D-STOP

Technical Report 148

An Application of a Rank Ordered Probit Modeling Approach to Understanding Level of Interest in Autonomous Vehicles

Research Supervisor
Chandra Bhat

Project Title: A New Microeconomic Theory-Based Model for Ranking Data

September 2018

Data-Supported Transportation Operations & Planning Center (D-STOP)

A Tier 1 USDOT University Transportation Center at The University of Texas at Austin

D-STOP is a collaborative initiative by researchers at the Center for Transportation Research and the Wireless Networking and Communications Group at The University of Texas at Austin.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
D-STOP/2018/148		

4. Title and Subtitle	5. Report Date / 6. Performing Organization Code
An Application of a Rank Ordered Probit Modeling Approach to Understanding Level of Interest in Autonomous Vehicles	5. Report Date: September 2018 6. Performing Organization Code:
7. Author(s) Gopindra S. Nair, Sebastian Astroza, Chandra R. Bhat, Sara Khoeini, Ram M. Pendyala	8. Performing Organization Report No. Report 148
9. Performing Organization Name and Address Data-Supported Transportation Operations & Planning Center (D-STOP) The University of Texas at Austin 3925 W. Braker Lane, 4th Floor Austin, Texas 78759	10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTRT13-G-UTC58
12. Sponsoring Agency Name and Address Data-Supported Transportation Operations & Planning Center (D-STOP) The University of Texas at Austin 3925 W. Braker Lane, 4th Floor Austin, Texas 787591	13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes

Supported by a grant from the U.S. Department of Transportation, University Transportation Centers Program.

Project Title: A New Microeconomic Theory-Based Model for Ranking Data

16. Abstract

Surveys of behavior could benefit from information about people's relative ranking of choice alternatives. Rank ordered data are often collected in stated preference surveys where respondents are asked to rank hypothetical alternatives (rather than choose a single alternative) to better understand their relative preferences. Despite the widespread interest in collecting data on and modeling people's preferences for choice alternatives, rank-ordered data are rarely collected in travel surveys and very little progress has been made in the ability to rigorously model such data and obtain reliable parameter estimates. This paper presents a rank ordered probit modeling approach that overcomes limitations associated with prior approaches in analyzing rank ordered data. The efficacy of the rank ordered probit modeling methodology is demonstrated through an application of the model to understand preferences for alternative configurations of autonomous vehicles (AV) using the 2015 Puget Sound Regional Travel Study survey data set. The methodology offers behaviorally intuitive model results with a variety of socio-economic and demographic characteristics, including age, gender, household income, education, employment and household structure, significantly influencing preference for alternative configurations of AV adoption, ownership, and shared usage. The ability to estimate rank ordered probit models offers a pathway for better utilizing rank ordered data to understand preferences and recognize that choices may not be absolute in many instances.

17. Key Words	18. Distribution Statement
rank ordered probit model, rank ordered data, travel demand modeling, autonomous vehicle adoption and usage	No restrictions. This document is available to the public through NTIS (http://www.ntis.gov): National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161

19. Security Classif.(of this report)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	22	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Disclaimer

Acknowledgements

The authors recognize that support for this research was provided by a grant from the U.S. Department of Transportation, University Transportation Centers.

AN APPLICATION OF A RANK ORDERED PROBIT MODELING APPROACH TO UNDERSTANDING LEVEL OF INTEREST IN AUTONOMOUS VEHICLES

Gopindra S. Nair
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Tel: 512-471-4535; Email: gopindra.s.nair@gmail.com

Sebastian Astroza
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Tel: +56-41-220-3618; Email: sastroza@utexas.edu
and
Departamento de Ingeniería Industrial, Universidad de Concepción,
Edmundo Larenas 219, Concepción, Chile

Chandra R. Bhat (corresponding author)
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Tel: 512-471-4535; Email: bhat@mail.utexas.edu
and
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Sara Khoeini
Arizona State University
School of Sustainable Engineering and the Built Environment
660 S. College Avenue, Tempe, AZ 85287-3005
Tel: 480-965-3589; Email: sara.khoeini@asu.edu

Ram M. Pendyala
Arizona State University
School of Sustainable Engineering and the Built Environment
660 S. College Avenue, Tempe, AZ 85287-3005
Tel: 480-727-4587; Email: ram.pendyala@asu.edu

1. INTRODUCTION

Travel demand forecasting models often involve the use of choice models that are estimated and calibrated based on data about a single alternative that an individual chose. For example, mode choice models predict the mode that will be used for a trip, destination choice models predict a single destination that will be visited, and route choice models predict the route that will be adopted. While these choice contexts lend themselves to modeling the choice of a single alternative as an absolute, there may be a number of instances where choice behavior is not as well-defined. People may often operate in a more gray area, where they exercise a choice of a single alternative, but are willing to consider the consumption of other alternatives in the choice set in a rank ordered preferential scheme. Individuals may choose an alternative from a menu of choices, but there may have been a number of other choices that were ranked second, third, fourth, and so on. If, for any reason, the first choice was not available, then the individual would have chosen the second ranked alternative. In the travel behavior context, individuals exhibiting a choice of a single alternative may consume other alternatives in the choice set that are ranked lower (but not eliminated from consideration). The interest in understanding and modeling the ranked preferences of various alternatives motivates this study.

In the econometric literature, there appears to be a perception that ranked data are not very reliable because of the cognitive demands placed on respondents in ranking several alternatives. This perception is based on consistent empirical findings of unstable coefficients based on the rank depth used in the typical rank-ordered logit (ROL) model (see, for example, Foster and Maurato, 2002). This, of course, leaves the impression that the decrease in coefficient magnitudes is a result of increasing variance of the kernel extreme-value error term as one goes down the sequencing hierarchy; that is, individuals are more precisely able to form their utilities for alternatives and translate those utilities into an equivalent choice at higher levels of rankings than at lower levels of ranking. Or, equivalently, individual responses at lower ranking levels are not reliable, calling into question the veracity of using ranking data (relative to traditional choice data) as a means to collect individual responses in the first place (see, for example, Caparros et al., 2008; Scarpa et al., 2011). There then is a perceived need for econometric techniques that address the increasing variance of the kernel extreme-value error terms of choice models as one progresses down the ranking chain (see, for example, Hausman and Ruud, 1987; Fok et al., 2012).

