003) in an attempt to analyse the driver injury severity levels at several roadway entities. Three separated models were developed for signalized intersections, roadway sections, and toll plazas. Older and male drivers, those not wearing a seat belt, drivers of passenger cars (i.e., vehicle type), vehicles struck at driver's side (i.e., point of impact/accident characteristics), and those who speed revealed a higher probability of a severe injury in all the models. Other variables were found significant only in specific models: alcohol, dark lighting conditions, and the existence of horizontal curvatures affected the likelihood of injuries in the roadway sections' model; a driver's error was significant in the model for signalized intersections; vehicles equipped with an electronic toll collection affected the likelihood of higher injury severity in the toll plazas' model; lastly, both signalized intersections and roadway sections models revealed higher probability of injuries in rural areas. Furthermore, the author tested other modelling approaches, namely a multinomial logit and a nested logit with different nesting structures, and compared the results produced by those models with the results provided by the ordered probit model. This comparison showed that the ordered probit model, besides being simpler, has also produced better results than the multinomial logit; by comparing the ordered probit model with the best nesting structure of the nested logit model, the author recommended the former to model driver injury severity because, in spite of the similarity found in the results provided by both models, the latter has introduced considerable complexity in the modelling process.

Using a combination of different methodologies is a common practice in accident severity studies (Torrão, 2013). Not only to compare the performance of different models, as in the case of Abdel-Aty (2003), but also to extract partner relationships between variables and to overcome data complexity, as presented, for example, by Kuhnert et al. (2000) and O'Donnell and Connor (1996).

O'Donnell and Connor (1996) assessed the probability of four levels of injury severity sustained by motor-

vehicles occupants as a function of occupant, vehicle and accident attributes. For that purpose they used two ordered response models, specifically the ordered probit model and the ordered logit model. The results showed that the probability of serious injuries and deaths widens with increases in speed, age of the vehicle, age of the occupant, and time of the accident. They also found that the left-rear seating position, female gender, alcohol levels greater than 0.08, non-use of seatbelt, driving light-duty trucks (i.e., vehicle type) and head-on collision (i.e., type of collision/accident) were factors related to increased severity outcomes.

Many of the results described in the previous studies were replicated in Christoforou et al. (2010), thus confirming the consistency of a number of factors contributing to the injury severity outcomes. Nevertheless, the factors that affect the frequency and the severity of the accidents depend on drivers' behaviour, as well as the vehicle fleet and the legal frame that differ from country to country. Therefore, the development of further investigation on this subject is still pertinent, namely if addressed to the specifications of each country.

3. Methodological approach

Accounting for the ordinal nature of injury data is an important consideration in road accident severity modelling (Savolainen et al., 2011). Unordered response models, such as the multinomial logit and probit models, or the nested logit model, regardless of accounting for the categorical nature of the injury data, would fail to account for the ordinal nature of the injury classes (Green, 2002). Furthermore, the multinomial logit model is associated to undesirable properties, such as the independence of irrelevant alternatives (Ben Akiva and Lerman, 1985) and the multinomial probit is related to a lack of a closed-form likelihood (Greene, 2002). Alternatively, ordered response models, namely the ordered probit (OP) or logit (OL) models have increasingly been employed for modelling injury severity when it is recorded in multiple ordinal categories (Eluru et al., 2008). Both formulations give very similar results, although the OP is more often selected than the OL. In this study, the modelling methodology used to analyse the injury severity sustained by non-motorists is the OP model. The OP model is especially appropriate to model injury severity because, besides identifying statistically significant relationships between explanatory variables and a dependent variable, it also discerns unequal differences between ordinal classes in the dependent variable (Duncan et al., 1998). Additionally, it also requires smaller samples when compared to unordered response models (Ye and Lord, 2014).

