ESTIMATING CONSUMERS' BEHAVIOUR IN MOTOR INSURANCE USING DISCRETE CHOICE MODELS

Cristian Mihai Dragos, Simona Laura Dragos

Introduction

As an effective tool for mitigating financial risk, insurance has nowadays become a key sector of a functioning modern society. Motor insurance constitutes the largest line of business of the non-life insurance sector in Europe, considering its fleet of around 334 million vehicles registered in 2013 (Insurance Europe, 2015). The total motor insurance premiums in Europe amounted to €123.5 billion in 2013, with a 28% share of the total Gross Written Premiums.

The motor insurance line of business has two components – the Motor Third Party Liability Insurance (MTPL) and the Motor Damage insurance. The motor damage policy covers the cost of repairing the owner's vehicle following an accident, while the MTPL insurance covers the cost of repairing the vehicle of the third party. The legal framework of those two components varies across Europe, the common point being the compulsory nature of the MTPL insurance. In some countries it is difficult to make an explicit distinction between those two components when it comes to the legal framework, because motor insurance is sold as comprehensive insurance (Insurance Europe, 2015). In other countries, motor damage insurance is a voluntary form of insurance, the premiums of this category amounting to around 40% of total gross written premiums in motor insurance policies.

Nowadays, when people tend to be more mobile in terms of car transport, an increase in the risk exposures affecting motor insurance can be perceived. The annual mileage covered by cars increases because of the road infrastructure and the migration of potential consumers' residence towards low traffic rural areas. This phenomenon, initially perceived only in developed countries, tends to generalise also towards emerging economies, mostly for those implied in an integration process to a certain community (Busu & Gyorgy, 2016). This significant diversification of risks covered by the motor insurance market must alert the insurance providers to be extremely cautious when developing an offer or when assessing the risk profile of potential customers.

The aim of this article is to evaluate the key determinants of the decision to subscribe to voluntary motor damage insurance. Our study wishes to fill a need for academic applied studies in this area, especially considering the fact that motor damage insurance is legally compulsory in many countries. Still, for insurers from countries where these products are voluntary, including Romania, our study provides useful information for product design.

By using logit models (binary, multinomial and nested), this empirical study emphasizes some significant factors affecting consumer behaviour when purchasing such an insurance policy: the educational background, the distance travelled by car, the risk profile of the insurer, the ratio between the owner's income and the value of the car. As control variables, we have retained a number of socio-demographic factors. The main contribution of our article is to define two behavioural factors in order to explain more thoroughly the decision to purchase motor damage insurance. The first factor measures the risk profile of the individual in terms of savings and loans, and the second one is defined as the ratio between the estimated price of the car and the revenue of the owner.

This study could be relevant for other countries where motor damage insurance is placed under a voluntary legal framework, as long as there are no major behavioural differences in the reasoning mechanism of deciding whether to purchase such insurance or not.

The remainder of the paper has the following structure: Section 1 reviews previous studies on

 DOI: 10.15240/tul/001/2017-4-007

insurance consumers' behaviour and formulates the research hypothesis. Section 2 explains the econometric models and the methodology used to identify and evaluate the main determinants of voluntary motor insurance consumption on the Romanian market. Section 3 provides the results and discussions of the empirical study, and Section 4 draws conclusions and provides recommendations and policy implications.

1. Literature Review and Hypothesis Development

Motor insurance and its components (compulsory insurance, damage or combined policies) are studied from various perspectives in the literature. Depending on continent or country, the factors considered when establishing premiums, as well as the insurance distribution channels vary considerably. In some North American states, the insurance premium is determined by taking into account the driver's history (number of accidents), the annual mileage and the driving licence holding period (Dwight & Russell, 1995). The insurance system based on the number of miles driven annually, PAYD (pay-as-you-drive), has also been adopted in Europe, in countries such as the UK and Italy. Insurers from Ireland, Sweden and UK incorporate a credit score of the individual into the premium in order to capture the individual's risk aversion (European Commission, 2009). Another insurance system frequently used in most European countries is the "bonus malus" system (European Commission, 2009) applied by insurance companies mainly to urge the policyholders to become more responsible.

In the academic literature, the theoretical research on motor insurance demand is rather scarce. For instance, Awunyo-Vitor (2012) has examined motor damage insurance drivers in Ghana by using a binary logit model. In his study, the considered factors were the income of the insured, the age, education level and gender of the car owner, the age and value of the vehicle, as well as the perception of the premium and of the claim procedure. He has found that the age of the vehicle has a significant negative influence over the demand of motor damage insurance, which means that new vehicles are more likely to be insured than old ones. The income of the car owners and the value of the vehicles were found to be positively related to motor damage insurance demand at 1% level of significance. Analysis of the persons who perceived the premium as satisfactory revealed a 70% probability of their buying insurance policies, and the probability was 51% in the case of those who perceived the claim procedure as satisfactory.

By means of a mixed Logit model, Hsu et al. (2014) have studied the degree to which the automobile insurance claims are affected by the characteristics of policyholders and insurance policies. They show that replacement value and the age of the vehicle are important factors in deciding to purchase voluntary motor insurance in Taiwan. Peng et al. (2016) have shown that the liberalization of automobile insurance market in Taiwan in 2009 increased the rate competition among insurers and prompted them to lower their rates.

Motor insurance consumers reveal a multivalent behaviour. In order to explain it, some studies focus on identifying the stimuli influencing individuals to subscribe to such policies, while other studies deal with the processes taking place inside the consumer's mind. As a type of social behaviour, consumption is influenced by psychological factors as well. Consequently, the third categories of studies are those that analyse the consumer data in terms of revealed behaviours: repurchase, preference, recommendation etc. (Query et al., 2007; Parvatiyar & Sheth, 2000). To put it differently, the literature stipulates that, much like other services, motor insurance must also contribute to maximizing the value felt by customers so that they may build a lasting relationship with the company and reveal a loyal behaviour – recommendation, repurchase, preference (Kotler & Armstrong, 2001; Gencer & Akkucuk, 2017).

