A mixed logit model for the sensitivity analysis of Greek drivers' behaviour towards enforcement for road safety

George Yannis[1] and Constantinos Antoniou

National Technical University of Athens,
Department of Transportation Planning and Engineering,
5 Iroon Polytechniou Street, 157 73 Zografou, Athens, Greece

ABSTRACT

Traffic violations are among the leading causes of road accidents. In this research, the sensitivity of Greek drivers to a hypothetical intensification of police enforcement for speed violations and improper overtaking is analyzed, using stated preference data. Under the assumption of increased police enforcement, drivers were presented with the option to maintain their unsafe driving patterns (and risk getting fined) or comply with the traffic laws (and experience longer trip duration). A parsimonious mixed logit model has been estimated and sensitivity analysis is performed with respect to the main variables. The model explicitly captures the (unobserved) heterogeneity in the sample, and reflects the fixed random parameter across observations from the same respondent. The behaviour of the surveyed drivers depends on socioeconomic characteristics and trip characteristics. Based on the presented sensitivity analysis, it can be argued that while the "typical" Greek driver may not be particularly risk-prone, there are segments of the population that show a tendency to violate traffic laws. This is a useful finding that could be used by policy makers e.g. to develop targeted police enforcement campaigns (or targeted media campaigns, special education initiatives, etc.), aimed at the demographic segments with a higher tendency for traffic violation.

Keywords: road safety, police enforcement, mixed logit model, panel data, stated-preference

[1] Corresponding author. Tel.: +30.210.7721326, Fax.: +30.210.7721454
Email addresses: geyannis@central.ntua.gr (G. Yannis), antoniou@central.ntua.gr (C. Antoniou)

1. INTRODUCTION

Road safety is one of the most important issues throughout the world. For example, according to the European Commission CARE database for 2002, the number of fatalities from more than 1,250,000 road accidents in the (then fifteen) European countries was 38,637, while another 1,700,000 were injured. In its White Paper on European transport policy (EC, 2001) the Commission has therefore proposed that the European Union (EU) should set itself the target of halving the number of road deaths by 2010 (a proposal that has since been adopted).

Research in the field of road safety has shown that traffic violations constitute one of the most important factors of road accidents. Specifically, it has been shown (Rothengatter and Harper, 1991) that a large proportion of road accidents inside the EU is the result of one or more traffic violations. In the same study it is concluded that the cumulative traffic violations is the most important factor leading to road accidents. Furthermore, according to the European Council of Road Safety (ETSC, 1999), it is calculated that significant improvement in road safety (as high as 50%) could be achieved through measures for the prevention of traffic violations (such as intensification of police enforcement). It is worth mentioning that while drivers commit traffic violations, they believe that police enforcement of road networks should be intensified to the benefit of road safety. According to the SARTRE EU research project, up to 70% of the drivers believe that police enforcement should be intensified in order for traffic violations to be reduced (SARTRE, 1998).

As it has been suggested from various researchers (e.g. Bjornskau and Elvik, 1992, Zaal, 1994, Newstead, Cameron, Mark and Leggett, 2001, Tay, 2005), one way of reducing the number of traffic violations is the intensification of police enforcement. This is supported by empirical results reported by Holland and Conner (1995).

It is clear from the literature that intensification of police enforcement is expected to result in an improvement in road safety. The goal of this paper is to analyze the sensitivity of Greek drivers towards the intensification of police enforcement (targeted primarily at speed limit violations and illegal overtaking). Stated-preference data are collected from a specially designed questionnaire. The collected data includes socioeconomic data, as well as the surveyed drivers' response to a number of hypothetical scenarios. For each scenario, the driver is asked to choose between two alternatives, with different attributes, such as police enforcement intensity, probability of getting fined (and/or being involved in an accident with injury) and trip duration. (For an overview of road safety in Greece, and analysis of factors affecting road safety in Greece c.f. e.g. Matsoukis et al., 1996, Golias et al., 1997, Kanellaidis et al., 1999). Policy makers as well as road safety practitioners could benefit from this research, as they can better support their choices and decisions through the use of the proposed methodologies.

The remainder of this paper is structured as follows. An overview of the techniques that are used in this paper (namely stated-preference surveys and discrete choice analysis using mixed-logit models) is presented in Section 2. The survey design and data collection procedure is described in Section 3. Model specification and estimation results are presented in Section 4, while sensitivity analysis using the estimated model coefficients is presented in Section 5. Conclusions and directions for further research are outlined in Section 6.

2. BACKGROUND

2.1 Stated-Preference Techniques

Stated preference techniques are an attractive tool for researching non-existing situations (Louviere et al., 2000). The analysis of stated preference data originated in mathematical psychology with the seminal paper by Luce and Tukey (1964). Stated preference methods were further developed in marketing research in the early seventies and over several decades have had several applications. Stated preference techniques have been used, for example, to examine the effect of travel information on mode choice (Abdel-Aty et al., 1997, Khattak et al., 1996, Polydoropoulou et al., 1996). More recently, Rizzi and Ortuzar (2003) used stated preference techniques in the context of road-safety for the estimation of the value of attributes (travel time, toll and annual accident rate) for the valuation of road accident fatalities.

Stated-preference techniques have also been used specifically for the assessment of drivers' preference with respect to enforcement. Yannis et al. (2005) investigated the behavioral parameters that influence drivers' choices in order to reduce accident risk, using stated-preference techniques and logistic regression models. SARTRE (2004) describes the third wave of a large-scale stated-preference survey across Europe, that was collected data in various aspects of road-safety, including perception and response to enforcement. Kanellaidis et al. (1999) used stated-preference surveys to assess the attitude of Greek drivers towards road safety.

