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 z_l^*, and we write z_l^* as a linear function of covariates:

$$z_l^* = \boldsymbol{\alpha}_l' \boldsymbol{w} + \eta_l, \tag{1}$$

where $\boldsymbol{w}$ is a $(D\times 1)$ vector of observed covariates (excluding a constant), $\boldsymbol{\alpha}_l$ is a corresponding $(D\times 1)$ vector of coefficients, and η_l is a standard normally distributed random error term. For future use, we also define the $(L\times D)$ matrix $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ..., \boldsymbol{\alpha}_L)'$, and the $(L\times 1)$ vectors $\boldsymbol{z}^* = (z_1^*, z_2^*, ..., z_L^*)'$ and $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3, \ldots, \eta_L)'$. In matrix form, we may write Equation (1) as:

$$\boldsymbol{z}^* = \boldsymbol{\alpha w} + \boldsymbol{\eta}. \tag{2}$$

We consider a multivariate normal correlation structure for $\boldsymbol{\eta}$ to accommodate interactions among the unobserved latent variables: $\boldsymbol{\eta} \sim MVN_L[\mathbf{0}_L, \boldsymbol{\Gamma}]$, where $\mathbf{0}_L$ is an $(L\times 1)$ column vector of zeros, and $\boldsymbol{\Gamma}$ is $(L\times L)$ correlation matrix. Equation (2) constitutes the structural equations model (SEM) component of the model.

Of course, we do not observed the latent construct vector $\boldsymbol{z}^*$. However, we can consider the point values (say $c_{z_l^*}$ for each latent construct z_l^*) obtained from the confirmatory factor analysis as manifestations of the stochastic latent construct z_l^*. Define the $(L\times 1)$ vector $\boldsymbol{c} = (c_{z_1^*}, c_{z_2^*}, ..., c_{z_L^*})'$. Then, the first component of the measurement equation model may be written as $\boldsymbol{c} = \boldsymbol{z}^*$. 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 $(g = 1, 2, 3, ..., G)$. 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 I_g be the number of alternatives corresponding to the g^{th} variable ($I_g \geq 3$) and let i_g be the corresponding index $(i_g = 1, 2, 3, ..., I_g)$. In our analysis, I_g =3 for all $g = 1,\ 2,\ 3$ since all the variables have 3 alterntives each. Consider the g^{th} variable and assume the usual random utility structure for each alternative i_g:

$$U_{gi_g} = \boldsymbol{b}_{gi_g}' \boldsymbol{x} + \boldsymbol{\vartheta}_{gi_g}' (\boldsymbol{\beta}_{gi_g} \boldsymbol{z}^*) + \varsigma_{gi_g}, \tag{3}$$

