**Online supplement to**

**“**A Socio-Technical Model of Autonomous Vehicle Adoption Using Ranked Choice Stated Preference Data”

By Katherine E. Asmussen, Aupal Mondal, and Chandra R. Bhat (corresponding author)

**TABLE 1. Stated Choice Experimental Design**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Experimental Design Attribute and Levels** | | | | | |
| ***Regular Vehicles (RV)*** | | ***Autonomous Vehicle (AV)*** | | ***SAV*** | |
| **Fixed cost per month** | **Variable travel cost per mile** | **Fixed cost per month** | **Variable travel cost per mile** | **Variable travel cost per mile** | **Waiting time** |
| $200  $300  $500 | $0.25  $0.50  $0.75 | $150 $200 $225 $250 $300 $375  $500 $625 | $0.25  $0.50  $0.75 | $1.50  $2.25  $3.00 | 3 minutes  6 minutes  9 minutes |
| **Scenario Example** | | | | | |
| Suppose AVs are now available for purchase, lease/rent, or to use via automated ride-hailing services, and half of the vehicles on the streets are AVs. What would you do when faced with your next car purchase decision in each of the following scenarios? Please rank the alternatives based on your preference (1=most preferred; 3=least preferred). Please do not give the same rank to multiple alternatives. | | | | | |
| ***Option A*** | | ***Option B*** | | ***Option C*** | |
| Buy a regular vehicle | | Buy an AV | | Don’t buy a vehicle and use AV ride-hailing/rental services | |
| $200/month + $0.50/mile | | $350/month + $0.50/mile | | $0/month + $2.25/mile | |
| Average wait time: 0 minutes | | Average wait time: 0 minutes | | Average wait time: 6 minutes | |

**Individual-level Sample Demographic Characteristics**

The sample descriptive statistics of the individual-level characteristics are presented in Table 2 (see left panel), and compared, whenever possible, with the census population of the Austin-Round Rock, TX Metro Area, as estimated by the U.S. Census Bureau (2018). The table indicates a clear over-representation of women in our sample, relative to the 50-50 split as reflected in the Census data for the Austin-Round Rock region. Not surprisingly, given our social media-based recruitment efforts and University-based efforts, the sample is skewed toward younger individuals (58.4% of adults 18 years or over in the age group of 18-29 years in our sample, relative to 23.7% of adults over the age of 18 years in this age group according to the Census). The Census does not report the number of students in the region, which makes it rather difficult to compare employment rates between our sample and that from the Census, especially given that a number of students both characterize themselves as being a student as well as being employed. In terms of education levels, again, our sample shows a markedly lower percentage of individuals who have completed high school or less (13.7% compared to 29.0% from the Census) and a higher percentage of individuals who have completed some college or technical school (35.4% relative to 25.0% from the Census). However, the distributions of those with an undergraduate degree or a graduate degree are very comparable to those from the Census.

As for household characteristics (right panel of Table 2), our sample is definitively skewed toward low income households. While 43.4% of our sample live in households that make less than $50,000 a year, and 28.1% of our sample live in households with an annual income of $100,000 or more, the corresponding percentages from the Census data are 31% and 38%, respectively. This lower income bias in our sample is consistent with the fact that many respondents were young and/or students. The average household size of sample respondents is close to three, while the corresponding figure from the Census data is 2.7 persons per household (the Census does not provide a breakdown by number of individuals in the household, and only provides an average household size value). Our sample and the Census align fairly well with regard to households with no children (83.1% compared to 81.3%). Finally, the Census provides no information on number of vehicles per household, though the low percentage of zero-vehicle households in our sample is to be expected.