The model, originally proposed by McKelvey and Zavoina (1975) as an alternative to the ordinary linear regression, is usually built around the notion of a latent underlying injury risk propensity occurring from a road accident that determines the observed ordinal injury severity level. The general specification is (Washington et al., 2003):

$$z_i = \beta X_i + \varepsilon_i \tag{1}$$

where z_i denotes the latent injury risk propensity for accident victim i, β is the vector of parameters to be estimated, X_i is the vector of observed non-random explanatory variables measuring the attributes of accident victim i, and ε_i is the random error term following standard normal distribution. Accordingly, the mean and the variance of ε_i are normalized to zero and one, respectively. Since the dependent variable, z_i, is unobserved, standard regression techniques cannot be applied to compute Eq. (1). Yet, as suggested by O'Donnell and Connor (1996), one can reasonably assume that a high risk of injury, denoted by z_i, is related to a high level of observed injury, denoted by y_i. This relationship can be translated as follows (Ye and Lord, 2014):

$$y_i = \begin{cases} 1, & \text{if } z_i \le \mu_1 \\ k, & \text{if } \mu_{k-1} < z_i \le \mu_k \\ K, & \text{if } z_i > \mu_{K-1} \end{cases} \tag{2}$$

where $\mu = \{\mu_1, \ldots, \mu_k, \ldots, \mu_{K-1}\}$ are the threshold values for all injury severity levels that define y_i, corresponding to integer ordering, and K is the highest ordered injury severity level. In turn, the probability that accident victim i

faces an injury severity level k is equal to the probability that the latent injury risk propensity, z_i, assumes a value between two fixed thresholds. In other words, given the value of X_i, the probability that the injury severity faced by accident victim i belongs to each injury severity level is:

$$\begin{cases} P(y=1) = \Phi(-\beta X_i) \\ P(y=k) = \Phi(\mu_{k-1} - \beta X_i) - \Phi(\mu_{k-2} - \beta X_i) \\ P(y=K) = 1 - \Phi(\mu_{K-1} - \beta X_i) \end{cases} \quad (3)$$

where Φ is the cumulative normal distribution function. For estimation, Eq. (3) can be written as (Washington et al., 2003):

$$P(y=k) = \Phi(\mu_k - \beta X_i) - \Phi(\mu_{k+1} - \beta X_i) \quad (4)$$

where μ_k and μ_{k+1} denote the lower and upper thresholds for the injury severity level k, respectively. For all the probabilities to be positive, the thresholds values must satisfy the restriction $\mu_1 < \ldots < \mu_k < \ldots < \mu_{K-1}$. Computation of these probabilities allows the understanding of the effect of individual estimated parameters. Indeed, a positive value of β_i implies that an increase in X_i will clearly generate the increase (respectively, decrease) of the probabilities of the highest (respectively, lowest) ordered injury severity levels. However, it is not obvious what effect a positive or negative β_i will generate on the probabilities of the intermediate levels. For this reason, the computation of marginal effects for each level is suggested (Washington et al., 2003). These marginal effects provide the direction of the probability for each level as follows:

$$P(y=k)/\partial X = \left[\Phi(\mu_k - \beta X_i) - \Phi(\mu_{K-1} - \beta X_i)\right]\beta \quad (5)$$

In the specific case of a categorical variable, the computation of Eq. (5) is not appropriate since it only makes sense if the variable is continuous. Instead, the probabilities that result when the variable changes from 0 to 1, holding all other variables values at their means, should be compared. That is, for a categorical variable X_i, the corresponding marginal effect is (Green, 2007):

$$X_i = P(y=k \mid X_i = 1) - P(y=k \mid X_i = 0) \quad (6)$$

The thresholds μ are unknown parameters to be estimated jointly with the model parameters β. Here, both are estimated through the maximum likelihood method (see, for example, Green (2002) for detailed information on the method), using the LIMDEP software (Green, 2007).

