Investigation of Greek driver behavior during the approach to suburban un-signalized intersections

Constantinos Antoniou and Konstantinos Papoutsis

National Technical University of Athens
School of Rural and Surveying Engineering
Laboratory of Transportation Engineering

Abstract

Intersections are one of the key components of the road network, accounting for a disproportionally high ratio of delays and road safety incidents. The objective of this paper is to present a methodology for the investigation of driver behavior during the approach to an intersection and apply it to the case of Greek drivers approaching a suburban un-signalized intersection. The data collection uses two regular, consumer-grade video-cameras positioned at approximately 300m upstream of the intersection and immediately upstream of the intersection. Two models were finally developed using the sample of 250 collected observations: a complete model with all statistically significant parameters, as well as a simpler, parsimonious model. Several observations can be made, e.g. that drivers in the age group 25-65 reduce their speed by about 8km/h more than younger drivers, while drivers above 65 years of age reduce their speed by more than 11km/h. Similarly, compared to the base case of all other vehicles, drivers tend to reduce their speed by more than 8km/h more if the preceding vehicle is a truck, while they reduce their speed by about 3km/h less if the preceding vehicle is a motorcycle. The presented results confirm the prior expectations of the behavior of drivers approaching an un-signalized intersection and provide a quantification of the impact of the various factors, which could be utilized to steer public policy and road safety campaigns.

Keywords: speed drop, driver behavior, intersection approach, road safety

INTRODUCTION

Intersections are one of the key components of the road network, accounting for a disproportionally high ratio of delays and road safety incidents. The large number and variability of intersection and interchange types makes a general analysis impractical. A significant body of research has tackled issues related to driver behavior during the approach to both signalized and unsignalized intersections.

El-Shawarby et al. [1] study the deceleration rates of different drivers as they approach a high-speed signalized intersection during the yellow-phase transition. Gates et al. [2] evaluated the stopping dilemma faced by drivers as they approach signalized intersections at the beginning of the yellow phase. Drivers were more likely to violate a read light in the absence of side-street traffic and if nearby vehicles also did run the light. Furthermore, heavy vehicles were found more likely to violate a red light. Rakha et al. [3] also studied the same phenomenon and found that the driver perception-reaction times are influenced primarily by the driver's time to the intersection (a finding consistent with Gates et al. [2]). Yan et al. [4] evaluate the impact of a pavement marking measure that could improve traffic safety through the reduction of the phenomenon of driver's dilemma and find that the device can reduce the occurrence of conservative decisions to stop and risky decisions to cross.

Heterogeneity across drivers, e.g. in terms of aggressiveness, becomes particularly relevant when dealing with intersections. For example, Kaysi and Abbany [5] investigate aggressive driver behavior using a model that predicts the probability of performing an aggressive performance as a function of several driver and traffic attributes.

The objective of this paper is to present a methodology for the investigation of driver behavior during the approach to an intersection and apply it to the case of Greek drivers approaching a suburban un-signalized intersection. In particular, the speed well upstream of the intersection (i.e. before the deceleration has started) is used as a base to be compared with the speed at which the vehicles approach the intersection before crossing it. The remainder of the paper is organized as follows. The methodology is first presented, with an emphasis in the data collection and data analysis components. The model development and presentation of results is presented next, followed by a concluding section and recommendations.

METHODOLOGY

Two main methodological components are employed in this work: (i) data collection and (ii) statistical data analysis. Figure 1 outlines the main steps of the methodology. The data collection uses two regular, consumer-grade video-cameras positioned at approximately 300m upstream of the intersection and immediately upstream of the intersection. A pilot study was first performed, in order to resolve issues related to the data collection as well as gain experience and obtain the necessary background in order to be able to correctly design the main data collection experiment. The dependent variable used for the analysis is computed as the difference between the two recorded speeds, i.e. the drop of the speed of the vehicle due to the approach to the intersection. The footage from the cameras was analyzed (frame-by-frame) to compute the speed of the vehicles at the two locations, as well as collect a number of other variables including: type of vehicle, as well as preceding and following vehicles, headway with preceding and following vehicles, (approximate) age and sex of the drivers, number of adult passengers, number of underage passengers, number of pedestrians waiting to cross the intersection. Furthermore, the movement of the current vehicle (as well as the preceding and following vehicles) was also recorded, i.e. whether the vehicle was continuing straight across the intersection or turning left or right. The data were analyzed using linear regression, as statistical testing indicated that the underlying Gauss-Markov assumptions were satisfied and there was no need to seek a more advanced modeling technique.