As far as consumers' perspective is concerned, Rundmo and Moen (2005) and Rundmo and Nordfjaern (2013) have found that the most significant predictor of risk perception involves evaluations of the probability of occurrence of an event, as well as assessments of the severity of the event, if it should happen. Rundmo and Nordfjaern (2013) have concluded that „risk awareness significantly predicts risk perception and that risk awareness is directly associated with a demand for risk mitigation in transport".

Ioncică et al. (2012) believe that the behaviour of insurance consumers is driven by different factors such as: marketing-related factors (e.g. advertising of insurance products, distribution channels), cultural factors, economic, demographic and legislative factors and the individual characteristics of potential insurers. The authors establish that, of all these factors, education plays the essential role in the decision of purchasing insurance. Sapelli and Vial (2003) have found a positive relationship between the education level and the probability of purchasing voluntary insurance, which means that better educated persons are more likely to understand the benefits of insurance and to protect themselves by subscribing to a policy.

Dewar (1998) and Currie (1995) have found that gender is also significant for the insurance demand. Nevertheless, the result may be correlated with income since, given the same educational background, women earn on average less than men. Liu et al. (2011) prove that risk preferences, along with income and education, are important predictors of voluntary insurance demand. The differences between the socio-demographic characteristics are correlated with risk perception, thus influencing the insurance demand and the risk-taking behaviour (Ioncică et al., 2012; Kašćelan et al., 2016).

Depending on customers' socio-demographic characteristics (age, income, education, gender, driving licence holding period etc.) and behaviour (number of accidents or claims, number of indemnity payments, type of deductible chosen by the policyholder), insurance companies segment customers to avoid adverse selection (Barone & Bella, 2004). Adverse selection can be avoided by discriminating through prices, which means that the same insurance service will generate different costs for different consumers because they represent heterogeneous risks (Dragos, 2007). Companies generally avoid customers with a bad reputation or those who had a large number of accidents in the past. Instead, they pursue and win the loyalty of customers who demonstrate exemplary behaviour. Saito (2009) tests the Japanese auto insurance market for the presence of adverse selection in the case of bodily injury liability insurance. He finds little evidence that informational asymmetry leads to the inefficiency of the auto insurance market, even if this is a significant factor for the annuity market (Finkelstein & Poterba, 2004) and the health insurance market (Höfter, 2006).

An element of overwhelming importance to insurance companies is the proper assessment of the drivers' risk profile. The literature abounds in theoretical and empirical approaches using econometric, statistical and mathematical models to explain the phenomenon of 'adverse selection in insurance' (Shi et al., 2012; Jindrová & Jakubínský, 2015). By asking their customers to answer various questions, fill out their profile or take part in role-playing, insurers attempt to determine as precisely as possible the extent to which their customers are either more or less inclined to take certain risks (in other words, the risk aversion degree of their clients). Based on such scenarios, the individuals who are willing to take increased risks will fall into the category of policyholders who pay a higher premium. At the same time, due to their increased risk preference, such individuals are likely to purchase extra insurance or buy higher-premium policies. Cohen (2005) found that, on the Israeli motor insurance market, new customers with more than three years of driving experience and choosing low deductibles for motor insurance are associated with more accidents and higher losses for the insurer.

Relying on quoted literature and logical reasoning, we are in a position to outline the following hypothesis:

H1: The risk profile of the individual influences the purchasing of voluntary motor insurance.

Generally, the economic literature accepts the influence of the risk profile of the individual on buying any type of voluntary insurance. Nevertheless, this mechanism is hard to distinguish in applications, because we do not have objective values of the risk profile at individual level. The empirical studies measure this profile by different proxy variables like age, gender, education, stability at work and type of job. In our study we are attempting a novel approach, relating the risk profile to the individual's behaviour regarding saving versus borrowing.

H2: The annual distance traveled by car influences the purchase of voluntary automobile insurance.

There is an inner risk profile of each individual, which comes from his human nature. Beyond this, there is a specific risk associated to intense car usage. This behaviour is easily noticed by the driver, influencing his decision to either buy a voluntary motor insurance or not. Basically, if the insured considers this risk, he or she can generate adverse selection in insurance.

Tab. 1: Variables description

Variable	Variable description
Endogenous variables (for binary logit model)	
Y = 1 (Motor Damage)	If the person has Motor Damage insurance
Y = 0 (No Motor Damage)	If the person does not have Motor Damage insurance
Endogenous variables (for multinomial logit model and nested logit model)	
Y = 1 (Motor Damage)	If the person has No Motor Damage insurance
Y = 2 (No Motor Damage)	If the person does not have Motor Damage insurance. Main motivation: owner's income allows him/her to cover easily the cost of potential damage
Y = 3 (No Motor Damage)	If the person does not have Motor Damage insurance. Main motivation: the likelihood of accident is assessed as being low and purchasing an insurance does not seem profitable
Exogenous variables	
RISK_PROFILE	1 if total savings exceed by far total value of loans 2 if savings are comparable in amount (max ± 20%) with loans 3 if total loans do not exceed 40% of the current value of real estate[1] 4 if total credits exceed 40% of the current value of real estate
CAR_PRICE_RANGE	1 if the estimated car price does not exceed 2,000 Euros 2 if the estimated car price is between 2,000 Euros and 4000 Euros[2] 3 if the estimated car price exceeds 4,000 Euros
KM/YEAR	Mileage covered by the car over the last 12 months (thousands km)
WAGE_RANGE	1 if the net wage does not exceed 800 RON 2 if the net wage is between 800 and 1,700 RON[3] 3 if the net wage exceeds 1,700 RON
WAGE/CARPRICE_RANGE	from -2 to 2. Difference between the variables *WAGE_RANGE* and *CAR_PRICE_RANGE*
BACHELOR	1 if the owner holds a bachelor's or higher degree 0 if not
URBAN	1 if the owner's main residence is in urban area 0 if not
AGE	Owner's age
GENDER	1 if the owner is male 0 if the owner is female

Source: own

Notes:
1 The cutoff (40%) was taken from bank credit reports. This value allows roughly equal shares for RISK_PROFILE=3 and RISK_PROFILE=4.
2 The cutoffs (2,000 Euros and 4,000 Euros) ensure roughly equal shares for the three values of the variable CAR_PRICE_RANGE. The data on car value distribution in Romania are taken from statistics of insurance companies and from car sales websites.
3 The cutoffs 800 RON and 1,700 RON ensure a balanced distribution among the three groups, in accordance with income breakdown in Romania.