The primary drawback to stated preference data is that they may not be congruent with actual behaviour (for example due to biases). This phenomenon can be critical under certain circumstances, when for example the results are not verified with results from the literature, or revealed-preference data. Additionally, particular attention should be given to the results' interpretation, because respondents show the tendency to exaggerate when they conceive that they take part in some experiment (Lin et al., 1986, van der Hoorn et al., 1984).

2.2 The Mixed-Logit Model for Panel Data

Discrete choice analysis is a well established approach for analyzing individual behavior (Ben-Akiva and Lerman, 1985). In the case of repeated observations (such as the case of stated-preference surveys with multiple responses) one often needs to capture the correlation across observations from the same individual. In general, pooling data across individuals while ignoring correlation across observations and unobserved heterogeneity among responses from different individuals (when it is present) will lead to biased and inconsistent estimates of the effects of pertinent variables (Hsiao, 1986). Several approaches have been developed to incorporate these effects in the model formulation. One is to estimate a constant term for each individual and each choice, which is referred to as a "fixed-effects" approach (Chamberlain, 1980). Perhaps the main drawback to this approach is the large number of parameters (and consequently large number of required observations per individual). A more tractable approach is to replace the fixed term with some probability distribution, which is referred to as a random effects specification (Heckman, 1981, Hsiao, 1986). The most common assumptions for this distribution are the normal and the lognormal. One drawback to this approach, however, is that it does not allow for a closed-form expression for the choice probabilities, thus leading to numerical complications, which will be detailed below.

Mixed logit is a highly flexible model that can approximate any random utility model (McFadden and Train, 2000). It obviates the three limitations of standard logit by allowing for random taste variation, unrestricted substitution patterns, and correlation in unobserved factors over time (e.g. in the case that data from the same individuals are collected at different times). Unlike probit, it is not restricted to normal distributions. Its derivation is straightforward, and simulation of its choice probabilities is computationally simple. Like probit, the mixed logit model has been known for many years but has only become fully applicable since the advent of simulation (Train, 2003). Some indicative applications from the literature follow. Han et al. (2001) develop a mixed logit model to accommodate the random heterogeneity across drivers and to cope with the correlation between repeated choices. Hess et al. (2004) use mixed logit models that allow for random taste heterogeneity for the computation of value-of-time. Bierlaire et al. (2006) present a mixed binary logit model with panel data to analyze the drivers' decisions when traffic information is provided during their trip.

3. SURVEY DESIGN AND DATA COLLECTION

The necessary data were collected through a stated preference survey using a specially developed questionnaire. The final sample comprises 251 questionnaires that were completed by drivers familiar with driving in the Greek national road network. The majority of the respondents were from the city of Halkida, where the data-collection took place. Halkida has a population of approximately 100,000 and is located 85 km northeast of Athens, Greece. The surveys were administered at road-side rest areas along the national freeway connecting Halkida with Athens. One potential impact of the survey execution is that interviewed drivers were likely to perceive accident risk related to highway travel, which in principle may be somewhat different than that related to non-highway trips.

The days and hours for the administration of the field survey were chosen so that the sample covers a wide spectrum of driver characteristics (e.g. in terms of age and education). In order to ensure that the sample would be representative and unbiased, further sampling approaches were used. For example only a subset of the drivers stopping at the rest area (1 out of 7) was randomly interviewed to avoid correlation issues. Before starting the questions, the interviewer presented the framework of the survey and made as clear as possible to the interviewees the meaning of the options proposed in the questionnaire; e.g. the current accident risk level was explained allowing to better understand what a 20% risk decrease means.

The questionnaire consisted of three parts and can be completed in approximately five minutes. The first part questions collected demographic characteristics of the subject, such as gender, age, residence area, educational level, occupation and annual income. The second part questions aimed to expose the subject to the road safety problem, and in particular the probability of being involved in an accident. In the third part, a subset of which is shown in Table 1, each respondent was asked about the duration of their usual highway trip, and was subsequently presented with scenarios for this duration. The scenarios were based on the assumption that the current trip duration is shorter than it should, since lack of continuous and effective police enforcement allows for speed violations and illegal overtaking.

[INSERT TABLE 1 ABOUT HERE]

If police enforcement was intensified, then the respondents would face the following dilemma: either continue to violate traffic laws and get fined (or get involved in a traffic accident with injury) or comply with traffic laws, not get fined but be subjected to increased trip duration. The assumption is made that in the hypothetical scenario of intensified police enforcement all traffic violations will be recorded and all violators will have to pay a fine (which for the purposes of this research was set at €120 or approximately US$150). The proposed options take into account the fact that higher compliance leads to both longer trip duration and lower risk.

4. MODEL ESTIMATION

Most of the data has been coded as categorical variables. While the order in which the levels are coded (e.g. ascending age groups) follow some logic, assuming that the behavioural patterns of the individuals would follow the same trend is overly restrictive. For example, using a single variable Age in the model, resulting in the estimation of a single coefficient for Age, would imply that the impact of age is a linear function of the age group. To overcome this issue, dummy variables for each level have been introduced. Naturally, for a categorical variable with m levels, only m-1 dummy variables can be defined (while the remaining level serves as the base).

