Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Investigation of the correlation between stated and revealed driving behaviour using data collected from on-board diagnostics (OBD) devices

Dimitrios I. Tselentis[a]*, Katerina Folla[a], Natalia Vittoratou[a], George Yannis[a], John C. Golias[a]

[a]*Department of Transportation Planning and Engineering, National Technical University of Athens (NTUA), 5 Heroon Polytechniou Str., Athens GR-15773,Greece*

Abstract

The objective of this research is the examination of the correlation between stated and revealed driving behaviour using vehicle on-board diagnostics (OBD) data. To this end, a large data set derived from a driving behaviour experiment is exploited, during which, the driving behaviour of 17 drivers is monitored for a period of three months. These data concerned the number of harsh acceleration and braking events occurred, the average driving speed and the mileage travelled and were recorded at a 1Hz frequency. Drivers' stated behaviour is investigated using the results of a survey administered to all users participated in the experiment. A linear regression model is developed to perform data analysis and study the factors influencing a number of different driving behaviour indicators that were recorded in real time. The results demonstrated a strong correlation between the number of harsh braking and acceleration events on one hand and the number of accidents, the annual income and the declared frequency of harsh braking on the other. This research contributes towards the improvement of driving behaviour by performing behavioral analysis of driving characteristics and resulting to models that could potentially serve as a platform that provides feedback to drivers on how to improve their behaviour and become less risky.

Keywords: stated - revealed behaviour; on-board diagnostics; harsh braking; harsh acceleration; linear regression; number of accidents

* Corresponding author. Tel.: +30-210-772-2210; fax: +30-210-772-1454.
E-mail address: dtsel@central.ntua.gr

1. Introduction

1.1. Background

The scientific analysis of accidents started approximately 100 years ago. Since then, there have been studies that investigate the cause of accidents. According to the World Health Organization (WHO, 2015), every year approximately 1.25 million people around the world die on a road accident and from 20 to 50 million people suffer from non-fatal injuries, many of which can lead to disability. Groups of people that are exposed to the dangers of a road accident vary depending on the social and financial status, age and gender.

In 2014, the annual number of deaths in road accidents was up to around 1.2 million people worldwide, to 26,000 in the European Union and to 800 in Greece, which reinforce the fact that road accidents are one of the most important problems in modern society. In the last decade, the number of road accident in Greece has been greatly reduced and more specifically within the financial crisis period (2008-2014). Greece reported a substantial reduction of 49% in the number of deaths in road accidents, whereas road accidents decreased by 21%. However, this trend seems to be disrupted in 2015, as for the first time since 2004 a rise of road accidents deaths is observed.

The factors influencing road accidents (Dingus et al., 2016) are the observed disorder (1.92 %), the errors related to the driver's performance (4.81%), the instantaneous mistake in driver's judgment (aggressive/speedy driving) (4.22%) and driving distraction (51.93%). These percentages illustrate the percentage of time that each factor is present during driving under normal circumstances.

1.2. Stated and revealed driver's behaviour

As mentioned earlier, human behaviour constitutes a determining factor for road accidents. People have the innate tendency to overestimate their abilities and this can among others affect their driving behaviour. The perception of danger is essential considering that young drivers, compared to other age groups, are likely to underestimate the existence of dangers in a road system and overestimate their driving ability (Brown et al., 1975). The perception of traffic behaviour that a driver has relies not only on his age but also on the gender, annual salary, personality and driving experience (Torfs et al., 2016).

It is therefore vital to analyze and continuously monitor the stated behavior's deviation from the revealed behaviour. The objective of this research is to correlate stated and revealed driver's behaviour with the use of vehicle on-board diagnostics (OBD) data as well as with the answers given in the survey administered, collected from a driving behaviour experiment organized by the Department of Transportation Planning and Engineering of the National Technical University of Athens.

This paper examines the degree in which all the different elements that compose driving behaviour (harsh acceleration/ braking events occurred, number of trips per day, total number of accidents, annual salary etc.) interact with each other and define traffic behaviour. In addition to this, the degree in which the drivers perceive their driving in a correct manner is examined by comparing drivers' answers to the questionnaires with the outputs of the vehicle on-board diagnostics (OBD). It is estimated that the results arising assist in the further understanding of the level at which the vehicle on-board diagnostics and the questionnaires express the stated and revealed driver's behaviour.

