**Online supplement to**

# Adoption of Partially Automated Vehicle Technology Features and Impacts on Vehicle Miles of Travel (VMT)

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

## Mathematical Formulation of the GHDM for the Current Study

Given all the main outcomes are either ordinal or grouped, and all the indicators are ordinal in nature, the GHDM model is formulated with only ordinal outcomes.[1]

Consider the case of an individual $q \in \{1,2,...,Q\}$. Let $l \in \{1,2,...,L\}$ be the index of the latent constructs and let $z_{ql}^{*}$ be the value of the latent variable $l$ for the individual $q$. $z_{ql}^{*}$ is expressed as a function of its explanatory variables as,

$$z_{ql}^{*} = \boldsymbol{w}_{\boldsymbol{ql}}^{\mathrm{T}}\boldsymbol{\alpha} + \eta_{ql}, \tag{1}$$

where $\boldsymbol{w}_{\boldsymbol{ql}}$ $(D\times 1)$ is a column vector of the explanatory variables of latent variable $l$ and $\boldsymbol{\alpha}$ $(D\times 1)$ is a vector of its coefficients. $\eta_{ql}$ is the unexplained error term and is assumed to follow a standard normal distribution. Equation (1) can be expressed in the matrix form as,

$$\boldsymbol{z}_{\boldsymbol{q}}^{*} = \boldsymbol{w}_{\boldsymbol{q}}\boldsymbol{\alpha} + \boldsymbol{\eta}_{\boldsymbol{q}}, \tag{2}$$

where $\boldsymbol{z}_{\boldsymbol{q}}^{*}$ $(L\times 1)$ is a column vector of all the latent variables, $\boldsymbol{w}_{\boldsymbol{q}}$ $(L\times D)$ is a matrix formed by vertically stacking the vectors $(\boldsymbol{w}_{\boldsymbol{q1}}^{\mathrm{T}}, \boldsymbol{w}_{\boldsymbol{q2}}^{\mathrm{T}}, ..., \boldsymbol{w}_{\boldsymbol{qL}}^{\mathrm{T}})$ and $\boldsymbol{\eta}_{\boldsymbol{q}}$ $(D\times 1)$ is formed by vertically stacking $(\eta_{q1}, \eta_{q2}, ..., \eta_{qL})$. $\boldsymbol{\eta}_{\boldsymbol{q}}$ follows a multivariate normal distribution centered at the origin and having a correlation matrix of $\boldsymbol{\Gamma}$ $(L\times L)$, i.e., $\boldsymbol{\eta}_{\boldsymbol{q}} \sim MVN_L(\boldsymbol{0}_{\boldsymbol{L}}, \boldsymbol{\Gamma})$, where $\boldsymbol{0}_{\boldsymbol{L}}$ is a vector of zeros. The variance of all the elements in $\boldsymbol{\eta}_{\boldsymbol{q}}$ is fixed as unity because it is not possible to uniquely identify a scale for the latent variables. Equation (2) constitutes the SEM component of the framework.

Let $j \in \{1,2,...,J\}$ denote the index of the outcome variables (including the indicator variables). Let $y_{qj}^{*}$ be the underlying continuous measure associated with the outcome variable $y_{qj}$. Then,

$$y_{qj} = k \text{ if } t_{jk} < y_{qj}^{*} \le t_{j(k+1)}, \tag{3}$$

where $k \in \{1,2,...,K_j\}$ denotes the ordinal category assumed by $y_{qj}$ and $t_{jk}$ denotes the lower boundary of the $k^{\text{th}}$ discrete interval of the continous measure associated with the $j^{\text{th}}$ outcome.

[1] A grouped outcome can be treated as an ordinal outcome with fixed thresholds, so that the only difference is that the thresholds are observed for a grouped variable. This allows the estimation of a scale for the error term in the grouped outcome.

$t_{jk} < t_{j(k+1)}$ for all $j$ and all $k$. Since $y_j^*$ may take any value in $(-\infty, \infty)$, we fix the value of $t_{j1} = -\infty$ and $t_{j(K_j+1)} = \infty$ for all $j$. Since the location of the thresholds on the real-line is not uniquely identifiable, we also set $t_{j2} = 0$ (except for the grouped outcome, for which all thresholds are observed). $y_{qj}^*$ is expressed as a function of explanatory variables and, as appropriate, the observed dummy variable values of other endogenous outcomes (though strictly only in a recursive fashion):

$$y_{qj}^* = \boldsymbol{x}_{\boldsymbol{qj}}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_{\boldsymbol{q}}^{*\mathrm{T}}\boldsymbol{d}_{\boldsymbol{j}} + \xi_{qj}, \tag{4}$$

