MODELING EFFECTS OF TRAVEL-TIME RELIABILITY ON MODE CHOICE USING PROSPECT THEORY

Sepehr Ghader, Graduate Research Assistant (Corresponding Author)
Department of Civil and Environmental Engineering, Maryland Transportation Institute
1173 Glenn Martin Hall, University of Maryland
College Park, MD 20742
Phone: (703)638-4046, Email: sghader@umd.edu

Aref Darzi, Graduate Research Assistant
Department of Civil and Environmental Engineering, Maryland Transportation Institute
1173 Glenn Martin Hall, University of Maryland
College Park, MD 20742
Phone: (240)760-0845, Email: adarzi@umd.edu

Lei Zhang, Herbert Rabin Distinguished Professor
Director, Maryland Transportation Institute
Department of Civil and Environmental Engineering
1173 Glenn Martin Hall, University of Maryland
College Park, MD 20742
Phone: (301)405-2881, Email: lei@umd.edu

ABSTRACT

This paper utilizes the well-known prospect theory to study how travel mode choice is affected by travel-time reliability. Prospect theory is developed to model decision making under risk. Travel-time reliability is related to uncertainties in travel-time, and its effect on travel behavior is a good candidate for applying prospect theory. The prospect theory is combined with a discrete choice model to build a mode choice framework. The choice model parameters, in addition to prospect theory parameters, are estimated using a combination of revealed preference household travel survey data and empirically observed reliability data for a real-world application. This application showcases how real-world observational data can be used in a prospect theory-based mode choice model. The proposed model's estimated parameters are discussed and goodness-of-fit is compared with the utility-based mean-variance model. The paper focuses on mode choice, but its extension to other choice dimensions is discussed.

KEYWORDS: Prospect Theory, Mode Choice, Travel-Time Reliability, GPS Data

INTRODUCTION

In transportation literature, there are two main approaches to considering reliability within utility maximization theory. The first is called mean-variance approach, in which a dispersion measure of travel-time distribution, such as variance or standard deviation, is added to the utility function by linear summation. The second is called scheduling approach, in which expected earliness penalty and lateness penalty are added to the utility function by linear summation.

Travel choice with alternatives that have unreliable attributes is a type of decision making under risk. Prospect theory (PT, see [1]) is a widely used theory in psychology and behavioral economics for decision making under risk. In PT, it is assumed that under uncertainty, the decision maker chooses the alternative that has the highest prospect value. The prospect value is evaluated relative to a reference point, using a probability weighting function and a value function. The theory states that gain or loss of each prospect is evaluated by comparison with a reference point. The role of value and weighting function is to model perceived uncertainty, risk-aversion and diminishing sensitivity.

In transportation literature, a number of studies have used PT to explain travel behavior [2]. De Palma et al. [3] described the theory of integrating risk and uncertainty into discrete choice models and utilizing non-expected utility-based models within discrete choice framework. Avineri and Prashker [4] conducted one of the earliest studies that used CPT to find the market share between two hypothetical unlabeled routes. In a numerical example, Avinery [5] used PT to study the choice between two bus lines with different headway distributions. Gao et al. [6] used simulated choice data and parameters estimated by Kahneman and Tversky [7] to compare expected utility theory (EUT) and PT for route choice in a risky network . Michea and Polak [8] studied train travelers' behavior using different models such as CPT. Senbil and Kitamura [9] collected a survey asking respondents about three consecutive commuting days, and used PT in the departure-time choice context. Connors and Sumalee [10] modeled network equilibrium on a hypothetical network using PT.

The authors see decision making under unreliable travel times as a good context to apply PT, as the theory is designed for decision making under uncertainty. The review of transportation literature revealed some valuable studies incorporating PT in travel behavior, but most of the studies were in the experimental setting. They had to collect the required data, and use limiting assumptions on the reference point, in addition to value and weighting function parameters. Besides, focus on mode choice is very limited.

This paper proposes a general framework that uses revealed-preference survey data and historical travel-time data to model effect of perceived reliability on mode choice based on PT. It estimates value and weighting function parameters, and the reference point from the data, using few assumptions. The proposed framework's data requirement is similar to the requirement of widely-used mean-variance model. It does not require any further data collection or experiment.

