Online supplement to

Investigating Autonomous Vehicle Impacts on Individual Activity-Travel Behavior

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

TABLE 1 Loadings of Latent Variables on Indicators

Attitudinal Indicators	Loading of Indicators on Latent Constructs							
	Tech-Savviness		Safety Concern		Variety-Seeking Lifestyle (VSL)		IPTT	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
I like to be among the first to have the latest technology	0.851	6.13						
Learning how to use new technologies is often frustrating for me	-0.336	-5.36						
Having internet connectivity everywhere I go is important to me	0.329	5.36						
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			0.459	13.69				
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 trying things that are new and different					0.704	14.94		
I like the idea of having store, restaurants, and offices mixed among the homes in my neighborhood					0.397	6.73		
I make good use of the time I spend traveling							0.372	5.15
The level of congestion during my daily travel bothers me							0.522	5.28
I would make more long-distance trips when AVs are available because I wouldn't have to drive							0.345	4.73

Methodology for Developing Continuous Latent Constructs

The four latent constructs correspond to a total of 12 indicators (three for tech-savviness, four for AV safety concerns, two for variety-seeking lifestyle, and three for IPTT). To make the modeling exercise more tractable, without also losing any information, we first reduce the group of indicators for each construct to a single continuous "factor" using the traditional confirmatory analysis results (see Moore *et al.,* 2020 for a similar procedure). In order to ensure that all the indicators for each latent construct are appropriately scaled, we make the following normalization:

$$\boldsymbol{\tau}_{z_l^*} = \frac{\boldsymbol{\varpi}_{z_l^*} - \boldsymbol{\mu}_{z_l^*}}{\boldsymbol{\sigma}_{z_l^*}}, \tag{1}$$

where $\boldsymbol{\mu}_{z_l^*}$ is the sample mean vector of the indicators and $\boldsymbol{\sigma}_{z_l^*}$ is the sample standard deviation vector. Then, the factor analysis is undertaken as $\boldsymbol{\tau}_{z_l^*} = \boldsymbol{R}_{z_l^*} z_l^* + \boldsymbol{\vartheta}_{z_l^*}$, where $\boldsymbol{R}_{z_l^*}$ is a vector of the z_l^* factor's (latent construct's) loadings on each of its indicators, and $\boldsymbol{\vartheta}_{z_l^*}$ is a vector of error terms to recognize that the indicator vector $\boldsymbol{\varpi}_{z_l^*}$ (and, equivalently, $\boldsymbol{\tau}_{z_l^*}$) is obtained only for a sample of the population. The loading vector $\boldsymbol{R}_{z_l^*}$ is essentially estimated by capturing as much of the variance-covariance of the original $\boldsymbol{\tau}_{z_l^*}$ elements through the variance-covariance of the loading vector $\boldsymbol{R}_{z_l^*}$ (see Mueller and Hancock, 2001). In doing so, the elements of the $\boldsymbol{\vartheta}_{z_l^*}$ vector are assumed independent of z_l^*, and the scale of the factor z_l^* itself is normalized to the standard deviation of one with a mean value of zero (this is an innocuous normalization). Once the loading vector $\boldsymbol{R}_{z_l^*}$ is estimated for each latent construct, the single continuous indicator value for each of the latent constructs is computed as $h_{z_l^*} = (\hat{\boldsymbol{R}}_{z_l^*}^{-1})' \boldsymbol{\tau}_{z_l^*}$. Of course, these are point values for a particular sample, and are considered as manifestations of the underlying stochastic latent construct z_l^*. That is, we write $h_{z_l^*} = \boldsymbol{\alpha}_l' \boldsymbol{w}$ in our econometric model, and then write z_l^* itself as a linear function of covariates:

$$z_l^* = h_{z_l^*} + \eta_l = \boldsymbol{\alpha}_l' \boldsymbol{w} + \eta_l, \tag{2}$$

where $\boldsymbol{w}$ is a $(D \times 1)$ vector of observed covariates (including 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. 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 our empirical case, L=4, corresponding to the four latent constructs. In matrix form, we may write Equation (2) as:

$$\boldsymbol{z}^* = \boldsymbol{\alpha} \boldsymbol{w} + \boldsymbol{\eta}. \tag{3}$$

We consider a multivariate normal correlation structure for $\boldsymbol{\eta}$ to accommodate correlations 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. As a first stage of estimation, we then perform a multivariate regression analysis on this system of latent construct equations using the maximum likelihood approach to obtain estimates for the coefficients in vector $\boldsymbol{\alpha}$ for the observed covariates. Based on the estimates obtained in our multivariate regression model, we construct the estimated continuous values for each of the latent constructs for each individual in the sample. Therefore, we can write the single continuous factor for each construct as $c_{z_l^*} = \hat{z}_l^* = \hat{\boldsymbol{\alpha}}_l' \boldsymbol{w}$, where $\hat{\boldsymbol{\alpha}}_l$ are the estimated coefficients and $\hat{z}_l^*$ is the estimated continuous value for latent construct *l*. In our second stage model (discussed in the following framework), these estimated latent values appear on the right side of the main outcome utilities as exogenous variables (along with other individual and household variables).