H3: The relationship between the individual's income and the value of the car influences the purchase of voluntary motor insurance.

We believe that treating the income of the individual and the value of the car separately cannot lead to conclusive results. Essentially, it is the connection between the two measurements that counts. An individual with a high income compared to the value of his car can easily support his own potential damages. On the contrary, an expensive car connected to an owner with a modest income generates a high risk against which it is better to be insured.

After briefly mentioning previous studies focusing on motor insurance demand, the following paragraphs describe and explain the econometric tools to estimate the consumers' behaviour when buying motor damage insurance, using a sample of Romanian car owners.

2. Data and Methodology

2.1 Data

The data was collected in Cluj County, Romania, over the period between September – and December 2015, and the tools used in the survey were the face-to-face and self-administered questionnaires. To avoid any systematic sampling bias, the respondents were chosen randomly. However, they had to meet two requirements simultaneously, namely, to possess a car and to have a driving licence. The questionnaire was run during the periodic roadworthiness tests carried out at accredited centres/garages. According to current Romanian legislation, the roadworthiness test is compulsory for all vehicles, regardless of the owner's socio-demographic characteristics (income, age, gender, workplace, domicile etc.), the value or age of the car, its annual mileage or technical condition etc. The sample comprised only persons who are car owners and their own driving time in the car exceeds 80% of the car's overall driving time. If the driver had not bought an insurance policy, the authors decided to limit the policy choice decision to two possible major motivations (Y = 2 and Y = 3, see Tab. 1). The drivers falling outside both of these defined motivations were excluded from the sample. As a result of the imposed prerequisites, the sample contains 311 fully-completed questionnaires providing values for all studied variables (see Tab. 1).

2.2 Methodology

Three models are used to model the individual choice: binary logit, multinomial logit and nested logit. The ROC curve is computed in order to compare the predictive performance of the models.

The Binary Logit Model

The individual's likelihood to choose (Y = 1) or not to choose (Y = 0) a Motor Damage insurance policy is modelled by the formula:

$$\Pr(Y_i = 1) = \frac{\exp(x_i b)}{1 + \exp(x_i b)} \tag{1}$$

where $i = \overline{1,N}$ Index of each individual; x_i the vector of the exogenous variables; b the the vector of the coefficients.

The Multinomial Logit Model

The individual's likelihood to choose one of the three alternatives (Y = 1, Y = 2, Y = 3, see Tab. 1) is modelled by the formula:

$$\Pr(Y_i = j) = \frac{\exp(x_i b_j)}{\sum_{j=1}^{m} \exp(x_i b_j)} \tag{2}$$

$i = \overline{1,N}$ index of each individual; $j = 1,2,3$ index of each alternative. x_i the vector of the exogenous variables. In the present application, all variables vary with respect to individual (i) but remain constant with respect to alternatives (j). b_j the vector of the coefficients. Coefficients vary with respect to alternatives (j). The values of b_j coefficients are interpreted with respect to a reference alternative.

The Nested Logit Model

The individual's likelihood to choose one of the three alternatives is modelled by grouping the alternatives. L represents the number of groups in which the alternatives are included. In each group l there are J_l possible choices, indexed by $j(l)$. The total number of possible alternatives is $J = J_1 + J_2 + ... + J_l$. The decision process takes place at two levels: inside each group and between groups. The variables x_l describing qualities that are common to choices of the same group take values that may vary between groups, but not inside the group. The variables $x_{j(l)}$ vary across alternatives. McFadden (1973) has proved that a discrete choice model can be developed to the utility maximization, based on the hypothesis that error terms follow a Weibull distribution. The probability $\Pr(j/x)$ of choosing the alternative *j* conditioned by the vector of explanatory variables *x* can be written as:

$$\Pr(j/x) = \Pr(l/x)\cdot \Pr(j(l)/x_{j(l)}) \tag{3}$$

$P(l/x)$ is the probability of choosing one of the *L* groups:

$$\Pr(l/x) = \frac{\exp(x_l b + \lambda_l I_l)}{\sum_{l=1}^{L}\exp(x_l b + \lambda_l I_l)} \tag{4}$$

$I_l = \ln\sum_{j=1}^{J_l}\exp(x_{j(l)}b_l)$ is the inclusion variable

$Pr\,(j(l)/x_{j(l)})$ is the probability of choosing an alternative belonging to the group l:

$$\Pr(j(l)/x_{j(l)}) = \frac{\exp(x_{j(l)}b_l)}{\sum_{j=1}^{J_l}\exp(x_{j(l)}b_l)} \tag{5}$$

For each individual, the probability of choosing each alternative can be estimated for both the multinomial logit model and the nested logit model. In this study all econometric estimations are performed using LIMDEP 10 and its extension NLOGIT 5.