**Table 2. Sample Distribution of Exogenous Variables: Socio-Demographic and Household Related Characteristics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Count | % | Variable | Count | % |
| *Individual Demographics* |  |  | *Household Characteristics* |  |  |
| **Gender** |  |  | **Household annual income** |  |  |
| Female | 658 | 64.4 | Less than $25,000 | 266 | 26.1 |
| Male | 363 | 35.6 | $25,000 to $49,999 | 177 | 17.3 |
| **Age** |  |  | $50,000 to $74,999 | 158 | 15.5 |
| 18 to 29 | 597 | 58.4 | $75,000 to $99,999 | 133 | 13.0 |
| 30 to 39 | 118 | 11.6 | $100,000 to $149,999 | 156 | 15.3 |
| 40 to 49 | 101 | 9.9 | $150,000 to $249,999 | 92 | 9.0 |
| 50 to 64 | 104 | 10.2 | $250,000 or more | 39 | 3.8 |
| 65 or older | 101 | 9.9 | **Household Size** |  |  |
| **Employment Type** |  |  | Live alone | 254 | 24.9 |
| Student | 530\* | 51.9 | 2 people | 283 | 27.7 |
| Employed | 623\* | 61.0 | 3 people | 150 | 14.7 |
| Unemployed and not a student | 138 | 13.5 | 4 or more people | 334 | 32.7 |
| **Education** |  |  | **Children (<18 years) in Household** |  |  |
| Completed high-school or less | 140 | 13.7 | Yes | 172 | 16.9 |
| Completed some college or technical school | 361 | 35.4 | No | 849 | 83.1 |
| Completed undergraduate degree | 348 | 34.1 | **Vehicles per Household** |  |  |
| Completed graduate degree | 172 | 16.8 | No vehicles | 84 | 8.2 |
|  |  |  | 1 vehicle | 250 | 24.5 |
|  |  |  | 2 vehicles | 337 | 33.0 |
|  |  |  | 3 vehicles | 211 | 20.7 |
|  |  |  | 4 or more vehicles | 139 | 13.6 |

\*270 respondents were both employed and students

**TABLE 3. Distribution of Attitudinal Indicators**

| **Indicators of…** | **Attitudinal Indicator** | **Response Category** | | | | | **Total** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Strongly disagree*** | ***Somewhat disagree*** | ***Neutral*** | ***Somewhat agree*** | ***Strongly agree*** |
| Frequency (Percent) | Frequency (Percent) | Frequency (Percent) | Frequency (Percent) | Frequency (Percent) |
| **Driving Control** | I will never ride in an AV | 361  (35.4) | 266  (26.0) | 234  (22.9) | 100  (9.8) | 60  (5.9) | 1021  (100.0) |
| AVs will eliminate my joy of driving | 192  (18.8) | 243  (23.8) | 267  (26.2) | 235  (23.0) | 84  (8.2) | 1021  (100.0) |
| When traveling in a vehicle, I prefer to be a driver rather than a passenger | 140  (13.7) | 209  (20.5) | 242  (23.7) | 207  (20.3) | 223  (21.8) | 1021  (100.0) |
| AVs would make traveling by car less stressful for me | 117  (11.5) | 186  (18.2) | 289  (28.3) | 284  (27.8) | 145  (14.2) | 1021  (100.0) |
| **Mobility Control** | I definitely like the idea of owning my own car | 26  (2.5) | 63  (6.2) | 112  (11.0) | 206  (20.2) | 614  (60.1) | 1021  (100.0) |
| Ride-hailing services allow me to live with fewer or no cars | 248  (24.3) | 210  (20.6) | 321  (31.4) | 166  (16.3) | 76  (7.4) | 1021  (100.0) |
| I will use AV ride hailing services alone or with coworkers, friends, or family | 140  (13.7) | 209  (20.5) | 242  (23.7) | 207  (20.3) | 223  (21.8) | 1021  (100.0) |
| **Safety Concern** | I would feel comfortable having an AV pick up/drop off children without adult supervision | 341  (33.4) | 301  (29.5) | 201  (19.7) | 128  (12.5) | 50  (4.9) | 1021  (100.0) |
| I am concerned about the potential failure of AV sensors, equipment, technology, or programs | 41  (4.0) | 76  (7.5) | 132  (12.9) | 431  (42.2) | 341  (33.4) | 1021  (100.0) |
| I would feel comfortable sleeping while traveling in an AV | 284  (27.8) | 277  (27.1) | 192  (18.8) | 179  (17.6) | 89  (8.7) | 1021  (100.0) |
| AVs would make me feel safer on the street as a pedestrian or as a cyclist | 156  (15.3) | 293  (28.7) | 291  (28.5) | 193  (18.9) | 88  (8.6) | 1021  (100.0) |
| **Technology Savviness** | I like to be among the first to have the latest technology | 58  (5.7) | 185  (18.1) | 202  (19.8) | 416  (40.7) | 160  (15.7) | 1021  (100.0) |
| Learning how to use new technologies is often frustrating for me | 361  (35.4) | 363  (35.5) | 132  (12.9) | 141  (13.8) | 24  (2.4) | 1021  (100.0) |

