Comparative Evaluation of Driving Efficiency Using Smartphone Data

Dimitrios I. Tselentis
Research Associate
National Technical University of Athens,
5 Iroon Polytechniou St., GR-15773, Athens, Greece
Tel: +30 210 772 4715; Fax: +30 210 772 1454
E-mail: dtsel@central.ntua.gr

Eleni I. Vlahogianni, PhD
Assistant Professor
National Technical University of Athens
Department of Transportation Planning and Engineering
5 Iroon Polytechniou St., GR-15773, Athens, Greece
Tel: +302107721369, Fax: +302107721454
E-mail: elenivl@central.ntua.gr

George Yannis, PhD
Professor
National Technical University of Athens
Department of Transportation Planning and Engineering
5 Iroon Polytechniou St., GR-15773, Athens, Greece
Tel: +302107721326, Fax: +302107721454
E-mail: geyannis@central.ntua.gr

Word Count: 6468 (Tables:3, Figures:1)

Submission Date: 01/08/2017

ABSTRACT

The objective of this paper is to provide a solid framework for the comparative evaluation of driving efficiency based on Data Envelopment Analysis (DEA). The analysis considers each driver as a Decision Making Unit (DMU) and aims to provide a relative efficiency measure to compare different drivers based on their driving performance. The last is defined based on a set of driving analytics (e.g. distance travelled, speed, accelerations, braking, cornering and smartphone usage) collected using an innovative data collection scheme, which is based on the continuous recording of personalized driving behavior analytics in real time, using smartphone device sensors. Efficiency is examined in terms of speed limit violation, driving distraction, aggressiveness and safety on urban, rural and highway road and in an overall model. DEA models are identifying the most efficient drivers that lie on the efficiency frontier and act as peers for the rest of the non-efficient drivers. The proposed methodological framework is tested on data from fifty-six (56) drivers during a 7-months driving experiment. Findings help distinguish the most efficient drivers from those that are less efficient. Moreover, the efficient level of inputs and outputs that should be reached by each one of the less-efficient and non-efficient drivers to switch to the efficiency frontier and become efficient is identified. Results also provide a potential for classification of the driving sample based on drivers' comparative efficiency. The main characteristics of the most and less efficient drivers are consequently analyzed and presented herein. The impact of this methodology lies on the fact that most common inefficient driving practices are identified (aggressive, risky driving etc.) and driving behavior is comparatively evaluated and analyzed.

Keywords: Driving Behavior, Driving Efficiency, Data Envelopment Analysis, Smartphone Data

INTRODUCTION

Measuring driving efficiency has been the focus of many studies in driving behavior literature in the past *(1, 2, 3)* From a traffic safety perspective, it is a matter of great significance to identify the parameters that influence driving behavior and therefore traffic risk. Several studies have been carried out regarding mobile phone usage distraction and methodologies for collecting and analyzing *(4)* driving behavior data. The most common methodology applied included driving simulators *(5, 6)*, questionnaires *(2)* combined with simulators and naturalistic driving experiments *(7,8)*, while the most common method of monitoring driving measures included recorders that relate to the car engine *(9,10)* and smartphones *(11)*.

Regarding mobile phone usage while driving, literature has shown that it has a significant effect on driver behavior. Cell phone use causes drivers to have higher variation in accelerator pedal position, drive more slowly with more variation in speed and report a higher level of workload *(12)* regardless of conversation difficulty level. Drivers tend to select larger vehicle spacing *(13)*, and longer time headways *(14)* suggesting possible risk compensatory behavior *(12, 15)*. Furthermore, the participants' reaction times *(16)* increase significantly when conversing, but no benefit of hands-free units over handheld units on rural roads/motorways were found *(17, 18)*. Thus, with regards to mobile telephones, the content of the conversation was far more important for driving and driver distraction.

Speeding is also recognized as one of the most important factors in driving risk since it influences the accident probability (e.g. decreased reaction distance, loss of control) as well as the crash impact. According to *(19)* speeding has been a contributory factor in 10% of the total accidents and more than 30% in fatal accidents. According to *(20, 13)* the probability of a crash involving an injury is proportional to the square of the speed, the probability of a serious crash is proportional to the cube of the speed and the probability of a fatal crash is related to the fourth power of the speed. Moreover, *(21)* depicts the relationship between speed and driving risk via an exponential curve, showing that the driving risk is not proportional to the speed.

Harsh acceleration, harsh breaking and harsh cornering events are three significant indicators for driving risk assessment *(22, 23)* especially for evaluating driving aggressiveness. This is because they are strongly correlated with unsafe distance from adjacent vehicles, possible near misses, lack of concentration, increased reaction time, poor driving judgement or low level of experience and involvement in situations of high risk (e.g. marginal takeovers). The correlation between HA and HB events with driving risk has been highlighted in the scientific papers published by *(4, 23)* and it has been widely recognized by the insurance and telematics industry.

Driver's efficiency on a microscopic level has been studied in a great extent but never by making use of DEA techniques. This paper proposes a methodological framework to address the issue of measuring driver's efficiency and categorize the drivers of the sample used in three groups i.e. non-efficient, weakly efficient, most efficient. The main characteristics of each group are presented in order to draw important conclusions on the features of each driving group and provide recommendations for drivers on how to improve their driving efficiency. For the purposes of this study, drivers will be considered as DMUs, which is deemed to be rational since a driver is a unit that makes decisions for a given mileage range about the number of events occurring and the time of mobile phone usage and speed limit violation. Driving attributes (metrics and distance recorded) will be considered the inputs and outputs of the DEA program. More details on the structure of the DEA implemented will be given below.

The concept of DEA is to minimize inputs (input-oriented model) or maximize the outputs of a problem (output-oriented model). More specifically in the case study examined herein, a driver should either drive more kilometers maintaining the same number of harsh braking or accelerating events or reduce the number of harsh braking/accelerating events for the same mileage. The same applies to the rest of the metrics recorded for each driver. From a road safety perspective, increasing mileage increases crash risk *(4)* and, therefore, an input-oriented DEA program is being developed aiming to minimize inputs (recorded metrics) maintaining the same number of outputs (recorded distance). Although a trip cannot literally behave as a decision making unit, it can be evaluated as a DMU and, therefore, it will be considered as such for the purpose of this research. This is deemed to be a correct assumption on a trip basis since a) all variables used are continuous quantitative variables as those used in previous DEA studies *(24, 25, 26, 27)* and b) a driver should reduce his mileage *(4)* and the frequency of some of his driving characteristics such as harsh acceleration and braking, mobile phone usage and speeding *(4, 28, 29)*. The proposed methodology is applied to a real-life case study of 34,060 recorded trips collected from fifty-six (56) drivers. More details on the procedure are given below.