4. Dataset characterization

In Portugal, the National Road Safety Authority (ANSR – Portuguese abbreviation) is responsible for managing the database of the national road accidents. Whenever a road accident results in at least one injury, police officers are in charge of filling an accident bulletin, so called *Boletim Estatístico de Acidentes de Viação* (BEAV), which is then sent to the ANSR to be validated. The BEAV is organized into six information categories, namely identification of the accident, external circumstances, nature of the accident, vehicles involved, drivers involved, and severity outcomes. The bulletins are well structured, which gives the possibility of extracting information that has proved to be very useful on road accident severity modelling.

For this study, the ANSR supplied information about all road accidents occurred in the municipality of Coimbra from 2007 to 2011. During this period, a total of 3158 road accidents were recorded, involving 5021 drivers, 1183 passengers, and 548 pedestrians. Since the main purpose of this study relies on the analysis of the injury severity sustained by motor-vehicle occupants, the records about pedestrians, cycles, moped, and motorcycle occupants were ignored. In the end, a total sample of 4528 observations was considered.

The observed injury severity sustained by motor-vehicle occupants is distributed as follows: no injury (41.63%); slight injury (56.85%); serious injury (1.10%); and fatal injury (0.42%). It is worth noting that removing the most vulnerable road users from the original dataset has decreased the shares of the two most severe injury outcomes. For that reason, the serious and the fatal injuries were aggregated into a single injury severity level. Hence, the dependent variable considered in this study is the *level of injury sustained by motor-vehicle occupants*, divided into three categorical levels: no injury (y=1); slight injury (y=2); and serious or fatal injury (y=3).

Along with the level of injury sustained by motor-vehicle occupants (dependent variable), each observation comprises a number of attributes (explanatory variables) related to the accident victim, vehicle, and roadway and environmental conditions. Table 1 presents the definition of each explanatory variable together with its mean (M) and standard deviation (SD) values. All of these variables, with the exception of the variable *age of the motor-vehicle occupant*, are binary/dummy variables with means between 0 and 1. Since OP models may not converge if the variables have not similar scales (Green, 2007), the variable *age of the motor-vehicle occupan*t (continuous variable) has been scaled (dividing by 100) to have mean with the same scale as those of the binary/dummy variables.

Table 1 – Explanatory variables

Explanatory variables	**Type**	**Coding**	**Mean**	**SD**
Age of motor-vehicle occupant	Continuous	Continuous variable	0.395	0.170
Gender of motor-vehicle occupant	Binary	1 if male, 0 if female	0.575	0.494
Vehicle type	Binary	1if light-vehicle, 0 if heavy-vehicle	0.972	0.166
Worn restraint system	Binary	1 if worn, 0 if not worn	0.970	0.170
Seating position				
Driver *vs.* Passenger	Binary	1 if driver, 0 if passenger	0.785	0.411
Front *vs.* Rear	Binary	1 if front, 0 if rear	0.922	0.267
Day type	Binary	1 if weekday, 0 if weekend	0.753	0.431
Road environment	Binary	1 if urban, 0 if rural	0.765	0.424
Lighting conditions	Binary	1 if daylight, 0 if darkness	0.761	0.427
Road surface conditions	Binary	1 if dry, 0 if wet	0.647	0.478
Traffic type	Binary	1 if two-way, 0 if one-way	0.616	0.487
Roadway alignment	Binary	1 if segment, 0 if curve	0.634	0.482
Accident type				
Head-on-collision	Dummy	1 if head-on collision	0.127	0.333
Rear-end collision	Dummy	1 if rear-end collision	0.248	0.432
Side collision	Dummy	1 if side collision	0.244	0.429
Rollover	Dummy	1 if rollover	0.053	0.225
Run-of-road	Dummy	1 if run of road	0.143	0.350
Fixed-object collision	Dummy	1 if fixed-object collision	0.086	0.280
Multivehicle collision	Dummy	1 if multivehicle collision	0.072	0.259
Other type of collision	Dummy	1 if other type of collision	0.027	0.163