METHODOLOGY

The proposed methodology introduces two new PT-based terms to capture effect of travel-time and travel-time variation in the utility function:

$$PU_i = \beta_{RTT} * RTT_i + \beta_{TTP} * TTP_i + \epsilon_i \qquad (1)$$

In Equation 1, PU_i is the perceived utility of alternative i, RTT_i is the reference travel-time of alternative i, and TTP_i is defined as the prospect value for alternative i's travel-time, obtained from travel-time distribution of alternative i using value and weighting functions. TTP captures effect of perceived travel-time variation or perceived reliability. ϵ_i is the error term assumed to be distributed as extreme value type 1 Gumbel in order to have logit formulation for choice probabilities. The following items must be assumed in order to calculate TTP from travel-time distribution at time of departure:

- The reference point
- The form of value function
- The form of probability weighting function

The assumption of reference point can have a significant effect on the results. The reference travel-time might be different for each traveler. In this paper it is assumed that the reference travel-time is a traveler's expected travel-time, which can be the traveler's average experienced travel-time between the origin and destination. In cases where an individual's experienced travel-time data is not available, mean value of travel-time distribution at time of departure between the origin and the destination can be used for RTT of all individuals traveling between the origin and the destination at the same departure-time.

Considering that the reference point (RTT) is defined as the expected travel-time, individuals schedule their travels based on the reference point. Any travel-time not equal to the reference travel-time leads to either early or late arrival (disutility). Therefore, any variation from the reference travel-time is considered a loss. As a result, value function is only defined for the loss domain. However, as early and late arrivals may have different effects, parameters of the value function are assumed to be different for travel-times bigger than the reference travel-time and travel-times smaller than that the reference.

Value and weighting functions are assumed to follow the form suggested by Kahneman and Tversky [1]. Equation 2, (with λ (degree of loss aversion) set to 1, and different β parameters for travel-times bigger and smaller than the reference travel-time (β_{early} and β_{late}) is assumed to be the value function for the loss domain.

$$V(X) = (-\lambda)(-X)^{\beta} \quad X < 0 \tag{2}$$

Equation 3 is assumed to be the form of probability weighting function,

$$W(p_m) = \frac{(p_m)^{\gamma}}{[(p_m)^{\gamma} + (1-p_m)^{\gamma}]^{\frac{1}{\gamma}}} \tag{3}$$

where $W(p_m)$ is the weighted probability of the mth outcome, and γ is the probability weighting parameter.

If the travel-time density function at time of departure (showing day-to-day variation of travel-time at time of departure) is available as a continuous function over travel-time domain, then TTP is obtained by the following equation:

$$TTP = \int_{tt \in TT} V(-|tt - RTT|) * W(\ f(tt))\ d(tt) \tag{4}$$

In this equation, TT is travel-time domain, $V()$ is the value function, $W()$ is the probability weighting function, and $f()$ is travel-time density function. Equation 4 is applying Kahneman & Tversky's PT formulation over a continuous range of values.

In some real-world applications, continuous function of travel-time distribution is not available. In many cases, only a limited number of travel-time observations are available for the time of departure. In such cases, the modeler can estimate the continuous travel-time distribution function using the observed data, or approximate the discrete version by forming a set of discrete bins for travel-time.

As the proposed model's data requirement is similar to the mean-variance model, it is interesting to compare their performance and goodness-of-fit using one dataset.

EMPIRICAL APPLICATION

Drivers tend to dislike high travel-time variations resulting from accidents, bad weather, roadwork, fluctuation in demand, etc. Compared to driving, rail usually has more reliable travel-times since it follows a fixed schedule of operation. It would be interesting to explore how this difference in travel-time reliability would affect travelers' choice between these two modes. Two models based on Equations 5 and 7 were estimated for this mode choice problem, and their performances were compared. The estimated PT-based parameters showed interesting interpretation.