DATA ENVELOPMENT ANALYSIS: BRIEF THEORY AND APPLICABILITY TO DRIVING EFFICIENCY

The terms “efficiency” and “productivity” are widely used in economics and refers to the optimal way a production unit can make use of its available resources *(30)*. More specifically *(31)*, a Decision-Making Unit (DMU) is “technically efficient” when the amount of outputs produced is maximized for a given amount of inputs, or for a given output the amount of inputs used is minimized. Thus, when a DMU is technically efficient, it operates on its production frontier and therefore DMUs lie on the efficiency frontier. Based on the assumptions that will be stated below, in this study drivers are considered as DMUs and DEA applicability on the field of driver’s assessment based on microscopic behavioral characteristics is investigated.

Efficiency can be defined as the ratio of input and output in a theoretical scenario of units that have a single input and output but in a real case scenario where typical organizational unit have multiple and incommensurate inputs and outputs a more scientific approach is needed. Data Envelopment Analysis (DEA) is an approach for efficiency and productivity analysis of production units with multiple inputs to produce multiple outputs mostly used thus far in business, economics, management and health. The rationale for using DEA is its applicability to the multiple input–output nature of DMUs provision and the simplicity of the assumptions underlying the method. It is a methodology of several different interactive approaches and models used for the assessment of the relative efficiency of DMU and for the assessment of the efficiency frontier. It assists in drawing important conclusions on operational management of the efficient and inefficient units.

DEA is a technique of mathematical programming problem with minimal assumptions that determines of a unit’s efficiency based on its inputs and outputs, and compares it to other units involved in the analysis. It is a data-oriented methodology that effects performance evaluations and other conclusions drawn from the analysis directly from the observed data. The efficiency of a DMU is comparatively measured and analyzed relatively to the rest of the DMUs considering that all DMUs lay on or below the efficiency frontier. No assumption is required about functional form (e.g. a regression equation, a production function, etc.) or the statistical distribution of data sample and as a result DEA is classified as a non-parametric method. It is a frontier analysis, a process of extremities, not driven by central tendencies in contrast to all statistical procedures. Each DMU is analyzed separately and the real and optimal performance that can be achieved for each unit is estimated.

DEA has become one of the most popular fields in operations research, with applications involving a wide range of context *(32)*. It has been applied in great extent in literature *(24, 33, 25)* to measure and compare the productivity performance of a group of DMUs. It is one of the most popular fields in operations research *(33, 34)* to say the least. *(35)* presented the ample possibilities for using DEA for evaluating among others the performance of banks, schools, university departments, farming estates, hospitals and social institutions, military services and entire economic systems. Since the introduction of CCR model *(36)* in 1978, the number of publications where DEA is implemented has exponentially grown. DEA has also been implemented in transport fields in assessing public transportation system performance *(26)*, as well as traffic safety studies *(27, 37)* where it was proved to be equally useful as in the fields stated above.

DEA is a non-parametric approach that does not require any assumptions about the functional form of a production function and a priori information on importance of inputs and outputs. DEA allows each DMU to choose the weights of inputs and outputs which maximize its efficiency. The DMUs that achieve efficiency equal to unit are considered efficient while the other DMUs with efficiency scores between zero and unit are considered as inefficient. The first DEA model proposed by *(36)* is the CCR model that assumes that production exhibits constant returns to scale i.e. outputs are increased proportionally to inputs. DEA models can also be distinguished based on the objective of a model; that can be either outputs maximization (output-oriented model) or inputs minimization (input-oriented model).

It is assumed that this study should adopt an input-oriented (IO) DEA model, since the objective is to minimize the number of harsh accelerations, harsh brakings etc. that occur per driving distance unit rather than to maximize driving distance for given metrics since the latter would increase the exposure of a driver and therefore driving risk. It is also implicitly assumed that the driving efficiency problem is a constant returns to scale (CRS) problem and that the sum of all metrics (inputs) recorded such as the number of harsh accelerations and brakings occurred in each $trip_i$ changes proportionally to the sum of driving distance (output).

Let us use X and Y to represent the set of inputs and outputs, respectively. Let the subscripts i and j to represent particular inputs and outputs respectively. Thus x_i represents the i^{th} input, and y_j represent the j^{th} output of a DMU. The input-oriented CCR model evaluates the efficiency of DMU_o by solving the following (envelopment form) linear program *(38)* and its mathematical formulation is formulated as:

$$
\begin{aligned}
&min\theta_B \\
&subject\ to\ the\ following\ constraints: \\
&\theta_B * x_o - X * \lambda \geq 0 \\
&Y * \lambda \geq y_o \\
&\lambda_i \geq 0\ \forall\ \lambda_i \in \lambda
\end{aligned}
\tag{1}
$$

where λ_i is the weight coefficient for each DMU_i that is an element of set λ, X is the set of Inputs, Y is the set of Outputs and θ_B is a scalar representing the efficiency of reference DMU_o. The objective function of this linear programming problem (DEA) is min θ_ι i.e. minimize the efficiency of DMU_i (in this case $trip_i$). In order to benchmark the efficiency of all trips (of each DMU) of the database, this linear programming problem should be solved for each DMU_i (i.e. each $trip_i$).

EXPERIMENTAL DATA COLLECTION

An integrated system for recording, collection and storage of driving behavior data using smartphone applications and advanced Machine Learning algorithms developed by OSeven Telematics. The system developed integrates a data collection, transmission and processing procedure using Smartphones, the main features of which are outlined in the next paragraphs.

Data recording and transmission from smartphone

A developed mobile App is employed for the purposes of this study to record user's behaviour exploiting the hardware sensors of the smartphone device and a variety of APIs to read sensor data and transmit it to a central database. Recorded data come from various smartphone sensors and data fusion algorithms provided by Android (Google) and iOS (Apple). The frequency of the data recording varies depending on the type of the sensor with a minimum value of 1Hz.