where $\boldsymbol{x}_{\boldsymbol{qj}}$ is an $(E \times 1)$ vector of explanatory variables (including a constant) as well as the observed values of any other observed dummy variable endogenous outcomes, $\boldsymbol{\beta}$ $(E \times 1)$ is a corresponding column vector of coefficients associated with $\boldsymbol{x}_{\boldsymbol{qj}}$, and $\boldsymbol{d}_{\boldsymbol{j}}$ $(L \times 1)$ is the vector of coefficients of the latent variables for outcome $j$ (in the current paper, the observed dummy variable endogenous outcomes included in $\boldsymbol{x}_{\boldsymbol{qj}}$ correspond to the presence of partial automation feature combinations that affect the grouped ln(VMT) equation). $\xi_{qj}$ is a stochastic error term that captures the combined effects of unobserved variables on $y_{qj}^*$. $\xi_{qj}$ is assumed to follow a standard normal distribution, except for the grouped variable (the logarithm of VMT in the current paper) for which a scale can be computed. Jointly, the continuous measures of the $J$ outcome variables may be expressed as:

$$\boldsymbol{y}_{\boldsymbol{q}}^* = \boldsymbol{x}_q\boldsymbol{\beta} + \boldsymbol{d}\boldsymbol{z}_{\boldsymbol{q}}^* + \boldsymbol{\xi}_{\boldsymbol{q}}, \tag{5}$$

where $\boldsymbol{y}_{\boldsymbol{q}}^*$ $(J \times 1)$ and $\boldsymbol{\xi}_{\boldsymbol{q}}$ $(J \times 1)$ are the vectors formed by vertically stacking $y_{qj}^*$ and $\xi_{qj}$, respectively, of the $J$ dependent variables. $\boldsymbol{x}_{\boldsymbol{q}}$ $(J \times E)$ is a matrix formed by vertically stacking the vectors $\left(\boldsymbol{x}_{\boldsymbol{q1}}^{\mathrm{T}}, \boldsymbol{x}_{\boldsymbol{q2}}^{\mathrm{T}}, ..., \boldsymbol{x}_{\boldsymbol{qJ}}^{\mathrm{T}}\right)$ and $\boldsymbol{d}$ $(J \times L)$ is a matrix formed by vertically stacking $\left(\boldsymbol{d}_{\boldsymbol{1}}^{\mathrm{T}}, \boldsymbol{d}_{\boldsymbol{2}}^{\mathrm{T}}, ..., \boldsymbol{d}_{\boldsymbol{J}}^{\mathrm{T}}\right)$. $\boldsymbol{\xi}_{\boldsymbol{q}}$ follows a multivariate normal distribution centered at the origin. The covariance matrix of $\boldsymbol{\xi}_{\boldsymbol{q}}$ is a diagonal matrix with values of one in the first ($J$-1) diagonal positions (for identification purposes) and an estimable variance in the last diagonal position corresponding to the grouped variable (assuming the grouped variable is positioned as the last outcome). Let $\mathrm{Cov}(\boldsymbol{\xi}_{\boldsymbol{q}}) = \boldsymbol{\Omega}$. The reader is referred to Bhat (2015) for further nuances regarding the identification of coefficients in the GHDM framework.

Substituting Equation (2) in Equation (5), $\boldsymbol{y}_{\boldsymbol{q}}^*$ can be expressed in reduced form as

$$\boldsymbol{y}_{\boldsymbol{q}}^* = \boldsymbol{x}_q\boldsymbol{\beta} + \boldsymbol{d}\left(\boldsymbol{w}_q\boldsymbol{\alpha} + \boldsymbol{\eta}_{\boldsymbol{q}}\right) + \boldsymbol{\xi}_{\boldsymbol{q}}, \tag{6}$$

$$= \boldsymbol{x}_q\boldsymbol{\beta} + \boldsymbol{d}\boldsymbol{w}_q\boldsymbol{\alpha} + \boldsymbol{d}\boldsymbol{\eta}_{\boldsymbol{q}} + \boldsymbol{\xi}_{\boldsymbol{q}}. \tag{7}$$

On the right side of Equation (7), $\boldsymbol{\eta}_q$ and $\boldsymbol{\xi}_q$ are random vectors that follow the multivariate normal distribution and the other elements are non-random. Therefore, $\boldsymbol{y}_q^*$ also follows the multivariate normal distribution with a mean of $\boldsymbol{b} = \boldsymbol{x}_q\boldsymbol{\beta} + \boldsymbol{d}\boldsymbol{w}_q\boldsymbol{\alpha}$ (all the elements of $\boldsymbol{\eta}_q$ and $\boldsymbol{\xi}_q$ have a mean of zero) and a covariance matrix of $\boldsymbol{\Sigma} = \boldsymbol{d}\boldsymbol{\Gamma}\boldsymbol{d}^{\mathrm{T}} + \boldsymbol{\Omega}$.

