Accident Analysis and Prevention 113 (2018) 244–256

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Applying crash data to injury claims - an investigation of determinant factors in severe motor vehicle accidents

Darren Shannon[a,*], Finbarr Murphy[a], Martin Mullins[a], Julian Eggert[b]

[a] University of Limerick, Ireland
[b] Honda Research Institute Europe GmbH, 63073 Offenbach/Main, Germany

ARTICLE INFO

Keywords:
RTAs
Linear regression
ADAS
Expected compensation costs
Claims reserving

ABSTRACT

An extensive number of research studies have attempted to capture the factors that influence the severity of vehicle impacts. The high number of risks facing all traffic participants has led to a gradual increase in sophisticated data collection schemes linking crash characteristics to subsequent severity measures. This study serves as a departure from previous research by relating injuries suffered in road traffic accidents to expected trauma compensation payouts and deriving a quantitative cost function. Data from the National Highway Traffic Safety Administration's (NHTSA) Crash Injury Research (CIREN) database for the years 2005–2014 is combined with the Book of Quantum, an Irish governmental document that offers guidelines on the appropriate compensation to be awarded for injuries sustained in accidents. A multiple linear regression is carried out to identify the crash factors that significantly influence expected compensation costs and compared to ordered and multinomial logit models. The model offers encouraging results given the inherent variation expected in vehicular incidents and the subjectivity influencing compensation payout judgments, attaining an adjusted-R^2 fit of 20.6% when uninfluential factors are removed. It is found that relative speed at time of impact and dark conditions increase the expected costs, while rear-end incidents, incident sustained in van-based trucks and incidents sustained while turning result in lower expected compensations. The number of airbags available in the vehicle is also a significant factor. The scalar-outcome approach used in this research offers an alternative methodology to the discrete-outcome models that dominate traffic safety analyses. The results also raise queries on the future development of claims reserving (capital allocations earmarked for future expected claims payments) as advanced driver assistant systems (ADASs) seek to eradicate the most frequent types of crash factors upon which insurance mathematics base their assumptions.

1. Introduction

Road Traffic Accidents (RTAs) result in an estimated 1.25 million deaths worldwide each year, and are the leading cause of death among young adults aged 15–29 (World Health Organisation 2013). The economic cost of RTAs and their impact on human society in general have led to concerted efforts from governments and corporations alike to enhance traffic safety, largely by way of a continuous stream of safety campaigns and regulatory mandates governing the safety features of vehicles. Road collision data collected by the automotive industry and government agencies have allowed researchers the opportunity to pinpoint the specific characteristics and contextual features that determine the severity of an incident. Statistical analyses using this data have proved critical in understanding the risks facing road users, as large volumes of literature have been dedicated to determining both the likelihood of entering into an incident, and the expected severity of the ensuing impact. Although the statistical techniques that are used vary widely, traditional literature has generally focused on predicting the damage sustained in an impact by means of an ordinal or probabilistic scale. A recent emphasis on improving the quality of data collection has led to a surge of databases with a wealth of information that offer more precise insights into the factors involved in RTAs. This study makes use of such data by generating a risk measure that associates the specific crash characteristics with the expected compensatory costs from the sustained injuries, while also exploring possible practical applications of this risk measure. As a result, the methodology presented in this article serves as a departure from categorical injury severity factors by utilising a proxy variable in order to quantify the expected severity of an incident.

The motivation for this research is prompted by recent developments in prospective ADAS technology, namely predictive systems, and offers further clarity on the influence of unmeasurable contextual

* Corresponding author.
E-mail address: Darren.Shannon@ul.ie (D. Shannon).

https://doi.org/10.1016/j.aap.2018.01.037
Received 28 July 2017; Received in revised form 26 January 2018; Accepted 28 January 2018

factors in determining crash severity. Predictive systems operate by attaching multiple cameras and sensors to a vehicle. The systems then provide warnings of oncoming potential hazards to the driver, who in turn will take corrective actions. In some cases, these corrective actions may be taken autonomously, as is the case with an Automatic Emergency Braking (AEB) system. Forecasting accurate trajectories are essential for an effective predictive system, and the implementation of forecasting frameworks have attracted significant attention in traffic safety literature. For example, the 'Intelligent Driver Model' is a traffic flow simulation technique used by on-board systems to alert drivers to oncoming perils (An and Harris 1996). A recent extension to this framework, the 'Foresighted Driver Model', has sought to improve its practical application by assuming that the user will attempt to balance risk-averseness with expedient driving. These on-board predictive models have made use of constant streams of continuous data to provide accurate results and have introduced cost functions. These cost functions attempt to capture the risks faced by the driver of the vehicle which arise in form of potential abnormal events (Eggert et al. 2015, Klingelschmitt and Eggert et al., 2015, Damerow et al. 2016). This study introduces an alternative view of these cost functions by directly linking the risks faced by the vehicle with a compensation cost outcome, which could potentially be adapted to an on-board system as their development continues.