After the end of the trip, all recorded data are transmitted to the central database via a Wi-Fi network or cellular network based on the user preference. After data is stored in the cloud server for central processing and data reduction, it is converted into meaningful behavioral and safety related indicators as a result of big data handling and processing. This is achieved by using the two Big data processing methods which include two families of techniques, Big Data mining techniques and Machine Learning (ML) algorithms.

Machine learning methods (filtering, clustering and classification methods) are used to clean the data from existing noise and errors, and to identify repeating patterns within the data. The methods applied allow for data filtering and outlier detection, data smoothening, speeding regions, harsh acceleration events, harsh braking events, harsh cornering events, mobile usage, risky hours driving and driver or passenger recognition. Subsequently, these data patterns will be processed by means of big data mining techniques, to calculate the necessary parameters and derive behaviour indicators to be used in the analysis.

After the ML process is completed, a variety of different indicators are calculated that are useful to the evaluation of driving behavior. These indicators are divided into two distinct categories, risk exposure and driving behavior indicators. The main risk exposure indicators are total distance travelled, driving duration, type(s) of the road network used, time of the day driving, combined with other data sources. The main driving behavior indicators are speeding, mobile phone use, number and severity of harsh events such as harsh braking, harsh acceleration and harsh cornering.

Aggregated Data are analyzed and filtered to retain only those indicators that will be used as inputs and outputs for the DEA problem. The procedure how inputs and outputs are selected will be described in the next section. Data filtering and DEA improvement algorithms are performed in Python programming language and several scripts are written for this reason. Python packages used include Pandas and Numpy for numeric calculations and transformations and Pulp for linear programming problem construction. More details on the algorithm implementation are given below.

Experiment design

A naturalistic driving experiment is implemented in this research by recording personalized driving behaviour analytics in real time, exploiting data collected from smartphone device sensors using a smartphone application developed by OSeven Telematics. Two hundred and thirty six (236) drivers participated in the designed experiment that took place between 25/08/2016 and 04/04/2017 and a large database of 50,741 trips was created. The first criterion chosen by the authors for specifying the driver's sample were adopted from study *(39)* which proved that sampling less than 100 driving hours per driver does not result in a reliable measure for analyzing driving patterns and changes in the behavior of drivers over time. On the top of that, all drivers should have positive mileage on all three types of road network. The third criterion was that drivers with zero input attributes (i.e. zero harsh acceleration, braking, speed limit violation, mobile phone usage) should be eliminated from the sample, which is a limitation of DEA. The business equivalent of a zero input could be a factory that is producing a product without making use of any material and/or workforce, which practically cannot occur. For the same reason, harsh cornering events were finally eliminated as a DEA input as in most cases there are no such events in highways and therefore results would not be comparable. This procedure resulted to 56 drivers who fulfilled these 3 criteria and were kept for the analysis that was conducted and 180 drivers were eliminated from this study. The total number of trips that took place by the 56 drivers chosen was 34,060 constructing thus a large database.

IMPLEMENTATION AND RESULTS

Input and output Selection

Models representing driving behavior in all road types and in total are developed with multiple inputs and outputs. A critical process for DEA is input and output selection. Thus, selection process should be linked to the conceptual specifications of each problem. Several issues that should be taken into consideration before applying DEA to a dataset are discussed in *(40)*. One of the pitfalls is that the efficiency score might be miscalculated when input and output variables are in the form of percentiles and/or ratios simultaneously with raw data *(41)*. Taking this into account the specific data used in this study are metrics recorded in the form of raw data i.e. the number of harsh brakings, harsh accelerations, harsh cornerings, seconds driving over the speed limit and seconds used the mobile phone and not as ratios or percentiles e.g. percentage of distance driving over the speed limit. Literature review revealed that all these indicators are influencing the most accident risk that is the reason why they are used in the models implemented. All indicators along with distance travelled by drivers are recorded per road type (urban, rural, highway) and in total e.g. number of harsh accelerations that occurred in urban road, time violating speed limits etc. Variables used in the analysis along with their description are shown in Table 1.

TABLE 1: Variables recorded during the experiment

Variable name	Variable short description
ha_X	number of harsh acceleration events in X road type
ha_{urban}	number of harsh acceleration events in urban road
ha_{rural}	number of harsh acceleration events in rural road
$ha_{highway}$	number of harsh acceleration events in highway
hb_X	number of harsh braking events in X road type
hb_{urban}	number of harsh braking events in urban road
hb_{rural}	number of harsh braking events in rural road
$hb_{highway}$	number of harsh braking events in highway
$speeding_X$	total seconds of speed limit violation in X road type
$speeding_{urban}$	total seconds of speed limit violation in urban road
$speeding_{rural}$	total seconds of speed limit violation in rural road
$speeding_{highway}$	total seconds of speed limit violation in highway
$mobile_X$	total seconds of mobile phone usage in X road type
$mobile_{urban}$	total seconds of mobile phone usages in urban road
$mobile_{rural}$	total seconds of mobile phone usage in rural road
$mobile_{highway}$	total seconds of mobile phone usage in highway
$distance_X$	total distance driven in X road type
$distance_{urban}$	total distance driven in urban road
$distance_{rural}$	total distance driven in rural road
$distance_{highway}$	total distance driven in highway

The driver is deemed to be a DMU with an aggregate performance for the entire monitoring period. Moreover, his driving behavior is considered equivalent to the sum of the driving characteristics that were recorded for the period examined. For instance, the total distance travelled in urban network is equivalent to the sum of the distance travelled in urban network in each $trip_{ij}$ (where i is the index of $driver_i$ and j the index of $trip_j$ of $driver_i$) by the specific $driver_i$ (DMU_i). In general, the same applies for all indicators of $driver_i$, which are calculated aggregately as shown in the following formula (2):

$$indicator_i = \sum_{j=1}^{N_i} indicator_{ij} \quad (2)$$

recorded $\forall\ trip_j,\ j \in (1, N_i)$ that took place by $driver_i$. As described above, each driver is treated as a distinct DMU to be analyzed in DEA and therefore the linear program constructed (see equation (1)) has 70 variables (λ_i, θ_B) that is equal the number of drivers plus the efficiency for $driver_o$. The number of constraints on the other hand is equal to the sum of a) the number of inputs ($\theta_B * x_o – X * \lambda \geq 0$) b) the number of outputs ($Y * \lambda \geq y_o$) and c) the number of drivers ($\lambda_i \geq 0$). The DEA procedure described by equation (1) is followed separately for each of the three different road types (urban, rural, highway) and aggregately in an overall model as described in Table 2.

