An Investigation of human Factors Impact on Tehran Urban Crash Severity

Amirreza Mamdoohi, Fatemeh Nazari, and Mohamad Hossein Noruzoliaee
Tarbiat Modares University / Transportation Planning, Tehran, Iran
Email: {armamdoohi, fatemeh.nazari, oliaee_hossein}@modares.ac.ir

***Abstract*—Due to the considerable social and economic burden traffic accidents imposed on societies, strategies to reduce crash severity and potential for crash (frequency) are of interest to transportation agencies. As injury-severity data are generally represented by discrete categories such as injury and property damage only, a variety of discrete choice methodological techniques can be applied to analyze crash-severity data. This paper presents a binary model for predicting severity of Tehran urban car-car collisions which can be used in safety planning and enforcements. Human impact and collision type variables are employed to act as surrogates for point of impact. Results indicate that fastening seat belt decreases the probability of accidents resulting in injury. Furthermore, disregarding regulations, as a human reason of an accident, results in the most severe consequence (injury/fatality) compared to other human reasons. On the other extreme, as a consequence of accidents occurring due to non-human reasons, property damage only is the most probable outcome. Finally, drivers involved in front to front collision types are most prone to injury. Other factors in decreasing order are: front to rear, front to side, other types of collision, rear to side, and side to side.**

***Index Terms*—human factor impacts, crash severity, binary logit, car-car crash**

I. Introduction

Strategies to reduce crash injury severity and crash frequency are of interest to transportation agencies because of the considerable social and economic burden traffic accidents impose on societies. These strategies, as how they approach the issue, can be categorized in two groups: strategies to lessen the potential for accidents (i.e. reducing crash frequency) and strategies to decrease the severity level of a crash. This paper deals with the severity side of strategies.

According to the Global Status Report on Road Safety, nearly 1.3 million people are killed annually and between 20 and 50 million people get injured every year around the globe in roadway crashes. The estimated cost of highway crashes to governments worldwide is estimated to be 518 billion US dollars [1].

In this respect, even worse conditions are confronted in developing countries. These countries, due to problems such as lack of proper safety actions in road infrastructure, vehicle design and lack of specialized human forces in transport safety planning suffer from more accidents compared with developed countries [2].

Factors affecting accidents can be compiled of five broad categories: driver characteristics (including driver gender, age, restraint system use, alcohol consumption and drug use), vehicle characteristics (including vehicle type and vehicle age), roadway design and operational attributes (including roadway class, speed limit, types of intersection and traffic control device), environmental factors (including time of day and road surface condition), and crash characteristics representing the crash situation and events (including driver ejection, vehicle rolled over, air bag deployment, manners of collision and collision location) [3].

Injury-severity data are generally represented by discrete categories such as fatal injury or killed, incapacitating injury, non-incapacitating, possible injury, and property damage only. Hence, a variety of discrete choice methodological techniques have been applied to analyze crash-severity data: binary outcome models, ordered discrete outcome models, and unordered multinomial discrete outcome model. The statistical methods employed by researchers have primarily relied on the nature of the dependent variable and various methodological issues associated with the data [4].

This paper presents a binary outcome model for predicting severity of urban car-car collisions which can be used in safety planning and enforcements. We employ the human impact and collision type variables to act as surrogates for point of impact. In the final specification of model, statistically insignificant variables were removed. The overall process was guided by intuition, judgment and insights from previous literature.

The paper structure is as follows: In the next section, a concise review on the research conducted on the issue is presented. The third section deals with the proposed methodology and the database. The analysis of results and model validation is presented in section 4. Finally, conclusions on the research are briefly discussed in the last section.

II. Literature

Within the area of crash severity research, discrete outcome models have become the prevailing analytical technique. Such models allow for an examination of crash severity outcomes while treating injury level as either an ordinal or a nominal variable. Recent advances in

Manuscript received January 5, 2014; revised March 25 2014.

doi: 10.12720/jtle.2.2.100-103

computational efficiency, as well as in the development and application of simulation methods that can be used to approximate integration, have allowed for the development of more flexible models that allow for new insights into those factors affecting severity outcomes [4]. Studies that have looked at binary injury-severity outcomes such as injury vs. non-injury crashes or fatal vs. non-fatal crashes have used common discrete outcome models such as the binary logit and binary probit models. However, because of the characteristics of crash injury data and the variety of methodological issues associated with these characteristics, a number of variations of the simple binary logit and probit have been employed [4].