**TABLE 4. Loadings of Latent Variables on Indicators**

| **Attitudinal Indicators** | **Loading of Indicators on Latent Constructs** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Driving Control | | Mobility Control | | Safety Concern | | Tech-Savviness | |
| Coeff. | t-stat | Coeff. | t-stat | Coeff. | t-stat | Coeff. | t-stat |
| I will never ride in an AV | 0.782 | 17.74 |  |  |  |  |  |  |
| AVs will eliminate my joy of driving | 0.632 | 14.74 |  |  |  |  |  |  |
| When traveling in a vehicle, I prefer to be a driver rather than a passenger | 0.422 | 8.75 |  |  |  |  |  |  |
| AVs would make traveling by car less stressful for me | -0.826 | -18.44 |  |  |  |  |  |  |
| I definitely like the idea of owning my own car |  |  | 0.676 | 10.04 |  |  |  |  |
| Ride-hailing services allow me to live with fewer or no cars |  |  | -0.686 | -9.42 |  |  |  |  |
| I will use AV ride hailing services alone or with coworkers, friends, or family |  |  | 0.410 | 7.96 |  |  |  |  |
| I would feel comfortable having an AV pick up/drop off children without adult supervision |  |  |  |  | 0.872 | 23.65 |  |  |
| I am concerned about the potential failure of AV sensors, equipment, technology, or programs |  |  |  |  | -0.483 | -14.69 |  |  |
| I would feel comfortable sleeping while traveling in an AV |  |  |  |  | 0.886 | 22.04 |  |  |
| AVs would make me feel safer on the street as a pedestrian or as a cyclist |  |  |  |  | 0.796 | 21.73 |  |  |
| I like to be among the first to have the latest technology |  |  |  |  |  |  | 0.341 | 8.62 |
| Learning how to use new technologies is often frustrating for me |  |  |  |  |  |  | -0.845 | -11.29 |

**Mathematical formulation of GHDM model for jointly modeling continuous, nominal, and ranked outcomes**

Let *l* be the index for the latent constructs (*l*=1,2,…*L*; *L*=4 in our analysis). Let the underlying stochastic latent construct be denoted by , and we write  as a linear function of covariates:

 (1)

where ***w*** is a  vector of observed covariates (excluding a constant),  is a corresponding  vector of coefficients, and  is a standard normally distributed random error term. For future use, we also define the  matrix , and the  vectors  and  In matrix form, we may write Equation (1) as:

. (2)

We consider a multivariate normal correlation structure for **η** to accommodate interactions among the unobserved latent variables: , where  is an  column vector of zeros, and **Γ** is correlation matrix. Equation (2) constitutes the structural equations model (SEM) component of the model.

Of course, we do not observed the latent construct vector . However, we can consider the point values (say  for each latent construct ) obtained from the confirmatory factor analysis as manifestations of the stochastic latent construct . Define the vector Then, the first component of the measurement equation model may be written as . This component, in our model system, comprises four continuous dependent outcome variables. Next, let there be *G* nominal and rank-ordered dependent outcome variables for an individual, and let *g* be the index for these variables . For our analysis, *G=*3(one unordered nominal outcome corresponding to the duration to adoption or DAD choice and two rank-ordered outcomes corresponding to the responses to the two questions related to AV adoption).Also, let *Ig* be the number of alternatives corresponding to the *g*th variable (*Ig*3) and let  be the corresponding index . In our analysis, *Ig* =3 for all since all the variables have 3 alterntives each. Consider the *g*th variable and assume the usual random utility structure for each alternative :

 (3)

where ***x*** is an  vector of exogenous variable (including a constant),  is an  column vector of corresponding coefficients, and is a normal error term.  is an -matrix of variables interacting with latent variables to influence the utility of alternative , and  is an -column vector of coefficients capturing the effects of latent variables and their interaction effects with other exogenous variables. If each of the latent variables impacts the utility of the alternatives for each nominal variable purely through a constant shift in the utility function, will be an identity matrix of size *L*, and each element of  will capture the effect of a latent variable on the constant specific to alternative  of nominal variable *g*. Let   vector), and . Taking the difference with respect to the first alternative, the only estimable elements are found in the covariance matrix  of the error differences,  (where .Further, the variance term at the top left diagonal of   is set to 1 to account for scale invariance.  is constructed from  by adding a row on top and a column to the left. All elements of this additional row and column are filled with values of zero. In addition, the usual identification restriction is imposed such that one of the alternatives serves as the base when introducing alternative-specific constants and variables that do not vary across alternatives (that is, whenever an element of ***x*** is individual-specific and not alternative-specific, the corresponding element in is set to zero for at least one alternative  To proceed, define   vector),   matrix), and   matrix. Also, define the matrix , which is initially filled with all zero values. Then, position the  row vector  in the first row to occupy columns 1 to  , position the  row vector  in the second row to occupy columns +1 to  and so on until the  row vector  is appropriately positioned. Further, define  matrix), ,   vector),  vector),  matrix), matrix), and  (that is,  is a column vector that includes all elements of the matrices ). Then, in matrix form, we may write Equation (1) as:

 where .