ROC (Receiver Operating Characteristics) Curve for Discrete Choice Models

Models are generally compared in the literature by means of purely econometric criteria such as pseudo R^2, parameters significance etc. A highly appreciated practitioners' tool is the ROC curve, which assesses the predictive power of binary models. Recently the methodology has also been developed for Discrete Choice Models (Dragos, 2010). The following notation is used:

$N_{11(j)}$ the number of individuals choosing alternative *j* and that the model predicts will choose alternative *j.*

$N_{1T(j)}$ the number of individuals choosing alternative *j.*

$N_{00(j)}$ the number of individuals *not* choosing alternative *j* and that the model predicts will *not* choose alternative *j.*

$N_{0T(j)}$ the number of individuals *not* choosing alternative *j.*

Therefore, the following definitions are used:

- the proportion of correct predictions for individuals choosing *j*

$$\frac{N_{11(j)}}{N_{1T(j)}} = sensitivity_{(j)} \tag{6}$$

- the proportion of correct predictions for individuals not choosing *j*:

$$\frac{N_{00(j)}}{N_{0T(j)}} = specificity_{(j)} \tag{7}$$

This allows the ROC curve to be plotted for each alternative *j.* To plot the global ROC curve (for all alternatives), *global sensitivity* and *global specificity* are defined (Dragos, 2010) as follows:

$$\begin{aligned} sensitivity_{(G)} &= \frac{\sum_{j=1}^{J} N_{11(j)}}{\sum_{j=1}^{J} N_{1T(j)}} = \frac{\sum_{j=1}^{J} N_{11(j)}}{N} = \\ &= \sum_{j=1}^{J}\frac{N_{11(j)}}{N} = \sum_{j=1}^{J}\left(\frac{N_{11(j)}}{N_{1T(j)}}\times\frac{N_{1T(j)}}{N}\right) = \\ &= \sum_{j=1}^{J}\left(sensitivity_{(j)}\times\frac{N_{1T(j)}}{N}\right) \end{aligned} \tag{8}$$

$$\begin{aligned} specificity_{(T)} &= \frac{\sum_{j=1}^{J} N_{00(j)}}{\sum_{j=1}^{J} N_{0T(j)}} = \frac{\sum_{j=1}^{J} N_{00(j)}}{\sum_{j=1}^{J}\left(N - N_{1T(j)}\right)} = \\ &= \frac{\sum_{j=1}^{J} N_{00(j)}}{(J-1)N} = \sum_{j=1}^{J}\left(\frac{N_{00(j)}}{N_{0T(j)}}\times\frac{N_{0T(j)}}{(J-1)N}\right) = \\ &= \sum_{j=1}^{J}\left(\frac{N_{00(j)}}{N_{0T(j)}}\times\frac{N_{0T(j)}}{(J-1)N}\right) = \sum_{j=1}^{J}\left(specificity_{(j)}\times\frac{N_{0T(j)}}{(J-1)N}\right) \end{aligned} \tag{9}$$

By comparing the ROC curves of two competing models, arguments may be advanced with respect to their comparative performance. Because the ROC curve for Discrete Choice Models has only recently been developed, it is not yet implemented in the econometric software. All the computations and figures concerning the ROC curve are performed using Microsoft Excel.

3. Results and Discussions

The use of the binary logit model (the results reported in Tab. 2) points to several preliminary results concerning the behaviour of car owners. The alternatives where the individual does not

Tab. 2: Results of binary logit model (Y = 0 No Motor Damage is the base outcome)

Exogenous variable	Coefficient	t-value
RISK_PROFILE	** -0.899	5.88
WAGE/CARPRICE_RANGE	0.080	0.60
KM/YEAR	0.014	0.48
BACHELOR	* 0.656	2.25
URBAN	0.146	0.53
AGE	-0.008	0.77
GENDER	-0.150	-0.52
Constant	*1.450	1.97
N = 311; Pseudo R^2 = 0.129		

Source: own

Note: **, * : significant at 1% and 5%

have a Motor Damage insurance are grouped into a single alternative where *Y* = 0 (without Motor Damage).

The individual's risk profile and the level of education are the only statistically significant variables. The coefficient of the variable *RISK_PROFILE* has the negative sign, meaning that car owners with low risk aversion are less incited to subscribe to insurance policies. The positive value of the variable *BACHELOR* indicates that university-educated drivers reveal an increased level of preference for insurance, this fact

Tab. 3: Results of multinomial logit model (Y = 3 is the base outcome)

Exogenous variable	Coefficient	t-value
RISK_PROFILE (Y = 1)	** -0.781	-5.41
RISK_PROFILE (Y = 2)	-0.213	-1.68
WAGE/CARPRICE_RANGE (Y = 1)	0.263	1.76
WAGE/CARPRICE_RANGE (Y = 2)	** 0.394	2.88
KM/YEAR (Y = 1)	** 0.086	2.84
KM/YEAR (Y = 2)	** 0.075	2.64
BACHELOR (Y = 1)	* 0.773	2.39
BACHELOR (Y = 2)	0.101	0.31
URBAN (Y = 1)	0.242	0.80
URBAN (Y = 2)	-0.115	-0.42
AGE (Y = 1)	0.010	1.18
AGE (Y = 2)	-0.005	-0.56
GENDER (Y = 1)	0.047	0.15
GENDER (Y = 2)	-0.146	-0.50
N = 311; Pseudo R^2 = 0.090		