Data

Two datasets were used in this application: INRIX data and the 2007-2008 TPB-BMC household travel survey. INRIX data was used to obtain origin-destination (OD) level travel-time observations, and 2007-2008 TPB-BMC household travel survey data was used to provide information on observed trips and their chosen mode.

INRIX Data: INRIX collects road speed information from millions of mobile phones, connected cars, trucks, delivery vans, and other fleet vehicles with GPS devices and provides these real-time and historical speed and travel-time data to users. Traffic Message Channels (TMC) location codes are used in INRIX data as location indices. Each road segment in the INRIX network has one unique TMC location code. INRIX data can provide minute-by-minute travel-time and speed information throughout the day for selected road segments, which can be used to calculate travel-time and travel-time reliability for any time-of-day for a specific road by using variation among different days. Kim and Coifman validated the quality of INRIX data by comparing it with loop detector data. In this study, INRIX data was used to obtain historical travel-time observations between selected OD pairs [11].

2007-2008 TPB-BMC Household Travel Survey (HHTS): This survey was conducted by the Transportation Planning Board (TPB) from February 2007 to April 2008 in order to gather information about demographics, socioeconomic and trip-making characteristics of residents in Washington and Baltimore metropolitan areas.

In this application example, OD pairs in the Washington, D.C. area that have both rail and driving trips in the travel survey were selected and studied. It was assumed that in these OD pairs, both travel modes are available and are competing with each other. In total, there were 160 OD pairs with both rail and driving trip records and available INRIX data. The reported travels between these 160 OD pairs formed a major component of this application example. In these 160 OD pairs, 179 rail trips, 193 driving trips, and only 37 trips of other travel modes were observed.

Due to the small number of trips in other modes, it was assumed that rail and driving are the only available alternatives. Therefore, only driving and rail was considered in the mode choice model. 179 rail trips and 193 driving trips (372 in total) formed the observations in the mode choice problem.

The INRIX historical travel-time data between the origin and the destination at the time of departure was collected for these 372 survey observations. One year of INRIX data corresponding to the year 2012 was obtained, and day to day travel-time variation (after removing outliers) at time of departure was used to get travel-time distribution for each observation. It should be noted that 2012 was the year closest to 2008 with the proper coverage on the studied OD pairs. The inconsistency between the survey year and the travel-time year is one of the limitations of this application example. TMC level INRIX data had to be processed to obtain OD-level travel-time observations. The process for obtaining OD-level travel-time observations from TMC-based data is explained in Tang et al. and Mishra et al.

INRIX data was used for driving travel-times to calculate the following:

- $TTR_{driving}$: Travel-time reliability measure of the driving mode for mean-variance model. It was calculated by taking the standard deviation of INRIX observations between the origin and the destination at the time of departure.
- $RTT_{driving}$: Reference travel-time of the driving mode for the PT-based model. It was calculated by taking the mean of the INRIX observations between the origin and the destination at the time of departure.
- $TTP_{driving}$: Prospect value of the driving travel-time. It was calculated by Equation 8 using INRIX observations between the origin and the destination at the time of departure.

INRIX does not have information on rail travel-time. Due to the selection method of studied OD pairs, they all had both rail and driving trip records in the survey. Therefore, the travel-time of the rail mode between each OD pair could be obtained from the survey. It was obtained by averaging the reported travel-time of all the rail trips between the origin and the destination. To use consistent dataset, driving travel-times were also obtained from the survey similar to rail travel-times. These reported survey travel-times \ were used to calculate the following:

- TT_{rail}: Mean travel-time of the rail mode for the mean-variance model. It was calculated by averaging the reported travel-time of all the rail trips between the origin and the destination.
- $TT_{driving}$: Mean travel-time of the driving mode for the mean-variance model. It was calculated by averaging the reported travel-time of all the driving trips between the origin and the destination.
- RTT_{rail}: Reference travel-time of the rail mode. Similar to TT_{rail}, it was calculated by averaging the reported travel-time of all the rail trips between the origin and destination.