Savolainen and Mannering used logit models for analyzing one and multi vehicle involving accidents and came to conclusion that factors such as age, type of accident, alcohol, wearing helmet, and speeding can affect accident severities [5].

Abay and colleagues contributed to the injury severity modeling literature by developing a multivariate probit model of injury severity and seat belt use decisions of both drivers involved in two-vehicle crashes. The modeling approach enables the joint modeling of the injury severity of multiple individuals involved in a crash, while also recognizing the endogeneity of seat belt use in predicting injury severity levels as well as accommodating unobserved heterogeneity in the effects of variables. The proposed model is applied to analyze the injury severity of drivers involved in two-vehicle road crashes in Denmark [6].

Yasmin and Eluru empirically compared the ordered response and unordered response models in the context of driver injury severity in traffic crashes. The alternative modeling approaches considered for the comparison exercise include: for the ordered response framework-ordered logit (OL), generalized ordered logit (GOL), and for the unordered response framework- multinomial logit (MNL), nested logit (NL), and ordered generalized extreme value logit (OGEV) model [3].

III. Methodology

Accident prediction models are widely used to estimate the severity and/or frequency of accidents for a given spatial unit over a certain period of time [7]. One of the most important practical applications of these models aims to identify factors involved in accident severity and their underlying quantitative effect.

Accidents are mutually exclusive events. In other words, an accident is in, and can only be in, one category of severities (i.e. injury or property damage only). Such data, involving two types of discrete outcomes (i.e. count and discrete choice), can be modeled using a discrete model like binary logit.

A. Model Structure

Binary logit models are based on observed choices made by individual units. To represent the attractiveness of the choices (alternatives), the concept of utility, which is a convenient theoretical construct defined as what the individual seeks to maximize, is used. Alternatives, per se, do not produce utility; this is derived from their characteristics and those of the individual. For example, the observable utility is usually defined as a linear combination of variables.

Binary logit models have some useful properties: the explanatory variables included in the model can have explicitly estimated coefficients. Binary logit models allow for a flexible representation of the policy variables considered relevant to the study. The coefficients of the explanatory variables have a direct marginal utility interpretation (i.e. they reflect the relative importance of each attribute). The general form of utility function is as follows:

$$V_i = U_i + e_i \quad (1)$$

where:

V_i: random utility of choice i (accident severity level i)

U_i: deterministic component of i

e_i: random error term of i

The utility of each alternative to a specific decision maker can be expressed as a function of the observed attributes of the alternatives and the observed characteristics of the decision maker. Once the distribution of the error term, e_i, is specified, the distribution of the utilities can be determined, and the choice function can be calculated explicitly.

Logit model, however, can be derived from the concepts of random utility and utility maximization by assuming that the random terms of each utility function are independently and identically distributed with a Gumbel (double exponential) distribution function [8, 9].

There are different kinds of Logit Models which depend on the utility function and number of deciding alternatives. In case of selecting among two choices, binary logit models can be applied for estimating the probability of selecting/happening each choice. The general structure of this model is as follows:

$$P_A = \frac{exp\ (U_A)}{exp\ (U_A) + exp\ (U_B)} = \frac{1}{1 + \exp(U_B - U_A)} = \frac{1}{1 + \exp(\Delta U)} \quad (2)$$

$$\Delta U = U_B - U_A = (a_j - b_j)X_j \quad (3)$$

where:

P_A : probability of choice A

U_A : utility function of choice A

U_B : utility function of choice B

X_j : independent variable j

a_j : coefficient of independent variable j in U_A

b_j : coefficient of independent variable j in U_B

The utility functions usually include a set of parameters that are statistically estimated from observed choices. For estimation of these models, the common method is Maximum Likelihood Estimation (MLE). Stated simply, a maximum likelihood estimator is the value of the parameters for which the observed sample is most likely to have occurred [10].

B. Data

A statistical sample was obtained from car-car crash reports in Tehran metropolitan area during 15 months (2009-2010). Table I shows the names of records and their respective descriptive statistics. As indicated, the data contain information about accident severity level (injury/fatality and property damage only), human impacts (fastening or not fastening seat belt, and human reasons of an accident), and collision types (direction of two colliding cars).