The off-diagonal elements of the **Λ** matrix capture the correlations of the unobserved factors across the alternatives of the various nominal variables.

To proceed further, we may write the components of the joint model as follows:

, (4)

 (MEM component), (5)

 (MEM component), (6)



**Ω** in the equation above represents the () correlation elements between the **η** and **ε** error elements (this recognizes the endogeneity of the latent constructs in the system). To develop the reduced form equations, replace the right side of the SEM component into the MEM components to obtain the following system:

 (7)

 (8)

Now, consider the  vector  Let , an -matrix. Define

 and . (9)

Then  is the multivariate joint distribution of the main outcomes and the latent factor continuous variables.

**GHDM Model estimation**

In the context of the nominal unordered variable in our analysis, i.e. the outcome related to the DAD dimension, assume that the individual under consideration chooses alternative corresponding to the *gth* nominal outcome. Under the utility maximization theory, must be less than zero for all  corresponding to the *gth* nominal variable, since the individual chose alternative . Let , and stack the latent utility differentials into a vector . However, for the case of a rank-ordered nominal variable (along the AVD dimensions), the utility differentials are arrived at based on the order of the ranking. In particular, let ***rg*** be a specific rank ordering of the alternatives corresponding to the *gth* nominal variable. That is,  is the first-ranked alternative,  is the second-ranked alternative and so on.  denotes the event that the alternatives are ranked in the order ***r*** by the individual. According to the random utility maximization framework, the following relationship must hold for ,



The latent utility differentials for the rank-ordered nominal outcomes are stacked in a similar fashion as the unordered nominal outcome. Now, define , where the utility differentials can either be based on unordered nominal outcomes or rank-ordered nominal outcomes. We now need to develop the distribution of the vector from that of . To do so, define a matrix **M** of size . Fill this matrix with values of zero. Then, insert an identity matrix of size *L* into the first *L* rows and *L* columns of the matrix **M**. Next, consider the rows from , and columns from  These rows and columns correspond to the first nominal variable. If this nominal variable is a pure unordered (single choice) variable, insert an identity matrix of size  after supplementing with a column of ‘-1’ values in the column corresponding to the chosen alternative. Next, rows  through  and columns  through correspond to the second nominal variable. Again position an identity matrix of size  after supplementing with a column of ‘-1’ values in the column corresponding to the chosen alternative for the second nominal variable (if this variable is again an unordered single choice variable). However, if any of the nominal variables is a rank-ordered decision variable, then undertake the following method to fill in each of such sub-matrices: place a value of ‘–1’ at the column corresponding to the first ranked alternative and ‘1’ at the column corresponding to the second ranked alternative. Similarly, in the second row, place a value of ‘–1’ at the column corresponding to the second ranked alternative and ‘1’ at the column corresponding to the third ranked alternative. Continue this procedure for  rows (if the *gth* nominal variable happens to be a rank-ordered variable). Therefore, based on whether the sub-matrix within the matrix **M** corresponds to an unordered nominal variable or a rank-ordered nominal variable, undertake one of the two respective ways as described to fill in these sub-matrices. Continue this procedure for all *G* nominal variables (again, nominal variables here include both, unordered and rank-ordered variables). With the matrix **M** as defined, we can write  where  and . Next, partition the vector  into components that correspond to the mean of the vectors ***c*** (for the continuous latent variables) and ***u*** (for the nominal outcomes), and the matrix  into the corresponding variances and covariances:

 vector, and  vector

The conditional distribution of ***u***, given ***c***, is MVN with mean  and variance . Then the likelihood function may be written as (where  is a -column vector of zeros):

(10)



where the integration domain  is simply the multivariate region of the elements of the ***u*** vector determined by the range  for the utility differences for the nominal outcomes.  is the MVN density function of dimension ***L*** with a mean of  and a covariance of , and evaluated at ***c***. The likelihood function for a sample of *Q* decision-makers is obtained as the product of the individual-level likelihood functions. The above likelihood function involves the evaluation of a -dimensional upper-truncated integral for each decision-maker, which can be computationally expensive. However, Bhat’s (2018) matrix-based approximation method for evaluating multivariate normal cumulative distribution (MVNCD) function was employed to evaluate this integral, which provides an efficient and tractable formulation to approximate high dimensional MVNCD integral.

**References**

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

U.S. Census Bureau, 2018. American Community Survey 1-year estimates. *Census Reporter Profile page for Austin-Round Rock, TX Metro Area.* <https://censusreporter.org/profiles/31000US12420-austin-round-rock-tx-metro-area/>. Accessed June 20, 2020.