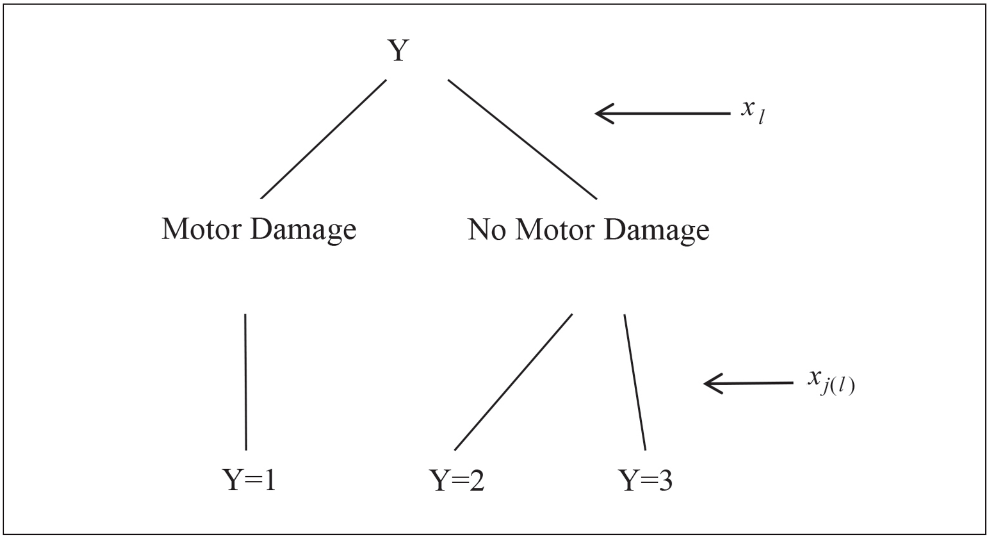
Source: own

Note: **, * : significant at 1% and 5%

accounting at least partially for their decision to subscribe or not to a Motor Damage insurance policy. However, if the decision is not to buy an insurance policy, the structure of the binary model doesn't reveal anything about the motivations behind this option. To better understand the mechanism, the range of alternatives is extended (*Y* = 1, 2, or 3, see Tab. 1). Because there isn't any ordered connection between the three alternatives, the econometric estimation is performed using discrete choice models (such as multinomial and nested logit).

Unlike in the case of the binary model, for the multinomial logit model (the results reported in Tab. 3, the variables *WAGE/CARPRICE_RANGE* and *KM/YEAR* significantly influence the individual's decision concerning an insurance policy. However, the multinomial logit model cannot determine which of the variables better explains the two options: a) the decision to subscribe or not to an insurance policy, or b) in case of non-subscription, what the appropriate motivation is? (*Y* = 2 or *Y* = 3). To solve this problem, the decision must take the form of a tree-like pattern (Fig. 1). The coefficients of different possible specifications of the nested logit model are estimated.

Fig. 1: The tree structure of the nested logit model



Source: own

According to the estimates of the binary and multinomial logit models, the variables *RISK_PROFILE* and *BACHELOR* belong to the attributes x_l and account for the decision between groups (nests). The variables *URBAN*, *AGE* and *GENDER* belong to the attributes $x_{j(l)}$ and account for the choice of the alternative in the group "without Motor Damage". By contrast, previous models cannot establish a definite conclusion for the variables *WAGE/CARPRICE_RANGE* and *KM/YEAR*. Consequently, each variable must be considered in two ways: a) as belonging to x_l attributes to explain the decision to subscribe or not an insurance policy (Nested Logit 1) and b) as belonging to $x_{j(l)}$ attributes to explain, in case of non-subscription, what the appropriate motivation may be (Nested Logit 2). The estimations are presented in Tab. 4.

Among all models used for estimations, Nested Logit 2 is the most adequate in terms of classical econometric criteria: the Pseudo R^2 and the significance of coefficients. A recently developed tool (Dragos, 2010), the Receiver Operating Characteristic Curve for Discrete Choice Models, allows us to estimate comparatively the predictive power of this

Tab. 4: Results of nested logit model (Y = 3 is the base outcome)

	Nested Logit 1		Nested Logit 2	
Exogenous variable	***Coeff.***	***t-value***	***Coeff.***	***t-value***
Attributes of the utility function				
WAGE/CARPRICE_RANGE (Y = 1)	…	…	-0.2980	-1.21
WAGE/CARPRICE_RANGE (Y = 2)	…	…	** 0.3930	2.99
KM/YEAR (Y = 1)	…	…	-0.0350	-0.71
KM/YEAR (Y = 2)	…	…	* -0.0530	2.28
URBAN (Y = 1)	0.2910	0.77	0.3020	0.80
URBAN (Y = 2)	-0.1130	-0.42	-0.1310	-0.48
AGE (Y = 1)	-0.0060	-0.53	0.0003	0.03
AGE (Y = 2)	-0.0026	0.44	-0.0010	-1.49
GENDER (Y = 1)	-0.0440	-0.12	0.2100	0.51
GENDER (Y = 2)	-0.1050	-0.40	-0.3610	-1.28
Attributes of Branch Choice Equations (G1 = reference)				
RISK_PROFILE (G2)	** 0.9330	6.02	** 0.9300	6.01
BACHELOR (G2)	* -0.6230	-2.19	* -0.6270	-2.20
WAGE/CARPRICE_RANGE (G2)	-0.0650	-0.48	…	…
KM/YEAR (G2)	-0.0140	-0.49	…	…
Inclusive Value Parameters				
G1	1.0000	fix. param.	1.0000	fix. param.
G2	* -2.1990	-2.04	* -2.0770	-2.00
	N = 311 Pseudo R^2 = 0.151		N = 311 Pseudo R^2 = 0.169	

Source: own

Note: **, * : significant at 1% and 5%

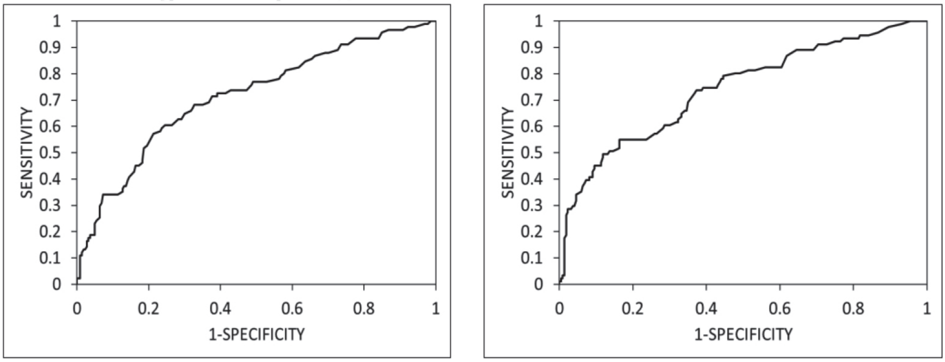
group of models for a single alternative (Fig. 2), as well as for all alternatives (Fig. 3).

According to the results of the econometric estimations, the ROC curve confirms that the best specification of a model is provided by Nested Logit 2. The tree structure of the model allows the separate identification of factors explaining: a) the decision to subscribe or not to an insurance policy (attributes of branch choice equations) and b) in case of non-subscription, what the appropriate motivation may be (attributes of the utility function).