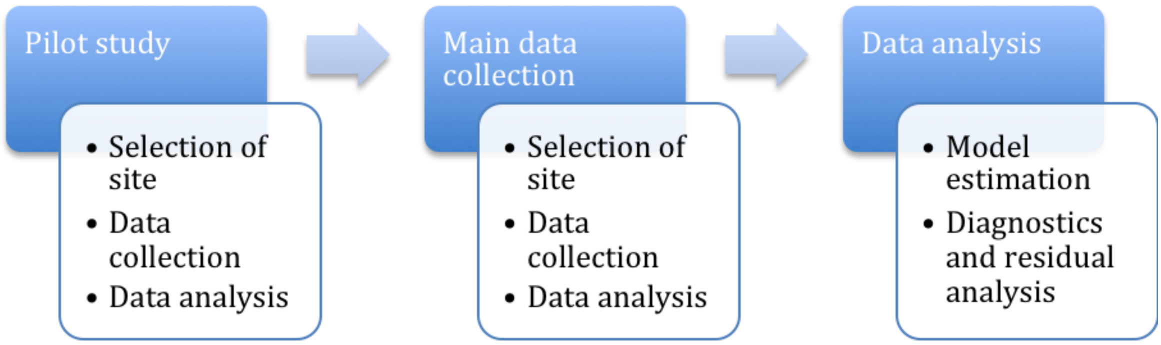


Figure 1. Main steps of the methodology

Data collection

Figure 2 provides an overview of the study area (top subfigure; the available map is somewhat outdated and shows an older version of the intersection, which has since been reconstructed), as well as views of the two data collection locations (bottom left subfigure: location 300m upstream of the intersection; bottom right subfigure: approach to the intersection). The data collection effort focused on the uncontrolled approach of the intersection and was performed in the morning and resulted in 250 usable observations. The following data were collected for each observation:

- Type of vehicle
- Speed 300m. upstream of the junction
- Speed at the junction approach

- Driver's age
- Driver's gender
- Number of adult passengers
- Children passengers
- Total pedestrians at intersection
- Children pedestrians at intersection
- Time headway with following and preceding vehicle (upstream of intersection and at intersection approach)
- Type of following and preceding vehicle (upstream of intersection and at intersection approach)
- Direction of vehicle (straight / turn left / turn right) at the intersection
- Direction of following and preceding vehicles (straight / turn left / turn right) at the intersection

Figure 2: Top: study area map (source: Google maps), bottom left: camera view for data collection upstream of the junction, bottom right: camera view for data collection at the junction

Driver's age and gender, as well as the number of passengers were determined through visual analysis of the video frames, resulting in the possibility of misclassifications. More information on the data collection effort is available in Papoutsis [6]. An analysis of the accuracy of the speed measurements using this frame-based method preceded the data-collection effort. For this analysis a car was driven at a known speed through the intersection a number of times and its speed was compared to that computed using the video camera footage. Through this process it was established that the accuracy of the speed measurements was satisfactory for this task.

Of course, there are many more parameters that affect the behavior of drivers, and many of them have not been recorded. This unobserved information could be partly responsible for some variation in the speed during the approach that cannot be attributed to the collected data and therefore explained by the developed models.