2. Literature Review

2.1. Driving behaviour monitoring

This sub-chapter introduces the key features of studies that are related to driving behaviour, recorded by in-vehicle installed devices with the aim to compare the results of the present research with that of other studies with relative content. Nadeem et al. (2004) aimed to exchange information between vehicles regarding driving speed and coordinates. As a vehicle enters the transmission range of another vehicle (connected vehicles), the system allows the driver to be informed about the real-time traffic volume, weather conditions and alternative route options for the specific journey, which are all proved to be useful. The methodology pursued is referenced as a propagation mechanism where each vehicle transmits information about itself and/or other vehicles, while when information is received, its memory is upgraded and promoted to the next period of transmittals. The recordings of the vehicle contain a unique identification number, the exact location, speed and time of the transmittal. The main objective of this study was the development of a suitable algorithm that processes and analyses all data resulting from the experiment. The results demonstrated that the algorithm ratio-based is the best choice concerning the monitoring and accuracy of the parameters that are used. Although the rest of the methods are more convenient in adjusting the parameters, the operating cost makes them unbeneficial compared to the previously mentioned.

On the other hand, the study of Huang et al. (2005) had the objective of investigating the opinion of truck drivers on the use of recording systems of driving behaviour that are installed inside the vehicle and provide information regarding road safety parameters. An additional objective was to identify which is the best method for a driver to receive such information. The study was focused on drivers that perform long distance journeys, as they are driving for more time than the rest of the truck drivers and hence they will be the ones that will benefit the most from this advanced technology. Overall, 198 drivers were interviewed and 66 people were involved in the collection of quality data related to their view on technology and information feedback. The results indicated that drivers of long distance journeys show willingness to receive mostly positive information on their driving behaviour. Nonetheless, they would prefer this information to be provided from supervisors or managers rather than technological devices. As far as the method or the frequency of feedback is concerned, they did not show to have a strong preference towards a specific one, while they did point out the significance of adapting a technological program to the specific demands of each driver.

2.2. Stated and Revealed Preference methodologies

Literature review revealed two methods of recording people’s preferences. The first method is the stated preference methodological approach that attempts the recording of preferences of people being interviewed against a new hypothetical situation. The second method is the revealed preference methodology, where driving behaviour and people’s choices are being recorded in alternative scenarios and therefore it is based on measurements and observation as it refers to real life situations.

The mathematical models of transport demand developed, are based on data collected either by direct measurements and observations, or from studies that record people’s opinions, by which preferences of passengers are detected. Therefore, the utility function of every alternative transport mode can be calculated as well as the probability of using each of them. In these cases, the revealed preference method is considered the most appropriate for the development of mathematical models that estimate transport demand (Kroes & Sheldon, 1988). Nonetheless, it entails several limitations that restrict its broad and general usage. More specifically, this method lacks flexibility and as a result not all variables that are of interest for a study can be analysed. Moreover, a strong correlation between explanatory variables that are of interest (time, journey cost) can appear, which can complicate the calculation of the coefficients of the mathematical model. On the other hand, this method cannot be used for the evaluation of hypothetical scenarios. Finally, it implies that explanatory variables can be expressed as absolute values and hence their usage can be limited to the exploitation of original interest variables (i.e. time, journey cost) and not for the evaluation of the effect of changes in the secondary variables (i.e. station facilities, positioning in a transport mode).

Therefore, the stated preference method shows a flexibility in terms of modifying questions towards the field of interest of the researcher and its ability of analysing large set of variables, as well as the low cost and short time of its actualization. Consequently, this method is considered more appropriate to be applied in traffic studies compared to the method of revealed preference.

However, the relevant studies demonstrated that the disadvantage of the stated preference method is that there are many times when the respondents’ answers- in the hypothetical scenario that is administered to them- differ from their actions. This enhances the importance of verification of the results of one study with another one depending on their objective. According to bibliography (Lin et al., 1986), attention should be given to the interpretation of the results of these types of studies as it has been proven that residents of western society tend to overreact, when being involved in an experiment. However, this is mitigated in the occasion when the objective of the study is the estimation of the relative importance of specific coefficients, and not the calculation of the absolute values.