5. Model estimation results

The modelling procedure started with the assessment of the presence of multicollinearity in the dataset. Indeed, the correlation matrix (which is not presented here due to space limitations) showed that the two variables associated to the seating position of motor-vehicle occupants (i.e., *driver vs. passenger* and *front vs. rear*) were strongly correlated to each other (presenting a r ca. 0.6). In fact, since both the driver and most of the individual passengers travel in front seats, this is not an independent relationship. In order to warrant meaningful model estimations, not only in terms of the statistically significance of the estimated coefficients, but also to ensure reasonable magnitudes

and signs for those coefficients, it is advisable that one of those variables should be excluded from the model. The selection of the variable to be removed from the model was based on the condition that the model fit will not vary significantly. Thus, the variable *front vs. rear* was chosen to be removed. The same procedure was followed for the pair *head-on collision* - *side collision* ($r \sim 0.5$). In this case, the variable *side collision* was the one excluded from the model. At this point, it is worth mentioning that the prior analysis of the correlation matrix contributed to the final coding of some explanatory variables. For example, as a first attempt, the use of a restraint system has been divided into two variables: (1) *worn seat belt* and (2) *worn child restraint system*. However, these variables were found to be strongly correlated to each other ($r \sim 0.7$); instead, both variables were aggregated into a single variable (*worn restraint system*). On the other hand, accident types were disaggregated into more categories than those initially thought aiming at minimizing the degree of correlation amongst some pairs.

After the multicollinearity problem solved, all the variables figuring on Table 1, excluding the variables *front vs. rear* and *side collision*, were included in the model, and the backward selection process was used in such manner that only the variables significant at the 0.05 significance level were retained. The resulting model is shown in Table 2. The variables such as *age* and *worn restraint system*, which were consistently pointed out in several earlier studies as factors that were related to increased and decreased injury severity outcomes, respectively, were not found to be statistically significant in this study. Also, the variables *roadway alignment* (i.e., straight or curve) and *day type* (i.e., weekday or weekend) were not significant.

Along with the maximum likelihood estimates of the model parameters β and the thresholds values μ (together with their statistical significance), two goodness-of-fit measures are provided – the adjusted log likelihood ratio index (adjusted ρ^2) and the classification accuracy (CA). The former establishes how well the model fits the data and the latter examines the prediction accuracy of the model. For a detailed description of these goodness-of-fit measures see Ben-Akiva and Lerman (1985). The calculated adjusted ρ^2 is 0.224, which is a notable value when compared to the values reported in similar studies (e.g., Kockelman and Kweon (2002) reported an adjusted ρ^2 value of 0.045; Abdel-Aty (2003) reported adjusted ρ^2 values ranging from 0.093, to 0.15, and to 0.20). In terms of prediction accuracy, the model predicts correctly 74% of the overall injury severity outcomes, which is similar to the CA value reported by Saleh and Pai (2008). It is worth mentioning that when dealing with unbalanced datasets, it is expected that the less represented outcome tends to be predicted very poorly.

Table 2 - Ordered probit model estimation results

			Marginal effects		
Variable	**Coefficient**	**t-Ratio**	No injury	Slight injury	Serious/fatal injury
Constant	1.241	8.270			
Gender of motor-vehicle occupant	-0.476	-11.291	0.1754	-0.1714	-0.0039
Vehicle type	0.416	3.278	-0.1629	0.1611	0.0017
Seating position					
Driver *vs.* Passenger	-1.449	-22.755	0.4187	-0.3753	-0.0434
Road environment	-0.289	-5.788	0.1053	-0.1027	-0.0026
Lighting conditions	-0.182	-3.881	0.0673	-0.0658	-0.0015
Road surface conditions	0.117	2.747	-0.0443	0.0435	0.0008
Traffic type	0.121	2.788	-0.0458	0.0450	0.0008
Accident type					
Head-on collision	0.520	8.186	-0.1785	0.1718	0.0067
Rollover	1.221	12.467	-0.3239	0.2798	0.0441
Run-off road	0.949	15.103	-0.2944	0.2746	0.0198
Fixed-object collision	0.634	8.667	-0.2085	0.1985	0.0100
Multivehicle	-0.206	-2.544	0.0793	-0.0782	-0.0011
Other type of collisions	0.388	3.230	-0.1341	0.1294	0.0047
Threshold parameters					