Figure 3 presents some indicative data: (i) speed drop for the vehicles approaching the intersection (speed of the vehicle immediately upstream of the intersection minus the speed 300m upstream of the intersection) and (ii) time headway 300m upstream of the intersection. These data have reasonable distributions. However, some variables presented less desirable properties. For example, the sample was heavily biased towards male drivers (88% of the drivers), while a pedestrian was present only at in four out of 250 observations of the observations. The former issue could raise concerns for bias in the results; however, if the actual population of drivers indeed is skewed in this respect, then corrective actions (such as enriched sampling) could actually adversely bias the results away from the truth. The latter issue (low presence of pedestrians) essentially precludes the use of this variable in the models. The left subfigure of Figure 3 indicates that a fraction of the drivers actually increased their speeds during the approach to the intersection. This observation needs to be further investigated, so that the underlying reasons (and contributing factors) for this can be better understood.

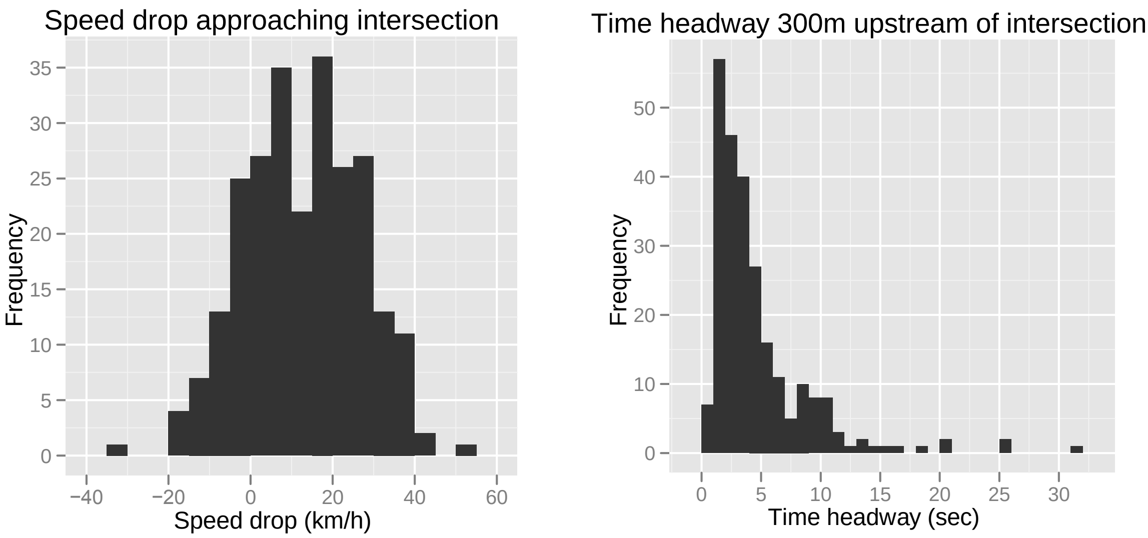


Figure 3. Collected data sample (left: speed drop for the vehicles approaching the intersection; right: time headway 300m upstream of the intersection)

Data analysis

The linear regression model is simple (to run and interpret), elegant and efficient, it is subject to the fairly stringent Gauss-Markov assumptions [7]. If these assumptions hold, it can be shown that the solution obtained by minimizing the sum of squared residuals ('least squares') is BLUE, i.e. best linear unbiased estimator. In other words, it is unbiased and has the lowest total variance among all unbiased linear estimators.

In particular, the basic Gauss-Markov assumptions require:

- Linearity (in the parameters; nonlinearity in the variables is acceptable);
- Homoscedasticity;
- Exogenous independent variables;
- Uncorrelated disturbances; and
- Normally distributed disturbances

These assumptions, however, are often violated in practice. Therefore, it is important to verify them through the use of the appropriate diagnostics and –if they are violated- use a different appropriate model. Generalized linear models (GLM), a generalization of the linear regression, can be used to overcome the restriction on the normality of the error structure [8-10]. The objective of GLM is to allow for more flexible error structures, besides the Gaussian which is assumed by –linear and nonlinear– regression. In order to provide flexibility, the GLM framework was adopted in this study. However, the analysis of the residuals and other diagnostics of the models suggested that the Gaussian distribution adequately models the collected data and therefore the investigation of other model families was not required.

