LJAST
Libyan Journal of Applied Science and Technology
مجلة ليبيا للعلوم التطبيقية والتقنية

Volume 11 Issue 01
Joune 2023
ISSN 2958-6119

مجلة ليبيا للعلوم التطبيقية والتقنية

Comparison Between Logistic Regression And Spike Models For The Probability Of Prediction Travelling Behavior

Aldukali Salem I. Almselati[1], Tarek Muftah D. Gnefeda[2] and Nagi Salem Ali Tueger[3]

[1,2,3] High Institute For Comprehensive Professions-Tripoli

Abstract

The objective of the present empirical study was to test “bridging assumptions,” that’s, to see how closely the analysis of the objective macro-structural conditions determining the travel mode choice comply with the situation as perceived by the M respondents. The adopted methodology is based on a contingent valuation (CV) survey. Stated Preference (SP) surveys, also called self-stated preferences for research or services, have been widely applied in the areas of marketing and travel demand modelling, separately or jointly with RP surveys with observed choices of product purchase or service use. Stated preference (SP) survey is essential for, and helpful in, evaluating the willingness-to-pay for mode changing, and investigating people’s acceptability and perceptions. Two SP surveys (attitudinal and hypothetical choices) were implemented on car users. An attitudinal survey asked the respondents to express their responses to various situations (e.g., if they would shift to public transport if the service improved). The two models were compared based on their predictability and accuracy indicators and the results revealed that spike model manifested higher accuracy than logit model for predication.

Keywords: willingness-to-pay, logit model, travel time.

1. Introduction

A review of applications of soft transport policy measures (Friman et al. 2010, Hsieh, H. S. 2020) shows their usefulness in motivating private car users to freely shift their travel modes. Long-term effects, synergies between soft and hard policy measures, assessments of individual differences, differences in determinants of participation and of changes in travel, generalizability, and non-quantifiable benefits are known to be the prevailing gaps (Richter et al. 2011; Piras et al. 2020)

The data from attitudinal surveys were used to estimate the potential impacts of improving public transportation services on travel mode shift and to quantify respondents’ preferences for the improvement, as well as the effects of the three policies proposed (reduction travel time of public transportation) on increasing transit use.

Two common modelling approach including logistic regression and spike modelling was applied to investigate the travelling behavior among private car user in terms of some policies such as increase reduction travel time for shifting toward public transportation.

Volume 11 Issue 01
Joune 2023
ISSN 2958-6119

2. Logistic Regression for Prediction Travelling Behaviour in Kajang City

Logistic regression is a statistical method for analysing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Logistic regression generates the coefficients (and its standard errors and significance levels) of a formula to predict a logit transformation of the probability of presence of the characteristic of interest. According to one above mentioned policies, this method was applied to predict the private car user's behavior in term of shifting toward public transportation.

2.1 Reducing the Travel Time

The private car users were asked about their behaviors regarding to different situation of reducing travel time at 5 different scenarios (rate) including 15%, 30%, 45%, 60% an75% reduction rate in travel time. The result shows that if the public transport traveling time can be reduced by 75%, 5.7% of the respondents would take it instead. If the reduction were 60%, 20.8% would shift and if the reduction were 45%, 30.8% would shift. Similarly, if the reduction were 30%, 24.7% will be shift, and if the reduction were 15%, 18.1% would like to shift (Fig 1.1).