TABLE 2: Inputs and Outputs of DEA models used

	Per road type model		Overall model	
	Set of Inputs used	**Set of Outputs used**	**Set of Inputs used**	**Set of Outputs used**
Model type 1	1) $speeding_X$	1) $distance_X$	1) $speeding_{urban}$ 2) $speeding_{rural}$ 3) $speeding_{highway}$	1) $distance_{urban}$ 2) $distance_{rural}$ 3) $distance_{highway}$
Model type 2	1) $mobile_X$	1) $distance_X$	1) $mobile_{urban}$ 2) $mobile_{rural}$ 3) $mobile_{highway}$	1) $distance_{urban}$ 2) $distance_{rural}$ 3) $distance_{highway}$
Model type 3	1) ha_X 2) hb_X	1) $distance_X$	1) ha_{urban} 2) ha_{rural} 3) $ha_{highway}$ 4) hb_{urban} 5) hb_{rural} 6) $hb_{highway}$	1) $distance_{urban}$ 2) $distance_{rural}$ 3) $distance_{highway}$
Model type 4	1) ha_X 2) hb_X 3) $speeding_X$ 4) $mobile_X$	1) $distance_X$	1) ha_{urban} 2) ha_{rural} 3) $ha_{highway}$ 4) hb_{urban} 5) hb_{rural} 6) $hb_{highway}$ 10) $speeding_{urban}$ 11) $speeding_{rural}$ 12) $speeding_{highway}$ 13) $mobile_{urban}$ 14) $mobile_{rural}$ 15) $mobile_{highway}$	1) $distance_{urban}$ 2) $distance_{rural}$ 3) $distance_{highway}$

This results to 16 different models of which 12 are per road type and 4 overall. The variables' combinations for structuring the four models of each category was based on literature review. Model 1 and 2 represents the speed limits violation and mobile phone distraction. Model 3 incorporates the three most significant explanatory driving indicators for driving aggressiveness, while model 4 is the overall model that includes all traffic safety parameters found in literature review and accounts for the overall safety profile of the driver.

Figure 1 illustrates the results of model 3 per road type where as it appears there is only one efficient DMU for urban and rural road, whereas for highway there are two, which confirms the results of the DEA LPs. In every plot of Figure 1, $distance_X/ha_X$ and $distance_X/hb_X$ is plotted in axis Y and X respectively along with the envelopment line accounting for the efficiency frontier. Extending the line joining the origin and DMU_i, it crosses the efficiency frontier at a point where virtual DMU_i' is created which represents the optimal performance which the specific DMU_i can achieve. The closer a DMU is to the efficiency frontier, the higher its efficiency index is. In urban and rural road subplots, the influence of outliers to the DEA solution is obvious since most DMUs appear to be near the origin. Nonetheless, the solution still remains reliable as the efficiency index calculated is comparable to that of the rest of the DMU set. It should be highlighted that models incorporating two-inputs/ one output or one-input/ two outputs can only be visualized in 2-D figures.