RESULTS

Two models were finally developed using the sample of 250 collected observations: a complete model with all statistically significant parameters, as well as a simpler, parsimonious model, with the main parameters only. The model estimation results are presented in Table 1, while the estimated values of the model parameters are visualized in Figure 4, along with the ranges indicated by the 90% (thicker lines) and 95% (thinner lines) confidence intervals. Considering that the dependent variable is the reduction in speed due to the approach to the intersection (speed well upstream of the intersection minus speed at the approach) and that it is therefore expected to have a positive value, the interpretation of the coefficient estimates is straightforward. For example, the intercept indicates that without considering other parameters, Model 1 shows an average/overall speed reduction of 12.2km/h (for Model 2 this number falls to -9.5km/h, perhaps due to the fact that the additional parameters explain part of the reduction).

The interpretation of data coded as factors (such as the age group of the driver) has to be made relative to the base group. In this application, using younger drivers (ages 18-25) as the base level (i.e. assuming a zero speed reduction due to their age for drivers in this group), drivers in the age group 25-65 reduce their speed by about 8km/h more, while drivers above 65 years of age reduce their speed by more than 11km/h. Similarly, compared to the base case of all other vehicles, drivers tend to reduce their speed by more than 8km/h more if the preceding vehicle is a truck, while they reduce their speed by about 3km/h less if the preceding vehicle is a motorcycle.

The interpretation of the coefficients of the last group of parameters of Model 2 (those related to headway from the previous vehicle) is not as direct, but still straightforward. In each case, the estimated coefficient needs to be multiplied by the headway. For example, if the headway to the previous vehicle upstream of the intersection was 5 seconds, then the estimated coefficient (0.248) should be multiplied by that number so that it would be converted to km/h.

Table 1. Model estimation results (simple and full models)

	Model 1		Model 2	
	Estimate	**t-value**	**Estimate**	**t-value**
Intercept	-12.22	-2.761	-9.45	-2.035
Driver characteristics				
Driver's age (25-45)	7.88	2.745	7.74	2.691
Driver's age (45-65)	8.29	2.837	8.31	2.848
Driver's age (65+)	12.9	3.549	11.86	3.29
Vehicle characteristics				
Preceding vehicle is truck			8.44	2.133
Preceding vehicle is motorcycle			-3.31	-1.616
Direction of traffic				
Preceding vehicle moving straight	-3.12	-2.195	-3.61	-2.552
Vehicle turning right	-5.09	-2.114	-5.97	-2.449
Vehicle moving straight	-15.83	-8.497	-16.66	-8.831
Traffic characteristics				
Speed upstream of the intersection (km/h)	0.73	12.698	0.702	11.885
Headway from previous vehicle upstream of the intersection (sec)			0.248	1.595
Headway from previous vehicle at the intersection (sec)			-0.286	-1.776
Number of observations	250		250	
Null deviance	49119	(249 d.o.f.)	49119	(249 d.o.f.)
Residual deviance	21241	(242 d.o.f.)	20314	(238 d.o.f.)
AIC	1838.0		1834.9	