Hindering the widespread introduction of on-board predictive systems are the prevalence of unobserved factors that cannot currently be controlled for, also referred to as unobserved heterogeneity. Examples of this phenomenon facing road users are roadway design and detrition, weather conditions and lighting – all of which have been found to affect the severity of injuries sustained in motor vehicle accidents (Eluru et al. 2008; Buddhavarapu et al. 2013). Given the unpredictable variations that play a role in each incident, efforts to accurately predict the expected injury severity outcome have been hampered by the vast number of contextual factors that affect driving behaviour, road characteristics and environmental conditions. These studies attempted to categorise the extent of the incident along a five or six-level injury severity scale, resulting in a slight loss of granularity. This study introduces a continuous scalar measure of quantifying the injuries sustained in the accident by way of a 'compensation cost' loss function, found by relating injuries suffered in a road traffic accident to the expected trauma compensation. As a result, by offering a quantitative insight in to the outcome of RTAs, 'compensation cost' loss functions can potentially capture the large variations in near-identical incidents.

A potential solution for underlying heterogeneity can be found in the anticipated introduction of commercial ADAS-enabled and semi-autonomous vehicles, which is expected to be a significant step toward mitigating and controlling the risks brought about by human driving behaviour. It is surmised that a gradual shift toward full autonomy will substantially reduce the hazards introduced by human error (Kyriakidis et al. 2015). A recent designation stated that over 95% of vehicles manufactured for U.S. consumption are to have Autonomous Emergency Braking (AEB) available as standard by 2022, reducing the number of traffic incidents by an estimated 20% or 5 million crashes (US DOT/NHTSA 2016). Meanwhile, the European Commission has encouraged a sustained adoption of emerging Advanced Driver Assistance Systems (ADAS) such as the aforementioned AEB, Collision Avoidance Systems, and adverse-environment Control Systems. These systems are designed to mitigate potential risks that occur throughout journeys and take immediate evasive action if a perilous situation arises. The impact that such evolutionary changes will have on the non-life insurance industry as vehicle safety comes increasingly to the fore is addressed in this study. Although current models function well in quantifying the risk faced by a given road user, the factors that contribute to an accurate derivation of risk pricing may wholly alter as ADAS function become increasingly prevalent. As such, models typically used in underwriting mathematics may need to adapt over time. As well as offering an incisive view of the outcome of RTAs, the framework introduced as part of this research can adapt to shifting RTA statistics as regulatory changes are introduced.

Mannering and Bhat (2014) perform a comprehensive literature review of the current problem space and detail the evolution of the methodological approaches used in accident research. They also pay close attention to the issues that underlie the estimation of accident prediction. Of greatest concern is unobserved heterogeneity, i.e. the lack of unmeasurable variables that significantly affect the frequency and severity of RTAs. Mannering et al. (2016) also target this issue and offer statistical solutions that seek to minimise the impact of unobserved heterogeneity. One commonly used approach is the introduction of random parameter models, which allow for variations in underlying sensitivities (Eluru et al. 2008; Kim et al., 2008; Ye and Lord 2011; Cerwick et al., 2014). Another emerging approach is the use of latent class, or finite mixture, models. These models are used to capture injury severity levels by identifying underlying subgroups that affect known variables of interest and subsequently incorporating the probabilities of their occurrence into the model. Examples of studies adopting this approach are Cerwick et al. (2014), Yasmin et al. (2014a), and Yasmin et al. (2014b).

A further matter addressed by Mannering and Bhat (2014) is that of selective bias, which could allow erroneous conclusions to be drawn. For example, when analysing safety measures that are put in place solely at accident hotspots, the inferred results may underestimate the initiative's impact on the wider domain. Furthermore, the opposite effect could occur – the risk compensation that drivers take under adverse conditions makes accurate estimations of risk difficult (Winston et al. 2006). Finally, as is the case with many traffic safety analyses featuring crash data, the under-reporting of less severe accidents is raised (Yamamoto et al. 2008; Ye and Lord 2011). Given that the injuries reported on in this paper concern RTAs in which a severe injury is sustained, this is a factor that also affects this research.

In terms of empirical studies, Mannering and Bhat (2014) and Mannering et al. (2016) highlight the vast body of literature surrounding vehicle crash analysis, and detail a large number of methodological approaches used when modelling crash-related data. Beginning from univariate models finding associations between crash factors and specific injury severities, the complexity of statistical analyses quickly expanded in an attempt to capture the large variations that are inherent in RTAs. Shankar and Mannering (1996) and O'Donnell and Connor (1996), amongst others, developed discrete multivariate model approaches. Research typically focused on ordered probit models, as exemplified by Kockelman and Kweon (2002) and Eluru et al. (2013). O'donnell and Connor et al. (1996), Kim et al. (2008) and Lemp et al. (2011) acknowledge the inherent randomness of RTAs by investigating links between attributes of road users and the levels of injury endured in incidents using heteroskedastic ordered logit and probit models. More generally, the research conducted by O'donnell and Connor et al. (1996) highlight through the use of ordered probit and logit models that, amongst other factors, speed, age of vehicle, and seatbelt use are significant indicators when predicting crash severity. They achieve reasonably high pseudo–R^2 values (36% and 35.6% respectively) in doing so.

Abdel-Aty (2003) and Ye and Lord (2011) examine a number of models in order to identify worthwhile predictors of injury severity. While Ye and Lord (2011) make use of ordered probit, multinomial logit, and mixed logit models, Abdel-Aty (2003) uses ordered probit, multinomial logit and nested logit models to find significant relationships between injury severity and age, gender, seatbelt use, point of impact, vehicle type and speeding. These results were found by