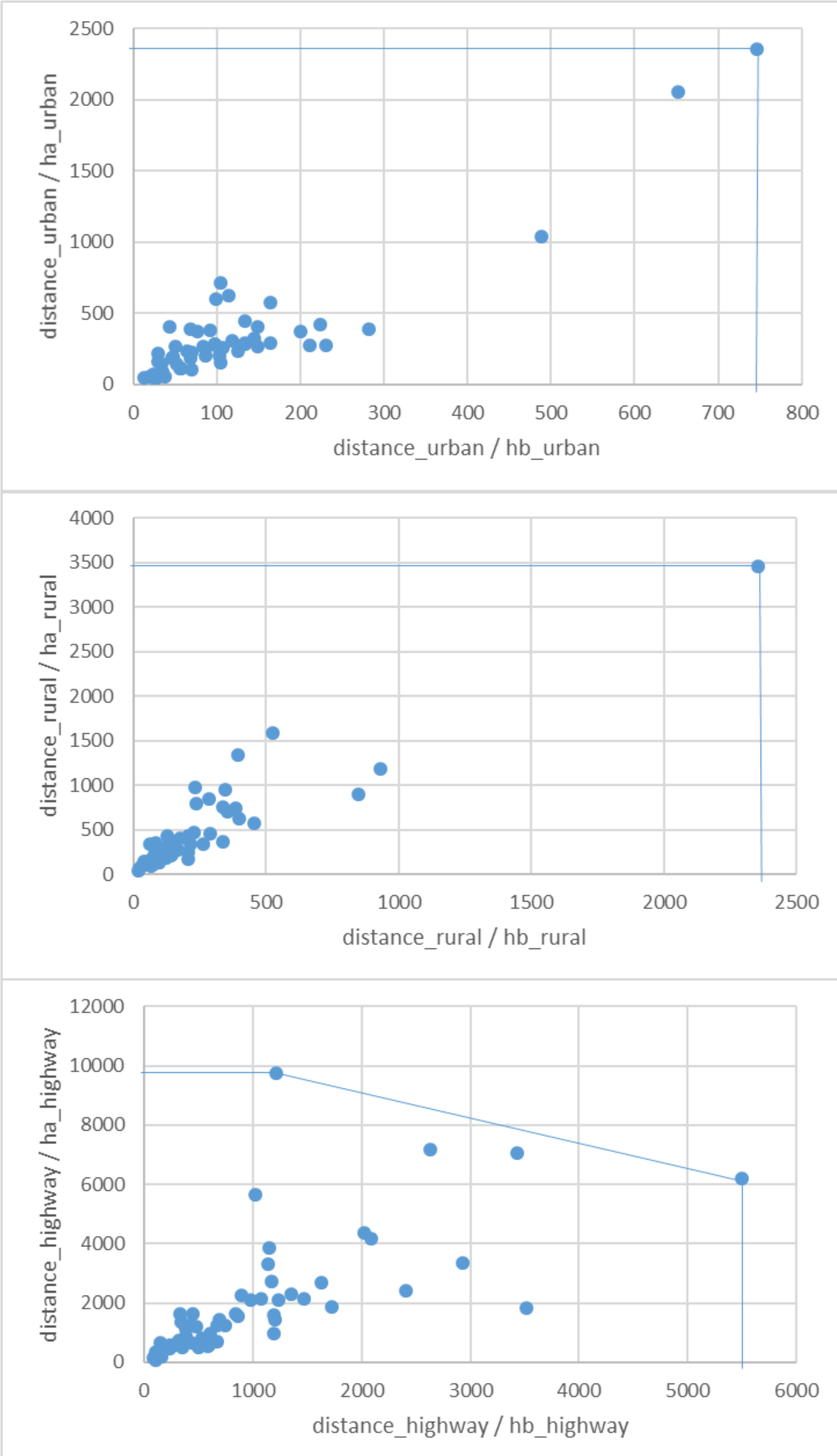

FIGURE 1: Efficiency frontier of drivers aggressiveness per road type

Drivers sample classification

The results of DEA are the efficiency index θ_B and coefficients λ_i for each DMU (driver). This allows for the classification of the whole set of DMUs to most efficient, weakly efficient and non-efficient. Since the absolute value of the efficiency index cannot be somehow interpreted unless it is compared to the efficiency index of the rest of the DMU set, the percentiles of the DMU set's θ_B are used to classify drivers. The percentile thresholds specified was 25% and 75%, which separate the subsets of non-efficient and weakly efficient as well as weakly efficient and most efficient DMUs respectively. The average of the attributes of each class arising, weighted on the distance (for harsh acceleration and braking) or driving time (for speeding and mobile usage) travelled by each driver, are shown in Table 3 where the models per type, road type and overall are presented based on the inputs that were used in each model. For brevity purposes, from here on class 1 drivers will be referred to as most efficient drivers despite the fact that only drivers with unit efficiency lie on the efficiency frontier.

For instance, in model $Rural_3$ (representing model 3 of rural road type) the average ha_{rural} and hb_{rural} per 100 km travelled (ha_x, hb_x are the inputs of model 3 for every road type as shown in table 2) of each class are illustrated. For better understanding, results are presented as a percentage of driving time for speeding and mobile usage and as events per 100 kilometers driven for harsh acceleration and braking.

Main characteristics of drivers efficiency classes

As expected for models 1, 2 and 3 in every road type the average of the attributes is reducing while a driver becomes more efficient. The reason why this is not valid for model 4 of urban and rural road types is probably because a) while the number of inputs and outputs increases, the number of efficient DMUs are increasing as well, especially for small scale samples as the one examined a which renders the classification of the DMUs to be more difficult and less accurate since many DMUs have unit efficiency and b) of DEA's sensitivity to outliers, which means that the model can sometimes be influenced by the extreme values of other inputs or outputs e.g. low values of speeding or mobile usage when estimating a DMU's efficiency.

Another observation is that the number of harsh events occurring in urban road is extremely higher than in rural and highway and that the number of harsh events in rural road is higher than in highway. The same is noticed for mobile usage but not for speeding where apparently drivers of all classes tend to drive over the speed limits in rural and highway at least the same or more than in urban. As for model 4 of all road types it should be highlighted that for a specific class some attributes appear to be increased compared to model 1, 2 or 3 and this is attributed to the fact that more parameters are taken into account in the model that might affect the final configuration of each class.

In general, it can be concluded from model 1 that speed limit violation does not fluctuate and is limited to less than 6.5% of driving time for most efficient drivers in all road types whereas for non-efficient drivers it ranges from 20% to over 32%. As for the set of weakly efficient drivers speed limit exceedance is around 12% - 14%. In terms of mobile usage distraction, it appears that non-efficient drivers use their mobile phone significantly more than the other two classes averaging at 16% while most efficient drivers use it less than 1.5% in average. Finally, weakly efficient group of drivers make mobile usage of less than 7%.