Multivariate Ordered-Response Probit (MORP) Framework for Modeling Outcomes

Let q be an index for individuals (q = 1, 2, …, Q), and let i be the index for emotion (i = 1, 2, …, I, where I denotes the total number outcomes of interest for each individual; in the current study, I = 5). Let the number of ordinal levels for the outcome variables be K + 1 (*i.e.*, the response of an emotional rating is indexed by k and belongs in {0, 1, 2, …, K}). There is no need to index K by i because all trip propensity variables are mapped to a five-point ordinal scale. Following the usual ordered response framework notation, the latent propensity (y_{qi}^*) for each trip propensity variable is written as a function of relevant covariates and this latent propensity is related to the observed count outcome (y_{qi}) through threshold bounds (McKelvey and Zavoina, 1975):

$$y_{qi}^* = \beta_i' x_{qi} + \varepsilon_{qi}, y_{qi} = k \text{ if } \theta_i^k < y_{qi}^* < \theta_i^{k+1}, \tag{4}$$

where x_{qi} is a $(L \times 1)$ vector of exogenous variables (not including a constant) which also includes the estimated continuous latent scores for each latent constructs as discussed in Section 3.2.2, β_i is a corresponding $(L \times 1)$ vector of coefficients to be estimated, ε_{qi} is a standard normal error term, and θ_i^k is the lower bound threshold for count level k of AV trip propensity variable i $(\theta_i^0 < \theta_i^1 < \theta_i^2 ... < \theta_i^{K+1};\ \theta_i^0 = -\infty,\ \theta_i^{K+1} = +\infty$ for each AV trip propensity i). The ε_{qi} terms are assumed independent and identical across individuals (for each and all i). Due to identification restrictions, the variance of each ε_{qi} term is normalized to 1. However, correlations are allowed in the ε_{qi} terms across the AV trip propensity variables i for each individual q. Specifically, define $\varepsilon_q = (\varepsilon_{q1}, \varepsilon_{q2}, \varepsilon_{q3}, \ldots, \varepsilon_{qI})'$. Then, ε_q is multivariate normal distributed with a mean vector of zeros and a correlation matrix as follows:

$$\varepsilon_q \sim N\left[\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix},\begin{pmatrix}1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1I}\\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2I}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \rho_{I1} & \rho_{I2} & \rho_{I3} & \cdots & 1\end{pmatrix}\right], \text{ or} \tag{5}$$

$$\varepsilon_q \sim N\left[\mathbf{0},\boldsymbol{\Sigma}\right]$$

The off-diagonal terms of $\boldsymbol{\Sigma}$ capture the error covariances among the underlying latent continuous variables of the different trip propensity variables; that is, they account for the presence of common unobserved factors influencing the intensity outcome for each variable. Thus, if ρ_{12} is positive, it implies that individuals with a higher propensity to undertake a greater number of trips in an AV setting are also likely to travel further for shopping. If all correlation parameters (*i.e.*, off-diagonal elements of $\boldsymbol{\Sigma}$) stacked into a vertical vector, $\boldsymbol{\Omega}$, are identically zero, the model system in Equation (1) collapses to a series of independent ordered response probit models for each AV trip propensity variable.

The parameter vector of the multivariate probit model is $\delta = (\beta_1', \beta_2', ..., \beta_I';\ \theta_1', \theta_2', ..., \theta_I';\ \Omega')'$, where $\theta_i = (\theta_i^1,\ \theta_i^2,\ ...,\ \theta_i^K)'$ for $i = 1, 2, ..., I$. Let the actual observed AV trip propensity level for individual q and outcome variable i be m_{qi}. In that case, the likelihood function for individual q may be written as follows:

$$L_q(\delta) = \Pr(y_{q1} = m_{q1}, y_{q2} = m_{q2}, ..., y_{qI} = m_{qI})$$

$$L_q(\delta) = \int_{v_1=\theta_1^{m_{q1}}-\beta_1' x_{q1}}^{\theta_1^{m_{q1}+1}-\beta_1' x_{q1}} \int_{v_2=\theta_2^{m_{q2}}-\beta_2' x_{q2}}^{\theta_2^{m_{q2}+1}-\beta_2' x_{q2}} \cdots \int_{v_I=\theta_I^{m_{qI}}-\beta_I' x_{qI}}^{\theta_I^{m_{qI}+1}-\beta_I' x_{qI}} \phi_I(v_1, v_2,\ ..., v_I \mid \Omega)\, dv_1 dv_2 ... dv_I \tag{6}$$

$\phi_I(.,.,...)$ in the above expression represents the standard multivariate normal density function. Calculating the high-order I-dimensional rectangular integral in Equation (3) is computationally challenging. However, a recent efficient matrix-based approach devised by Bhat (2018), has been used to compute the rectangular integral shown above and estimate coefficients of the multivariate ordered response model. The mathematical formulations for the method have been omitted for brevity and may be found elsewhere (Bhat, 2018).