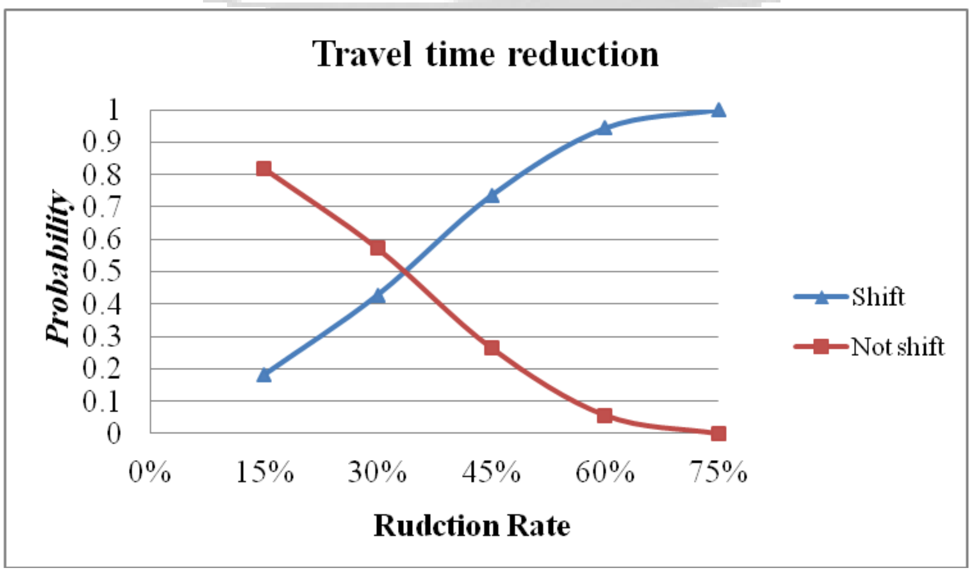


Fig 1.1 Shifting to public transport if the travel time improved

Volume 11 Issue 01
Joune 2023
ISSN 2958-6119

Table 1.1 shows reduction rate in public travel time with respect to survey result of private car users and probability of prediction (P) values. P value is derived from equation 1.1 which involve constant and alpha (α) values to verify the logistic prediction model used in this study.

$$P = \frac{1}{1 + De^{\alpha\,(X)}} \qquad (1.1)$$

Where P = Probability prediction of shift to PT
D = constant
α = coefficient of x
x = level of travel time reduction
e = the base of natural logarithms (approximately 2.718)

Table 1.1 Survey results and data calibration

Reduction Travel Time	Cumulative survey result (P)	1-P	1-P/P	Ln(1-P/P)
15%	0.181	0.819	4.525	1.510
30%	0.428	0.572	1.337	0.290
45%	0.735	0.265	0.361	-1.020
60%	0.943	0.057	0.061	-2.806
75%	0.999	0.001	0.001001	-6.907

To calculate the probability prediction of shift to PT, results of table 1-1 were used for model calibration process by using in Regression analysis between reduction travel time and Ln(1-P/P) to determine the alpha (α) and (D) values. According to results of regression analysis, Table 1.2, the alpha (α) is equal to regression weight of (X) "reduction travel time" and the intercept of regression equation is considered as Ln D.

Table 1.2 The result of Regression analysis

Model parameter	**Intercept**	**X (reduction rate travel time)**
Coefficients	4.192	-1.992
Standard Error	1.101	0.332
t Stat	3.808	-6.005
P-value	0.032	0.009
Lower 95%	0.689	-3.049
Upper 95%	7.695	-0.937
Lower 95.0%	0.689	-3.049
Upper 95.0%	7.695	-0.937

Ln D = 4.192

α = -1.993

D = 66.152

R2 = 0.975

Considering the P value equal to 0.009, which was significant because P value was less than 0.05 levels the model was significant. R2 (R square) as another criteria for evaluating the power of regression analysis (close value to one indicates the model's is more strong and accurate), in the current analysis the R2 value was 0.975 which shows the high level of predictability of model.

Appling "D "and "α" in equation 1.2, the logit model for reduction travel time defined as:

$$P = 1/(1 + 66.152e^{-1.993\,(X)}) \quad (2.2)$$

Table 1.3 shows the respective survey data and the prediction of logit model corresponding to the traveling reduction time. Respective results can be substituted into the calibrated logit model to validate its functionality. If we take travel time rate15%, cumulative P value of 0.100, which had maximum 8.1%, score differently from the survey data. Then the probability prediction to shift by travel time reduction rate 15% rate is 18.1% to public transport.

Table 1.3 Survey results and logit model results

Travel time reduction	Survey result (P) Cumulative	Result from logit model (P)
15%	0.181	0.100
30%	0.428	0.449
45%	0.735	0.857
60%	0.943	0.978
75%	0.999	0.997

Fig 1.2 shows the graph plotted with the proportion of reduction travel time by private car users that is based on the logit model and survey results.