TABLE I. DESCRIPTIVE STATISTICS ANALYSIS OF DATA

Name	Frequency	Relative frequency (%)
Severity (injury)	335	17.7
Severity (property damage only)	1559	82.3
Seat belt (fastened)	1517	80.1
Seat belt (not fastened)	377	19.9
Human reason (disregarding regulations)	324	17.1
Human reason (driver in a hurry)	1080	57.0
Human reason (no human reason)	278	14.7
Human reason (other human reasons)	212	11.2
Collision type (front to front)	161	8.5
Collision type (front to rear)	647	34.2
Collision type (front to side)	537	28.4
Collision type (rear to side)	57	3.0
Collision type (side to side)	270	14.3
Collision type (other types)	222	11.7

C. Model Variables

The dependent variable (severity of accident) is a dichotomous variable with the value 0 for property damage only and 1 for injury. The independent variables account for human and accident impacts on accident severity (Table II).

TABLE II. INDEPENDENT VARIABLES OF THE MODEL

Variable name	Description	Variable type
SBT_yes	Seat belt (fastened)	Dummy (human)
HUR_hurry	Human reason (hurry)	Dummy (human)
HUR_no	Human reason (no human reason)	Dummy (human)
HUR_else	Human reason (else)	Dummy (human)
COLL_FR	Collision type (Front to Rear)	Dummy (accident)
COLL_FS	Collision type (Front to Side)	Dummy (accident)
COLL_RS	Collision type (Rear to Side)	Dummy (accident)
COLL_SS	Collision type (Side to Side)	Dummy (accident)
COLL_else	Collision type (else)	Dummy (accident)

IV. RESULTS

Table III shows the results of modeling severity of accidents in Tehran metropolitan area. The contribution of each variable to accident severity is as follows.

1) Seat belt

As expected, fastening seat belt (SBT_yes), compared to not fastening it, decreases the probability of an accident resulting in injury with a negative coefficient of -0.43535.

2) Human reason

The human reasons with decreasing impact on human injury are HUR_hurry, HUR_else, and disregarding regulations. Whereas, the non-human reason has the highest impact on reducing human injury.

3) Collision type

Model results indicate that drivers involved in front to front collision types are most prone to injury. Other factors in decreasing order are: front to rear, front to side, other types of collision, rear to side, and side to side.

TABLE III. BINARY LOGIT MODEL RESULTS OF ACCIDENT SEVERITY

Var	Coef	Std error	t-stat
SBT_yes	-0.43535	0.12687	-3.432*
HUR_hurry	-0.94878	0.13143	-7.219*
HUR_no	-1.48463	0.23183	-6.404*
HUR_else	-0.39352	0.19195	-2.050*
COLL_FR	-0.30809	0.15530	-1.984*
COLL_FS	-0.32197	0.16127	-1.996*
COLL_RS	-0.91152	0.40014	-2.278*
COLL_SS	-1.33904	0.24173	-5.539*
COLL_else	-0.55668	0.22370	-2.489*

* Significance at %5

** Property damage only is base choice (severity level)

The Log Likelihood (LL) functions of the estimated model and the two base models are calculated below. As seen, the estimated model improves the LL function. Moreover, model fit is examined by pseudo-R^2 which shows a relatively good fit of the model.

$$LL\,(\beta) = -845.35$$

$$LL\,(c) = -883.78 \quad LL\,(0) = -1312.82$$

$$\rho^2 = 1 - \frac{LL\,(\beta)}{LL\,(c)} = 0.356 \qquad \rho_c^2 = 1 - \frac{LL\,(\beta)}{LL\,(0)} = 0.043$$

To determine whether the model is statistically significant, the LL function of the estimated model should be compared to that of the base model. The test to compare the LL functions is called the LL ratio test (Equation 4). If the -2LL value exceeds the critical Chi-square value, the null hypothesis that the specified model is no better than the base comparison model is rejected [11], [12].

$$-2\,(LL_{base\ model} - LL_{estimated\ model}) \sim \chi^2_{(\alpha,k)} \tag{4}$$

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - \varepsilon_i)^2}{\varepsilon_i} \tag{5}$$

where:

n: total observations

O_i: real observations

ε_i: predicted value (Expected)

The LL ratio test executed in Nlogit software package, indicates a p-value of 0.0000000 which confirms the overall model significance. Table IV shows the cross tabulation of predicted severity levels (estimated by the proposed model) and the corresponding observed (real) data. Results indicate a percent correct value of 72, which means that 72 percent of the observations are correctly replicated (a rather good value for such a model).