TABLE 3: Driving characteristics of efficiency groups per road type and overall

		Efficiency classes		
Model		**1: 0 - 25 % percentile**	**2: 25 - 75 % percentile**	**3: 75 - 100 % percentile**
Urban	**1**	20.08 % $speeding_{urban}$	11.95 % $speeding_{urban}$	6.51 % $speeding_{urban}$
	2	19.48 % $mobile_{urban}$	6.80 % $mobile_{urban}$	2.31 % $mobile_{urban}$
	3	45.97 $ha_{urban}/100km$ 17.38 $hb_{urban}/100km$	27.40 $ha_{urban}/100km$ 8.99 $hb_{urban}/100km$	10.71 $ha_{urban}/100km$ 5.08 $hb_{urban}/100km$
	4	41.06 $ha_{urban}/100km$ 16.75 $hb_{urban}/100km$ 17.77 % $mobile_{urban}$ 15.79 % $speeding_{urban}$	22.85 $ha_{urban}/100km$ 8.43 $hb_{urban}/100km$ 6.78 % $mobile_{urban}$ 13.02 % $speeding_{urban}$	24.72 $ha_{urban}/100km$ 6.81 $hb_{urban}/100km$ 4.05 % $mobile_{urban}$ 8.66 % $speeding_{urban}$
Rural	**1**	23.79 % $speeding_{rural}$	14.21 % $speeding_{rural}$	6.33 % $speeding_{rural}$
	2	15.10 % $mobile_{rural}$	5.69 % $mobile_{rural}$	1.64 % $mobile_{rural}$
	3	23.65 $ha_{rural}/100km$ 11.43 $hb_{rural}/100km$	14.28 $ha_{rural}/100km$ 6.96 $hb_{rural}/100km$	6.36 $ha_{rural}/100km$ 3.00 $hb_{rural}/100km$
	4	20.31 $ha_{rural}/100km$ 8.71 $hb_{urban}/100km$ 10.28 % $mobile_{rural}$ 20.58 % $speeding_{rural}$	12.32 $ha_{rural}/100km$ 6.26 $hb_{urban}/100km$ 6.51 % $mobile_{rural}$ 14.49 % $speeding_{rural}$	13.62 $ha_{rural}/100km$ 7.13 $hb_{urban}/100km$ 4.81 % $mobile_{rural}$ 8.97 % $speeding_{rural}$
Highway	**1**	32.39 % $speeding_{highway}$	13.06 % $speeding_{highway}$	3.98 % $speeding_{highway}$
	2	12.34 % $mobile_{highway}$	3.73 % $mobile_{highway}$	0.74 % $mobile_{highway}$
	3	3.40 $ha_{highway}/100km$ 1.67 $hb_{highway}/100km$	1.74 $ha_{highway}/100km$ 1.02 $hb_{highway}/100km$	0.98 $ha_{highway}/100km$ 0.49 $hb_{highway}/100km$
	4	2.80 $ha_{urban}/100km$ 1.61 $hb_{highway}/100km$ 5.40 % $mobile_{highway}$ 29.31 % $speeding_{highway}$	1.91 $ha_{urban}/100km$ 1.05 $hb_{highway}/100km$ 5.61 % $mobile_{highway}$ 13.08 % $speeding_{highway}$	1.24 $ha_{urban}/100km$ 0.50 $hb_{highway}/100km$ 3.92 % $mobile_{highway}$ 7.01 % $speeding_{highway}$
Overall	**1**	17.12 % $speeding_{urban}$ 21.25 % $speeding_{rural}$ 24.24 % $speeding_{highway}$	12.50 % $speeding_{urban}$ 14.41 % $speeding_{rural}$ 14.26 % $speeding_{highway}$	8.37 % $speeding_{urban}$ 8.48 % $speeding_{rural}$ 9.72 % $speeding_{highway}$
	2	17.07 % $mobile_{urban}$ 13.30 % $mobile_{rural}$ 9.75 % $mobile_{highway}$	7.22 % $mobile_{urban}$ 5.99 % $mobile_{rural}$ 4.37 % $mobile_{highway}$	3.89 % $mobile_{urban}$ 2.85 % $mobile_{rural}$ 2.05 % $mobile_{highway}$
	3	36.94 $ha_{urban}/100km$ 19.26 $ha_{rural}/100km$ 3.12 $ha_{highway}/100km$ 12.42 $hb_{urban}/100km$ 9.33 $hb_{rural}/100km$ 1.44 $hb_{highway}/100km$	30.09 $ha_{urban}/100km$ 16.26 $ha_{rural}/100km$ 1.76 $ha_{highway}/100km$ 10.34 $hb_{urban}/100km$ 7.36 $hb_{rural}/100km$ 0.95 $hb_{highway}/100km$	17.13 $ha_{urban}/100km$ 8.46 $ha_{rural}/100km$ 1.32 $ha_{highway}/100km$ 7.87 $hb_{urban}/100km$ 4.85 $hb_{rural}/100km$ 0.87 $hb_{highway}/100km$
	4	-	-	-

It is also noticeable from model 3 that drivers of all ranges of aggressiveness have a 2-3 times larger number of harsh acceleration than braking events per 100km of driving. For instance, in urban roads, the number of harsh acceleration events ranges from 11 to 46 per 100km while the number of harsh braking events from 5 to 17.4 for most efficient to non-efficient drivers. The ranges become narrower for rural and highway. In terms of traffic safety, the conclusion that can be drawn from model 4 is that the overall driving profile of a “safer” driver in urban and rural road is not considerably influenced by the driver’s number of harsh events since it is much higher than in model 3 where it accounts for aggressiveness. On the other hand, in highway, mobile usage and speeding seems to be significantly higher than model 1 and 2 whereas the number of harsh acceleration and braking events appears to be more critical since they are kept at a much lower level. The same is observed in highways for weakly efficient drivers but not for non-efficient who tend to have a lower mobile usage rate than in model 2, which accounts for distraction. Additionally, weakly efficient drivers in urban and rural road have a lower average number of harsh acceleration event and in average, the same driving characteristics for the rest of the attributes investigated. Finally, for non-efficient drivers of urban and rural road, it was found that all driving attributes were reduced compared to model 1, 2 and 3 probably due to the interaction among variables.

As stated above, as the number of inputs and outputs increases while the number of DMUs remains low, the number of efficient DMUs that are found to be efficient is radically increased. This is the case of the overall model, model 4, where 38 drivers with unit efficiency were found and this is the reason why the authors did not consider it to be significant enough to be presented.

When considering all road types together in table 3, in terms of speeding percentages a greater tolerance is noticed for drivers to be characterized as most efficient or weakly efficient than in per road type models, which appear to be from slightly in class 2 rural to more than 100% more in class 3 highway model. The same is observed for model 2 and 3 as well for class 2 and 3 drivers except for $hb_{highway}$ which are slightly lower in the overall model. On the other hand, non-efficient drivers have lower speeding percentages in all road types and especially in highway where the difference is higher. The same can be highlighted for model 2 and 3 in highway.

Efficient level of inputs and outputs for non-efficient drivers

After DEA LPs of (1) are solved and the efficiency index θ_B and coefficients λ_i are estimated for each DMU the efficient level of inputs and outputs at which each DMU could optimally reach can be calculated. The efficient level of inputs for trip 1 can be calculated as the product sum of the lamdas and the input values of each of the identified peers whereas to find the efficient level of outputs for the same DMU, each output value should be divided by theta value. Considering $driver_i$ as the reference DMU and a set of m drivers, where $m \in \mathbb{N}$ as his peers, the efficient level of e.g. ha_{urban} can be estimated using following formula (3):

$$\text{Efficient Level of } ha_{urban_i} = \sum_{j=1}^{m} \lambda_j * ha_{urban_j} \tag{3}$$

The efficient level of e.g. $distance_{urban}$ is calculated from formula (4):

$$\text{Efficient Level of } distance_{urban_i} = distance_{urban_i} / theta_i \tag{4}$$

It should be noted that a DMU achieves its efficient level by reaching the efficient level of either its inputs or outputs. Additionally, a DMU is deemed to have achieved the efficient level when it reaches unit efficiency. For the purpose of brevity, lamdas and thetas calculated are not presented herein.