$$\boldsymbol{y}_q^* \sim MVN_J(\boldsymbol{b}, \boldsymbol{\Sigma}). \tag{8}$$

The parameters that are to be estimated are the elements of $\boldsymbol{\alpha}$, strictly upper triangular elements of $\boldsymbol{\Gamma}$, elements of $\boldsymbol{\beta}$, elements of $\boldsymbol{d}$, $t_{jk}$ for all $j$ (except for $j$=$J$) and $k \in \{3, 4, ..., K_j\}$ and the scale of the grouped dependent variable (note that the $t_{Jk}$ values are however, observed thresholds, because the $J$th outcome is the grouped outcome) . Let $\boldsymbol{\theta}$ be a vector of all the parameters that need to be estimated. The maximum likelihood approach can be used for estimating these parameters. The likelihood of the $q^{\text{th}}$ observation will be,

$$L_q(\boldsymbol{\theta}) = \int_{v_1 = t_{1y_{q1}} - b_1}^{v_1 = t_{1(y_{q1}+1)} - b_1} \int_{v_2 = t_{2y_{q2}} - b_2}^{v_2 = t_{2(y_{q2}+1)} - b_2} \cdots \int_{v_J = t_{Jy_{qJ}} - b_J}^{v_J = t_{J(y_{qJ}+1)} - b_J} \phi_J(v_1, v_2, \ldots, v_J \mid \boldsymbol{\Sigma}) dv_1 dv_2 \ldots dv_J \ , \tag{9}$$

where, $\phi_J(v_1, v_2, \ldots, v_J \mid \boldsymbol{\Sigma})$ denotes the probability density of a $J$ dimensional multivariate normal distribution centered at the origin with a covariance matrix $\boldsymbol{\Sigma}$ at the point $(v_1, v_2, \ldots, v_J)$. Since a closed form expression does not exist for this integral and evaluation using simulation techniques can be time consuming, we used the One-variate Univariate Screening technique proposed by Bhat (2018) for approximating this integral. The estimation of parameters was carried out using the *maxlik* library in the GAUSS matrix programming language.

**References**

Bhat, C. R., 2015. A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research Part B* 79, 50-77.

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.

**Table 1. Distribution of Attitudinal Indicators**

| Attitudinal Indicators | Loading of Indicators on Latent Constructs | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Driving Control | | Mobility Control | | Safety Concern | | IPTT | |
| | Coeff. | t-stat | Coeff. | t-stat | Coeff. | t-stat | Coeff. | t-stat |
| I will never ride in an AV | 0.861 | 21.44 | | | | | | |
| AVs will eliminate my joy of driving | 0.539 | 19.38 | | | | | | |
| When traveling in a vehicle, I prefer to be a driver rather than a passenger | 0.333 | 13.61 | | | | | | |
| I definitely like the idea of owning my own car | | | 0.468 | 12.95 | | | | |
| I will use AV ride hailing services alone or with coworkers, friends, or family | | | -0.398 | -13.22 | | | | |
| I would feel comfortable having an AV pick up/drop off children without adult supervision | | | | | -0.718 | -26.69 | | |
| I am concerned about the potential failure of AV sensors, equipment, technology, or programs | | | | | 0.461 | 21.17 | | |
| I would feel comfortable sleeping while traveling in an AV | | | | | -0.916 | -27.18 | | |
| I make good use of the time I spend traveling | | | | | | | 0.047 | 2.07 |
| The level of congestion during my daily travel bothers me | | | | | | | 0.078 | 3.46 |
| I would make more long-distance trips when AVs are available because I wouldn't have to drive | | | | | | | 2.017 | 28.21 |

**Table 2. IOP ATE and VMT Analysis for each PAF Combination across each Gender and Age Group**

| PAF Combination | Gender and Age Group ATEs | | | | | | Overall ATE for each PAF | Overall PATE |
|---|---|---|---|---|---|---|---|---|
| | 18-29 Years | | 30-64 Years | | 65 Years or Older | | | |
| | Male | Female | Male | Female | Male | Female | | |
| Only Backup Camera | 1,375 (8.4%) | 1,310 (11.8%) | 1,428 (8.0%) | 1,361 (13.0%) | 1,049 (10.2%) | 1,000 (20.9%) | 1,303 | 10.6% |
| Only Adaptive Cruise Control | 1,292 (7.9%) | 1,231 (11.0%) | 1,342 (7.5%) | 1,279 (12.2%) | 986 (9.6%) | 940 (19.6%) | 1,225 | 10.0% |
| Only Automatic Braking System | -204 (-1.2%) | -194 (-1.7%) | -212 (-1.2%) | -202 (-1.9%) | -156 (-1.5%) | -148 (-3.1%) | -193 | -1.6% |
| Only Backup Camera and Adaptive Cruise Control | 1,578 (9.7%) | 1,504 (13.5%) | 1,640 (9.2%) | 1,562 (14.9%) | 1,204 (11.7%) | 1,148 (24.0%) | 1,496 | 12.2% |
| Only Backup Camera, Adaptive Cruise Control and Automatic Braking System | 1,632 (10.0%) | 1,555 (13.9%) | 1,695 (9.5%) | 1,616 (15.4%) | 1,245 (12.1%) | 1,187 (24.8%) | 1,547 | 12.6% |
| Only Adaptive Cruise Control and Automatic Braking System | 926 (5.7%) | 883 (7.9%) | 962 (5.4%) | 917 (8.7%) | 707 (6.9%) | 673 (14.1%) | 878 | 7.1% |
| Only Backup Camera, Lane Keeping System, and Blind Spot Monitoring | 581 (3.6%) | 553 (5.0%) | 603 (3.4%) | 575 (5.5%) | 443 (4.3%) | 422 (8.8%) | 551 | 4.5% |