d.o.f.: degrees of freedom

AIC: Akaike Information Criterion

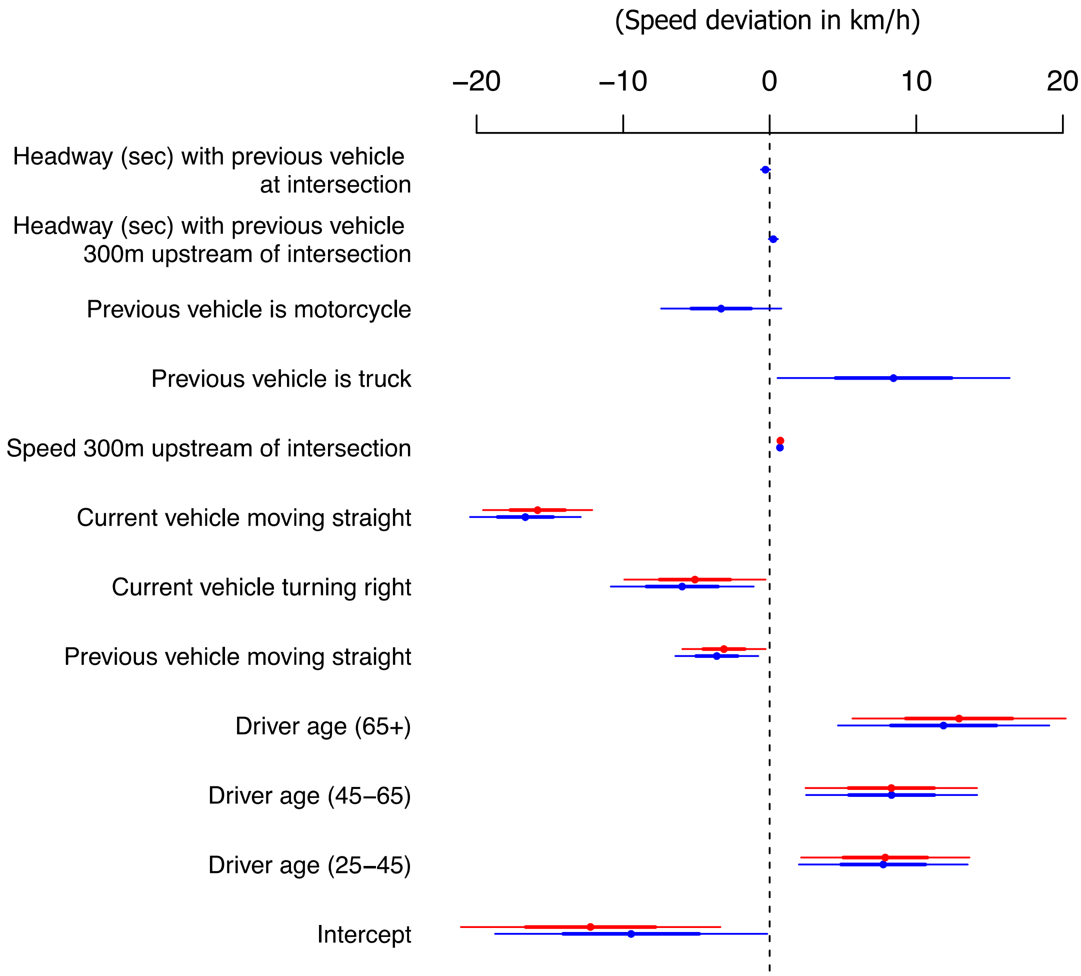


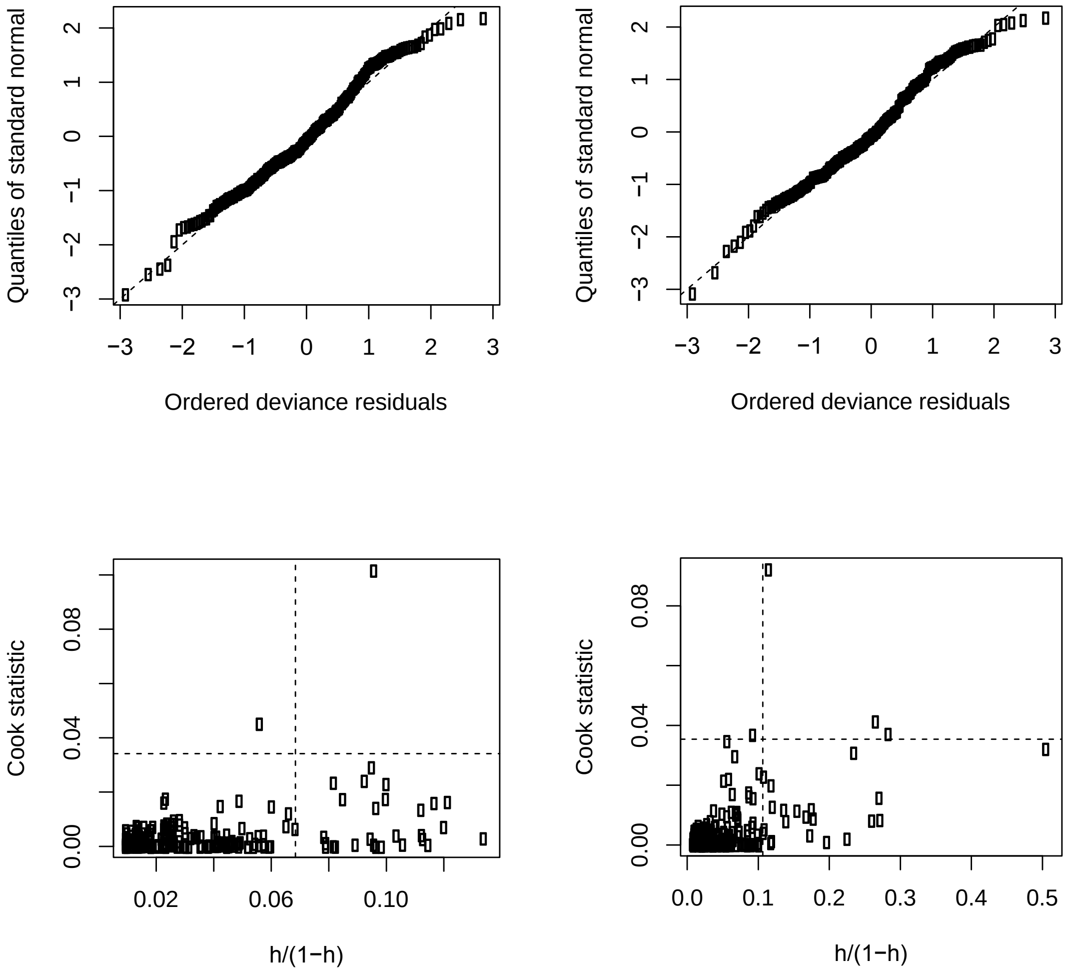
Figure 4. Visual inspection of estimated coefficients for both models (expressed as speed deviation in km/h).

In order to confirm that the normality assumption of the residuals (and the other Gauss-Markov assumptions) were satisfied by the model, a number of diagnostic tests were performed, some of which are presented in Figure 5 (left: model 1; right: model 2). Normal scores plots (QQ plot) of standardized deviance residuals are presented in the top subfigure of each figure. The x-axis represents the standardized deviance residuals, while the y-axis represents the quantiles of the standard normal. The dotted line in the QQ plot (top) is the expected line if the standardized residuals are normally distributed, i.e. it is the line with intercept 0 and slope 1. Indeed, the residuals of both models are essentially located along the desired line.

A plot of the Cook statistics against the standardized leverages is provided in the bottom of each subfigure. The standardized leverage of the i-th observation x_i can be computed as [11]:

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x}_i)}{(n-1)s_x^2} \qquad (1)$$

where n is the number of observations, the overbar indicates the predicted value, and s_x is the standard error. There are two dotted lines on each plot. The horizontal line is at 8/(n-2p) where n is the number of observations and p is the number of parameters estimated. Points above this line may be points with high influence on the model. The vertical line is at 2p/(n-2p) and points to the right of this line have high leverage compared to the variance of the raw residual at that point. Again, most points are to the bottom and left of the dotted lines; having said that, several leverage points exist.