The variables kept preserved in the model to explain the decision to subscribe or not are both significant. An individual's willingness to incur financial risks and a lower level of education increase the probability of his not purchasing a Motor Damage insurance policy. On the other hand, nor the risk profile, nor the education level can explain the motivation to subscribe to such insurance or not.

From the variables preserved in the model to explain the motivation for not subscribing to insurance, the annual mileage and the ratio between the estimated price of the car and the revenue of the owner are significant. There are two main motives behind the behaviour of the uninsured. The first motivation (*Y* = 2) consist in the possibility of easily providing coverage from the owner's available income in case of a car accident. The variable *WAGE/CARPRICE_RANGE* is statistically significant,

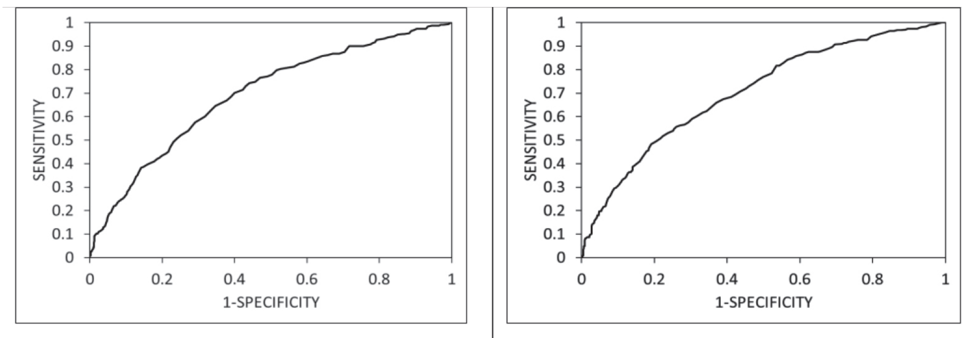
Fig. 2: Comparing ROC curves for the alternative Y = 1



2a) Multinomial Logit. AUROC = 0.714 2b) Nested Logit 2. AUROC = 0.741

Source: own

Fig. 3: Comparing ROC curves for all three alternatives



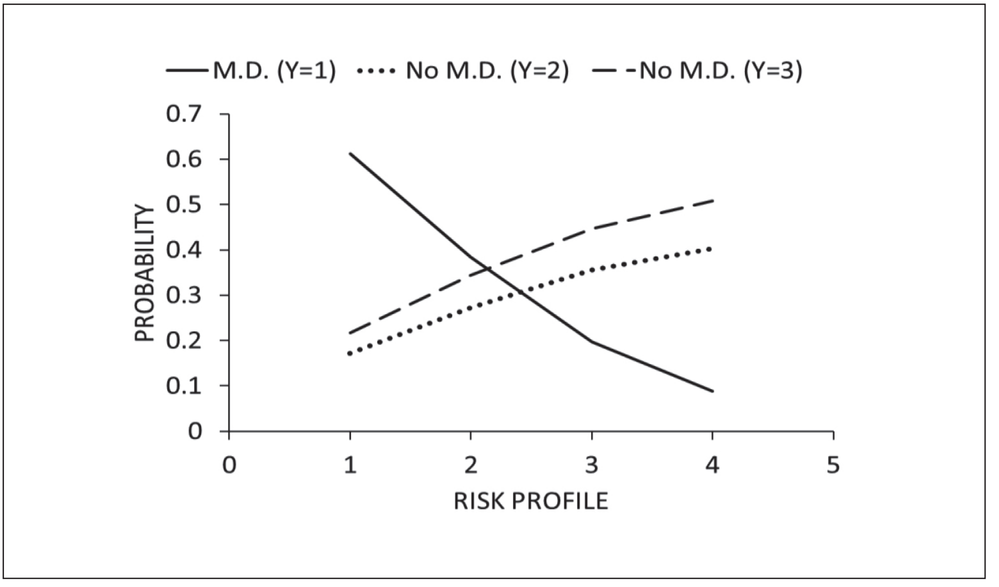
3a) Multinomial Logit. AUROC = 0.693 3b) Nested Logit 2. AUROC = 0.705

Source: own

accurately capturing the car owners' rationale. The second motivation (*Y* = 3) refers to the way in which drivers perceive the low likelihood of a car accident. Their perception is closely correlated with the car mileage (variable *KM/YEAR*). The sample includes people who drive between four and five thousand kilometres annually, sometimes only in rural areas where the volume of traffic is low. The fact that age, gender and the urban/rural environment are not statistically significant can be surprising, but is mainly due to the correlation, to some extent, with other regressors. For instance, women are less willing than men to incur risks of any kind, including car and traffic-related risks. This effect was expressed by the *RISK_PROFILE* variable. A similar effect was detected for the variable *AGE*. While the rural residence of the respondents does not have any significance, the values of variables like annual mileage, net income, and car price are significant. Consequently, the *H1*, *H2*, and *H3* hypotheses are accepted.