On the other hand, the combination of stated and revealed preference methods is recommended for the estimation of absolute values in order to eliminate the main disadvantages of each method individually and produce consistent results. The method of stated preference is broadly used for the evaluation of non-economic goods. It is a method widely implemented nowadays in the field of economics as well as in the sectors of health and traffic. This method was developed as a tool for scientific studies of marketing at the beginning of the decade of 1970 (Kroes & Sheldon, 1988; Persson, 1992) and since 1978 (Green & Srinirasan, 1978), it has expanded its application.

2.3. Summary

The most beneficial studies related to the evaluation of driver's behaviour with the use of data monitoring systems located in the vehicle are presented in this chapter along with the contribution of the usage of new technological devices in road safety. The analysis of the studies mentioned above lead to the conclusions listed below:

- The efficient function of data recording systems requires connection with a system platform, to directly save and process data
- These data are related to the driving behaviour (speed, acceleration, braking, distance etc.) and the mechanical features of the vehicle (fuel consumption, tire pressure etc.)
- When a driver (experienced or inexperienced) is notified with an audible alert installed inside the vehicle in the case of a traffic offense, he/she tends to improve driving behaviour immediately
- These systems not only improve road safety and driver's behaviour, but also contribute in a timely prevention and handling of road accidents
- Data analysis is optimum when stated and revealed preference methods are combined, due to the cross-elimination of the main disadvantages of each one and the production of consistent results

3. Theoretical background

This chapter discusses the theoretical background of linear regression and the selection criteria for the optimum mathematical model among those resulting from the analysis.

3.1. Linear Regression

The statistical method, which examines the relationship between two or more variables in order to predict the dependent one from the independent(s), is called regression analysis. Dependent variable is the variable whose value is being predicted, while the independent variable refers to the one used to predict the dependent variable. The development of a mathematical model is a statistical process that forms equations that describe the relationship between the independent variables and the dependent one. It is worth mentioning that the choice of the methodological approach for the development of the most appropriate mathematical model depends on whether the dependent variable has continuous or discrete values.

In the occasion where the dependent variable is continuous and follows normal distribution, regression analysis is used, the simplest form of which is simple linear regression. Simple linear regression makes use of only one independent variable X and one dependent variable Y that is a linear function of X. The value y_i of variable Y, for each value of xi of variable X, is given by the following equation:

$$y_i = a + \beta x_i + \varepsilon_i \tag{1}$$

The scope of the regression analysis is to identify the parameters α and β that denote the linear dependence of dependent variable Y on the independent variable X. Each pair of values (α, β) describes a different linear relationship that is expressed geometrically with a straight line and the two parameters are defined as below:

- The constant a is the value of y for x = 0
- The coefficient β of x is the slope of the straight line or the regression coefficient. It illustrates the change of variable Y when variable X changed by one unit.

- The random variable ε_i is called regression error and is defined as the difference between y_i and the conditional expected value $E(Y| X = x_i)$ where $E(Y| X = x_i) = \alpha + \beta x_i$.
- The following assumptions are made in the regression analysis:
- The values of variable X are known without a doubt and therefore it is controlled for the study undertaken.
- The dependence of Y from X is linear.
- The regression error has an average value of zero for each value of X and the its deviation is constant and does not depend on X, which means $E(\varepsilon_i)=0$ and $Var(\varepsilon_i) = \sigma\varepsilon^2$.

The assumptions made above for linear relationship and constant deviation are statistical characteristics of populations that follow a normal distribution. This justifies the assumption made that the conditional distribution of Y is normal.

The case where the dependent variable Y depends linearly from more than one independent variables X (X_1, $X_2,\ldots,X_v$) is called multiple linear regression. The equation that represents the relationship between the dependent and the independent variables has the general form of

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_v x_v + \varepsilon_i \quad (2)$$

The assumptions of multiple linear regression are the same with those of simple linear regression, hence the assumption that the regression errors εi (as well as the random variable Y for each value of X) follow a normal distribution with a constant deviation. This is true if the zero correlation of the independent variables is certified, which means that:

$$\rho(X_i, X_j)\ldots i \neq j \to 0 \quad (3)$$

As mentioned previously, each model that is developed has to follow some basic rules. The most important factor is normality, which means that the values of the variable should follow normal distribution.

3.2. Coefficient of determination R^2

The check of statistical significance is followed by the examination of the quality of the model. The quality of the model is defined by the coefficient of determination. This coefficient shows the percentage of variability of the variable Y explained by the variable X. It takes values in the [0,1] range. As the value of R^2 approaches to 1, a higher percentage of Y's variability is explained by X and the linear relationship between variables Y and X becomes more powerful. The coefficient R^2 a has relative importance, which implies that there is no specific value of R^2 which is accepted or rejected but when comparing two or more models the model chosen as most appropriate is the one with the largest value of the coefficient R^2.