DISCUSSION

This paper provides an innovative solid framework for benchmarking and evaluation of drivers' efficiency based on Data Envelopment Analysis (DEA). Data exploited were collected from smartphone device sensors, which continuously recorded real time personalized information on driving behavior from a sample of fifty-six (56) drivers during 7-months. Combinations of driving analytics collected are taken into consideration for driving assessment, including distance travelled, speed, accelerations, braking and smartphone usage, which serve as inputs and outputs in DEA to calculate a comparative efficiency index for each driver in the sample. Efficiency is examined in terms of speed limit violation, driving distraction, aggressiveness and safety on urban, rural and highway road and in an overall model. An additional value of the methodology proposed is that it enables the estimation of the optimal level of inputs and outputs that should be reached by each driver to become efficient.

The impact of this methodology lies also on the fact that a potential for classifying driving sample based on drivers' comparative efficiency is identified. Drivers were divided into three categories (non-efficient, weakly efficient and most efficient) based on the 25% and 75% percentile thresholds specified. The highlights of the analysis conducted for each category indicated considerable differences in driving characteristics between inefficient drivers and the classes of weakly efficient and most efficient drivers with the difference of the two latter to be less significant. Concerning aggressiveness, harsh braking events appeared to be 2-3 times less than harsh acceleration events in all models indicating a higher significance of this attribute for a driver to be characterized as aggressive. The same observation is made for harsh acceleration events in overall safety models (model 4) of all road types where percentage of speeding and mobile usage was identified as key factors for safety efficiency index estimation.

Further research should center to larger samples of trips with a representative sample of drivers population. It is a fact that models become more representative of the average characteristics of each class as more trips and drivers are aggregated. As the sample grow bigger the high proportion of efficient DMUs to the total number DMus will be reduced. Other DEA's limitations should also be addressed which among others include DEA's sensitivity to outliers and that drivers with zero input attributes should be eliminated from the sample.

ACKNOWLEDGEMENT

The authors would like to thank OSeven Telematics, for providing all necessary data to accomplish this study.

REFERENCES

1. Matthews, Gerald, Timothy J. Sparkes, and Helen M. Bygrave. Attentional overload, stress, and simulate driving performance. Human Performance Vol. 9, No. 1 1996, pp. 77-101. http://dx.doi.org/10.1207/s15327043hup0901_5
2. Matthews, Gerald, Lisa Dorn, Thomas W. Hoyes, D. Roy Davies, A. Ian Glendon, and Ray G. Taylor. Driver stress and performance on a driving simulator. Human Factors Vol. 40, No. 1, 1998, pp. 136-149. https://doi.org/10.1518/001872098779480569
3. Young, Mark S., Stewart A. Birrell, and Neville A. Stanton. Safe driving in a green world: A review of driver performance benchmarks and technologies to support 'smart'driving. Applied ergonomics Vol. 42, No. 4, 2011, pp. 533-539.

https://doi.org/10.1016/j.apergo.2010.08.012
4. Tselentis, Dimitrios I., George Yannis, and Eleni I. Vlahogianni. Innovative motor insurance schemes: a review of current practices and emerging challenges. Accident Analysis & Prevention Vol. 98, 2017, pp. 139-148. https://doi.org/10.1016/j.aap.2016.10.006
5. Desmond, Paula, Peter Hancock, and Janelle Monette. Fatigue and automation-induced impairments in simulated driving performance. Transportation Research Record: Journal of the Transportation Research Board No. 1628, 1998, pp. 8-14. https://doi.org/10.3141/1628-02
6. Lenné, Michael G., Thomas J. Triggs, and Jennifer R. Redman. Time of day variations in driving performance. Accident Analysis & Prevention Vol. 29, No. 4, 1997, pp. 431-437. https://doi.org/10.1016/S0001-4575(97)00022-5
7. Toledo, Tomer, Oren Musicant, and Tsippy Lotan. In-vehicle data recorders for monitoring and feedback on drivers' behavior. Transportation Research Part C: Emerging Technologies Vol. 16, No. 3, 2008, pp. 320-331. https://doi.org/10.1016/j.trc.2008.01.001
8. Birrell, Stewart A., Mark Fowkes, and Paul A. Jennings. Effect of using an in-vehicle smart driving aid on real-world driver performance. IEEE Transactions on Intelligent Transportation Systems Vol. 15, No. 4, 2014, pp. 1801-1810. https://doi.org/10.1109/TITS.2014.2328357
9. Zaldivar, Jorge, Carlos T. Calafate, Juan Carlos Cano, and Pietro Manzoni. Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones. In Local Computer Networks (LCN), 2011 IEEE 36th Conference on, pp. 813-819. IEEE, 2011. https://doi.org/10.1109/LCN.2011.6115556
10. Backer-Grøndahl, Agathe, and Fridulv Sagberg. Driving and telephoning: Relative accident risk when using hand-held and hands-free mobile phones. Safety Science Vol. 49, No. 2, 2011, pp. 324-330. https://doi.org/10.1016/j.ssci.2010.09.009
11. Vlahogianni, Eleni I., and Emmanouil N. Barmpounakis. Driving analytics using smartphones: Algorithms, comparisons and challenges. Transportation Research Part C: Emerging Technologies Vol. 79, 2017, pp. 196-206. https://doi.org/10.1016/j.trc.2017.03.014
12. Haque, Md Mazharul, and Simon Washington. The impact of mobile phone distraction on the braking behaviour of young drivers: a hazard-based duration model. Transportation research part C: emerging technologies Vol. 50, 2015, pp. 13-27. https://doi.org/10.1016/j.trc.2014.07.011
13. Nilsson G. The effects of speed limits on traffic accidents in Sweden. Sartryck, Swedish National Road and Transport Research Institute, 1982.
14. Saifuzzaman, Mohammad, Md Mazharul Haque, Zuduo Zheng, and Simon Washington. Impact of mobile phone use on car-following behaviour of young drivers. Accident Analysis & Prevention Vol. 82, 2015, pp. 10-19. https://doi.org/10.1016/j.aap.2015.05.001
15. Törnros, Jan, and Anne Bolling. Mobile phone use–effects of conversation on mental workload and driving speed in rural and urban environments. Transportation Research Part F: Traffic Psychology and Behaviour Vol. 9, No. 4, 2006, pp. 298-306. https://doi.org/10.1016/j.trf.2006.01.008
16. Patten, Christopher JD, Albert Kircher, Joakim Östlund, and Lena Nilsson. Using mobile telephones: cognitive workload and attention resource allocation. Accident

analysis & prevention Vol. 36, No. 3, 2004, pp. 341-350. https://doi.org/10.1016/S0001-4575(03)00014-9