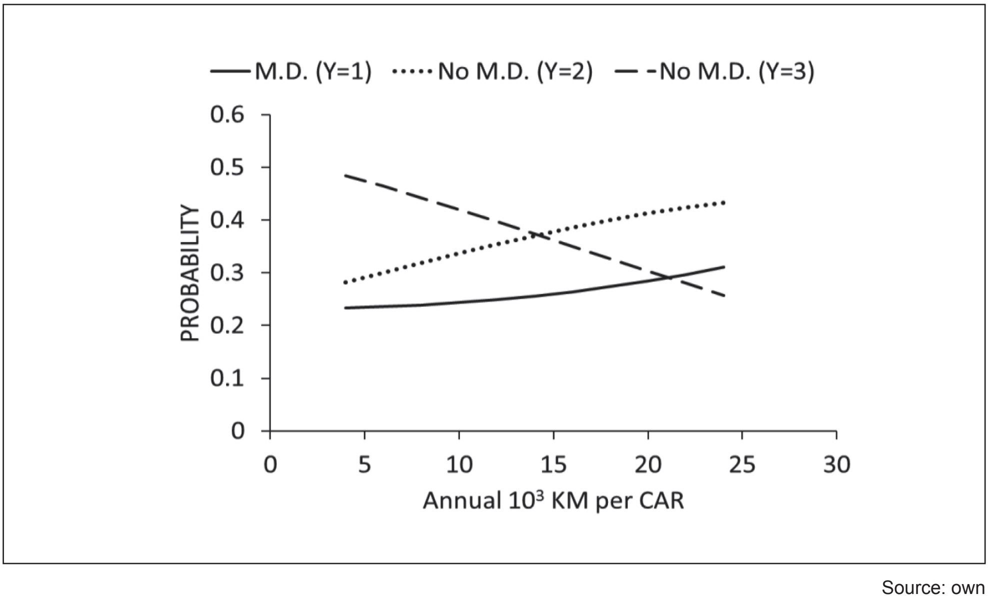
The most performant econometric model of our study, Nested Logit 2, may well lie behind the simulation of commercial policies

Fig. 4: Probability of choosing alternatives with respect to RISK_PROFILE



Source: own

Fig. 5: Probability of choosing alternatives with respect to annual mileage



Source: own

on insurance. We simulated the effect exerted by the variation of each explanatory variable on the probability of choosing one out of the three alternatives. The remaining regressors are supposed to be constant, taking the average value in the sample (Fig. 4 and Fig. 5).

The variation of the driver's risk profile has the strongest effect on the probability of subscribing an insurance policy (Fig. 4). By contrast, the perception of the probability of being involved in an accident and therefore of choosing the third alternative is most affected by the annual mileage (Fig. 5).

Conclusions

Meeting and satisfying customers' changing preferences, expectations and needs represent a major challenge for insurance companies. Therefore, all insurance providers must innovate on a regular basis and come up with new ways of approaching the target segments and of promoting their offers. Companies should also attempt to increase customers' awareness of risk exposure when driving irresponsibly, as well as their understanding of the negative impact of this on themselves and on the society at large.

The international literature on the individual's behaviour when purchasing motor insurance appears rather scarce, not least because in many countries the two insurance components – the compulsory (MTPL) and the optional (Motor Damage) – are sold together as one package. The relevance of this analysis only for voluntary insurance – where the customer decides to subscribe or not to a motor insurance policy – points to this state of affairs. While previous studies estimated the risk profile only through proxy variables such as income, age or gender, without accounting for any behavioural aspects, our study has successfully integrated the risk profile of the policyholders as a self-standing explanatory variable.

In emerging economies, the upward trend of the workforce mobility entails more frequent and longer commutes between home and workplace. The awareness of risk exposures (the perception of the probability of being involved in a car accident) will increase given the heavy road traffic. Even if, from a legal point of view, motor damage insurance remains voluntary, its share of the non-life segment is expected to increase significantly. The percentage of university-educated people is also an upward trend, with the educational aspect instrumental in rendering these individuals more aware of the advantages of an insurance system compared to self-insurance. The other variables in the study (income, age, gender) are either statistically insignificant or relatively constant in time. Therefore, they are not expected to exert any impact on motor insurance consumption in the near future.

To improve the profitability of this insurance category and because of the high number of road casualties, we consider the introduction of a common database for all motor damage insurers imperative. This database should provide, for the benefit of all potential insureds, the individuals' characteristics such as their risk perception, the properties of their car and the history of their road casualties. The database would help reduce the phenomenon of migration in the case of "high risk" policyholders. It refers to individuals with a bad history of past claims who try to avoid penalties charged for repeated damage by changing the insurer.

This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-II-RU-TE-2014-4-0745: Study of Romanian Life Insurances in International Context: Innovation, Spatial and Behavioral Modelling; Impact of Institutional Factors.

References

Awunyo-Vitor, D. (2012). Comprehensive Motor Insurance Demand in Ghana: Evidence from Kumasi Metropolis. *Management, 2*(4), 80-86. doi:10.5923/j.mm.20120204.01.

Barone, G., & Bella, M. (2004). Price-elasticity based customer segmentation in the Italian auto insurance market. *Journal of targeting, measurement and analysis for marketing, 13*(1), 21-31. doi:10.1057/palgrave.jt.5740129.

Busu, M., & Gyorgy, A. (2016). Real Convergence, Steps from Adherence to Integration. Countries from Central and Eastern Europe. *Amfiteatru Economic*, *18*(42), 303-316.

Cohen, A. (2005). Asymmetric information and learning: Evidence from the automobile insurance market. *Review of Economics and Statistics, 87*(2), 197-207. doi:10.1162/0034653053970294.

Currie, J. (1995). Gender gaps in benefits coverage. In D. Lewin, D. Mitchell, & M. Zaidi

(Eds.), *Handbook of Human Resources*. Greenwich CT: JAI Press.

Dewar, D. M. (1998). Do those with more formal education have better health insurance opportunities? *Economics of Education Review, 17*(3), 267-277. doi:10.1016/S0272-7757(97)00034-4.

Dragos, C. (2010). ROC curve for discrete choice models an application to the Romanian car market. *Applied Economics Letters, 17*(1), 75-79. doi:10.1080/13504850701719793.

Dragos, S. (2007). *Life Insurance Pricing*. Cluj-Napoca: University of Cluj Press, Romania.

Dwight, D., & Russell, T. (1995). The causes and consequences of rate regulation in the auto insurance industry. In D. Bradford (Ed.), *The Economics of Property Casualty Insurance* (pp. 81-112). University of Chicago Press.

European Commission. (2009). *Retail Insurance Market Study*. [MARKT/2008/18/H– Final Report, Europe Economics].

Finkelstein, A., & Poterba, J. (2004). Adverse selection in insurance markets: Policyholder evidence from the UK annuity market. *Journal of Political Economy, 112*(1), 183-208. doi:10.1086/379936.