Not all levels of all categorical variables have a significant contribution to the model, however. Therefore, based on formal statistical significance tests, some levels have been grouped together. For example, the two lower levels of the Age factor (18-24 and 25-34) have been grouped together to provide a single level (18-34), which is used as a base for the age groups. Similarly, the two highest levels of the Age factor (55-64 and 65+) have also been grouped together. The specification table for the binary and mixed logit models is shown in Table 2. The available data (including the alternative specific constant) have been used to construct the utility function for the option that the users would be likely to comply with the increased police enforcement and experience a higher travel time (and reduction of their probability of getting involved in an accident with injury) in order not to be fined. As no variables are used for the specification of the alternative option (choosing to violate the speed limit and/or perform illegal overtaking manoeuvres at the risk of getting fined), the utility function of that alternative option is constant and equal to zero. Since only the difference in utilities can be captured in discrete choice models, using one alternative as a reference case in this way does not affect the estimation of the model. More formally, the systematic utility specification for the two alternatives can be expressed as:

$$\begin{aligned} V_{\text{Compliance, ij}} = \; & \beta_{\text{Compliance}} * 1 + \beta_{\text{Age 35-44}} * X_{\text{age dummy (35-44),ij}} + \\ & + \beta_{\text{Age 45-54}} * X_{\text{age dummy (45-54),ij}} + \beta_{\text{Age 55+}} * X_{\text{age dummy (55+),ij}} + \\ & + \beta_{\text{Low income}} * X_{\text{low income dummy,ij}} + \beta_{\text{High income}} * X_{\text{high income dummy,ij}} + \\ & + \beta_{\text{Low education}} * X_{\text{low education dummy,ij}} + \beta_{\text{Trip duration}} * X_{\text{trip duration,ij}} + \\ & + \beta_{\text{Trip duration increase}} * X_{\text{trip duration increase,ij}} + \beta_{\text{Risk change}} * X_{\text{Risk change,ij}} \end{aligned}$$

$$V_{\text{Non-compliance}} = 0$$

The following notation is used in the utility specification:

- $X_{\text{age dummy (35-44)}}$ Binary dummy variable, taking the value 1 if the age of the individual is between 35 and 44, and 0 otherwise
- $X_{\text{age dummy (45-54)}}$ Binary dummy variable, taking the value 1 if the age of the individual is between 45 and 54, and 0 otherwise
- $X_{\text{age dummy (55+)}}$ Binary dummy variable, taking the value 1 if the age of the individual is above 55, and 0 otherwise
- $X_{\text{low income dummy}}$ Binary dummy variable, taking the value 1 if the income of the individual is low, and 0 otherwise
- $X_{\text{high income dummy}}$ Binary dummy variable, taking the value 1 if the income of the individual is high, and 0 otherwise
- $X_{\text{low education dummy}}$ Binary dummy variable, taking the value 1 if the education level of the individual is low, and 0 otherwise

- $X_{trip\ duration}$ Numerical explanatory variable, taking a value equal to the trip duration of the individual (in minutes)
- $X_{trip\ duration\ increase}$ Numerical explanatory variable, taking a value equal to the trip duration increase of the individual (in minutes)
- $X_{Risk\ change}$ Numerical explanatory variable, taking a value equal to the percentage of assumed risk change (e.g. if risk change is 20% then this variable is equal to 20).

The utility specification of the binary logit model is given by:

$$U^{binary}_{compliance,ij} = V_{compliance,ij} + \varepsilon_{ij}$$

where ε_{ij} is a zero-mean, random error term that is iid (independently and identically distributed) extreme value.

A random error term has been added in the utility specification of the mixed logit model to account for the presence of serially correlated repeated responses from the same respondent (panel data):

$$U^{mixed}_{compliance,ij} = V_{compliance,ij} + \sigma_{panel}\xi_i + \varepsilon_{ij}$$

where σ_{panel} is an unknown parameter to be estimated, and ξ_i is a standardized normal random parameter $\xi_i \sim N(0,1)$. In the field of transport economics this is often referred to as a compound error term.

[INSERT TABLE 2 ABOUT HERE]

The probability that an individual i chooses the first alternative (comply with the increased enforcement) in experiment j is given by:

$$P_i^{binary}\left(compliance \mid \{compliance, non-compliance\}\right) = \frac{e^{V_{compliance,ij}}}{e^{V_{compliance,ij}} + e^{V_{non-compliance,ij}}}$$

while for the mixed logit model, the same probability is given by:

$$P_i^{mixed}\left(compliance \mid \{compliance, non-compliance\}\right) = \int_{\xi_i} \prod_j \frac{e^{V_{compliance,ij} + \sigma_{panel}\xi_i}}{e^{V_{compliance,ij} + \sigma_{panel}\xi_i} + e^{V_{non-compliance,ij}}} f(\xi_i) d\xi_i$$

where the product ranges over all experiments j of individual i, σ_{panel} is an unknown parameter to be estimated, and ξ_i is a standardized normal random parameter $\xi_i \sim N(0,1)$, so that

$$f(\xi_i) = \frac{1}{\sqrt{2\pi}} e^{-\xi_i^2/2}$$

Since the sum of the probabilities to choose all alternatives equals to one, the probability that the second (base) alternative is chosen can easily be obtained by subtracting the probability that the first alternative is chosen from one.

The model estimation was performed using the Biogeme software package (Bierlaire, 2003, 2005). Binary logit and mixed-logit models were estimated. The mixed-logit specification differs from the logit in the addition of a zero-mean, normally distributed random component, capturing the unobserved heterogeneity between individuals. The normality assumption for the random component is commonly found in the literature. In the absence of strong evidence suggesting a different distributional assumption (e.g. log-normal), it has been used in this research. Furthermore, the correlation between choices made by the same respondent is explicitly incorporated by recognizing that responses from the same individual are correlated. This is taken into account by estimating the same value of the random parameter for all observations by the same respondent. Unlike logit, mixed-logit model estimation requires simulation, which can be based e.g. on pseudo-random numbers and draws from a Halton sequence (Train, 2003, Sandor and Train, 2004, Sivakumar et al., 2005). For the mixed-logit estimates, draws from a Halton sequence have been used instead of pseudo-random numbers, as they are more efficiently spread over the unit interval. Two hundred draws have been used, which empirically was found to be an adequate number (estimated coefficients had already stabilized well below one hundred Halton draws).