where $\boldsymbol{x}$ is an $(A\times 1)$ vector of exogenous variable (including a constant), $\boldsymbol{b}_{gi_g}$ is an $(A\times 1)$ column vector of corresponding coefficients, and ς_{gi_g} is a normal error term. $\boldsymbol{\beta}_{gi_g}$ is an $(N_{gi_g} \times L)$-matrix of variables interacting with latent variables to influence the utility of alternative i_g, and $\boldsymbol{\vartheta}_{gi_g}$ is an $(N_{gi_g} \times 1)$-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, $\boldsymbol{\beta}_{gi_g}$ will be an identity matrix of size L, and each element of $\boldsymbol{\vartheta}_{gi_g}$ will capture the effect of a latent variable on the constant specific to alternative i_g of nominal variable g. Let $\boldsymbol{\varsigma_g} = (\varsigma_{g1}, \varsigma_{g2}, ... \varsigma_{gI_g})'$ ($I_g \times 1$ vector), and $\boldsymbol{\varsigma_g} \sim MVN_{I_g}(\mathbf{0}, \mathbf{\Lambda_g})$. Taking the difference with respect to the first alternative, the only estimable elements are found in the covariance matrix $\breve{\mathbf{\Lambda}}_g$ of the error differences, $\breve{\boldsymbol{\varsigma}}_{\boldsymbol{g}} = (\breve{\varsigma}_{g2}, \breve{\varsigma}_{g3}, ..., \breve{\varsigma}_{gI_g})$ (where $\breve{\varsigma}_{gi} = \varsigma_{gi} - \varsigma_{g1},\ i \neq 1$). Further, the variance term at the top left diagonal of $\breve{\mathbf{\Lambda}}_g$ $(g = 1, 2, ..., G)$ is set to 1 to account for scale invariance. $\mathbf{\Lambda_g}$ is constructed from $\breve{\mathbf{\Lambda}}_{\mathbf{g}}$ 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 $\boldsymbol{x}$ is individual-specific and not alternative-specific, the corresponding element in $\boldsymbol{b}_{gi_g}$ is set to zero for at least one alternative i_g). To proceed, define $\boldsymbol{U}_g = (U_{g1}, U_{g2}, ..., U_{gI_g})'$ ($I_g \times 1$ vector), $\boldsymbol{b}_g = (\boldsymbol{b}_{g1}, \boldsymbol{b}_{g2}, \boldsymbol{b}_{g3}, ..., \boldsymbol{b}_{gI_g})'$ ($I_g \times A$ matrix), and $\boldsymbol{\beta}_g = (\boldsymbol{\beta}'_{g1}, \boldsymbol{\beta}'_{g2}, ..., \boldsymbol{\beta}'_{gI_g})'$ $\left(\sum_{i_g=1}^{I_g} N_{gi_g} \times L\right)$ matrix. Also, define the $\left(I_g \times \sum_{i_g=1}^{I_g} N_{gi_g}\right)$ matrix $\boldsymbol{\vartheta_g}$, which is initially filled with all zero values. Then, position the $(1 \times N_{g1})$ row vector $\boldsymbol{\vartheta}'_{g1}$ in the first row to occupy columns 1 to N_{g1}, position the $(1 \times N_{g2})$ row vector $\boldsymbol{\vartheta}'_{g2}$ in the second row to occupy columns $N_{g1}+1$ to $N_{g1} + N_{g2}$, and so on until the $(1 \times N_{gI_g})$ row vector $\boldsymbol{\vartheta}'_{gI_g}$ is appropriately positioned. Further, define $\boldsymbol{\varpi}_g = (\boldsymbol{\vartheta}_g \boldsymbol{\beta}_g)$ ($I_g \times L$ matrix), $\vec{G} = \sum_{g=1}^{G} I_g$, $\tilde{G} = \sum_{g=1}^{G} (I_g - 1)$, $\boldsymbol{U} = \left(\boldsymbol{U}'_1, \boldsymbol{U}'_2, \dots, \boldsymbol{U}'_G\right)'$ ($\vec{G} \times 1$ vector), $\boldsymbol{\varsigma} = (\boldsymbol{\varsigma}_1, \boldsymbol{\varsigma}_2, ... \boldsymbol{\varsigma}_G)'$ ($\vec{G} \times 1$ vector), $\boldsymbol{b} = (\boldsymbol{b}'_1, \boldsymbol{b}'_2, ..., \boldsymbol{b}'_{\boldsymbol{G}})'$ ($\vec{G} \times A$ matrix), $\boldsymbol{\varpi} = (\boldsymbol{\varpi}'_1, \boldsymbol{\varpi}'_2, ..., \boldsymbol{\varpi}'_G)'$ ($\vec{G} \times L$ matrix), and $\boldsymbol{\vartheta} = \text{Vech}(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, ..., \boldsymbol{\vartheta}_G)$ (that is, $\boldsymbol{\vartheta}$ is a column vector that includes all elements of the matrices $\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, ..., \boldsymbol{\vartheta}_G$). Then, in matrix form, we may write Equation (1) as:

$$\boldsymbol{U} = \boldsymbol{bx} + \boldsymbol{\varpi}\, \boldsymbol{z}^* + \boldsymbol{\varsigma}, \qquad \text{where } \boldsymbol{\varsigma} \sim MVN_{\vec{G}}(\mathbf{0}_{\vec{G}}, \mathbf{\Lambda}).$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_1 & \boldsymbol{\Lambda}_{12} & \boldsymbol{\Lambda}_{13} & \boldsymbol{\Lambda}_{14} & \cdots \boldsymbol{\Lambda}_{1\boldsymbol{G}} \\ \mathbf{0} & \boldsymbol{\Lambda}_2 & \boldsymbol{\Lambda}_{23} & \boldsymbol{\Lambda}_{24} & \cdots \boldsymbol{\Lambda}_{2\boldsymbol{G}} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Lambda}_3 & \boldsymbol{\Lambda}_{34} & \cdots \boldsymbol{\Lambda}_{3\boldsymbol{G}} \\ \vdots & \vdots & \vdots & \vdots & \ldots \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots \ldots \boldsymbol{\Lambda}_{\boldsymbol{G}} \end{bmatrix} (\vec{\boldsymbol{G}} \times \vec{\boldsymbol{G}} \text{ matrix}).$$