In a recent paper, Yan and Yoo (2014) have indicated that correlating the finding of unstable coefficients in an ROL to intrinsic unreliability associated with ranking data may be misplaced. Specifically, they show through analytic computations and simulations that the attenuation of coefficients is a natural consequence of translating a ranking into a sequence of choice decisions. At the lower ranks, the systematic utilities of the remaining alternatives are likely to be closer to one another as individuals become more and more indifferent among the remaining choices, naturally pushing coefficients toward zero and leading to attenuation and seeming unreliability. Thus, any modeling approach that uses an explosion scheme for ranking data will naturally manifest coefficient attenuation.

At the same time, the ROL proposed by Beggs et al. (1981) is the only known utility maximizing model that is also consistent with a decision construction sequence in which the rank-ordered response may be exploded into pseudo-choice observations and viewed as a collection of sequential (and independent) decision-making processes in the same vein as the top-down psychological model of Luce (1959). The basic reason is that the conditional probability that an alternative is chosen at each rank is independent of the probability that another alternative has already been chosen at the earlier rank, a simple manifestation of the IIA (independence of irrelevant alternatives) property (see Beggs et al., 1981 for the derivation). But this explosion comes at the cost that, in the ROL, rankings from best to worst are not compatible with rankings from worst to best, as identified by Luce and Suppes's (1965) "impossibility theorem" (Theorem 51, page 357). In other words, this is another way to state that the ROL is an "impossible" structure. If individuals do not necessarily sequence from best to worst, the rank-ordered probit (ROP), introduced as a generalization of the Multinomial Probit model in Hajivassiliou and Ruud (1994), constitutes a more flexible behavioral structure to deal with rank-ordered data. Besides, the ROL maintains independence across the utilities of the ranked alternatives, while the ROP allows a full covariance structure across the alternatives (subject to identification considerations).

One reason for the continued use of the ROL for ranked data, despite its many limitations, is that the ROP can be difficult to estimate in the presence of many alternatives. However, recent analytic approximation techniques for estimation of probit-based models, as proposed by Bhat (2011) and Bhat (2018), resolve this issue. The objective of this study is to offer a robust methodological approach to model rank-ordered data while overcoming the limitations of past approaches. It is envisioned that the development of a computationally tractable methodological approach for modeling rank-ordered data would motivate behavioral researchers to increasingly collect such data, which presumably contains more information about *relative preferences* than regular single-discrete choice data. Among publicly available travel survey data sets, there are few – if any – instances where rank-ordered data is included. It would be desirable for the profession to collect rank-ordered data to a greater degree so that relative preferences for various choice alternatives could be better understood. The modeling methodology presented in this paper offers a significant step in facilitating the effective use of rank-ordered data and may provide the much needed impetus to increase the collection of such data.

The methodology presented in this paper is applied to a data set derived from the 2015 Puget Sound Regional Travel Study in which respondents were asked to rate their level of interest in alternative AV technologies and service modes. The ratings furnished by respondents were converted to rank-ordered data for ROP model estimation purposes. The application of the methodology presented in this paper is intended to serve as a demonstration of the ability of the ROP modeling approach to effectively utilize information contained in rank-ordered data when analyzing choice behaviors. Given the significant implications that autonomous vehicle (AV) technologies could have on the future of transportation systems, there is widespread interest in understanding and modeling possible adoption pathways in the marketplace. However, there are a number of different technologies and ways in which the technologies may be deployed, owned, and used. Because there is considerable uncertainty in how the technology will manifest itself in the market, it is difficult to identify well-defined alternatives from a survey design perspective and it is difficult for the respondent to choose a single AV technology or mode from among a set of alternatives. In fact, the technology may enter the marketplace in multiple formats, and people may be willing and interested to leverage alternative configurations in which the technology becomes available. It is therefore of interest to understand the *relative* preferences of individuals towards different AV technology forms to see which modes may gain traction faster than others, and identify policy instruments and information campaigns that could help communities achieve desired mobility outcomes.

The remainder of the paper is organized as follows. The next section presents a data description, the third section offers a description of the ROP modeling methodology, and the fourth section presents model estimation results. The fifth and final section offers a discussion and interpretation of the results together with concluding thoughts.

2. DATA AND SAMPLE DESCRIPTION

The data used for this study is derived from the Puget Sound Regional Travel Study that was conducted in 2014 and 2015. A comprehensive survey was conducted as part of the study, with respondents asked to provide detailed socio-economic and demographic information and complete a 24-hour travel diary that includes detailed attributes about all trips undertaken over the course of a day. The survey was conducted in the four county region of Puget Sound, including the counties of King, Kitsap, Pierce, and Snohomish. Travel diary days were limited to Tuesdays, Wednesdays, or Thursdays in order to collect travel information for days that are more typical weekdays.

A rather unique element of the survey is that it included a battery of questions to obtain detailed information about attitudes, values, and preferences as well as technology ownership and use behavior of individuals. For example, the survey collected information on smartphone ownership and the respondent's use of smartphone apps or websites to obtain travel information. There were a number of questions that gathered information on the frequency of use of car-share and ride-sourcing services, whether an individual had bike- or car-share subscription, and the importance that individuals attach to various considerations or criteria when making residential location decisions (e.g., proximity to highways, work place, transit, and local activities).

Besides all of these questions, the survey also included a set of questions to elicit information about people's preferences and level of interest in AV technologies and service modes. There were a number of questions that also elicited information about the extent to which individuals are concerned about various issues in relation to the adoption and implementation of autonomous vehicle technology. These issues include insurance, legal liability, safety, cybersecurity, and performance in poor weather or unexpected conditions. Individuals were asked to rate their level of concern with the technology on each of these issues. Thus, the survey has a rich amount of information related to people's preferences, level of interest, and concerns in the context of autonomous vehicle technologies and service modes.