Gencer, Y. G., & Akkucuk, U. (2017). Measuring Quality in Automobile Aftersales: AutoSERVQUAL Scale. *Amfiteatru Economic*, *19*(44), 110-123.

Höfter, R. H. (2006). Private health insurance and utilization of health services in Chile. *Applied Economics, 38*(4), 423-439. doi:10.1080/00036840500392797.

Hsu, Y. C., Chou, P. L., Chen, Y. M. J., & Lin, J. J. (2014). Mixed Logit Model of Voluntary Selection of Automobile Insurance. *Journal of Information and Optimization Sciences, 35*(5-6), 503-528. doi:10.1080/02522667.2014.961823.

Ioncică, M., Petrescu, E. C., Ioncică, D., & Constantinescu, M. (2012). The role of education on consumer behaviour on the insurance market. *Procedia-Social and Behavioural Sciences, 46*, 4154-4158. doi.10.1016/j.sbspro.2012.06.217.

Insurance Europe. (2015). *European Motor Insurance Markets*. Brussels.

Jindrová, P., & Jakubínský, R. (2015). Significance and possibilities of major accident insurance. *E&M Ekonomie a Management, 18*(4), 121-131. doi:10.15240/tul/001/2015-4-009.

Kašćelan, V., Kašćelan, L., & Novović Burić, M. (2016). A nonparametric data mining approach for risk prediction in car insurance: a case study from the Montenegrin market. *Economic Research-Ekonomska Istraživanja, 29*(1), 545-558. doi:10.1080/1331677X.2016.1175729.

Kotler, P., & Armstrong, G. (2001). *Principles of Marketing* (4th ed.). NJ: Prentice Hall.

Liu, H., Gao, S., & Rizzo, J. A. (2011). The expansion of public health insurance and the demand for private health insurance in rural China. *China Economic Review, 22*(1), 28-41. doi:10.1016/j.chieco.2010.08.006.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behaviour. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp.105-142). NY: Academic Press.

Parvatiyar, A., & Sheth, J. N. (2000). The Domain and Conceptual Foundation of Relationship Marketing. In J. N. Sheth, & A. Parvatiyar (Eds.), *Handbook of Relationship Marketing* (pp. 3-38). Thousand Oaks, CA: Sage.

Peng, S. C., Li, C. S., & Liu, C. C. (2016). Deregulation, Pricing Strategies, and Claim Behaviour in the Taiwan Automobile Insurance Market. *Emerging Markets Finance and Trade, 52*(4), 869-885. doi:10.1080/1540496X.2015.1117869.

Query, J. T., Hoyt, R. E., & He, M. (2007). Service quality in private passenger automobile insurance. *Journal of Insurance Issues, 30*, 152-172.

Rundmo, T., & Moen, B. E. (2005). Demand for Risk Mitigation in Transport. In T. Rundmo, & B. E. Moen (Eds.), *Risk Judgement and Safety in Transport* (pp. 31-45). Trondheim: Rotunde Publ. no. 87.

Rundmo, T., & Nordfjærn, T. (2013). Predictors of demand for risk mitigation in transport. *Transportation research part F: traffic psychology and behaviour. 20*, 183-192. doi:10.1016/j.trf.2013.04.004.

Saito, K. (2009). Does asymmetric information matter in the early insurance market? Evidence from the auto insurance market. *Applied Economics, 41*(21), 2653-2666. doi:10.1080/00036840701335546.

Sapelli, C., & Vial, B. (2003). Self-selection and moral hazard in Chilean health insurance. *Journal of health economics, 22*(3), 459-476. doi:10.1016/S0167-6296(02)00121-2.

Shi, P., Zhang, W., & Valdez, E. A. (2012). Testing Adverse Selection with Two-Dimensional Information: Evidence from the Singapore Auto Insurance Market. *Journal of Risk and Insurance, 79*(4), 1077-1114. doi:10.1111/j.1539-6975.2012.01472.x.

Prof. Cristian Mihai Dragos, Ph.D.
Babes-Bolyai University
of Cluj Napoca
Faculty of Economics
and Business Administration
Department of Statistics, Forecasting
and Mathematics
cristian.dragos@econ.ubbcluj.ro

Assoc. Prof. Simona Laura Dragos, Ph.D.
Babes-Bolyai University
of Cluj Napoca
Faculty of Economics
and Business Administration
Department of Finance
simona.dragos@econ.ubbcluj.ro

Abstract

ESTIMATING CONSUMERS' BEHAVIOUR IN MOTOR INSURANCE USING DISCRETE CHOICE MODELS

Cristian Mihai Dragos, Simona Laura Dragos

Insurance is a financial service in which consumption is highly affected by the characteristics of the potential buyer and his perceptions about the offered product. Motor insurance with its two components – the Motor Third Party Liability Insurance (MTPL) and the Motor Damage insurance – constitutes the largest line of business of the non-life insurance sector in Europe. The present study models the voluntary motor damage insurance consumer behaviour using discrete choice models, hypothesizing a hierarchical and a non-hierarchical decision. The sample consists of 311 car owners from Cluj County, Romania. The econometric estimations use binary logit, multinomial logit and nested logit models. The predictive power of these models is compared by means of the Receiver Operating Characteristic curve for discrete choice models. The results reveal that the main factors affecting the purchase of a voluntary motor insurance policy are risk preference/aversion, the distance travelled by car, the driver's education level and the ratio between the driver's income and the car price. In contrast to previous studies who estimated the risk profile only through proxy variables without accounting for any behavioural aspects, our study has successfully integrated the risk profile of the policyholders as a self-standing explanatory variable. Since the explanatory variables are representative not only for a particular geographical area, the highlighted behaviour may be applied to all cases where motor damage insurance is voluntary.

Key Words: *Motor insurance, risk profile, consumer behaviour, discrete choice models, ROC curve.*

JEL Classification: *C25, G22, M31.*

DOI: *10.15240/tul/001/2017-4-007*