The off-diagonal elements of the $\boldsymbol{\Lambda}$ 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:

$$\boldsymbol{z}^* = \boldsymbol{\alpha w} + \boldsymbol{\eta} \quad \text{(SEM component)}, \tag{4}$$

$$\boldsymbol{c} = \boldsymbol{z}^* \quad \text{(MEM component)}, \tag{5}$$

$$\boldsymbol{U} = \boldsymbol{bx} + \boldsymbol{\varpi}\, \boldsymbol{z}^* + \boldsymbol{\varsigma}, \quad \text{(MEM component)}, \tag{6}$$

with $Cov \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\varsigma} \end{pmatrix} = \boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Omega} \\ \boldsymbol{\Omega}' & \boldsymbol{\Lambda} \end{bmatrix} (E \times E \text{ matrix}),\ E = L + \vec{G}.$

$\boldsymbol{\Omega}$ in the equation above represents the $(L \times \vec{G})$ correlation elements between the $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ 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:

$$\boldsymbol{c} = \boldsymbol{\alpha w} + \boldsymbol{\eta} \tag{7}$$

$$\boldsymbol{U} = \boldsymbol{bx} + \boldsymbol{\varpi}\,(\boldsymbol{\alpha w} + \boldsymbol{\eta}) + \boldsymbol{\varsigma} = \boldsymbol{bx} + \boldsymbol{\varpi\alpha w} + \boldsymbol{\varpi\eta} + \boldsymbol{\varsigma} \tag{8}$$

Now, consider the $[(E \times 1)]$ vector $\boldsymbol{cU} = \begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{U} \end{pmatrix}$ Let $\boldsymbol{d} = (\boldsymbol{\varpi}, \boldsymbol{IDEN_G})'$, an $(E \times \vec{G})$ -matrix. Define

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_c \\ \boldsymbol{B}_U \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha w} \\ \boldsymbol{bx} + \boldsymbol{\varpi\alpha w} \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_c & \boldsymbol{\Sigma}_{cU} \\ \boldsymbol{\Sigma}'_{cU} & \boldsymbol{\Sigma}_U \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Omega} + \boldsymbol{\Gamma}\boldsymbol{\varpi}' \\ \boldsymbol{\Omega}' + \boldsymbol{\varpi}\boldsymbol{\Gamma}' & \boldsymbol{d}'\Psi\boldsymbol{d} \end{bmatrix}. \tag{9}$$

Then $cU \sim \boldsymbol{MVN_E}(\boldsymbol{B}, \Sigma)$ 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 m_g corresponding to the g^{th} nominal outcome. Under the utility maximization theory, $U_{gi_g} - U_{gm_g}$ must be less than zero for all $i_g \neq m_g$ corresponding to the g^{th} nominal variable, since the individual chose alternative m_g. Let $u_{gi_g m_g} = U_{gi_g} - U_{gm_g}\ (i_g \neq m_g)$, and stack the latent utility differentials into a vector $\boldsymbol{u_g} = \left[\left(u_{g1m_g}, u_{g2m_g}, ..., u_{gI_g m_g}\right)'; i_g \neq m_g\right]$. 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 $\boldsymbol{r_g}$ be a specific rank ordering of the alternatives corresponding to the g^{th} nominal variable. That is, $r_g^{\,1}$ is the first-ranked alternative, $r_g^{\,2}$ is the second-ranked alternative and so on. $R_{\boldsymbol{r}}$ denotes the event that the alternatives are ranked in the order $\boldsymbol{r}$ by the individual. According to the random utility maximization framework, the following relationship must hold for $R_{\boldsymbol{r}}$,

$$R_{\boldsymbol{r},g} : U_{i_g r^2} - U_{i_g r^1} < 0, U_{i_g r^3} - U_{i_g r^2} < 0, ..., U_{i_g r^{I_g}} - U_{i_g r^{I_g - 1}} < 0$$