The questions that provided the data for this study were those that asked individuals to rate their level of interest in alternative AV modes. The level of interest is not exactly a rank order (because an individual can express a high level of interest or the same level of interest for multiple alternatives), but the data were converted to a rank ordered data set for purposes of this study. The dependent variable in this study is the level of interest expressed on a five-point scale (very uninterested to very interested) for the following alternatives:

- Taking a taxi ride in an autonomous car with no driver present
- Taking a taxi ride in an autonomous car with a backup driver present
- Owning an autonomous car
- Participating in an autonomous car-share system for daily travel

These four alternatives were rated by each respondent on a five-point level-of-interest scale and these levels were converted to a rank ordered variable indicative of preferences for alternative service modes. Autonomous vehicles were defined to the respondents as follows: "Autonomous cars, also known as “self-driving” or “driverless” cars, are capable of responding to the environment and navigating without a driver controlling the vehicle. Advantages of autonomous car usage include the potential for reduced congestion, increases in parking capacity, and faster travel times.” (RSG, 2014).

Data was collected from 4,786 individuals aged 18 years and above. The analysis in this paper was limited to the adult respondents who did not have a proxy provide responses on their behalf. Individuals who indicated that they do not know their level of interest in any of the AV service modes were removed from the sample. In addition, individuals who rated the same level of interest for all four alternatives were removed from the sample because it is not possible to identify a rank-ordered preference for such individuals. At least one alternative needs to be ranked higher or lower than the others for a rank ordering to be derived. After filtering the data set and removing all records that have missing data for explanatory variables of interest, the final analysis sample included 1,365 persons. A summary of the analysis sample is furnished in Table 1. Details about the survey and the entire sample may be found in RSG (2014).

Table 1. Sample Characteristics (N=1,365 persons)

Characteristic	Categories	Distribution (%)
Gender	Male	54.5
	Female	45.4
Age	Age 18 - 35 years	26.6
	Age 36 - 65 years	52.5
	Age above 65 years	20.9
Number of Children in Household	Belongs to household with no children	79.5
	Belongs to household with a single child	10.6
	Belongs to household with multiple children	9.9
Employment Status	Unemployed	35.8
	Employed	64.2
Household Income	In household with income less than $25,000	11.8
	In household with income between $25,000 and $49,999	18.8
	In household with income between $50,000 and $74,999	15.8
	In household with income between $75,000 and $99,999	15.8
	In household with income $100,000 or greater	37.9
Education Attainment	Does not have a Bachelor's or Graduate degree	28.9
	Has a Bachelor's Degree but no Graduate degree	40.7
	Has a Graduate Degree	30.5
Household Size	Household size = 1	30.6
	Household size = 2	43.3
	Household size = 3	13.3
	Household size = 4 or more	12.8
Vehicle Count	Vehicle ownership = 0	11.9
	Vehicle ownership = 1	39.6
	Vehicle ownership = 2	36.7
	Vehicle ownership = 3	11.8

It is found that 54.5 percent of the sample is comprised of males. The sample shows a distribution of individuals across age groups, with 26.6 percent falling into the younger 18-35 year age bracket and 20.9 percent falling into the over 65 years of age range. A majority of the individuals (64.2 percent) are employed. Nearly 93 percent of the sample has a driver’s license, 40.7 percent have a Bachelor’s degree, and 30.5 percent have attained graduate education. This indicates that the sample is fairly well educated. Whereas 30.6 percent of the individuals live alone, 20.5 percent live in households with children. Nearly three-quarters reported smartphone ownership. The income distribution shows that 30.6 percent reside in households that make less than $50,000, but 37.9 percent reside in households that make $100,000 or more per year. Among respondents who commute to work (comprising 58.3 percent of the overall sample), it is found that 61.1 percent drive alone or carpool, 13.7 percent walk or bike, and 25.3 percent used transit. Thus, the level of transit usage is quite high in the analysis sample of this study.

The response distribution of level of interest in AV technology service modes and adoption is shown in Table 2.

Table 2. Distribution of Responses by Level of Interest in AV Service Modes

Characteristic	Very Interested	Somewhat Interested	Neutral	Somewhat Uninterested	Not at all Interested
Age					
Age 18-35 years					
AV as Taxi without Backup Driver	23.4	30.6	18.2	9.6	18.2
AV as Taxi with Backup Driver	13.2	37.2	25.3	11.9	12.4
AV Ownership	24.2	25.3	16.5	11.3	22.6
AV for Carshare	24.0	28.7	18.7	11.0	17.6
Age 35-65 years					
AV as Taxi without Backup Driver	14.4	29.3	19.1	12.4	24.8
AV as Taxi with Backup Driver	11.4	32.8	24.3	16.2	15.3
AV Ownership	15.6	22.6	18.4	12.4	31.0
AV for Carshare	13.4	22.2	21.1	13.5	29.8
Age > 65 years					
AV as Taxi without Backup Driver	8.4	25.3	13.0	11.2	42.1
AV as Taxi with Backup Driver	14.0	35.1	22.5	17.2	11.2
AV Ownership	8.1	20.0	15.1	8.8	48.1
AV for Carshare	5.3	8.1	14.0	9.5	63.2
Employment Status					
Employed					
AV as Taxi without Backup Driver	17.7	30.7	19.6	11.0	21.0
AV as Taxi with Backup Driver	11.3	35.2	24.8	13.7	15.1
AV Ownership	18.8	24.0	17.9	12.3	26.9
AV for Carshare	17.7	25.5	20.1	13.0	23.7
Unemployed					
AV as Taxi without Backup Driver	11.7	25.4	13.9	12.3	36.8
AV as Taxi with Backup Driver	14.5	33.1	23.1	18.0	11.3
AV Ownership	11.9	20.7	15.9	9.6	41.9
AV for Carshare	8.8	12.9	17.0	10.2	51.1

The table shows the percent of individuals in different age groups and employment status indicating various levels of interest for the alternative service modes. In each row of the table, percentages add up to 100 because individuals could only indicate one level of interest for each AV mode. However, numbers in columns are not likely to add up to 100 percent because individuals could give the same level of interest to multiple modes of AV technology adoption.