The estimation results for the final models are shown in Table 3. For each parameter, the estimated coefficient value and the robust t-test value are provided. The robust statistics allow for non-severe misspecification errors related with the characteristics of the postulated distributions of the error terms (Bierlaire et al., 2005). For example, the use of robust t-tests alleviates the potential impact of a non-severe misspecification due to the choice of the normal distribution for the random component. Aggregate goodness of fit measures (testing the adequacy of the entire model specification) are presented first (Bierlaire, 2005, Washington et al., 2003). At the individual coefficient level, both informal tests (sign and magnitude of coefficient estimates) as well as more formal tests (robust t-test) have been performed. A parsimonious model specification has been sought.

Summary goodness of fit statistics indicate that the mixed logit model provides a superior fit (a lower final log-likelihood, a higher ρ^2 and corrected ρ^2, only at the expense of a single additional estimated parameter, i.e. the random coefficient). The random coefficient is also very significant, which suggests that including it in the model specification was appropriate. The magnitude of the coefficients (which is higher for the mixed logit model) also suggests that this model more accurately captured the drivers' behaviour.

Most of the coefficients are significant at the p=5% level. The coefficients for the income variables and the low education variable are significant at the p=10% level. These coefficients have been retained in the model since the informal specification tests (sign and relative magnitude) indicate that these coefficients have been estimated according to prior expectations and would therefore provide intuitive results for the sensitivity analysis.

The inclusion of the variables in the final model implies that the behaviour of the surveyed drivers in relation to an intensification of enforcement that would result in a travel time increase and risk reduction (for complying drivers), or a monetary fine (for non-complying drivers) depends on the following variables:

- Age group
- Income level
- Education level
- Trip duration (prior to intensification of enforcement)
- Trip time increase (due to intensification of enforcement)
- Risk change (due to intensification of enforcement)

A discussion of the sign and magnitude of the estimated coefficients (also called informal specification tests) is presented next. It is important that signs and relative magnitudes of estimated coefficients agree with a priori expectations (Ben-Akiva and Lerman, 1985, pp. 157-160). The positive alternative specific constant suggests that there is some a priori tendency of the drivers to choose the conservative option of compliance to the increased police enforcement and not risk getting fined to decrease travel time. This finding is intuitive and consistent with expectations.

Age has been incorporated into the model as three coefficients (using age group 18-34 as the base). All estimated parameters are positive, indicating that the younger base group is less likely to comply with police enforcement and thus implicitly more risk-prone, which is again consistent with expectations. With the exception of the age group 45-54, compliance increases with age. Drivers in the age group 45-54 show the second highest tendency towards non-compliance. It is useful at this point to revisit the choice of individual dummy variables for each age group, instead of using a single ordinal variable (with values e.g. 1 through 4, where 1 would be the age group 18-34 and 4 would be 55+). Such a model might give a significant coefficient (and indeed it does, capturing the overall trend of increasing risk aversion with age) but would miss the fact that the drivers in the age group 45-54 actually appear to exhibit more risk-prone behaviour than drivers in the age group 35-44. This misspecification would have significantly altered the sensitivity analysis results presented in the following sections (and hence the conclusions drawn from this research).

Income has been modeled using two dummy variables (one for low and one for high income) with medium income serving as the base. The coefficient for the low-income dummy variable has a positive sign, indicating that drivers with low income are more likely to comply and not risk paying a fine. High-income drivers, on the other hand, are more likely to risk non-compliance. This is an intuitive finding, since the cost of the fine (~□120 or ~US$150 according to the survey setup) is less of a disincentive for higher-income drivers. Similarly to the Age variable, modeling income levels as an ordinal variable with three levels would have resulted in erroneous results: even though the order of the levels is retained, the magnitude of the coefficients varies.

Education has also been modeled through two dummy variables (low and medium), while high education is used as the base. Note that low education in this level is the two first levels of education (elementary and junior/high-school) combined. However, only the "low education" variable has been retained in the model, as the coefficients for the others were not statistically significant. The negative estimated coefficient indicates that drivers with low education are less likely to comply with the increased police enforcement, thus risking to get fined and/or involved in an accident.

The duration of the trip is coded as a continuous variable (even though it only takes two values, i.e. 120min and 300min). A positive coefficient suggests that drivers are getting more compliant as the duration of the trip increases. This may be due to the fact that drivers associate longer trips with a higher probability of actually getting caught (or getting involved in an accident).

This issue deserves deeper investigation; ideally a follow-up study would include a more detailed mapping of the variation of the trip duration.

The coefficient associated with the increase in the duration of the trip is negative, implying that the drivers' tendency to comply with the intensified enforcement (and avoid getting fined and/or hurt) is inversely proportional to the additional travel time in which it would result. Furthermore, the coefficient associated with the risk change is positive, as expected.

Finally, the random error term capturing the intra-personal correlation in the responses is significant, thus confirming that the mixed logit model is effective in capturing the unobserved heterogeneity among respondents. The sign of this parameter is not relevant.