In this study, the coefficient adjusted R^2 is used as the selection criterion of the model. The adjusted R^2 is always smaller than R^2 because it isolates all random factors and focuses only on the variability factors of the independent variable.

Finally, the regression error of the model should fulfil the following three conditions:

- Follow normal distribution
- Have a constant deviation, $Var(\varepsilon_i) = \sigma\varepsilon^2 = c$
- Have a zero correlation, $\rho(\varepsilon_i, \varepsilon_j) = 0 \ \forall \ i \neq j$

It is mentioned that the deviation of the error depends on the coefficient of determination R2. The larger the R2 coefficient is, the smaller the deviation f of the error, in other words the better the prediction based on the linear regression model that arises.

4. Data collection and process

4.1. OBD device

All data that is essential for the investigation of the stated and revealed driver's behaviour is collected from a device located inside the vehicle and is connected with a central electronic control unit. The On Board Diagnostics (OBD) is a machine with dimensions of 27 x 48 x 49.5 cm that directly sends driving information to a central database (usually cloud) through the mobile network (with or without a mobile phone) for further process and analysis.

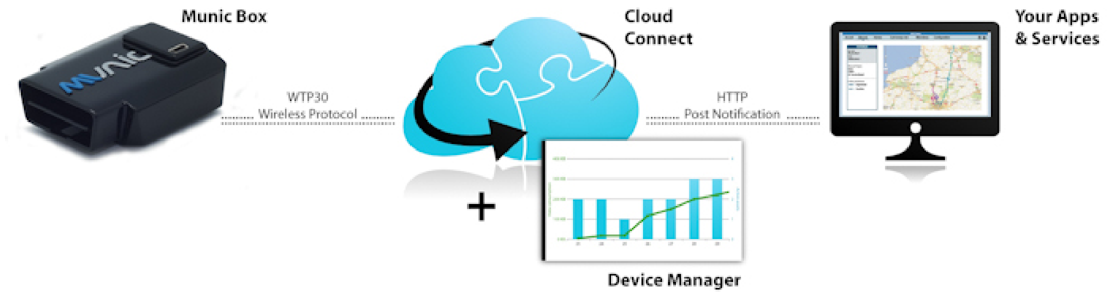

Fig. 1. Illustration of system operation [Source: munic.box].

The connection between the installed device inside the vehicle and Cloud Connect (data storage service on the Internet where access is gained via creation of account) is achieved via an advanced protocol, which allows constant and in real time data transfer, compress of all files and upgraded management of driver tracking data. Modern vehicles include a complex control system of the vehicle (OBD- On Board Diagnostics), which is able to offer whichever information related to the vehicle is needed, like seatbelts use, the emissions of pollutants the tire pressure etc. For the investigation of revealed driver's behaviour, probably the most important information collected from the device is the recording of an event. An event is the instant when the driver proceeds to a sudden change of speed (acceleration or braking) or to another sudden turn (lateral acceleration). Every time there was an event during driving, the device recorded it as well as the value of acceleration (or braking) at that point.

4.2. Questionnaire

The questionnaire designed was administered to the 17 drivers participated in the experiment and data were collected. Questionnaires included valuable information for the driving sample such as age, gender, annual income, historical accidents and traffic violations, driving periods, vehicle data, etc. Drivers answered regarding their driving perception, which was an important element to investigate driver's stated behaviour.

4.3. Summary Statistics

In order to investigate the stated and revealed behaviour of the driver, the most remarkable observations are highlighted below. These notes serve as a preliminary analysis that helps to better understand the results and will be used to draw qualitative conclusions.

- The 9 out of 17 drivers participated in the experiment are freelancers while only one is a student.
- Almost 50% of experimental participants have a high annual income.
- The reported frequency of harsh acceleration events generally corresponds to the corresponding detected frequency. There are also more harsh acceleration events per 100 km of high-income drivers, while two of them have given a conservative answer.