A few interesting patterns are seen in this table. It is found that younger individuals are more willing and interested, on average, in trying autonomous vehicles when compared with older age groups. Those who are aged 65 years or over show the lowest levels of interest in use or adoption of AV technologies and services. Among the various AV modes, the oldest age group appears more amenable to using AV as a taxi with a backup driver. On the other hand, younger individuals – who may be more comfortable with and trusting of technology – express a higher level of interest in use of AVs without backup drivers. Nearly one-half of those 65 years or over expressed no interest at all in AV ownership and 63 percent expressed no interest at all in AV use as part of a car-share system. The numbers are considerably different and lower (in the not interested at all category) for the younger age groups. In general, employed individuals express a greater level of interest in AV usage and adoption than unemployed individuals across all modes. Due to the high degree of correlation between age and employment status, it is not surprising to see that the patterns exhibited by those aged 65 years or over and the unemployed are very similar.

3. METHODOLOGY

This section presents an overview of the rank ordered probit (ROP) modeling methodology and formulation. Consider a consumer q (q = 1, 2,…, Q) who ascribes a utility U_{qi} to alternative i. For ease in presentation, assume that all alternatives are available for ranking for each individual, and that individuals provide a full ranking of the alternatives. These assumptions are innocuous and help in presentation. The utility function is written in a traditional fashion as follows:

$$U_{qi} = \boldsymbol{\beta}'_q \boldsymbol{x}_{qi} + \xi_{qi}\,;\ \boldsymbol{\beta}_q = \boldsymbol{b} + \tilde{\boldsymbol{\beta}}_q,\ \tilde{\boldsymbol{\beta}}_q \sim MVN_L(\mathbf{0}, \mathbf{\Omega}), \tag{1}$$

$\boldsymbol{x}_{qi}$ is an $(H \times 1)$-column vector of exogenous attributes (including a constant for each alternative, except one of the alternatives), and $\boldsymbol{\beta}_q$ is an individual-specific $(H \times 1)$-column vector of corresponding coefficients that varies across individuals based on unobserved individual attributes. Assume that the $\boldsymbol{\beta}_q$ vector is a realization from a multivariate normal distribution with a mean vector $\boldsymbol{b}$ and covariance matrix $\mathbf{\Omega} = \boldsymbol{LL}'$. Also assume that ξ_{qi} is independent and identically normally distributed across q, but allow a general covariance structure across alternatives for individual q. Specifically, let $\boldsymbol{\xi}_q = (\xi_{q1}, \xi_{q2}, ..., \xi_{qI})'$ ($I \times 1$ vector). Then, assume $\boldsymbol{\xi}_q \sim MVN_I(0, \mathbf{\Lambda})$. As usual, appropriate scale and level normalization must be imposed on $\mathbf{\Lambda}$ for identification of parameters. Specifically, only utility differentials matter in ranking choice models, just as in traditional discrete choice models (see Alvo and Yu, 2014). Taking the utility differentials with respect to the first alternative, only the elements of the covariance matrix $\mathbf{\Lambda}_1$ of $\tilde{\xi}_{qi1} = \xi_{qi} - \xi_{q1}$ $(i \neq 1)$ are estimable. However, the inference approach proposed here takes the difference in utilities in a specific way that is a function of the observed ranking (as discussed later). Thus, if individual q is observed to choose ranking θ_q, the covariance matrix $\mathbf{\Lambda}_{\theta_q}$ is desired for the individual. But, even though different differenced covariance matrices are used for various individuals in the sample, they must originate in the same matrix $\mathbf{\Lambda}$. To achieve this consistency, $\mathbf{\Lambda}$ is constructed from $\mathbf{\Lambda}_1$ by adding an additional row on top and an additional column to the left. All elements of this additional row and additional column are filled with values of zeros. An additional normalization needs to be imposed on $\mathbf{\Lambda}$ because the scale is also not identified. For this, normalize the element of $\mathbf{\Lambda}$ in the second row and second column to the value of one. Note that these normalizations are innocuous and are needed for identification. The $\mathbf{\Lambda}$ matrix so constructed is fully general. Also, as in multinomial probit (MNP) models, identification is tenuous when only individual-specific covariates are used (see Keane, 1992 and Munkin and Trivedi, 2008). In particular, exclusion restrictions are needed in the form of at least one individual characteristic being excluded from each alternative's utility in addition to being excluded from a base alternative (but appearing in some other utilities). However, these exclusion restrictions are not needed when there are alternative-specific variables.

The model above may be written in a more compact form by defining the following vectors and matrices: $\mathbf{U}_q = (U_{q1}, U_{q2}, ..., U_{qI})'$ ($I \times 1$ vector) and $\boldsymbol{x}_q = (\boldsymbol{x}_{q1}, \boldsymbol{x}_{q2}, \boldsymbol{x}_{q3}, ..., \boldsymbol{x}_{qI})'$ ($I \times H$ matrix). Then, Equation (1) can be written in matrix form as:

$$\mathbf{U}_q = \boldsymbol{x}_q \boldsymbol{b} + \boldsymbol{x}_q \tilde{\boldsymbol{\beta}}_q + \boldsymbol{\xi}_q, \tag{2}$$

Also, $\mathbf{U}_q \sim MVN_Q(\mathbf{V}_q, \tilde{\boldsymbol{\Xi}}_q)$, where $\mathbf{V}_q = \boldsymbol{x}_q \boldsymbol{b}$ and $\tilde{\boldsymbol{\Xi}}_q = \boldsymbol{x}_q \boldsymbol{\Omega} \boldsymbol{x}_q' + \boldsymbol{\Lambda}$.