5. SENSITIVITY ANALYSIS

The model application in the previous section provided some insight into the behaviour of the driver population in response to compliance with police enforcement intensification. A sensitivity analysis with respect to some of the main variables is presented in this section. This analysis is indicative of the breadth of similar analyses that can be performed using such models to illustrate the modeled behavior of the sampled population. Such analyses can be used to develop policies and strategies, estimating the potential impact of alternative scenarios.

In all cases, the dependent variable is the proportion of non-compliant drivers. Figure 1 presents the sensitivity with respect to age group for trips with a duration of 2 hours, a risk reduction of 10%, medium income level and high education. Both the ranking of the curves, and their concavity, are consistent with expectations. Furthermore, it is important to note that the proportion of drivers choosing to not alter their driving patterns in response to the hypothetical enforcement intensification and thus risking to be fined does not increase strictly with age group, as the age group 45-54 is less likely to comply to increased enforcement than the age group 35-44.

[INSERT FIGURE 1 ABOUT HERE]

Figure 2 presents the sensitivity with respect to income level for trips with duration of 2 hours, a risk reduction of 10%, and highly educated drivers in the 35-44 age group. Not surprisingly, drivers with high income are less likely to comply as the financial cost of getting fined affects them less.

[INSERT FIGURE 2 ABOUT HERE]

Figure 3 presents the sensitivity with respect to increase in trip duration for a risk reduction of 10% and young drivers (i.e. the 18-34 age group) with medium education and income level. For travel time increase up to almost 60% of the original trip duration, young drivers in shorter trips (2 hours) would be less likely to comply than for longer trips (5 hours). It should be stressed, however, that the x-axis in Figure 3 corresponds to percent increase in travel time with respect to the original trip duration. This means that a 30% increase for a 2-hour trip is 36 minutes, while a 30% increase for a 5-hour trip is 90 minutes. The difference in the trip duration may also be the primary reason for the lower curvature of the shorter trip curve, while the curve for the longer trip is clearly concave.

[INSERT FIGURE 3 ABOUT HERE]

6. CONCLUSION

A parsimonious mixed logit model that captures the unobserved heterogeneity between individual respondents thus modeling the correlation among several responses from the same individual (panel data) has been developed. Model parameters were estimated using results from a questionnaire-based survey among highway drivers in Greece. Based on the presented analysis, the behaviour of the surveyed drivers towards an intensification of police enforcement that would result in a travel time increase and accident risk reduction (for complying drivers, but also arguably indirectly for non-complying drivers) or a monetary fine (for non-complying drivers) depends on the following variables:

- Age group: younger drivers exhibit in general less compliant and consequently more risky behaviour (with the exception of the age group 45-54 showing more aggressive behaviour that the 35-44 group).
- Income level: wealthier drivers are less likely to comply, possibly due to their decreasing marginal utility of money, which makes them more indifferent to paying a fine.

- Education level: drivers with higher education show a higher tendency to comply with the traffic laws.
- Trip time increase (due to intensification of enforcement): as in all travel related models, travel time is typically considered as "impedance". Therefore, an increase in travel time due to a change in behaviour (in this case strict compliance to traffic enforcement) is translated to a trade-off: the drivers weigh the increase in travel time against the probability of getting fined and/or getting involved in a traffic accident.

The analysis also provides evidence that the existing trip duration (before the intensification of police enforcement) affects the driver's preferences. Ceteris paribus, drivers who intend to make a 5-hour trip are found to be more "patient" or compliant than those who intend to make a 2-hour trip, as long as the increase in time duration is below 60% of the original trip duration.

Based on these findings, it can be argued that while the "typical" Greek driver may be compliant, there are some segments of the population that show a higher tendency to violate traffic laws (even if they know that they can get fined and/or involved in an accident with injury). This is a useful and practical finding for road safety policy makers, which could use it to develop targeted police enforcement campaigns (or targeted media campaigns, special education initiatives), aimed at the demographic segments with a higher tendency for the violation of traffic laws. Especially, this sensitivity analysis allows policy makers to better define the effective range of the enforcement intensification for the various driver categories. The findings of this research can also be applicable to other cases, as long as the particularities of these cases are met through adequate adaptations of the proposed methodology.

An additional contribution of this research is that it demonstrates how a state-of-the-art modelling technique (mixed logit) can be used by road safety practitioners and policy makers in general, in their quest to identify critical characteristics of driver behaviour and develop the related road safety measures and programmes. New computing and software advances make the use of the more flexible mixed logit models more accessible, instead of the simpler binary logit models often preferred by practitioners.

This research has provided valuable insight into the behaviour of the Greek drivers in a situation where an intensification of police enforcement forces them to choose between complying to an intensified police enforcement scenario or being subject to a fine or a trip time increase. Similar initiatives in other countries of the EU (and the world) could show which of the underlying behavioural patterns are shared across populations and which are specific to the Greek population.

ACKNOWLEDGMENT

Special acknowledgment belongs to Mr. Stamatis Halapas, NTUA civil engineer for the coordination of the survey execution and for the data processing.

REFERENCES

Abdel-Aty, M., R. Kitamura, and P. Jovanis (1997). Using stated preference data for studying the effect of advanced traffic information on drivers' route choice. Transportation Research (C), Vol. 5, No. 1, pp. 39-50.

Ben-Akiva, M., Lerman, S., (1985). Discrete Choice Analysis. The MIT Press.

Bierlaire, M. (2003). BIOGEME: a free package for the estimation of discrete choice models, Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Bierlaire, M. (2005). An introduction to BIOGEME (Version 1.4). December 16, 2005. URL: roso.epfl.ch/mbi/biogeme/doc/tutorial.pdf

Bierlaire, M., Thémans, M. and Axhausen, K. (2006). Analysis of driver's response to real-time information in Switzerland, European Transport, Issue n° 34, December 2006.