- Likewise, regarding the number of harsh braking events, it is worth noting that high-income drivers declare a higher number than they perform, moderate-income drivers reflect their true behaviour, and the driver with the highest number of harsh braking events has a low annual income.
- The distribution of the total number of accidents as a function of annual income is incidental.
- Seven out of the 17 drivers who declared the largest number of traffic rules violations reported they were cautious drivers in contrast to the others who reported fewer traffic rules violations
- Drivers with the smallest number of harsh acceleration events per 100 km display the lowest number of traffic violations, which is rational.
- There does not seem to be a correlation between accident numbers with harsh acceleration events per 100 km.
- There is no correlation between alcohol driving frequency and the number of abrupt accelerations or braking

5. Methodological approach

This chapter includes a detailed description of the implementation of the methodology, as well as the presentation of the results of the work. Taking into account the related literature review, the theoretical background on the methods used to analyse data and the data collection and processing description, the appropriate methodology for the present work was selected.

5.1. Independent variables

Due to the high data volume, it was impossible to use every question and all metrics recorded by the specific device in the analysis and therefore, a selection was made as follows. In particular, the questions used concerned the sections below:

- General driver information,
- Driving experience - commuting,
- Accident history
- Demographics
- Driving performance,
- Vehicle data

Twenty-five out of a total of 77 questions, were empirically selected as more essential, which were deemed to have a higher impact on the driver's stated and revealed behaviour. The correlation of these variables are investigated afterwards having the limitation that there is no correlation (Correlation <0.4) among the independent variables selected for the analysis.

After the correlation coefficients of the variables are found, a selection of five basic variables is made that are not related to the purpose of being used as independent variables in the model to be analyzed. This option was made on the assumption that specific variables listed below are important and cannot be omitted as they reflect their influence on the stated and revealed behaviour of the driver.

The following key independent variables were selected:

- The total number of accidents to date that each driver has been involved in
- Annual family income
- Driver's perception on sharp braking events
- Number of traffic violations within the past two years
- Number of traffic violation fines received within the last two years

The correlation between the rest of the variables' set and these five basic variables is also investigated and those not associated are also considered in the independent variables set. These variables are:

- The number of trips per day

- The driving frequency under the influence of alcohol

It is worth noting that the age and gender variables are not taken into account for the data analysis conducted due to the sample characteristics. More specifically, most drivers fall in the 30-40 age category while a 24-year-old and a 51-year-old are also included. Two out of the 17 drivers are women.

5.2. Linear regression model

Apart from the dependent variables mentioned above, the following driver behaviour variables collected from the OBD device are used as dependent variables:

- Number of harsh braking events occurred per 100 km.
- Number of harsh acceleration events occurred per 100 km.
- Number of harsh braking events occurred with a driving speed of between 50 and 90 km/ h at the beginning of the event
- The number of harsh acceleration events occurred with an acceleration value between 0.22g and 0.26g m/s^2 at the peak of the event.

Among these variables, the number of harsh braking and acceleration events occurred per 100 km are continuous, and therefore receive any numerical value, while the number of harsh braking and acceleration events occurred at a speed between 50 and 90 km/h receive discrete values. This is the reason why the analysis was performed using linear regression modelling. Since the dependent variable is linearly dependent on more than one independent variable, the method used is multiple linear regression.

According to the theoretical background presented above, the following factors should be tested in each model:

- The constant term of the equation, which expresses all the parameters that have not been taken into account, should be as small as possible.
- The significance level (Sig-Significance) should be less than 5%.
- The Adjusted R square value should be as high as possible (ideally greater than 0.4). On the other hand, a very high value of R^2 indicates that similar elements are correlated so that no interesting results arise and as a result very high values of R^2 (less than 0.85) should be avoided.

In addition to the mathematical tests, the ultimate goal of each model is to accurately predict the phenomenon that attempts to describe.

6. Results

The objective of this research is to reveal the association between stated and revealed driving behaviour of the driver with the use of vehicle diagnostic tools. The collection of the data required to investigate the driver's stated and revealed behaviour was carried out using a vehicle diagnostic system that transmits data via a dedicated device that records data per second and collects them in a central database.

A high number of linear regression model tests are developed for a combination of the dependent variables mentioned before, a procedure that resulted in the final mathematical models that capture the correlation between dependent variables and the factors affecting them. It is noted that the relative influence of the independent variables of each model on the respective dependent variable is determined by the elasticity value. Relative influence is used to quantify the influence of each variable individually, thus enabling comparisons between the influences of the different variables of the same model.