To progress to estimation, define a contrast matrix for individual q based on the observed ranking r_q of alternatives for the individual. Specifically, let the first ranked alternative for individual q be r_q^1, the second r_q^2, and so on until the last-ranked alternative r_q^I. Then, the following $(I-1)$ inequalities should hold: $U_{qr_q^2} - U_{qr_q^1} < 0, U_{qr_q^3} - U_{qr_q^2} < 0, \ldots U_{qr_q^I} - U_{qr_q^{I-1}} < 0.$ To write these inequalities in vector notation, define a contrast matrix $\mathbf{M}_q$ with $(I-1)$ rows and I columns, each row representing one inequality and each column representing an alternative. Fill all the elements of the matrix with zeros. Then, in the first row, place an entry of '1' in the column corresponding to the second-ranked alternative, and a '-1' in the column corresponding to the first-ranked alternative (corresponding to the first inequality above). In the second row, place an entry of '1' in the column corresponding to the third-ranked alternative, and a '-1' in the column corresponding to the second-ranked alternative (corresponding to the second inequality above), and so on until placing an entry of '1' in the column corresponding to the penultimate-ranked alternative, and a '-1' in the column corresponding to the last-ranked alternative (corresponding to the last inequality above). Then, the inequalities in vector notation at the individual level take the form $\mathbf{u}_q = \mathbf{M}_q \mathbf{U}_q < \mathbf{0}_{I-1}$. and it can be seen that $\mathbf{u}_q \sim MVN_{(I-1)}(\boldsymbol{B}_q, \boldsymbol{\Xi}_q)$, where $\boldsymbol{B}_q = \mathbf{M}_q \mathbf{V}_q$ and $\boldsymbol{\Xi}_q = \mathbf{M}_q \tilde{\boldsymbol{\Xi}}_q \mathbf{M}_q'$. The likelihood of the observed sample (*i.e.*, individual 1 having the ranking r_1, individual 2 having the ranking r_2, …, individual Q having the ranking r_Q) may then be written succinctly as $\text{Prob}[\mathbf{u}_q < \mathbf{0}_{I\text{-}1}]$. The parameter vector to be estimated is $\boldsymbol{\theta} = (\boldsymbol{b}', \overline{\boldsymbol{\Omega}}', \overline{\boldsymbol{\Lambda}}')'$, where $\overline{\boldsymbol{\Omega}}$ is a column vector obtained by vertically stacking the upper triangle elements of the matrix $\boldsymbol{\Omega}$, and $\overline{\boldsymbol{\Lambda}}$ is another column vector obtained by vertically stacking the unique elements of the matrix. The likelihood function is:

$$L(\boldsymbol{\theta}) = \prod_{q=1}^{Q} \Phi_Q(\boldsymbol{\omega}_{\Xi_q}^{-1}(-\mathbf{B}_q), \boldsymbol{\Xi}_q^{\,*}), \tag{3}$$

where $\boldsymbol{\Xi}_q^{\,*} = \boldsymbol{\omega}_{\Xi_q}^{-1} \boldsymbol{\Xi}_q \boldsymbol{\omega}_{\Xi_q}^{-1}$, $\boldsymbol{\omega}_{\Xi_q}$ is the diagonal matrix of standard deviations of $\boldsymbol{\Xi}_q$.

The likelihood function above entails the evaluation of an $(I-1)$-dimensional integral, which can be easily evaluated using the Maximum Approximate Composite Marginal Likelihood (MACML) procedure. Details about this procedure and associated identification and estimation issues are provided in Bhat (2011).

As mentioned earlier, respondents were able to assign the same level of interest to alternative AV modes. Those who gave the same level of interest to all AV modes were eliminated from the original sample. Among the 1,365 respondents in the analysis sample, only 3.8 percent offered a unique level of interest to the four different AV modes. About 32 percent of respondents indicated three different levels of interest across the four AV modes, while 64 percent of respondents provided only two unique levels of interest among the four alternatives. This created a situation where a large segment of the sample essentially provided records where alternatives were ranked the same (i.e., had the same level of interest). To deal with the many cases where different alternatives had identical levels of interest, this study adopts a generalization of the likelihood function for ranking with ties based on the theoretical framework proposed in Allison and Christakis (1994) for the ROL model. It is assumed that even if the respondents provide identical ratings, they really have a preference ordering between the identically rated alternatives. However, since this preference ordering is unobserved, the likelihood is calculated as the probability of all utility values that can result in the rank ordering depicted by the respondent. For example, if an individual q assigns the first rank to alternative 3, second rank to two alternatives (say, 2 and 4), and third rank to alternative 1, the contrast matrix $\mathbf{M}_q$ is structured to represent the following four conditions:

$$U_{q2} - U_{q3} < 0, U_{q4} - U_{q3} < 0, U_{q1} - U_{q2} < 0, U_{q1} - U_{q4} < 0.$$

This is equivalent to the following:

$$\mathbf{M}_q \mathbf{U}_q = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} U_{q1} \\ U_{q2} \\ U_{q3} \\ U_{q4} \end{bmatrix}$$

Note that the number of rows in $\mathbf{M}_q$ varies depending on the number of ties at different rank levels.

It is recognized that there may be issues associated with converting a ratings-based set of questions into a rank ordering of alternatives when individuals are not entirely aware of the attributes of the alternatives presented to them (such as AV alternatives that are not yet present in the marketplace). To address this, level of interest variables, which constituted data in the preference rating format, were analyzed using the censored ranking approach described in Halvorsen and Layton (2006, Ch. 12). In models typically used to analyze ratings data, it is assumed that all respondents interpret the intervals of the rating scale in the same manner. This assumption may not be applicable in this study. Because AVs are not yet prevalent in the real world, there are likely to be individuals who do not have strong opinions about AVs. At the same time, there may be some respondents who are well-informed and thoroughly invested in the development of AV technologies and therefore have strong opinions about these emerging technologies. Individuals who are better informed about AVs may have a broader rating scale (i.e., distinguish between rating levels more clearly) than individuals who are less informed. Therefore, the censored ranking approach, where the interest ratings are converted to a ranking with possible ties, is used. That is, only the ordinal information that may be inferred from the ratings is used for analysis.