μ_2	3.183	45.298	
Log-likelihood with constant only	-3394.46	Overall CA	73.8%
Log-likelihood at convergence	-2621.328	CA for no injury	73.7%
Degrees of freedom	13	CA for slight injury	75.8%
Adjusted ρ^2	0.224	CA for serious/fatal injury	0%

In the following paragraphs, the magnitudes and in particular the signs of the estimated parameters are discussed. It is worth remembering that a positive (negative) value of a parameter β, combined with an increase in the corresponding variable X, will increase (decrease) the probability of the highest ordered injury severity level (i.e., serious or fatal injury) and decrease (increase) the probability of the lowest ordered injury severity level (i.e., slight injury). In other words: the estimated parameters that are greater (lesser) than zero imply that increases in the corresponding variables tend to augment (diminish) the injury risk propensity, z_i, as translated by Eq. (1). Given that, the explanatory variables whose parameters have a positive sign are: *vehicle type*, *road surface conditions*, *traffic type*, *head-on collision*, *rollover*, *run-off road*, *fixed-object collision*, and *other type of collisions*. It means that motor-vehicle occupants who travel in a light-vehicle, or on dry road surfaces, or at two-way roads are more likely to be involved in a serious or fatal road accident. In addition, the average injury risk is increasingly aggravated by rollover-type accidents, followed by run-off-road accidents, collisions against fixed-objects, and head-on collisions. In fact, the largest positive parameter listed in Table 2 is the one associated to the variable rollover. This means that a rollover-type accident is the most critical factor, increasing significantly the average risk of injury faced by a motor-vehicle occupant.

By comparing the estimated coefficients of all the variables mentioned above, one can rank the influence of each variable on the average injury risk, concluding that the rollover-type of accident apparently has the greatest impact on injury risk (β=1.221) while the road surface condition appears to have the smallest impact (β=0.117). In terms of magnitude, for example, the injury risk faced by a motor-vehicle occupant involved in a rollover-type accident is about 2.4 times higher than the injury risk faced by a motor-vehicle occupant involved in a head-on collision (β=0.520), all remaining variables being equal.

Moreover, the set of variables associated to negative parameters include: *gender*, *seating position*, *road environment*, *lighting conditions*, and *multivehicle*. It means that the average injury risk decreases when a motor-vehicle occupant is travelling on an urban environment, or at daylight, or is involved in a multivehicle accident. These results are in agreement with the known Portuguese tendencies. In reality, though the majority of the Portuguese road accidents occur within urban areas (about 70%) the seriousness of rural accidents is higher. The estimated parameters also show that the probability of suffering no injury increases whether the accident victim is a male or a driver. Actually, the variable *seating position* is associated to the largest negative parameter presented in Table 2, which means that the average risk of injury will decrease significantly if a motor-vehicle occupant is seated at the driver's position. In fact, this might mean that, in an imminent risk situation, the driver anticipates the occurrence of the accident. Consequently, the driver adopts earlier defensive measures, whilst the passenger is more easily surprised and produces later and less effective defensive reactions.

As one should expect, the computation of the marginal effects for each variable, also presented in Table 2, confirms the previous findings on the direction of the probabilities of the lower and upper injury severity levels. In addition, based on those indicators, some considerations concerning the effect of the explanatory variables on the intermediate injury severity level (i.e., slight injury) can be provided. For example, it can be found that men or drivers are less likely to be involved in a slight injury accident than women and passengers (in both cases the corresponding marginal effect is lesser than zero). On the opposite, the probability of a motor-vehicle occupant being involved in a slight injury accident will increase weather travelling at a two-ways road, or on a dry road surface when compared to one-way roads or wet road surfaces, respectively (in these cases, the marginal effects are greater than zero). The influence of the remaining variables on the slight injury outcome can be interpreted in this way.