Bjornskau, T., Elvik, R. (1992). Can road traffic law enforcement permanently reduce the number of accidents? Accident Analysis and Prevention, Vol. 24, pp 507-520.

Chamberlain, G, (1980). Analysis of Covariance With Qualitative Data". Review of Economic Studies, 47, pp. 225-238.

European Transport Safety Council (ETSC 1999). Police enforcement strategies to reduce traffic casualties in Europe.

European Commission (EC 2001). European transport policy for 2010: Time to decide (COM(2001) 370 final, 12 September 2001).

Golias, J., Matsoukis, E., Yannis, G., (1997). An analysis of factors affecting road safety: The Greek Experience. Journal of the Institute of Transportation Engineers, Vol. 67, No. II, pp. 26-31.

B. Han, S. Algers, and L. Engelson. Accommodating Drivers' Taste Variation and Repeated Choice Correlation in Route Choice Modeling by Using the Mixed Logit Model. Proceedings of the

80th Annual Meeting of the Transportation Research Board, January 7-11, 2001, Washington, D.C.

Heckman, J. (1981). The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process", In Structural Analysis of Discrete Data with Econometric Applications. C. Manski and D. McFadden, editors. MIT Press, Cambridge, MA.

Hess, S., Bierlaire, M. and Polak, J. (2005). Estimation of value of travel-time savings using Mixed Logit models, Transportation Research Part A, 39(3), pp. 221-236.

Holland, C. and Conner, M., (1995). Exceeding the speed limit: An evaluation of the effectiveness of a police intervention. Accident Analysis and Prevention, Vol. 28, pp. 587-597.

Hoorn, A. van der, Kroes, E., Meijer, E., (1984). Mode choice and mode capacity interlocal commuting: Two Dutch studies on modal split in congested transportation corridors. Transport policy decision-making, 2, pp. 373-388.

Hsiao, C, (1986). Analysis of panel data. Cambridge University Press, Cambridge, U.K.

Kanellaidis, G., Yannis, G., Harvatis, M., (1999). Attitude of Greek drivers towards road safety. Transportation Quarterly, Vol. 53, No.1, 1999, pp. 109-121.

Khattak, A., A. Polydoropoulou, and M. Ben-Akiva (1996). Modeling Revealed and Stated Pre-trip Travel Response to Advanced Traveler Information Systems. Transportation Research Record, 1537, 46.

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.

Louviere, J.J., D.A. Hensher and J.D. Swait (2000) Stated Choice Methods: Analysis and Application, Cambridge University Press.

Luce, R.D. and J.W. Tukey (1964). Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement, Journal of Mathematical Psychology 1, pp. 1-27.

Matsoukis, E.C., J.C. Golias and G.D. Yannis, (1996). "An Overview of Road Safety in Greece". Transport Reviews, vol. 16, no. 4, pp. 357-367.

McFadden, D. and K. Train (2000), ‘Mixed MNL models of discrete response’, Journal of Applied Econometrics 15, 447–470.

Newstead, S., Cameron, M., Mark, L., Leggett, W. (2001). The crash reduction effectiveness of a network-wide traffic police deployment system. Accident Analysis and Prevention, Vol. 33, pp. 393-406.

Polydoropoulou, A., M. Ben-Akiva, A. Khattak and G. Lauprete (1996). Modeling Revealed and Stated En-Route Travel Response to Advanced Traveler Information Systems, Transportation Research Record, 1537.

Rothengatter, J.A., Harper, J. (1991). The scope and design of automatic policing information system with limited artificial intelligence. Advanced Telematics in Road Transport, pp. 1499-1515. Amsterdam: Elsevier.

Rizzi, L., and J. de D. Ortuzar, (2003). Stated preference in the valuation of interurban road safety. Accident Analysis and Prevention, 35 (1), pp. 9-22.

Sándor, Z. and K. Train. Quasi-random simulation of discrete choice models. Transportation Research Part B: Methodological, Volume 38, Issue 4, May, Pages 313-327. (2004).

SARTRE, (1998). Social attitudes to road traffic risk in Europe - Contextual analysis. *SWOV*, Leidschendam, The Netherlands.

SARTRE (2004). Making our drivers and roads safer. Selected results from a European survey. INRETS, October 2004.

Sivakumar, A., C. R. Bhat, and G. Okten. Simulation Estimation of Mixed Discrete Choice Models with the Use of Randomized Quasi-Monte Carlo Sequences. Transportation Research Record: Journal of the Transportation Research Board 1921, pp. 112-122. (2005).

Tay, R. (2005). General and Specific Deterrent Effects of Traffic Enforcement. Do we have to Catch Offenders to Reduce Crashes? Journal of Transport Economics and Policy, Volume 39, Part 2, pp. 209–223.

Train, K. (2003). Discrete Choice Methods with Simulation, Cambridge University Press.

Washington S., Karlaftis M., Mannering F., (2003). Statistical and Econometric Methods for Transportation Data Analysis, Chapman & Hall / CRC.

Yannis, G., A.Kanellopoulou, K.Aggeloussi, D.Tsamboulas, "Modelling driver choices towards accident risk reduction", Safety Science, Vol.43/3, 2005, pp. 173-186.

Zaal, D., (1984). Traffic law enforcement: a review of the literature. Report Number 53. Monash University accident Research Center, Melbourne, Australia.