When the censored ranking approach is used, observations in which the same interest level is indicated against multiple alternatives is effectively not used in estimation (because all such observations have a likelihood of one). This also applies to observations in which individuals indicated that they have no opinion of AV services. However, if an individual provides a different rating to at least one alternative when compared with another, the ordinal information corresponding to this difference in interest levels is fully utilized by the ROP model. This is a substantial improvement over traditional discrete choice data where an individual chooses just one alternative from a choice set, thus providing virtually no insights about relative preferences for different alternatives. In fact, another alternative in a choice set may be a close second (in terms of preferences), and yet a traditional discrete choice data set would provide no information to this effect.

4. MODEL ESTIMATION RESULTS

This section presents model estimation results for the rank ordered probit (ROP) model proposed in this study. The model is applied to the Puget Sound Regional Travel Study survey data set to demonstrate the efficacy of the modeling methodology. The goal of this paper is to present the modeling methodology and underlying formulation with a view to offer a rigorous basis for the analysis and modeling of rank ordered data. Rank ordered data is generally not collected in travel behavior surveys, presumably because of the absence of a robust modeling methodology that can utilize the information contained in rank ordered data. It is envisioned that the availability of a computationally tractable methodology for analyzing and modeling rank ordered data may motivate the profession to increasingly collect such rich data on *relative preferences* among choice alternatives. The application presented in this section of the paper is merely intended to serve as an illustration of the efficacy of the modeling methodology and not necessarily shed significant new light on the behavioral factors influencing the adoption of alternative autonomous vehicle (AV) modes. As long as model estimation results are behaviorally intuitive, and findings are generally consistent with those reported in previous research, the modeling methodology may be considered appropriate for analyzing rank ordered data. It should, however, be recognized that the mere behavioral plausibility of the results does not necessarily prove the merits of the approach. Additional research aimed at validating the modeling methodology is needed to further demonstrate the merits of the ROP model and its ability to accurately predict choice behaviors. To do this, it would be preferable to use rank ordered data corresponding to choice alternatives that are currently in the marketplace and being used by consumers.

Table 3 presents model estimation results for the rank ordered probit (ROP) model estimated on the Puget Sound data. The model coefficients provide behaviorally intuitive interpretation and show the expected signs. The base alternative corresponds to the use of AV as a taxi without a backup driver. Older individuals are likely to rank AVs with backup drivers higher than other modes. Individuals above the age of 65 are also less inclined to use AV based car-share services. These individuals may not feel attuned to the idea of car-share services. Workers as well as individuals who own a smartphone favor the use of AVs with backup drivers less than the other modes, namely, AV ownership, AV as taxi without backup driver, and AV based car-share. It appears that these two groups have a high degree of comfort with the use of technology and therefore consider the presence of a backup driver unnecessary and undesirable. However, there is heterogeneity in this preference among individuals who are smartphone owners. This may be because of the higher stakes involved in allowing a computer to control a person’s vehicle, and there may be considerable heterogeneity in the degree of sensitivity to vehicle automation.

Table 3. Rank Ordered Probit Model Estimation Results

Alternative (*Base: AV as Taxi without Backup driver*)	AV as Taxi with Backup Driver		AV Ownership		AV Car-share	
Variables	**Coef**	**t-stat**	**Coef**	**t-stat**	**Coef**	**t-stat**
Constant	0.019	0.126	-0.869	-7.042	-0.583	-5.950
Age (*Base: < 35 years*)						
35-65 years	0.324	2.874	---	---	---	---
> 65 years	0.713	4.328	---	---	-0.424	-3.180
Worker	-0.250	-2.240	---	---	---	---
Household Income > $100,000	---	---	---	---	-0.229	-2.317
Education (*Base: Up to junior college, Graduate*)						
Bachelor's degree	---	---	---	---	0.311	3.293
Smartphone Ownership - Mean	-0.236	-2.207	---	---	---	---
Smartphone Ownership - Std. Dev.	0.972	6.403	---	---	---	---
Commute Mode (*Base: Non Commuters, Walk, Bike*)						
Drive Alone/Carpool - Mean	---	---	0.574	4.306	0.598	4.989
Drive Alone/Carpool - Std. Dev.	---	---	1.419	4.834	---	---
Transit	---	---	---	---	0.827	6.012
Male – Mean	0.550	5.326	0.245	2.370	---	---
Male - Std. Dev.	0.851	4.677	---	---	---	---
Household Structure (*Base: Single Person Household*)						
Multi-person Household without Children	---	---	0.529	4.402	---	---
Multi-person Household with Children	---	---	0.533	3.562	---	---
Error Correlation: *AV as Taxi with Backup Driver* and *AV as Taxi Without Backup Driver*	0.790 (t-stat = 10.338)					

(---) Coefficient not statistically significant and removed from the model

Individuals belonging to higher income households are less inclined to use AVs for car-share. This finding is aligned with the general trend of wealthier individuals placing a higher value on privacy. An unusual finding is that individuals whose maximum educational attainment is a Bachelor's degree prefer the use of car-share more than individuals who received lower or more education. Those who drive alone or carpool to work are more likely to prefer alternatives that involve greater control on the part of the user, namely, AV ownership or AV as a car-share. But the high standard deviation on the coefficient associated with AV ownership suggests a high degree of heterogeneity in their preference towards this mode. It is entirely possible that there are individuals who enjoy the driving experience and do not necessarily find the idea of handing over the driving task to a computer appealing. Transit users, who may not need to use a car on a very frequent basis, rank AV car-share more highly than others. Males prefer AV ownership and AV as taxi with backup driver, suggesting that females may be more comfortable using AV as taxi without backup driver (possibly, due to safety concerns). There is, however, significant heterogeneity in the preference of males for AV with a backup driver. Individuals in multi-person households favor AV ownership over other modes, presumably because they have higher trip-making needs – possibly involving the chauffeuring of other individuals in the household.