6. Conclusion

In this study, the ordered probit model was used to examine the influence of a number of factors on the injury severity faced by motor-vehicle occupants involved in road accidents. The model estimation results suggest that some types of road accidents, namely the rollover-type, run-off-road, collisions against fixed objects and head-on collisions, appear to be the major contributors for the most severe injury level. Also, those who travel in a light-vehicle, at a two-way road and on dry road surface tend to suffer more severe injuries than those who travel in a heavy-vehicle, at a one-way road, and on a wet road surface. In contrast, the driver´s seat is clearly the safest seating position, and urban areas, although presenting the highest accident occurrence frequency, are linked to decreased severity level. Also, women tend to be more likely to suffer serious or fatal injuries than men.

All the findings presented above are consistent with the ones from several former studies and, at the same time, seem to be intuitively reasonable. However, some of those findings merit further comments. Indeed, it is interesting to find that urban areas are associated to a decreased injury severity, not only because it is a trend systematically pointed out by other authors, but also because it is actually an observed trend in Portugal. There is no evidence to support the generally accepted tendencies that increased occupant's age increases the level of injury severity and, on the contrary, the use of restraint systems significantly decreases the level of injuries sustained by road users. Both variables were not statistically significant. The estimated results also show that the probability of suffering no injury increases when the accident victim is a male. At a first glance, this result might be controversial, since it is a fact that males drive more aggressively and take more risks than females, increasing the likelihood of being seriously injured or killed. Yet, it is also true that male drivers seem to be more technical proficient at certain driving situations, which may possibly lead them to take more effective defensive reactions in an imminent road accident and, consequently, reduce the injury outcome. It is also noteworthy that, for example, O'Donnell and Connor (1996), who have also focussed their analysis on motor-vehicle occupants, reported a similar result, supporting it on the greater ability of males to tolerate certain types and levels of physical traumas. Lastly, the results indicate that motor-vehicle occupants who travel on dry road surfaces are more likely to be involved in a serious or fatal road accident, which is consistent with the findings of, for example, Christoforou et al. (2010). This is perhaps a consequence of a more reckless behaviour when driving on dry pavements compared to wet or icy pavements where the drivers tend to be more careful, as concluded by Christoforou et al. (2010). Furthermore, dry pavements can encourage speeding, which is a factor strongly correlated to greater frequencies and also increased severity outcomes.

Apparently the model fits the data reasonably well. Yet, the interpretation of the goodness-of-fit measures used in this study calls for some caution (i.e., adjusted likelihood ratio index and classification accuracy). Prudence should also be used in the interpretation of the marginal effects. On the one hand, there is no unquestionable goodness-of-fit measure for ordered response models, which compromises the judgment of the explanation/prediction capacity of the model. On the other hand, the marginal effects can be misleading, mainly when the explanatory variable is a categorical variable.

Further replications of the ordered probit model to larger and more comprehensive samples, including exposure variables, such as traffic flow and speed at the time of accident, could be challenging.

Acknowledgments

This work has been conducted within the framework of the project "TICE.Mobilidade – Mobility System User Centered" – financed by the fund FEDER (*Programa Operacional Factores de Competitividade*).

References

Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research*, *34*, 597 - 603.

ANSR. (2009). *National Road Safety Strategy 2008-2015*.

ANSR. (2012). *Ano de 2011 – Sinistralidade rodoviária*. Observatório de Segurança Rodoviária.