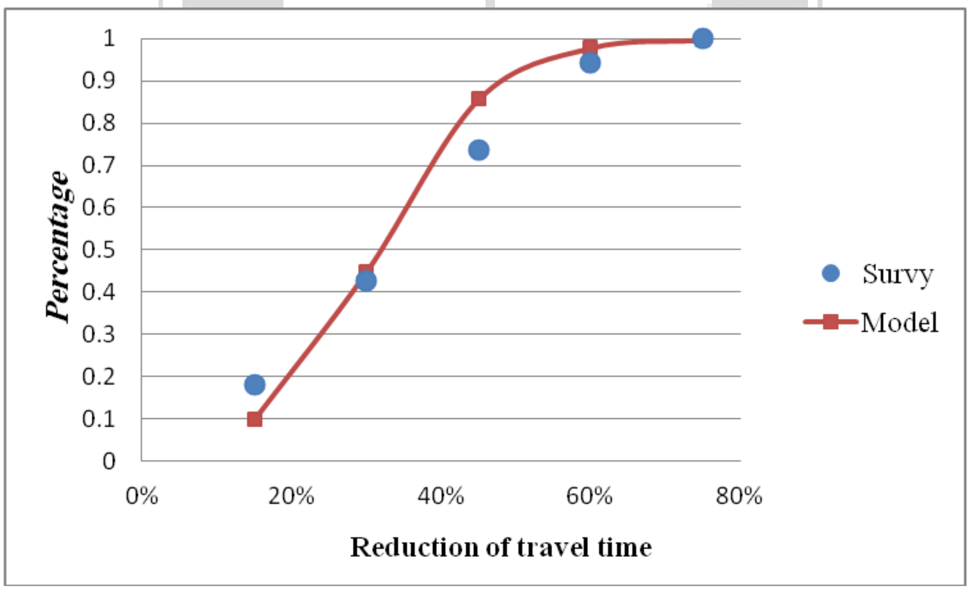


Fig 1.2 Improving travel time by private car

3. Spike Model for Prediction Travelling Behaviour in Kajang City

The spike model was proposed by Kristroöm (1997) to resolve the issues that arise when the reported price that users are willing to pay is zero or negative (Kriström, 1997). In contingent valuation studies respondents are often assumed to have positive willingness to pay the market price for buying certain goods. Popular distributional assumptions such as the lognormal, or Weibull along with other popular models such as the logit and the probit model, provide examples of cases where zero willingness to pay is excluded. Although zero willingness to pay is not allowed in these kinds of studies, zero or even negative responses are not uncommon in contingent valuation studies. Zero consumption of certain goods may arise when the goods in question do not contribute to the individual's utility. Moreover, there are goods that contribute negatively to some consumer's utility, like what meat means to a vegetarian.

In such a situation, the spike model has been proven superior to other traditional models, such as logit and probit. The major advantage of the spike model is that it can detect responses by respondents that are equal to zero or negative. In this study, many respondents (above 10%) are not willing to accept any percentage of travel time or travel cost improvement to shift from their respective private vehicles to public transportation. Moreover, the results of traditional models are always underestimated in comparison with those of the spike model, given that zero willingness to pay is not considered in the former.

The spike value can be defined as following the equation by setting $A_1 = 0$ in the cumulative distribution function of the respondent who is not willingness to pay or accept the amount A_1:

$$Spike = \frac{1}{1+\exp{(\alpha)}} \tag{1.3}$$

Where α is the marginal utility of improving travel conditions after adopting the model based on traveling time reduction.

3.1 Traveling Time Improvement for All Trips

Table 1.4 contains the estimated model parameters α and β, estimated average willingness to accept (WTA), and estimated spike value for different percentage of travelling time reduction rate. As we can see, 15% reduction rate has the lowest mean WTA 0.177, while the highest mean WTA 0.763 belongs to 75% reduction rate. Moreover, 45% reduction rate has the lowest spike value 0.190, which means 19% of people said no to 45% reduction rate and didn’t accept to shift to public transports, while 81% said yes and accepted to shift. On the other hand 75% reduction rate has the highest spike value 0.357 which means 35.7% of people said no to 75% reduction rate and did not accept to shift to public transports, while 64.3% said yes and accepted to shift. Total spike value 0.267 shows the probability of zero willingness to accept the traveling time reduction rate, which means 26.7% of people didn’t accept to shift to public transport in any traveling time reduction rate, and 73.3% accepted to shift to public transport. Total mean WTA 0.35 shows that 35% traveling time reduction rate encourages people to shift to public transport.

The spike value can be defined using the following equation by setting $A_1 = 0$ in the cumulative distribution function of the respondents who are not willing to accept the amount A_1:

$$Spike = \frac{1}{1+\exp(\alpha)} \tag{1.4}$$

Where α is the marginal utility of improving travel time after adopting the model based on traveling time reduction.