TABLE 2 Average Treatment Effect (ATE) for the TDS Dimension

Variable	Base Level	Treatment Level	% Contribution by mediation through				% Direct Effect	Overall ATE
			Tech-Savviness increase	Safety Concern decrease	Variety-Seeking Lifestyle increase	IPTT increase		
Socio-demographic								
Gender	Male	Female	0	-58	0	0	42	-0.202
Age	$\geq$65	18-29 years	0	40	4	19	37	0.229
Employment Status	Unemployed	Employed	0	52	0	48	0	0.09
Student status	Non-student	Student	0	0	-100	0	0	-0.021
Education	Less than graduate degree	Graduate degree	0	0	0	100	0	0.091
Income	<$100,000	$\geq$$250,000	0	47	23	0	-30	0.151
Presence of children	Not present	Present	0	-59	0	-41	0	-0.065
Built-environment effects								
Land use	Rural/suburban	Urban	-	-	-	-	-	-
Population density	Low/Medium	High	-	-	-	-	-	-
Land-use mix	25th percentile	75th percentile	-	-	-	-	-	-
Retail density	Low/Medium	High	0	0	0	0	-100	-0.16

TABLE 3 Average Treatment Effect (ATE) for the TDL Dimension

Variable	Base Level	Treatment Level	% Contribution by mediation through				% Direct Effect	Overall ATE
			Tech-Savviness increase	Safety Concern decrease	Variety-Seeking Lifestyle increase	IPTT increase		
Socio-demographic								
Gender	Male	Female	5	-60	0	0	35	-0.207
Age	≥65	18-29 years	-4	31	3	18	44	0.269
Employment Status	Unemployed	Employed	0	47	0	53	0	0.143
Student status	Non-student	Student	0	0	-100	0	0	-0.048
Education	Less than graduate degree	Graduate degree	0	0	0	100	0	0.112
Income	<$100,000	≥$250,000	-10	62	28	0	0	0.274
Presence of children	Not present	Present	0	-54	0	-46	0	-0.021
Built-environment effects								
Land use	Rural/suburban	Urban	-	-	-	-	-	-
Population density	Low/Medium	High	-	-	-	-	-100	-0.2
Land-use mix	25th percentile	75th percentile	-	-	-	-	-100	-0.047
Retail density	Low/Medium	High	-	-	-	-	-	-

TABLE 4 Average Treatment Effect (ATE) for the ALDT Dimension

Variable	Base Level	Treatment Level	% Contribution by mediation through				% Direct Effect	Overall ATE
			Tech-Savviness increase	Safety Concern decrease	Variety-Seeking Lifestyle increase	IPTT increase		
Socio-demographic								
Gender	Male	Female	0	-55	0	0	45	-0.062
Age	$\geq$65	18-29 years	0	29	3	18	50	0.416
Employment Status	Unemployed	Employed	0	46	0	54	0	0.155
Student status	Non-student	Student	0	0	-100	0	0	-0.021
Education	Less than graduate degree	Graduate degree	0	0	0	100	0	0.133
Income	<\$100,000	$\geq$\$250,000	0	68	32	0	0	0.17
Presence of children	Not present	Present	0	-53	0	-47	0	-0.076
Built-environment effects								
Land use	Rural/suburban	Urban	-	-	-	-		-
Population density	Low/Medium	High	-	-	-	-		-
Land-use mix	25th percentile	75th percentile	-	-	-	-		-
Retail density	Low/Medium	High	-	-	-	-	-	-

TABLE 5 Average Treatment Effect (ATE) for the CTT Dimension

Variable	Base Level	Treatment Level	% Contribution by mediation through				% Direct Effect	Overall ATE
			Tech-Savviness increase	Safety Concern decrease	Variety-Seeking Lifestyle increase	IPTT increase		
Socio-demographic								
Gender	Male	Female	0	-100	0	0	0	-0.023
Age	≥65	18-29 years	0	46	0	0	54	0.083
Employment Status	Unemployed	Employed	0	100	0	0	0	0.012
Student status	Non-student	Student	0	0	0	0	0	-
Education	Less than graduate degree	Graduate degree	0	0	0	100	0	0.013
Income	<$100,000	≥$250,000	0	100	0	0	0	0.027
Presence of children	Not present	Present	0	-100	0	0	0	-0.012
Built-environment effects								
Land use	Rural/suburban	Urban	-	-	-	-		-
Population density	Low/Medium	High	-	-	-	-		-
Land-use mix	25^{th} percentile	75^{th} percentile	0	0	0	0	-100	-0.001
Retail density	Low/Medium	High	-	-	-	-	-	-

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.

McKelvey, R.D., and Zavoina, W., 1975. A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, 4(1), 103–120.

Mueller, R.O. and Hancock, G.R., 2001. Factor analysis and latent structure: Confirmatory factor analysis. *International Encyclopedia of the Social & Behavioral Sciences*, 5239–5244.