List of Tables

Table 1. Subset of part three of the questionnaire

TRIP DURATION 2 HOURS

Choice	Compliance to increased enforcement is	Trip duration increase (min)	Risk probability reduction (%)
A	unlikely	0	0%
B	likely	+30	-10%

Choice	Compliance to increased enforcement is	Trip duration increase (min)	Risk probability reduction (%)
A	unlikely	0	0%
B	likely	+90	-10%

TRIP DURATION 5 HOURS

Choice	Compliance to increased enforcement is	Trip duration increase (min)	Risk probability reduction (%)
A	unlikely	0	0
B	likely	+ 60	-20%

Choice	Compliance to increased enforcement is	Trip duration increase (min)	Risk probability reduction (%)
A	unlikely	0	0%
B	likely	+180	-10%

Table 2. Model specification table

Model:	Binary logit		Mixed logit - random effects	
	Compliance to increased enforcement scenario is		Compliance to increased enforcement scenario is	
	Likely	Unlikely	Likely	Unlikely
$\beta_{Compliance}$	1	0	1	0
$\beta_{Age\ 35\text{-}44}$	Age dummy (35-44)	0	Age dummy (35-44)	0
$\beta_{Age\ 45\text{-}54}$	Age dummy (45-54)	0	Age dummy (45-54)	0
$\beta_{Age\ 55+}$	Age dummy (55+)	0	Age dummy (55+)	0
$\beta_{Low\ income}$	Low income dummy	0	Low income dummy	0
$\beta_{High\ income}$	High income dummy	0	High income dummy	0
$\beta_{Low\ education}$	Low education dummy	0	Low education dummy	0
$\beta_{Trip\ duration}$	Trip duration (min)	0	Trip duration (min)	0
$\beta_{Trip\ duration\ increase}$	Trip duration increase (min)	0	Trip duration increase (min)	0
$\beta_{Risk\ change}$	Risk change (%)	0	Risk change (%)	0

Table 3. Estimation results

Utility parameter name	Binary logit Value	Binary logit t-test	Mixed logit - random effects[a] Value	Mixed logit - random effects[a] t-test
$\beta_{Compliance}$	2.5589	8.2984	3.3262	6.3203
$\beta_{Age\ 35\text{-}44}$	1.1843	4.7182	1.6885	3.5026
$\beta_{Age\ 45\text{-}54}$	0.7346	3.4713	0.9725	2.6388
$\beta_{Age\ 55+}$	1.8336	5.1579	2.3494	3.4951
$\beta_{Low\ income}$	0.4946	2.2330	0.8390	1.8858
$\beta_{High\ income}$	-0.4783	-2.6156	-0.5975	-1.8013
$\beta_{Low\ education}$	-0.5316	-2.6463	-0.7255	-1.7882
$\beta_{Trip\ duration}$	0.0047	3.0714	0.0070	2.4780
$\beta_{Trip\ duration\ increase}$	-0.0075	-2.2456	-0.0108	-2.5328
$\beta_{Risk\ change}$	0.1136	6.6275	0.1466	6.7072
σ_{panel}	--- [b]	--- [b]	1.6487	7.7446

	Binary logit	Mixed logit - random effects
Number of Halton draws:	--- [b]	200
Number of estimated parameters:	10	11
Sample size:	1184	250[c]
Null log-likelihood $LL(0)$:	-820.69	-820.69
Final log-likelihood $LL(\beta)$:	-490.23	-454.18
Likelihood ratio test:	660.92	733.01
ρ^2 (Rho-square):	0.4027	0.4466
Adjusted ρ^2 (rho-square):	0.3905	0.4332

[a] 200 Halton draws have been used (results had stabilized well below 100 Halton draws)

[b] --- denotes not applicable

[c] Sample size for the mixed logit refers to individual respondents (each providing up to 5 responses). Number of observations is again 1184.