Model 1 Model 2

Figure 5. Model fit diagnostic data

CONCLUSION

Two models for the speed drop in the approach to a suburban, unsignalized intersection in Greece, are presented. However, no "best" model is identified. The use of the appropriate model may be dependent on the purpose of the application, as e.g. the complete model may provide a richer insight into the underlying parameters influencing the behavior of the drivers (and therefore could be used to e.g. compare different populations). On the other hand, the more parsimonious model might be better suited to inference and or prediction applications, as it requires fewer data.

The presented results confirm the prior expectations of the behavior of drivers approaching an un-signalized intersection and provide a quantification of the (already qualitatively understood) impact of the various factors. The quantification of the results could be utilized to target public policy and road safety campaigns; for example focusing on the fact that drivers apparently do not pay as much attention to motorcycles (on the contrary they understandably seem rather aware of trucks), or towards younger drivers, who apparently reduce their speed less during the approach to an intersection. When the youngest drivers (18-25) are used as the reference level, drivers older than 65 years appear to be more conservative during the approach to an intersection, followed by drivers belonging in the age groups 25-45 and 45-65. The direction in which the considered vehicle, as well as the preceding vehicle, are moving through the intersection also influences the speed reduction during the intersection approach. The interaction between the vehicle of interest and surrounding vehicles is a potentially decisive one, and one that deserves further elaboration.

While the presented research provides some insight into the problem, there are several enhancements that could be foreseen. For example, in this research only two points are used for the speed measurement. Future research could utilize more elaborate surveillance equipment that could allow the tracking of the complete speed trajectory of the vehicles during their approach to the intersection, thus providing valuable additional insight into the finer behavioral decisions that drivers undertake. Furthermore, the presented results are tied to the considered population of drivers and a single intersection. Further research in additional intersections could provide significant information that could strengthen the confidence in the findings of this work.

Acknowledgement

The authors would like to thank Prof. Andreas Georgopoulos for the help and support during the data collection used in this research.

References

1. El-Shawarby, I. H.A. Rakha, V.W. Inman and G.W. Davis (2007). Evaluation of Driver Deceleration Behavior at Signalized Intersections. Transportation Research Record: Journal of the Transportation Research Board, No. 2018, pp. 29-35.
2. Gates, T., D. Noyce, L. Laracuente and E. Nordheim (2007). Analysis of Driver Behavior in Dilemma Zones at Signalized Intersections. Transportation Research Record: Journal of the Transportation Research Board, No. 2030, pp. 29-39.
3. Rakha, H.A., A.M.M. Amer and I. El-Shawarby (2008). Modeling Driver Behavior Within a Signalized Intersection Approach Decision-Dilemma Zone. Transportation Research Record: Journal of the Transportation Research Board, No. 2069, pp. 16-25.
4. Yan, X., E. Radwan, D. Guo and S. Richards (2009). Impact of "Signal Ahead" pavement marking on driver behavior at signalized intersections. Transportation Research Part F Traffic Psychology and Behaviour, Volume: 12, Issue: 1, pp. 50-67.
5. Kaysi, I.A. and A.S. Abbany (2007). Modeling aggressive driver behavior at unsignalized intersections. Accident Analysis & Prevention, Volume 39, Issue 4, July 2007, pp. 671-678.
6. Papoutsis, K. (2010). Investigation of driver behavior during the approach to un-signalized intersections. Diploma thesis, Laboratory of Transportation Engineering, Department of Infrastructure and Rural Development, School of Rural and Surveying Engineering, National Technical University of Athens, Greece (in Greek).
7. Washington, S. P., Karlaftis, M. G., and F. L. Mannering (2003). Statistical and Econometric Models for Transportation Data Analysis. Chapman & Hall/CRC.
8. Dobson, A.J. (1990), *An Introduction to Generalized Linear Models*. Second edition, Chapman and Hall, London.
9. Gill, J. (2000), *Generalized Linear Models: A Unified Approach*, Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-134, Thousand Oaks, CA: Sage.
10. McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models*. Second edition. Chapman Hall, New York.
11. Belsley, D., Kuh, E., and Welsch, R. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley and Sons, New York.