ANSR. (2013). *Ano de 2012 – Sinistralidade rodoviária*. Observatório de Segurança Rodoviária

Bédard, M., Guyatt, G. H., Stones, M. J., & Hirdes, J. P. (2002). The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. *Accident Analysis and Prevention*, *34*, 717 - 727.

Ben-Akiva, M., & Lerman, S. (1985). *Discrete choice analysis: theory and application to travel demand*. MIT Press, Cambridge, MA.

Christoforou, Z., Cohen, S., & Karlaftis, M.G. (2010). Vehicle occupant injury severity on highways: an empirical investigation. *Accident Analysis and Prevention*, *42*, 1606 - 1620.

Das, A., Pande, A., Abdel-Aty, M., & Santos, J. (2008). Urban arterial crash characteristics related with proximity to intersections and injury severity. *Transportation Research Record*, *2083*, 137 - 144.

Delen, D., Sharda, R., & Bessonov, M. (2006). Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. *Accident Analysis and Prevention*, *38 (3)*, 434 - 444.

Duncan, C. S., Khattak, A. J., & Council, F.M. (1998). Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions. *Transportation Research Record*, *1635*, 63 - 71.

Eluru, N., Bhat, C.R., & Hensher, D.A. (2008). A mixed generalized response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis and Prevention*, *40 (3)*, 1033 - 1054.

European Commission (EC). (2011). *Roadmap to a Single Transport Area – Towards a competitive and resource efficient transport system*. White Paper. COM 144 final.

European Commission (EC). (2013). *Road safety trends, statistics and challenges in the EU 2011-2012*. Road Safety Vademecum.

Gray, R.C., Quddus, M.A., & Evans, A. (2008). Injury severity analysis of accidents involving young male drivers in Great Britain. *Journal of Safety Research*, *39*, 483 – 495

Greene, W.H. (2007). LIMDEP User's Manual: Version 9.0. Econometric software, Plainview, NY.

Greene,W. H. (2002). *Econometric Analysis* (5th ed). New Jersey: Prentice Hall.

Kockelman, K. M., & Kweon, Y.-J. (2002). Driver injury severity: an application of ordered probit models. *Accident Analysis and Prevention*, *34*, 313 - 321.

Kononen, D.W., Flannagan, C. A. C., & Wang, S. C. (2011). Identification and validation of a logistic regression model for predicting serious injuries associated with motorvehicle crashes. *Accident Analysis and Prevention*, *43 (1)*, 112 - 122.

Kuhnert, P. M., Do, K.-A., & McClure, R. (2000) Combining non-parametric models with logistic regression: an application to motor vehicle injury data. *Computational Statistics and Data Analysis*, *34(3)*. 371 – 386.

Manner, H., & Wünsch-Ziegler, L. (2013). Analyzing the severity of accidents on the German Autobahn. *Accident Analysis and Prevention*, *57*, 40 - 48.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, *4*, 103 - 120.

O'Donnell, C. J., & Connor, D. H. (1996). Predicting the severity using models of ordered multiple choice. *Accident Analysis and Prevention*, *28 (6)*, 739 - 753.

Pai, C.-W., & Saleh, W. (2008). Modelling motorcyclist injury severity by various crash types at T-junctions in the UK. *Safety Science*, *46*, 1234 - 1247.

Savolainen, P., Mannering, F., Lord, D., & Quddus, M. (2011). The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. *Accident Analysis and Prevention*, *43 (5)*, 1666 - 1676.

Torrão, G. (2013). Effect of vehicle characteristics on safety, fuel use and emissions. University of Aveiro, Portugal (PhD dissertation).

Ye, F., & Lord, D. (2014). Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. *Analytical Methods in Accident Research*, *1*, 72 - 85.

Washington, S.P., Karlafatis, M.G, & Mannering, F.L. (2003). *Statistical and Econometric Methods for Transportation Data Analysis.* requirements: Multinomial logit, ordered probit and mixed logit models. *Analytical Methods in Accident Research*, *1*, 72 - 85.