TABLE IV. COMPARISON OF MODEL RESULTS AND OBSERVATIONS

Observed \ Predicted	0	1	Sum
0	1291	267	1559
1	262	72	335
Sum	1554	339	1894

Table V indicates the marginal effect and elasticity of the variables. As seen, the variables with the highest and the lowest marginal effect are HUR_no (no human reason) and COLL_FR (Front to Rear collision), respectively. Also, the results of elasticity analysis of variables show that accident severity is inelastic to almost all independent variables with HUR_hurry (hurry as the crash human reason) and COLL_RS (Rear to Side collision) as the most and least elasticity, respectively.

TABLE V. MARGINAL EFFECT AND ELASTICITY OF VARIABLES

Variable	Marginal Effect	Elasticity
SBT_yes	-0.06495	-0.31802
HUR_hurry	-0.13636	-0.47532
HUR_no	-0.14340	-0.12866
HUR_else	-0.04852	-0.03320
COLL_FR	-0.04081	-0.08523
COLL_FS	-0.04205	-0.07288
COLL_RS	-0.09262	-0.01704
COLL_SS	-0.13299	-0.11589
COLL_else	-0.06585	-0.04718

V. CONCLUSIONS

Strategies to reduce crash injury severity and crash frequency are of interest to transportation agencies because of the considerable social and economic burden traffic accidents impose on societies. These strategies, as how they approach the issue, can be categorized in two groups: strategies to lessen the potential for accidents (i.e. reducing crash frequency) and strategies to decrease the severity level of a crash. This paper deals with the severity side of strategies. As injury-severity data are generally represented by discrete categories such as injury and property damage only, a variety of discrete choice methodological techniques can be applied to analyze crash-severity data. This paper presents a binary outcome model for predicting severity of Tehran urban car-car collisions.

Results indicate that, as expected, fastening seat belt, compared to not fastening it, decreases the probability of an accident resulting in injury. Furthermore, disregarding regulations, as a human reason of an accident, results in the most severe consequence (injury/fatality) compared to other human reasons. On the other extreme, as a consequence of accidents occurring due to non-human reasons, property damage only is the most probable outcome. Finally, the model shows that drivers involved in front to front collision types are most prone to injury.

Suggestions for further research can be as follows: Considering the dependent variable in 3 values (fatal/injury/property damage only) by using MNL models. Another suggestion can be simultaneous modeling of crash severity and frequency.

REFERENCES

[1] World Health Organization (WHO), Global status report on road safety: Time for action, Geneva, Switzerland, 2009.

[2] World Health Organization (WHO), World Report on Road Traffic Injury Prevention, 2003.

[3] Y. Shamsunnahar and N. Eluru, "Evaluating alternate discrete choice frameworks for modeling crash injury severity," Technical Paper, Department of Civil Engineering & Applied Mechanics, McGill University, CA, 2012.

[4] P. Savolainen, F. Mannering, D. Lord, and M. Quddus, "The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives," *Journal of Accident Analysis and Prevention*, vol. 43, pp. 1666-1676, 2011.

[5] P. Savolainen and F. Mannering, "Probabilistic models of motorcyclists' injury severities in single and multi-vehicle crashes," *Journal of Accident Analysis and Prevention*, vol. 39, pp. 955–963, 2007.

[6] K. A. Abay, R. Paleti, and C. H. R. Bhat, "The joint analysis of injury severity of drivers in two-vehicle crashes accommodating seat belt use endogeneity," Workshop at Technical University of Denmark, 8-9 November, 2012.

[7] C. Wang, A. Quddus, and G. Ison, "Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model," *The 90th Annual Meeting of the Transportation Research Board*, USA, 2011.

[8] J. D. Ortuzar and L. G. Willumsen, *Modeling Transport*, John Wiley and Sons, New York, USA, 2011.

[9] A. Kanafani, "Transportation demand analysis," *McGraw-Hill Book Company*, New York, USA, 1983.

[10] M. Ben-Akiva and S. R. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press, USA, 1985.

[11] S. Washington, M. Karlaftis, and F. Mannering, *Statistical and Econometric Methods for Transportation Data Analysis*, 2nd ed. Chapman and Hall/CRC, Boca Raton, FL, USA, 2011.

[12] D. A. Hensher, J. M. Rose, and W. H. Greene, *Applied Choice Analysis*, UK: Cambridge University Press, 2005.

Amir Reza Mamdoohi was born in Tehran, Iran in 1965. He received his PhD in Transportation Planning and Engineering from Sharif University of Technology, Tehran, Iran, in 2005. His major field of study includes Transportation Planning and Modeling and Transportation Demand Management.

He is currently an assistant professor of Civil and Environmental Engineering Faculty at Tarbiat Modares University.