The latent utility differentials for the rank-ordered nominal outcomes are stacked in a similar fashion as the unordered nominal outcome. Now, define $\boldsymbol{u} = \left(\left[\boldsymbol{u}_1\right]', \left[\boldsymbol{u}_2\right]', ..., \left[\boldsymbol{u_G}\right]'\right)'$, 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 $\boldsymbol{cu} = \left(\boldsymbol{c}', \boldsymbol{u}'\right)'$ from that of $\boldsymbol{cU} = \left(\boldsymbol{c}', \boldsymbol{U}'\right)'$. To do so, define a matrix **M** of size $\left[\boldsymbol{L} + \tilde{\boldsymbol{G}}\right] \times \left[\boldsymbol{L} + \ddot{\boldsymbol{G}}\right]$. 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 $\boldsymbol{L}+1$ to $\boldsymbol{L}+\boldsymbol{I}_1-1$, and columns from $\boldsymbol{L}+1$ to $\boldsymbol{L}+\boldsymbol{I}_1$. 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 $(I_1 - 1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative. Next, rows $\boldsymbol{L}+\boldsymbol{I}_1$ through $\boldsymbol{L}+\boldsymbol{I}_1+\boldsymbol{I}_2-2$ and columns $\boldsymbol{L}+\boldsymbol{I}_1+1$ through $\boldsymbol{L}+\boldsymbol{I}_1+\boldsymbol{I}_2$ correspond to the second nominal variable. Again position an identity matrix of size $(I_2 - 1)$ 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 $(\boldsymbol{I_g} - 1)$ rows (if the g^{th} 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 $\boldsymbol{cu} \sim MVN_{L+\tilde{G}}(\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Omega}})$, where $\tilde{\boldsymbol{B}} = \mathbf{M}\boldsymbol{B}$ and $\tilde{\Sigma} = \mathbf{M}\Sigma\mathbf{M}'$. Next, partition the vector $\tilde{\boldsymbol{B}}$ into components that correspond to the mean of the vectors $\boldsymbol{c}$ (for the continuous latent variables) and $\boldsymbol{u}$ (for the nominal outcomes), and the matrix $\tilde{\Sigma}$ into the corresponding variances and covariances:

$$\tilde{\boldsymbol{B}} = \begin{bmatrix} \tilde{\boldsymbol{B}}_c \\ \tilde{\boldsymbol{B}}_u \end{bmatrix}, \ (\boldsymbol{L}+\tilde{\boldsymbol{G}})\times 1 \text{ vector, and } \tilde{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_c & \tilde{\Sigma}_{cu} \\ \tilde{\Sigma}'_{cu} & \tilde{\Sigma}_u \end{bmatrix}, \ (\boldsymbol{L}+\tilde{\boldsymbol{G}})\times(\boldsymbol{L}+\tilde{\boldsymbol{G}}) \text{ vector}$$

The conditional distribution of $\boldsymbol{u}$, given $\boldsymbol{c}$, is MVN with mean $\ddot{\boldsymbol{B}}_u = \tilde{\boldsymbol{B}}_u + \tilde{\Sigma}'_{cu}\tilde{\Sigma}_c^{-1}\left(\boldsymbol{c} - \tilde{\boldsymbol{B}}_c\right)$ and variance $\ddot{\Sigma}_u = \tilde{\Sigma}_u - \tilde{\Sigma}'_{cu}\tilde{\Sigma}_c^{-1}\tilde{\Sigma}_{cu}$. Then the likelihood function may be written as (where $\mathbf{0}_{\tilde{G}}$ is a $\tilde{G}\times 1$-column vector of zeros):

$$\begin{aligned} \boldsymbol{L}(\boldsymbol{\delta}) &= \boldsymbol{f}_L(\boldsymbol{c} \mid \tilde{\boldsymbol{B}}_c, \tilde{\Sigma}_c) \times \Pr\left[\boldsymbol{u} \le \mathbf{0}_{\tilde{G}}\right], \\ &= \boldsymbol{f}_L(\boldsymbol{c} \mid \tilde{\boldsymbol{B}}_c, \tilde{\Sigma}_c) \times \int_{\boldsymbol{D}_r} \boldsymbol{f}_{\tilde{G}}(\boldsymbol{r} \mid \ddot{\boldsymbol{B}}_u, \ddot{\Sigma}_u)\boldsymbol{dr}, \end{aligned} \tag{10}$$

where the integration domain $\boldsymbol{D}_r$ is simply the multivariate region of the elements of the $\boldsymbol{u}$ vector determined by the range $(-\infty_{\tilde{G}}, \mathbf{0}_{\tilde{G}})$ for the utility differences for the nominal outcomes. $\boldsymbol{f}_L(\boldsymbol{c} \mid \tilde{\boldsymbol{B}}_c, \tilde{\Sigma}_c)$ is the MVN density function of dimension $\boldsymbol{L}$ with a mean of $\tilde{\boldsymbol{B}}_c$ and a covariance of $\tilde{\Sigma}_c$, and evaluated at $\boldsymbol{c}$. 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 $\tilde{\boldsymbol{G}}$-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.