List of Figures

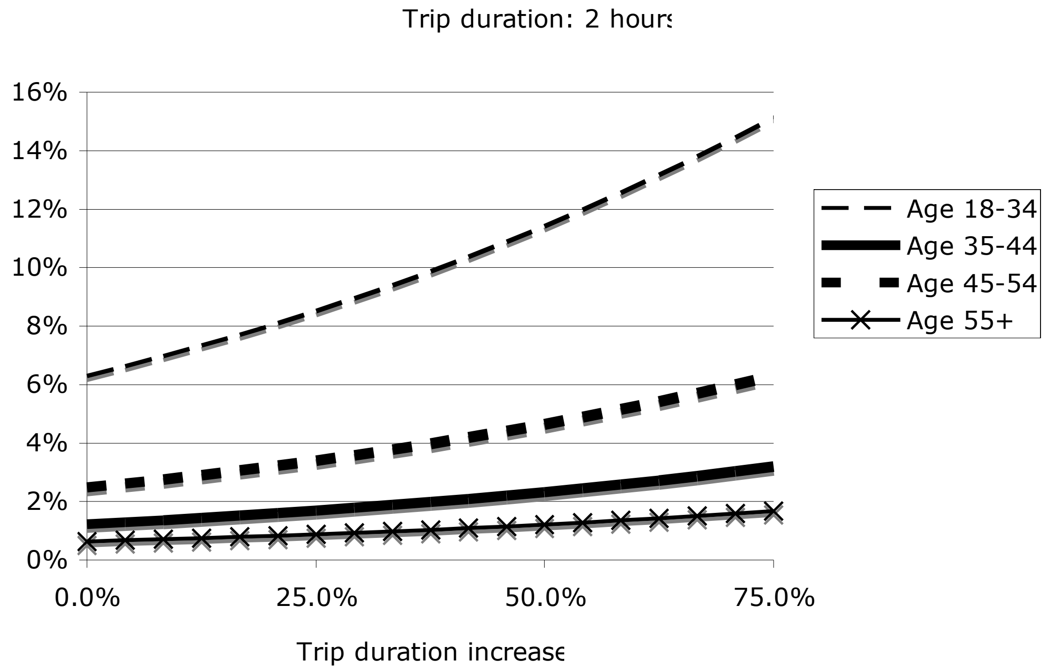

Figure 1. Sensitivity analysis w.r.t. age group

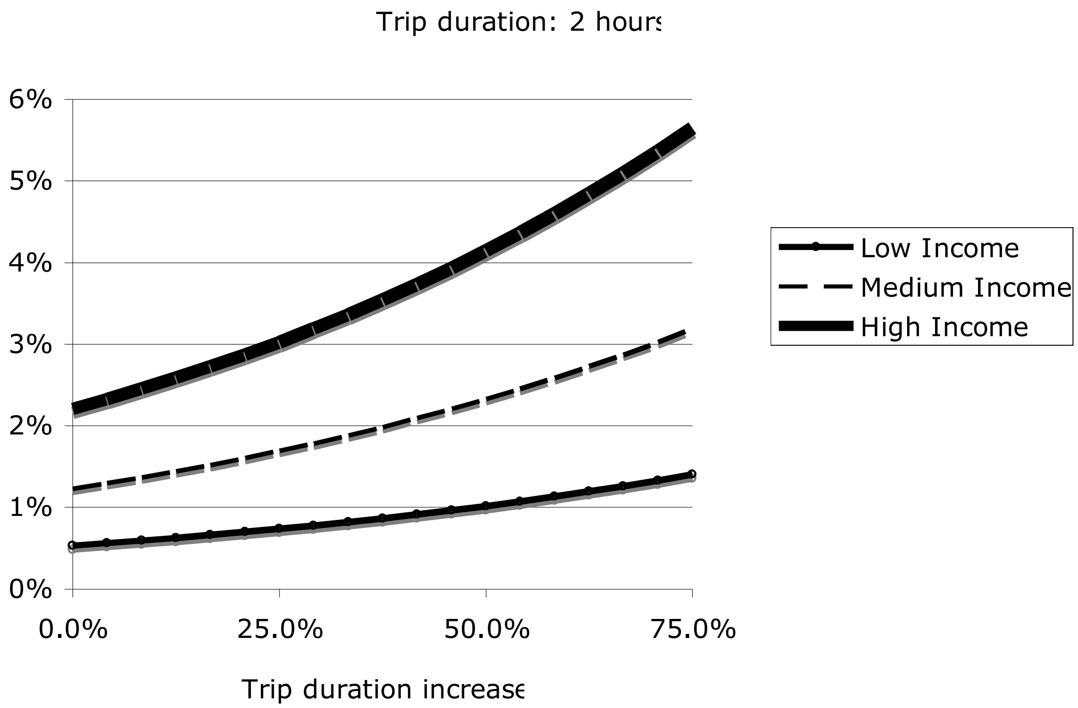

Figure 2. Sensitivity analysis w.r.t. income level

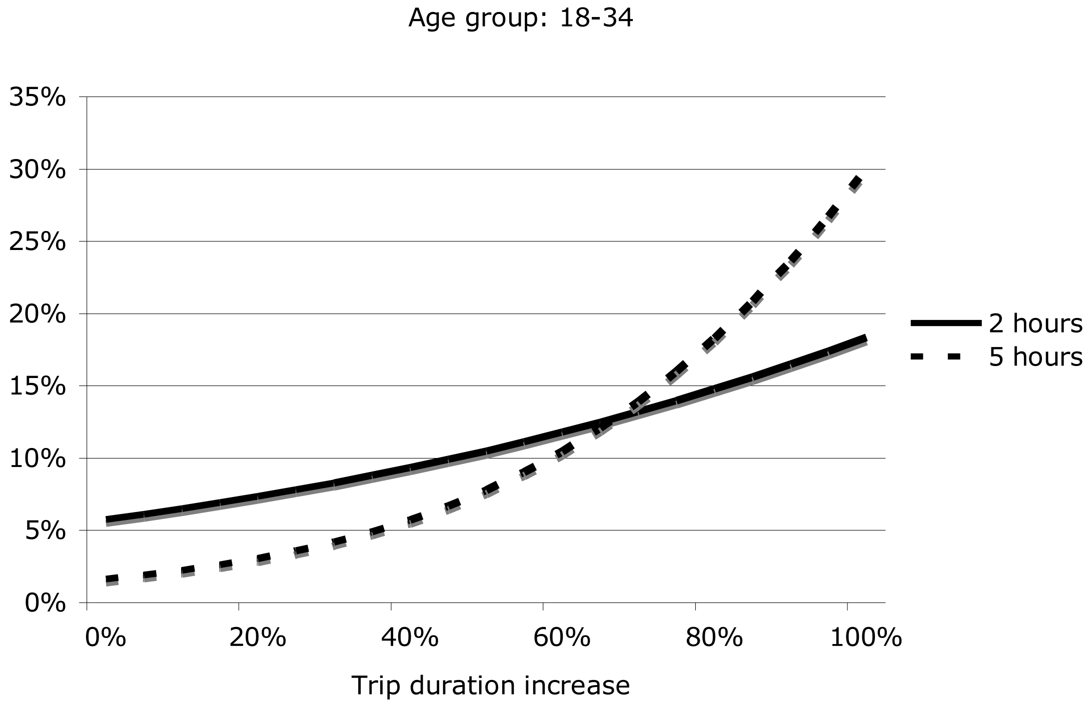

Figure 3. Sensitivity analysis w.r.t. trip duration