Table 1.4 Traveling time improvement

Reduction Rate	α	β	Mean WTA	Spike Value
15%	0.756	6.422	0.178	0.319
30%	1.262	5.258	0.287	0.221
45%	1.446	3.647	0.455	0.191
60%	0.989	2.119	0.616	0.271
75%	0.587	1.348	0.763	0.357
Total	1.008	3.759	0.351	0.267

The expected price users are WTA can be derived as $A_1 \rightarrow \infty$ in the cumulative distribution function of the respondent who are not willingness to accept the amount A_1 as in equation 1.5:

$$E(willingness\ to\ accept) = \frac{1}{\beta}\ln[1+\exp(\alpha)] \tag{1.5}$$

Where α is the marginal utility of improving the travel time, and β is the marginal utility of willingness to accept.

Table 1.5 Predicted values and actual values of spike model for travel time reduction

Reduction Rate	Mean WTA	Expected WTA
15%	0.178	0.15
30%	0.287	0.30
45%	0.455	0.45
60%	0.616	0.60
75%	0.763	0.75

Fig 1.3 shows the graph plotted with the proportion of reduction travel time by private car users based on the spike model and survey results.

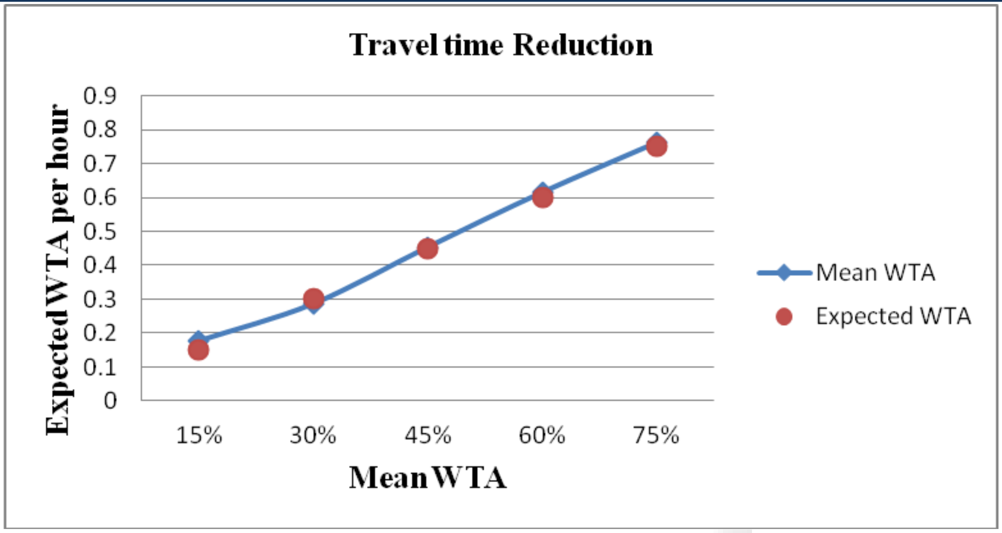


Fig 1.3 Fitting of spike model for traveling time improvement

4. Comparison Between Spike Model and Logistic Regression Model For Reduction Travel Time

The accuracy of the model can be evaluated by comparing predicted values to actual values (Musiek et al. 2011). The correlation between actual and predicted values for each mode can be specified. In theory, correlation coefficient can range from -1 to 1. Maximum positive would have a correlation coefficient (r) equal to +1, and maximum negative would have a correlation coefficient (r) equal to -1. Uncorrelated tests would have zero correlation (r= 0), which means that the predicted values identified by one model is not correlated with actual values.

R-squared is the square of the correlation between the model's predicted values and the actual values which is called “coefficient determination” and it ranges from 0 to 1. The greater value of R2 between the predicted values through the model and the actual values indicate the accuracy of model prediction. Some studies applied simple linear regression between actual values and predicted values to evaluate the accuracy of models (Oludolapo et al. 2012).

4.1 Travel time reduction

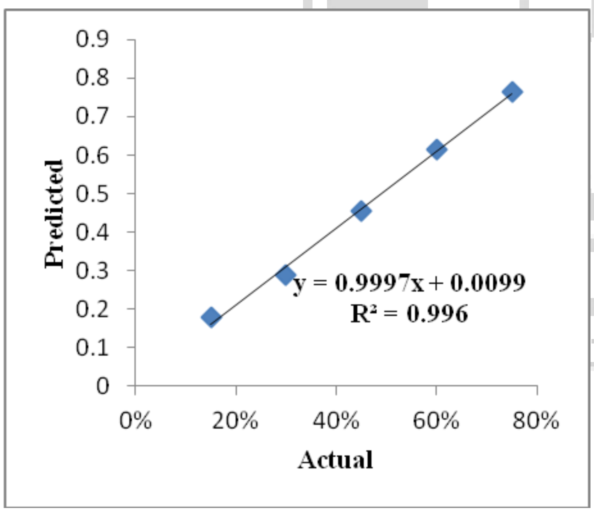
Predicted values were calculated (table 1-6) using both models (Spike – Logit). Evaluation was done to find the agreement between actual values and predicted values for travel time reduction as shown in the scatter plot and fitting line in Figure 1.4.