Table 1 summarizes the results of the four models including the coefficients βi and the values of the independent variables' relative influence. The models' dependent variable are:

- Model 1: Harsh braking events per 100 km

- Model 2: Harsh acceleration events per 100 km
- Model 3: Harsh braking events (driving speed = 50-90 km/h at the beginning of the event)
- Model 4: Harsh acceleration events (acceleration rate = 0.22-0.26 m/s² at the peak of the event)

Table 1. Results of the linear regression models developed

	Model 1			Model 2			Model 3			Model 4		
Independent variables	β_i	p-value	e_i	β_i	p-value	e_i	β_i	p-value	e_i	β_i	p-value	e_i
(Constant)	43.818	.000		70.521	.000		7.169	.000		16.645	.000	
Annual income < 10,000 Euro	-33.296	.052	1.171	-50.742	.001	1.392	-7.694	.000	1.135	-19.300	.003	1.288
2 accidents to date	44.115	.033	1.156	57.078	.002	1.357	13.856	.000	1.626	16.824	.022	1.280
1 traffic violation fine	31.718	.052	1.000	38.564	.007	1.276	8.601	.000	1.254	12.753	.026	1.000
3 trips/ day	40.961	.036	1.097									
Sometimes I brake harshly				-24.289	0.025	1.000	-2.752	.048	1.000			
I rarely brake harshly										10.628	.047	1.001
R^2 adjusted	0.442			0.645			0.818			0.518		
F test	3.697			7.818			17.827			5.026		
df	15			15			15			15		

The results of the linear regression models developed are compared with the descriptive statistics presented before. The most important observations arising from this comparison are set out below:

- Descriptive statistics showed that high-income drivers perform the highest number of harsh braking and accelerations events per 100 km, which is also shown in the models developed.
- High-income drivers perform a higher number of harsh braking events than that declared.
- The number of traffic violations does not appear to be associated with the number of harsh acceleration events per 100 km, which is in contrast with the results of the models showing that as the number of traffic violation fines received by the driver increases, the possibility of recording a steep accelerating or braking decreases.
- The initial statistical analysis showed that the number of accidents does not seem to affect the number of harsh acceleration events per 100 km. On the contrary, the linear regression models prove that the number of accidents is the most determinant factor.
- Both types of analysis revealed that drink-driving does not affect the number of recorded events.

7. Conclusions

The objective of this research is to examine the correlation between stated and revealed driver's behaviour using data collected from the OBD and a questionnaire administered. To this end, a large data set collected from a driving behaviour experiment is exploited, which recorded real time behavioral data from a sample of 17 drivers. There are many factors influencing driving behaviour, but analysis of the elasticity shows that the most determining factors are the number of accidents the driver has been involved in to date and the annual income. Other factors are the number of traffic violations that the driver has received over the past 2 years as factors as well as the perception of the harsh braking frequency and the number of trips performed per day.

Since the number of accidents and the number of traffic fines received by a driver is a representative indicator of the degree to which it is attentive, it is noticed that the higher the number of accidents a driver has been involved in, the lower the number of harsh events he performs while driving. It is therefore inferred that the higher the number of harsh braking or acceleration events, the more cautious a driver is. More specifically, drivers who brake harshly often also tend to reduce their speed due to obstacles, markings or speed limits often and accelerate often due to the low driving speed developed in order to adapt to normal traffic conditions. It is also possible to brake harshly often to avoid accidents efficiently and receive less speed limit fines. Drivers with low annual income appear to be more dangerous and careless than high-income drivers as fewer events are recorded and therefore they are developing higher driving speed. This might be due to the fact that high-income drivers' lives are characterized by a stability and security that is lacking from the lives of the former. Low-income drivers are likely to be young drivers according to literature tend to drive over the speed limits. Finally, low-income drivers may not be able to use the vehicle as much as high-income drivers resulting in less trained, and therefore riskier drivers.

The stated behaviour of the driver, imprinted through the independent variable of the stated number of harsh braking events, appears to be a less significant predictor than the rest of the variables in the set. This is probably because human perception involves many errors of capabilities overestimation and as a result it is more difficult for someone to assess his/her own traffic behaviour. Apart from that, the driver does not always have adequate knowledge regarding road safety rules to be self-assessed, is often overestimating or underestimating his capabilities and does not always remember events occurred to provide accurate information.