Rail was assumed to be very reliable; meaning that rail travel-time between each OD pair was assumed to be constant and equal to the average reported rail travel-time between the OD pair; therefore, rail's TTR for mean-variance model and TTP for the PT-based model were assumed to be equal to 0. Although rails usually are reliable and follow their schedule, but assuming them to be perfectly reliable is another limitation of this application example. Data on rail reliability can certainly improve this example.

Mode Choice Model Estimation Results

Mode-specific utility functions for the mean-variance-based model based on the previous notation can be seen in the following equations:

$$U_{driving} = \beta_{TT} * TT_{driving} + \beta_{TTR} * TTR_{driving} + \varepsilon_{car} \quad (10)$$

$$U_{rail} = \beta_{TT} * TT_{rail} + \varepsilon_{rail} \quad (11)$$

Equations 11 and 12 do not include any alternative specific constants as the mode share is almost equal between the two modes. As discussed earlier, rail travel-time was assumed to be reliable, so no reliability term was added to rail's utility function. Mode-specific perceived utility functions for the PT-based model based on the previous notation can be seen in the following equations:

$$PU_{driving} = \beta_{RTT} * RTT_{driving} + \beta_{TTP} * TTP_{car} + \varepsilon_{car} \quad (12)$$

$$PU_{rail} = \beta_{RTT} * RTT_{rail} + \varepsilon_{rail} \quad (13)$$

For the sake of simplicity, all error terms were assumed to be extreme value type 1 Gumbel (Multinomial Logit [12]). It should be noted that the simplest type of discrete choice model is assumed here to showcase the application of the proposed methodology. More advanced types of discrete choice models can also be used for comparing mean-variance and PT-based approaches.

Maximum Likelihood Estimation was used to estimate model parameters in the Multinomial Logit model described by utilities in Equations 10 and 11. The Maximum Likelihood Estimation was coded in R [13]. ***TABLE*** summarizes the estimation results for the mean-variance-based model.

TABLE 1. Estimation Results for the Mean-Variance-Based Model

Parameter	Coefficient	Std. Err.	t stat
β_{TT}(t stat)	-0.010	0.004	-2.331
β_{TTR}(t stat)	-0.112	0.039	-2.850
Log Likelihood	252.337		

In order to estimate the Multinomial Logit described by perceived utilities in Equations 12 and 13, TTP should be known. TTP is a function of three unknown parameters γ (probability weighting function parameter), β_{early} (value function parameter for early arrivals) and β_{late} (value function parameter for late arrivals). The likelihood function was maximized as a function of all unknown parameters (β_{RTT}, β_{TTP}, γ, β_{early}, β_{late}) using BFGS method [14]. $\gamma = 1.194$, $\beta_{early} = 0.597$, and $\beta_{late} = 0.931$ maximized the utility function. After keeping these three parameters fixed, Multinomial Logit model corresponding to Equations 12 and 13 was estimated. ***TABLE*** summarizes the estimation results of the PT-based model.

TABLE 2. Estimation Results for the PT-Based Model

Parameter	Coefficient	Std. Err.	t stat
β_{RTT}(t stat)	-0.013	0.005	-2.961
β_{TTP}(t stat)	0.584	0.182	3.199
Log Likelihood	251.803		

The following items should be discussed about the estimation results:

- The estimated coefficients for travel-time and travel-time unreliability in the mean-variance-based model were negative and statistically significant with $\alpha = 0.05$. It was consistent with the expectation, because higher travel-time and higher unreliability are both undesired. The estimated coefficient of reference travel-time was negative and statistically significant with $\alpha = 0.05$ as expected. Higher reference travel-time for a mode, similar to higher travel-time in the mean-variance-based model is undesirable, as travelers prefer to spend less time traveling between their origin and destination. The estimated coefficient for the travel-time prospect was positive and statistically significant with $\alpha = 0.05$. A higher travel-time prospect means more unreliability, which is undesirable. Considering that TTP is defined as a negative variable (The value function is only defined over loss domain), the estimated coefficient was expected to be positive. Therefore, the estimated coefficient for TTP had the expected sign.
- The value of γ should be positive by definition. A negative γ will result in a decreasing function, while the probability weighting function needs to be increasing. A γ smaller than 1 will result in a weighting function that increases low probabilities and decreases high probabilities. A γ value bigger than 1 has the reverse effect. The estimated value of γ was a positive value bigger than 1, which showed that travelers exaggerate high probabilities and play down low probabilities. It implied that travelers ignore travel-time conditions with very low probability. While this finding can be reasonable in this contexts, it contradicts the main body of PT literature in which γ is assumed to be smaller than 1. This finding highlights the possible variation of PT parameters in different contexts.
- The value function adds diminishing sensitivity to the model. The value of β must be between 0 and 1 for the value function to comply with the diminishing sensitivity property. A lower β value corresponds to more extreme diminishing sensitivity. The estimated values of β_{early} and β_{late} were both between 0 and 1, as expected. The estimated value for early

