Evaluating Individual Heterogeneity in the Probability of Crash Involvement

Mariane Bobermin[1], Sara Ferreira

Research Centre for Territory, Transports and Environment, University of Porto, Rua Doutor Roberto Frias s/n, Porto, Portugal, mariane.bobermin@fe.up.pt
Research Centre for Territory, Transports and Environment, University of Porto, Rua Doutor Roberto Frias s/n, Porto, Portugal, sara@fe.up.pt

Abstract

The Driver Behavior Questionnaire (DBQ) [1] has been one of the most applied tools to evaluate driver behavior and its relationship with crashes. In this questionnaire, drivers are associated with a score representing the frequency of committing a series of behaviors, usually divided into errors and violations. Several studies have investigated the hypothesis that errors and violations have a distinct effect on crash liability [2]; however, many studies applied zero-order correlations [3], assuming that parameters were fixed across observations. A way to overcome this issue is to use a random-parameter approach that accounts for the heterogeneity across occurrences, particularly relevant when several individuals are analyzed based on variables describing their behavior. Therefore, an analysis considering age, sex, driving exposure, and DBQ scores was performed using a random-parameter logit model to investigate their influence on crash involvement.

A total of 1,321 drivers answered a validated DBQ version with 20 items and three dimensions (Errors, Ordinary Violations, and Aggressive Violations) [4]. A binary logit random-parameter model was used to assess the relationship between drivers' demographic characteristics and DBQ subscale scores with the associated probability of a self-reported crash. Since there were only a few drivers involved in more than one crash, the discrete choice nature of the data is met. The variables mentioned above are gathered in vector *x*, and the probabilities are represented in equations 1 and 2.

$$\text{Prob}(Y=1|x) = F(x, \beta) \qquad (1)$$
$$\text{Prob}(Y=0|x) = 1\text{-}F(x, \beta) \qquad (2)$$

The set of parameters β reflects the impact of changes in vector x on the probability. The logistic distribution (logit model) was assumed for $F(x, \beta)$, giving rise to the logit model of equation 3.

$$\text{Prob}(Y = 1|x) = \frac{e^{\beta x}}{1+e^{\beta x}} = \Lambda(\beta x) \qquad (3)$$

In order to account for potential heterogeneity within the data, a random parameters approach is applied to allow estimated parameters to vary across observations. A continuous density function of β is introduced, $f(\beta|\varphi)$, with φ referring to a vector of parameters of that density function (mean and variance). Thus, the random-parameter binary logit model is formulated as equation 4 [5]–[7].

$$\text{Prob}(Y = 1|x) = \int \Lambda(\beta x) f(\beta|\varphi) d\beta \qquad (4)$$

The results indicate that age, sex, and exposure did not have a random effect and were set to be fixed parameters. There is a decline in the probability of crash involvement as age increases, being female, and drivers who drove less frequently. It is important to note that only 3% of the sample is over 60 years old, which likely prevented the model from predicting any motor impairment effects of aging. Moreover, the three DBQ scores produced statistically significant random parameters.

Considering a normal distribution with mean and standard deviation from **Table 1** (Estimated value and Standard Error), we can derive that, in just a few observations, Aggressive Violations contributed positively to the likelihood of crashes (P: 35.41%; N: 64.59%), while the opposite was true for Ordinary Violations (P: 83.14%; N: 16.86%).

[1] * Corresponding author. Tel.: +351-937-064-103;
E-mail address: mariane.bobermin@fe.up.pt

As a matter of fact, the separation between Aggressive and Ordinary Violation is not commonsense between studies as sometimes the AV are not considered or are considered together with the OVs. Furthermore, although the validated DBQ version used in this study was based on a widespread version of the DBQ, it contains some differences in its items, which must be considered when comparing it with other studies. Nevertheless, it would be expected that, in general, high AV would increase the likelihood of crashes; therefore, this result suggests that, in general, the hostile behavior towards other users might not imply higher involvement in risky situations or the risks may be compensated by other factors, for example, the driver skills.