The number of trips performed per day variable is found to be significant only in one of the four models developed and to a very limited extent. This is worth to be commented as it shows that the frequency of trips per day is a secondary factor when modelling driving behaviour. More specifically, the driving pattern does not seem to be adjusted to a specific type when driving but is likely to be influenced solely by the driver's personal character, his/her psychological condition or traffic conditions. The frequency of driving under the influence of alcohol did not appear to be significant in any model. This is explained by the fact that it has a greater impact on the response time that was not addressed in this research due to a lack of relevant data or because only a small portion of drivers stated drink-driving. The variable of traffic violation fines received in the last 2 years is also not found significant in all models. This is because traffic violation includes many driving actions that are not necessarily linked to driving speed and hence to the recording of harsh events. For instance, some of these actions are non-use of seatbelt, traffic light violation in a non-pedestrian or pedestrian area and mobile phone usage while driving. Consequently, drivers who responded to this question may have reported individual acts of traffic rules violation that cannot be adequately captured in the models of the overall revealed behaviour.

Finally, it is stated that under conditions it may be possible to generalize the results of this work so that it can be used in subsequent related research to investigate the reported and revealed driver behaviour through the use of diagnostic data. It is of course necessary to make the necessary adjustments in the choice of variables according to the object and purpose of the survey.

The results and conclusions of this paper could contribute towards the improvement of driving behaviour. Based on the present analysis, a platform that includes a massive number of drivers could be built to provide recommendations to drivers on the most significant factors of their driving that should be improved to become less risky. This may take place in the form of information campaigns across the media and the internet, training programs and road safety courses in school activities.

As for further research, a larger driving sample is proposed so that participants can be grouped and conclusions drawn based on different driver elements such as gender, driving age, age, etc. Particularly for the influence of the age variable, the revealed behaviour of the driver could be assessed using a wider range of age categories. In addition, since the behaviour of the driver varies depending on the psychological situation of the individual, this research should be carried out in coordination with special psychologists who will examine the participants before the experiment. Other exposure indicators that affect the driving behaviour and should also be considered are traffic and weather conditions. Future research could explore the influence of the driver's distraction factors, such as the presence of a passenger or other persons (particularly children) in the vehicle or the mobile phone usage. As a final suggestion, surveys should be administered before and after drivers have been recorded and received feedback on their behaviour. Thus, a comparison between the results could be made and investigate whether or not feedback provided had a significant impact on their behaviour.

References

Brown, I.D. & Copeman, A.K.,1975. Drivers' attitudes to the seriousness of traffic offences considered in relation to the design of sanctions, Accident Analysis and Prevention 7, 15–26.

Dingus T.A., Guo F., Lee S., Anthin J.F., Perez M., Buchanan-King M., Hankey J., 2016. Driver crash risk factors and prevalence evaluation using naturalistic driving data, Proceedings of the National Academy of Sciences of United States of America 113 (10), 2636-2641.

Green P., Srinivasan V., 1978. Conjoint analysis in consumer research: Issues and outlook. Journal of consumer research 5, 103-212.

Huang, Y., Roetting M., McDevitt J/R., Melton D., Smith S.G., 2005. Feedback by technology: Attitudes and opinions of truck drivers. Transportation Research Part F: Traffic Psychology and Behaviour 8.4-5, 277-297.

Kroes E.P., Sheldon R.J., 1988. Stated Preference Methods: An Introduction. Journal of Transport Economics and Policy 22.1, 11-25.

Lin L., Pioche A., Stander P., 1986. Estimating sales volume potential for innovative products with case histories, Proceeding of the 1986 ESOMAR congress, Monte Carlo.

Nadeem T, Dashtinezhad S., Liao C., Iftode L., 2004. TrafficView: traffic data dissemination using car-to-car communication. ACM SIGMOBILE Mobile Computing and Communications Review 8.3, 6-19.

Persson, U., 1992. Three economic approaches to valuing benefits of traffic safety measures, Lund University.

Sheldom J., Steer J., 1982. The use of conjoint analysis in transport research. Paper presented to the 1982 PTRC Summer Annual Meeting, Warwick.

Torfs, K., Meesmann, U., Van den Berghe, W., & Trotta, M., 2016. ESRA 2015 – The results. Synthesis of the main findings from the ESRA survey in 17 countries. ESRA project (European Survey of Road users' safety Attitudes). Brussels, Belgium: Belgian Road Safety Institute.

WHO, 2015. Global status report on road safety 2015. World Health Organization, Geneva.