Volume 11 Issue 01
Joune 2023
ISSN 2958-6119

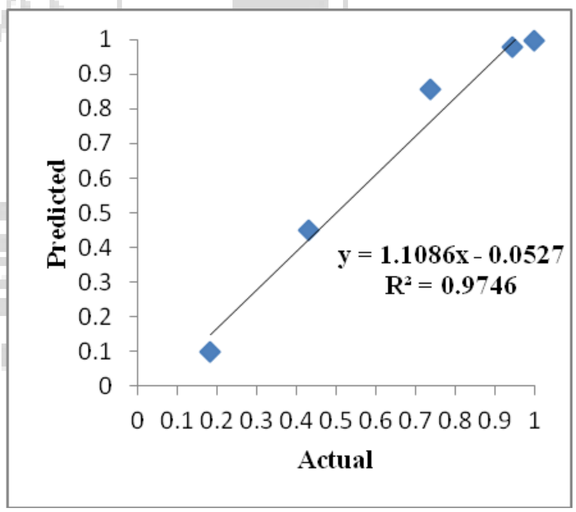
According to the results it can be seen that spike model had higher goodness of fitting level, R2 = .996 than logistic model R2 = 0.975 which means both models predictability are more than 95% but the spike model predictability was higher than logistic model. Therefore, it can be concluded that the spike model was more accurate than logit model for prediction of car user behavior regarding to travel time reduction.

Table 1.6 Predicted values and actual values for both models of travel time reduction

Travel Time Reduction Rate	**Spike Model**		**Logit Model**	
	Actual	predicted	Actual	predicted
15%	0.15	0.178	0.181	0.0998
30%	0.30	0.287	0.428	0.449
45%	0.45	0.455	0.735	0.857
60%	0.60	0.616	0.943	0.978
75%	0.75	0.763	0.999	0.997
R2	0.996		0.975	



Spike Model



Logit Model

Figure 1.4 Travel time reduction scatter plot and fitting line for both models

5. Conclusion

This study presented the results of data analysis. In the first part, the mode choice models for public transportation based on their trips including work and shopping trips were studied to find the most effective factors on respondent's mode choice. This was followed by validation of these models. The binary model examined the characteristics of private car vs. public transport such as the travel time, demographic and socio-economic characteristics to determine their influence on mode choice behavior.

Logit and spike models were developed on a hypothetical choice stated preference survey for car users in Kajang. It was carried out to predict the effect of possible policy measures on mode shifting behavior improvement of public transportation such as reduction travel time. In the last part, the two models were compared based on their predictability and accuracy indicators and the results revealed that spike model manifested higher accuracy than logit model for predication.

References

[1] Dargay, J., & Gately, D. (1999). Income's effect on car and vehicle ownership, worldwide: 1960–2015. *Transportation Research Part A: Policy and Practice*, *33*(2), 101-138.

[2] Farr, M., Stoeckl, N., & Beg, R. A. (2014). The non-consumptive (tourism)'value'of marine species in the Northern section of the Great Barrier Reef. *Marine Policy*, *43*, 89-103.

[3] Hsieh, H. S. (2020). Transport policy evaluation based on elasticity analysis with social interactions. *Transportation Research Part A: Policy and Practice*, *139*, 273-296.

[4] Kitamura, R., Golob, T. F., Yamamoto, T., & Wu, G. (1999). Accessibility and auto use in a motorized metropolis.

[5] Piras, F., Sottile, E., Tuveri, G., & Meloni, I. (2022). Does the joint implementation of hard and soft transportation policies lead to travel behavior change? An experimental analysis. *Research in transportation economics*, *95*, 101233.

[6] Richter, J., Friman, M., & Gärling, T. (2010). Review of evaluations of soft transport policy measures. *Transportation: Theory and application*, *2*(1), 5-18.

[7] Richter, J., Friman, M., & Gärling, T. (2011). Soft transport policy measures: Gaps in knowledge. *International journal of sustainable transportation*, *5*(4), 199-215.