departures were significantly lower than late arrivals. It showed that the sensitivity toward late departures do not diminish as acutely as the sensitivity toward early arrivals. It implied that travelers care more about the amount of their lateness than the amount of their earliness. The estimated β_{late} was very close to 1 which showed very weak diminishing sensitivity for late arrivals.

- The performances of the two models were relatively similar. The PT-based model had a slightly better likelihood.

SUMMARY AND CONCLUSION

This paper introduced a methodology to consider the effect of travel-time reliability on mode choice behavior using prospect theory. Prospect theory was introduced in the fields of psychology and behavioral economics to model decision making under risk. Mode choice with unreliable travel-time attribute was considered as a case of decision making under risk. As an application, proposed model was estimated in a mode choice between unreliable driving and reliable rail using a combination of revealed preference survey data and empirically observed travel-time data. The estimation results showed that the perceived travel-time distribution is different from the actual travel-time distribution. The results implied that travelers ignore conditions with low probability when they evaluate reliability of a mode. The results also showed that travelers' sensitivity toward being early decreases more rapidly than their sensitivity toward being late. The performance of the proposed model was compared with the widely used mean-variance model and showed slight improvement.

REFERENCES

1. Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society.* 1979:263-291.
2. Li Z, Hensher D. Prospect theoretic contributions in understanding traveller behaviour: a review and some comments. *Transport Reviews.* 2011;31(1):97-115.
3. De Palma A, Ben-Akiva M, Brownstone D, et al. Risk, uncertainty and discrete choice models. *Marketing Letters.* 2008;19(3-4):269-285.
4. Avineri E, Prashker JN. Sensitivity to travel time variability: travelers' learning perspective. *Transportation Research Part C: Emerging Technologies.* 2005;13(2):157-183.
5. Avineri E. A cumulative prospect theory approach to passengers behavior modeling: waiting time paradox revisited. Paper presented at: Intelligent Transportation Systems2004.
6. Gao S, Frejinger E, Ben-Akiva M. Adaptive route choices in risky traffic networks: A prospect theory approach. *Transportation research part C: emerging technologies.* 2010;18(5):727-740.
7. Tversky A, Kahneman D. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty.* 1992;5(4):297-323.
8. Michea A, Polak J. Modelling risky choice behaviour: Evaluating alternatives to expected utility theory. Paper presented at: 11th International Conference on Travel Behaviour Research, Kyoto2006.
9. Senbil M, Kitamura R. Reference points in commuter departure time choice: a prospect theoretic test of alternative decision frames. Paper presented at: Intelligent Transportation Systems2004.
10. Connors RD, Sumalee A. A network equilibrium model with travellers' perception of stochastic travel times. *Transportation Research Part B: Methodological.* 2009;43(6):614-624.
11. Kim S, Coifman B. Comparing INRIX speed data against concurrent loop detector stations over several months. *Transportation Research Part C: Emerging Technologies.* 2014;49:59-72.
12. Ben-Akiva ME, Lerman SR. *Discrete choice analysis: theory and application to travel demand.* Vol 9: MIT press; 1985.
13. RCore T. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. In: ISBN 3-900051-07-0, URL http://www. R-project. org; 2012.
14. Wright S, Nocedal J. Numerical optimization. *Springer Science.* 1999;35(67-68):7.