Table 1: Estimated Coefficients of the Logit Model with Random Parameters

Variable	Estimated value	Standard Error	P[Z>z]	Marginal Effects	Confidence Intervals
Nonrandom parameters					
Age	**-.023*** **	.006	.000	**-.003*** **	(-.034; -.012)
Sex	**-.362*** **	.125	.004	**-.052*** **	(-.607; -.118)
Exposure 02 - two to three times a week	.269	.216	.214	.038	(-.155; .692)
Exposure 03 - four to five times a week	**.357***	.213	.093	**.051***	(-.060; .774)
Exposure 04 - six or more times a week	**.394** **	.197	.046	**.056** **	(.008; .779)
Means for random parameters					
Constant	**-1.068*** **	.263	.000	--	(-1.585; -.552)
Score Errors	.065	.151	.669	.009	(-.232; .361)
Score Ordinary Violations	**.263*** **	.087	.003	**.037*** **	(.093; .433)
Score Aggressive Violations	**-.415*** **	.130	.002	**-.059*** **	(-.670; -.159)
Scale parameters for dists. of random parameters					
Constant	**.983*** **	.098	.000	--	(.0791; 1.175)
Score Errors	**.861*** **	.121	.000	--	(.0624; 1.097)
Score Ordinary Violations	**.274*** **	.055	.000	--	(.167; .381)
Score Aggressive Violations	**1.109*** **	.133	.000	--	(.848; 1.370)

***, **, * Significance at 1%, 5%, 10% level, respectively.

Concerning the Ordinary Violation, despite the general agreement of the outcome with the literature, i.e., the major positive contribution of the OV to crash likelihood, the model showed that OV could even be associated with a reduction in the probability of crashes in a few cases. The Error score represented the intermediate situation, meaning that in around half of the observations, it contributed negatively, and in the other half, positively, resulting in a non-significant mean, despite the confirmed heterogeneity (significant standard deviation) [9]. This result supports the theory that errors are unintentional, without a motivational component, and thus, may occur randomly among drivers.

Note that all these observations are not possible with a logit model with fixed parameters. The standard binary logit model (not shown), which included all the independent variables, had very similar goodness-of-fit results to the suggested model. However, the variables Errors and Aggressive Violations were not significant, and only the highest level of exposure (Exposure 04 – drive frequency of six or more times a week) was found statistically significant.

Overall, we conclude that the driver behavior, translated into DBQ scores, has a heterogeneous effect on crash involvement. This result may be due to factors influencing crash occurrences not included in the analysis; therefore, the same DBQ score results in different crash probabilities. Indeed, a limitation of this study is the few explanatory variables considered to explain the complexity of crashes. Nevertheless, using a random-parameter approach represents a step forward in the DBQ analysis. Another limitation is that a questionnaire was used to assess driving behavior, which may not describe real-world behavior. In order to mitigate this problem, a future study is planned, using a driving simulator to investigate the influence of driver profile (DBQ scores).

Keywords: Driver Behavior Questionnaire; Binary Logit Models; Random Parameters; Crashes

References

[1] J. Reason, A. Manstead, S. Stephen, J. Baxter, and K. Campbell, "Errors and violations on the roads: A real distinction?," *Ergonomics*, vol. 33, no. 10–11, pp. 1315–1332, 1990.

[2] J. C. F. de Winter, D. Dodou, and N. A. Stanton, "A quarter of a century of the DBQ: some supplementary notes on its validity with regard to accidents," *Ergonomics*, vol. 58, no. 10, pp. 1745–1769, Oct. 2015.

[3] A. E. af Wåhlberg, P. Barraclough, and J. Freeman, "The Driver Behaviour Questionnaire as accident predictor: A methodological re-meta-analysis," *J. Safety Res.*, vol. 55, pp. 185–212, Dec. 2015.

[4] M. Bobermin, S. Ferreira, and C. I. de Campos, "New Version of Driver Behaviour Questionnaire to investigate Driver Behaviour on Rural Roads," in *35º Congresso de Pesquisa e Ensino Em Transporte Da ANPET*, 2021.

[5] W. H. Greene, *Econometric Analysis*, SEVENTH ED. 2012.

[6] S. P. Washington, M. G. Karlaftis, and F. L. Mannering, *Statistical and Econometric Methods for Transportation Data Analysis*, Second edi. CRC Press, 2011.

[7] D. A. Schroeder, "Discrete choice models," in *A Dictionary of Transport Analysis*, 2010, pp. 77–95.

[8] K. E. Train, *Discrete Choice Methods with Simulation*, 2nd ed. Cambridge University Press, 2009.

[9] M. Sarrias, "Discrete choice models with random parameters in R: The rchoice package," *J. Stat. Softw.*, vol. 74, no. 10, 2016.