The correlation between error terms is statistically significant for the pair of alternatives that correspond to the use of AV as taxi with and without backup drivers. It is likely that there are unobserved attributes (attitudes and perceptions) that positively contribute to a preference for AV use as a taxi. Those who are comfortable using technology to hail rides are likely to favor future options in which the human driven vehicles are simply replaced by autonomous vehicles. Both options are similar to the current transportation network companies (TNC) that provide door-to-door reliable service at a reasonable cost, except for the presence of the backup driver. This finding further justifies the use of ROP over ROL as the ROL model assumes independence across alternatives.

Goodness-of-fit statistics were computed to assess model performance. The log-likelihood value for the model in Table 3 is –2340.24 and that for the constants-only model is –2532.34 (18 restrictions). The log-likelihood ratio test statistic is 384.2, which indicates that the two models are significantly different in the degree to which they explain preferences for different AV modes at any reasonable significance level. This demonstrates the ability of the rank ordered probit model to capture patterns in rank ordered data.

6. DISCUSSION AND CONCLUSIONS

This paper is concerned with modeling rank ordered data in which the ordinality of preferences among various alternatives were elicited from respondents of a survey. Although many surveys ask respondents to choose a single alternative from a choice set of alternatives, and travel demand models are often trained on such data sets, the emergence of new and transformative transportation technologies has provided renewed motivation for the collection of rank ordered data where relative preferences of individuals are revealed. In many instances, individuals may not necessarily be bound to a single choice with all other alternatives deemed unacceptable; rather, they may have a rank order of preferences with respect to various alternatives and the ability to model such rank ordered data may provide the ability to better predict relative market adoption rates for different modes or services. For example, consider the advent of autonomous technologies. There is considerable debate as to whether autonomous vehicles will be adopted in a shared mode or in an ownership mode or in some combination of both modes. And within a shared paradigm, there may be multiple alternatives including car-share services (akin to ZipCar or car2go) and ride-hailing services (similar to Uber and Lyft). The rate at which various modes and services will be adopted and the relative frequency with which they are used is going to depend on the preference ranking that individuals attach to these various modes or services. Unfortunately, the modeling of rank ordered data has been hindered due to methodological challenges that arise from the increasing level of indifference that people are likely to have with respect to the ranking of their least preferred alternatives.

This paper presents the formulation for a rank ordered probit modeling methodology that accounts for the possible presence of correlated unobserved attributes that affect the ranking of multiple alternatives. The methodology is applied to model the preference for adoption and usage of alternative autonomous vehicle (AV) modes and services using survey data collected from a sample of residents in the 2014-2015 Puget Sound Regional Travel Study. A sample of 1,365 respondents was used for the analysis in this paper. The rank ordered probit (ROP) model is estimated on this sample to understand preferences for use and adoption of AV modes, defined by four alternatives. The four alternatives include: AV use as a taxi with a backup driver, AV use as a taxi without a backup driver, AV ownership, and AV use in car-share mode. The model estimation results suggest that socio-economic attributes including age, employment status, household structure and composition, driver's license holding, and gender play a significant role in shaping preferences for AV alternatives. Through the application of the ROP model, it is possible to draw deeper insights into the differences in adoption preferences across groups.

The model estimation results could be used to inform public policy and marketing/information campaigns. The mode specific preferences predicted by the ROP model will be useful to identify market adoption pathways that are likely to play out in different socio-economic groups. Agencies interested in promoting certain evolutionary pathways can implement policies and information campaigns accordingly. For example, at the inception of an AV based taxi system, the model can be used to determine the fraction of taxis that should have a backup driver to enhance utility to the consumer. At a later stage, if an agency wishes to gradually phase out backup drivers, the model can be used to identify demographic groups that are most reluctant to make the switch to AVs without backup drivers, and information campaigns and demonstration sessions can be implemented and targeted accordingly. Based on data from the PSRC study, the demographic group most averse to AVs without backup drivers seems to be older individuals. An additional example is one in which an agency wishes to promote Mobility-as-a-Service (MaaS), i.e., the idea of shifting individuals away from privately owning a vehicle to subscription-based usage of a shared vehicle fleet. The agency would need to discourage vehicle ownership in favor of subscription-based ride sharing. Our model reveals that multi-person households have a preference for AV ownership. The agency could conduct targeted studies of multi-person household travel patterns and expenditures, and design MaaS options that accommodate intra-household interactions and joint activity-travel needs. Models of this nature can help agencies determine the specific markets that need to be studied and accommodated to help advance desirable outcomes.

Future research efforts should aim to further elucidate the value in collecting and using rank ordered preference data for travel demand forecasting. The methodology presented in this study may also be significantly extended to account for spatial dependency effects. This study ignored spatial dependency effects which may be quite important in the context of adoption of emerging transportation technologies. Diffusion effects play a key role in market adoption rates and patterns among different socio-economic groups and hence a spatial rank ordered probit (SROP) would provide the ability to explicitly account for such effects in travel forecasts.

ACKNOWLEDGEMENTS
This research was partially supported by the Center for Teaching Old Models New Tricks (TOMNET) (Grant No. 69A3551747116) as well as the Data-Supported Transportation Operations and Planning (D-STOP) Center (Grant No. DTRT13GUTC58), both of which are Tier 1 University Transportation Centers sponsored by the US Department of Transportation. The authors are grateful to Lisa Macias for her help in formatting this document. The authors thank four anonymous reviewers for their valuable comments and input that greatly improved the paper.

Conflict of interest statement: On behalf of all authors, the corresponding author states that there is no conflict of interest.

Authors' contributions:
G.S. Nair: Literature review, model specification and estimation, coding
S. Astroza: Model specification and estimation, coding, manuscript editing
C.R. Bhat: Conceptual development, methodology development, manuscript writing
S. Khoeini: Manuscript review and editing, model interpretation
R.M. Pendyala: Manuscript writing, model specification development

REFERENCES

Allison, P. and Christakis, N.A.: Logit model for sets of ranked items. *Sociological Methodology* 24, 199–228 (1994).