17. Handel, Peter, Isaac Skog, Johan Wahlstrom, Farid Bonawiede, Richard Welch, Jens Ohlsson, and Martin Ohlsson. Insurance telematics: Opportunities and challenges with the smartphone solution. IEEE Intelligent Transportation Systems Magazine Vol. 6, No. 4, 2014, pp. 57-70. https://doi.org/10.1109/MITS.2014.2343262
18. Yannis, G., E. Papadimitriou, and P. Papantoniou. Distracted driving and mobile phone use: Overview of impacts and countermeasures. NTUA Road Safety Observatory. In Proceedings of the Communication Technologies and Road Safety Conference, Abu Dhabi. 2014.
19. Speed management. Paris, OECD/ECMT Transport Research Centre (JTRC), 2006
20. Andersson, Gunnar, and Göran Nilsson. Speed management in Sweden: Speed, speed limits and safety. Swedish National Road and Transport Research Institute, 1997.
21. Nilsson, G. (2004) Traffic safety dimensions and the power model to describe the effect of speed on safety. Bulletin 221, Lund Institute of Technology, Lund http://lup.lub.lu.se/record/21612
22. Johnson, Derick A., and Mohan M. Trivedi. Driving style recognition using a smartphone as a sensor platform. In Intelligent Transportation Systems (ITSC), 14th International IEEE Conference on, 2011 pp. 1609-1615. https://doi.org/10.1109/ITSC.2011.6083078
23. Bonsall, Peter, Ronghui Liu, and William Young. Modelling safety-related driving behaviour—impact of parameter values. Transportation Research Part A: Policy and Practice Vol. 39, No. 5, 2005, pp. 425-444. https://doi.org/10.1016/j.tra.2005.02.002
24. Cook, W. D., & Seiford, L. M. Data envelopment analysis (DEA)–Thirty years on. European journal of operational research, Vol. 192, No. 1, 2009, pp. 1-17. https://doi.org/10.1016/j.ejor.2008.01.032
25. Hollingsworth, B., Dawson, P. J., & Maniadakis, N. Efficiency measurement of health care: a review of non-parametric methods and applications. Health care management science, Vol 2, No. 3, 1999, pp. 161-172. https://doi.org/10.1023/A:1019087828488
26. Karlaftis, M.G., Gleason, J.M., Barnum, D.T. Bibliography of Urban Transit Data Envelopment Analysis (DEA) Publications. Available at SSRN: http://ssrn.com/abstract=1350583, http://dx.doi.org/10.2139/ssrn.1350583, 2013
27. Egilmez, G., & McAvoy, D. Benchmarking road safety of US states: A DEA-based Malmquist productivity index approach. Accident Analysis & Prevention, Vol. 53, 2013, pp. 55-64. https://doi.org/10.1016/j.aap.2012.12.038
28. Aarts, L., & Van Schagen, I. Driving speed and the risk of road crashes: A review. Accident Analysis & Prevention, Vol. 38, No. 2, 2006, pp. 215-224. https://doi.org/10.1016/j.aap.2005.07.004
29. Young, K. & Regan, M. (2007). Driver distraction: A review of the literature. In: I.J. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter & J.D. Irwin (Eds.). Distracted driving. Sydney, NSW: Australasian College of Road Safety. Pages 379-405.
30. R. Shone, Applications in Intermediate Microeconomics, Martin Robertson,

Oxford, 1981.

31. M.J. Farrell, The measurement of productive efficiency, J. Royal Statistical Society Series A Vol. 120, No. 3, 1957, pp. 253–281. https://doi.org/10.2307/2343100

32. Thanassoulis, Emmanuel. Introduction to the theory and application of data envelopment analysis. Dordrecht: Kluwer Academic Publishers, 2001.

33. Emrouznejad, A., Parker, B. R., & Tavares, G. Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. Socio-economic planning sciences, Vol. 42, No 3, 2008, pp. 151-157. https://doi.org/10.1016/j.seps.2007.07.002

34. Seiford, Lawrence M. A bibliography for data envelopment analysis (1978-1996). Annals of Operations Research Vol. 73, 1997, pp. 393-438. https://doi.org/10.1023/A:1018949800069

35. Martić, Milan, Marina Novaković, and Alenka Baggia. Data envelopment analysis-basic models and their Utilization. Organizacija Vol. 42, No. 2, 2009, pp. 37-43. https://doi.org/10.2478/v10051-009-0001-6

36. Charnes, A., Cooper, W. W., & Rhodes, E. Measuring the efficiency of decision making units. European journal of operational research, Vol. 2, No. 6, 1978, pp. 429-444. https://doi.org/10.1016/0377-2217(78)90138-8

37. Alper, D., Sinuany-Stern, Z., & Shinar, D. Evaluating the efficiency of local municipalities in providing traffic safety using the Data Envelopment Analysis. Accident Analysis & Prevention, Vol. 78, 2015, pp. 39-50. https://doi.org/10.1016/j.aap.2015.02.014

38. Ramanathan, Ramu, ed. An introduction to data envelopment analysis: a tool for performance measurement. Sage, 2003.

39. Shichrur, Rachel, Adi Sarid, and Navah Z. Ratzon. Determining the sampling time frame for in-vehicle data recorder measurement in assessing drivers. Transportation research part C: emerging technologies Vol. 42, 2014, pp. 99-106. https://doi.org/10.1016/j.trc.2014.02.017

40. Dyson RG, Allen R, Camanho AS, Podinovski VV, Sarrico CS, Shale EA. Pitfalls and protocols in DEA. European Journal of Operational Research, Vol. 132, Issue 2, 2001, pp. 245–259. https://doi.org/10.1016/S0377-2217(00)00149-1

41. Cooper WW, Seiford LM, Tone K. Data envelopment analysis: a comprehensive Text with models, applications, references and DEA-Solver Software, Springer Science & Business Media, 2006.