Alvo, M., Yu, P.L.H.: Probit models for ranking data. In *Statistical Methods for Ranking Data*, Frontiers in Probability and the Statistical Sciences, pp. 171-198, Springer Science+Business Media New York (2014). https://doi.org/10.1007/978-1-4939-1471-5_9.

Beggs, S., Cardell, S., Hausman, J.: Assessing the potential demand for electric cars. *Journal of Econometrics* 17(1), 1–19 (1981).

Bhat, C.R.: The maximum approximate composite marginal likelihood (MACML) estimation of multinomial Probit-Based Unordered Response choice models. *Transportation Research Part B* 45(7), 923–939 (2011).

Bhat, C.R.: New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function. *Transportation Research Part B* 109, 238–256 (2018).

Caparrós, A., Oviedo, J.L., Campos, P.: Would you choose your preferred option? Comparing choice and recoded ranking experiments. *American Journal of Agricultural Economics* 90(3), 843–855 (2008).

Fok, D., Paap, R., Van Dijk, B.: A rank-ordered logit model with unobserved heterogeneity in ranking capabilities. *Journal of Applied Econometrics* 27(5), 831–846 (2012).

Foster, V., Mourato, S.: Testing for consistency in contingent ranking experiments. *Journal of Environmental Economics and Management* 44(2), 309–328 (2002).

Hajivassiliou, V.A., Ruud, P.A.: Classical estimation methods for LDV models using simulation. In: *Handbook of Econometrics*, pp. 2383–2441, Elsevier (1994).

Halvorsen, R., Layton, D.F.: *Explorations in Environmental and Natural Resource Economics*. Edward Elgar Publishing (2006).

Hausman, J.A., Ruud, P.A.: Specifying and testing econometric models for rank-ordered data. *Journal of Econometrics* 34(1-2), 83–104 (1987).

Keane, M.: A Note on identification in the multinomial probit model. *Journal of Business and Economic Statistics* 10(2), 193–200 (1992).

Luce, R.D.: *Individual Choice Behavior: A Theoretical Analysis*. John Wiley & Sons, Inc., New York (1959).

Luce, R.D., Suppes, P.: Preference, utility, and subjective probability. In R.D. Luce, R.R. Bush, and E. Galanter (Eds.), *Handbook of Mathematical Psychology*, Vol. 3, Ch. 19, pp. 249–410, John Wiley & Sons, Inc. (1965).

Munkin, M.K., Trivedi, P.K.: Bayesian Analysis of the Ordered Probit Model with Endogenous Selection. *Journal of Econometrics* 143(2), 334–348 (2008).

RSG: Puget Sound Regional Travel Study. Puget Sound Regional Council, Seattle, WA (2014). Available at: https://www.psrc.org/sites/default/files/2014-household-survey-tech-memo.pdf. Accessed July 15, 2017.

Scarpa, R., Notaro, S., Louviere, J., Raffaelli, R.: Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in alpine grazing commons. *American Journal of Agricultural Economics* 93(3), 813–828 (2011).

Yan, J., Yoo, H.I.: The seeming unreliability of rank-ordered data as a consequence of model misspecification. MPRA Paper No. 56285 (2014).

Author Bios

Gopindra Sivakumar Nair is currently pursuing his PhD in Transportation Engineering at The University of Texas at Austin. His research interests span the areas of travel demand forecasting, modeling of connected and autonomous vehicles, and machine learning. He received his B.Tech in Civil Engineering and M.Tech in Infrastructure Engineering from the Indian Institute of Technology Madras.

Sebastian Astroza finished his PhD in Civil Engineering (Transportation Engineering) at The University of Texas at Austin and recently joined the Departamento de Ingeniería Industrial at Universidad de Concepción in Concepcion, Chile. His primary research interests include discrete choice models and travel behavior, spatial econometrics, and discrete-continuous econometric systems. He received his MS degree in Transportation Engineering and his BS degree in Civil Engineering from Universidad de Chile in Santiago, Chile.

Chandra R. Bhat is the Director of the Data-Supported Transportation Operations and Planning (D-STOP) Tier 1 USDOT University Transportation Center and the Joe J. King Chair in Engineering at The University of Texas at Austin. He has contributed toward the development of advanced econometric techniques for travel behavior analysis, in recognition of which he received the 2015 Frank M. Masters Transportation Engineering Award from the American Society of Civil Engineers (ASCE), and the 2017 Lifetime Achievement in Transportation Research and Education Award (Academic) from the Council of University Transportation Centers (CUTC). He is the current chair of the Transportation Research Board Travel Analysis Methods Section.

Sara Khoeini is an Assistant Research Professor of Transportation Systems in the School of Sustainable Engineering and the Built Environment at Arizona State University (ASU) in Tempe, Arizona. She is also the Assistant Director of a USDOT Tier 1 University Transportation Center (UTC) called TOMNET (Teaching Old Models New Tricks). Before joining ASU, Dr. Khoeini was a Research Scientist I at Georgia Institute of Technology. Dr. Khoeini has her Ph.D. in Civil Engineering with a specialization in transportation systems engineering from Georgia Institute of Technology. Her areas of expertise include urban transportation systems planning, travel behavior analysis, transportation demand modeling and forecasting, geographic information systems, travel survey methods and data collection, sustainability and energy.

Ram M. Pendyala is a professor of transportation and Director of a University Transportation Center in the School of Sustainable Engineering and the Built Environment at Arizona State University. He specializes in activity-based travel behavior analysis, transportation demand forecasting, and development of agent-based integrated land use-transport microsimulation models. He is past chair of the International Association for Travel Behaviour Research and the

Transportation Research Board's Planning and Environment Group, Travel Analysis Methods Section, and Traveler Behavior and Values Committee. He has his Ph.D. in Civil Engineering (Transportation) from the University of California at Davis.