**Online supplement to**

**“Pooled Versus Private Ride-hailing: A Joint Revealed and Stated Preference Analysis Recognizing Psycho-Social Factors”**

By Shuqing Kang, Aupal Mondal, Aarti C. Bhat, and Chandra R. Bhat (corresponding author)

**Table 1. Loading of Latent Constructs on Indicators (MEM)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Indicators** | **Tech-savviness** | **Sharing Propensity** | **GLP** |
| Coeff. | t-stat | Coeff. | t-stat | Coeff. | t-stat |
| I like to be among the first to have the latest technology. | 0.537 | 9.059 |   |   |  |   |
| Learning how to use new technologies is often frustrating for me. (inverse scale) | 0.513 | 8.804 |   |   |  |   |
| Having internet connectivity everywhere I go is important to me. | 0.356 | 6.864 |   |   |  |   |
| I like trying things that are new and different. | 0.383 | 6.757 |   |   |  |   |
| I feel uncomfortable around people I do not know. (inverse scale) |   |   | 0.349 | 8.632 |  |   |
| Traveling with a driver I don’t know makes me feel uncomfortable. (inverse scale) |   |   | 1.793 | 8.612 |  |   |
| For shared ride-hailing (e.g., uberPOOL, Lyft Share), traveling with unfamiliar passengers makes me uncomfortable. (inverse scale) |   |   | 1.528 | 10.667 |  |   |
| Sharing my personal information or location via internet-enabled devices concerns me a lot. (inverse scale) |   |   | 0.226 | 5.399 |  |   |
| I am concerned that my travel logs and personal information stored in AVs could be leaked. (inverse scale) |   |   | 0.246 | 6.128 |  |   |
| The government should raise the gas tax to help reduce the negative impacts of transportation on the environment. |   |   |   |   | 0.554 | 10.671 |
| I am committed to an environmentally-friendly lifestyle. |   |   |   |   | 0.938 | 9.917 |
| I am committed to using a less polluting means of transportation (e.g., walking, biking, and public transit) as much as possible. |   |   |   |   | 1.302 | 8.031 |

**Table 2 ATE Table for Pooled RH -- Shopping Purpose**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Base Level** | **Treatment Level** | **% Contribution by mediation through** | **Overall ATE** |
| **RH familiarity direct effect** | **RH familiarity sharing propensity increase** | **Tech-savviness decrease** | **Sharing propensity increase** | **GLP increase** | **Pooled RH choice direct effect** |
|  |
| **Pooled RH interest for the shopping purpose** |  |  |  |  |  |  |  |  |
| *Socio-demographic* |   |   |   |   |   |   |   |  |
| Gender | Female | Male | 0 | 45 | -34 | 19 | -2 | 0 | 0.019 |  |
| Age | 18-24 | 55+ | -80 | 0 | 19 | 0 | -1 | 0 | -0.213 |  |
| Race/Ethnicity | Other races | Non-Hispanic/Non-Latino White | -37 | -14 | 0 | -4 | 0 | -45 | -0.086 |  |
| Education | High school or less | Graduate degree | 61 | 0 | 0 | 0 | 3 | 36 | 0.129 |  |
| Employment | Unemployed | Employed | 0 | 71 | 0 | 29 | 0 | 0 | 0.020 |  |
| Tenure | Own or other | Rent | 100 | 0 | 0 | 0 | 0 | 0 | 0.112 |  |
| Household income | < $150,000 | ≥ $150,000 | 65 | 0 | -30 | 0 | -5 | 0 | 0.026 |  |
| *Built environment* |  |  |   |   |   |   |   |  |
| Living environment | Urban/suburban | Rural | -100 | 0 | 0 | 0 | 0 | 0 | -0.084 |  |
| Transit access | Transit access | No transit access  | -100 | 0 | 0 | 0 | 0 | 0 | -0.067 |  |
| Population density | Low | High | 0 | 0 | 0 | 0 | 0 | 100 | 0.040 |  |
| *Trip level attributes* |   |   |   |   |   |   |   |  |
| Travel time  |  Current time | Decrease by 5 mins | - | - | - | - | - | 100 | 0.026 |  |
| Travel cost  |  Current cost | Decrease by $1 | - | - | - | - | - | 100 | 0.017 |  |
| Additional passenger |  Current scenario | 1 additional passenger | - | - | - | - | - | -100 | -0.032 |  |

**Table 3 ATE Table for Pooled RH -- Leisure Purpose**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Base Level** | **Treatment Level** | **% Contribution by mediation through** | **Overall ATE** |
| **RH familiarity direct effect** | **RH familiarity sharing propensity** | **Tech-savviness decrease** | **Sharing propensity increase** | **GLP increase** | **Pooled RH choice direct effect** |
| **increase** |
| **Pooled RH interest for the leisure purpose** |  |  |  |  |  |  |  |
| *Socio-demographic* |   |   |   |   |   |   |   |
| Gender | Female | Male | 0 | 36 | -50 | 10 | -4 | 0 | -0.006 |
| Age | 18-24 | 55+ | -72 | 0 | 26 | 0 | -2 | 0 | -0.171 |
| Race/Ethnicity | Other races | Non-Hispanic/Non-Latino White | -33 | -13 | 0 | -3 | 0 | -51 | -0.080 |
| Education | High school or less | Graduate degree | 55 | 0 | 0 | 0 | 6 | 39 | 0.125 |
| Employment Status | Unemployed | Employed | 0 | 80 | 0 | 20 | 0 | 0 | 0.015 |
| Tenure type | Own or other | Rent | 100 | 0 | 0 | 0 | 0 | 0 | 0.096  |
| Income | < $150,000 | ≥ $150,000 | 52 | 0 | -40 | 0 | -8 | 0 | 0.003 |
| *Built environment* |  |  |   |   |   |   |   |
| Living environment | Urban/suburban | Rural | -100 | 0 | 0 | 0 | 0 | 0 | -0.109 |
| Transit access | Transit access | No transit access  | -100 | 0 | 0 | 0 | 0 | 0 | -0.058 |
| Population density | Low | High | 0 | 0 | 0 | 0 | 0 | 100 | 0.045 |
| *Trip level attributes* |   |   |   |   |   |   |   |
| Travel time  |  Current time | Decrease by 5 mins | - | - | - | - | - | 100 | 0.021 |
| Travel cost  |  Current cost | Decrease by $1 | - | - | - | - | - | 100 | 0.023 |
| Additional passenger |  Current scenario | 1 additional passenger | - | - | - | - | - | -100 | -0.031 |

## Mathematical formulation of the GHDM for the current study

Since the main outcome variables are all binary models, they can be modeled as ordinal variables as well (with 0 and 1 as the ordered levels). Given all the indicators are ordinal in nature, the GHDM model is formulated with only ordinal outcomes.

Consider the case of an individual . Let  be the index of the latent constructs and let  be the value of the latent variable *l* for the individual *q*.  is expressed as a function of its explanatory variables as,

,

where  is a column vector of the explanatory variables of latent variable *l* and is a vector of its coefficients.  is the unexplained error term and is assumed to follow a standard normal distribution. Equation (1) can be expressed in the matrix form as,

,

where is a column vector of all the latent variables, is a matrix formed by vertically stacking the vectors  and  is formed by vertically stacking .  follows a multivariate normal distribution centered at the origin and having a correlation matrix of , i.e., , where  is a vector of zeros. The variance of all the elements in  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 denote the index of the outcome variables (including the indicator variables). Let  be the underlying continuous measure associated with the outcome variable . Then,

,

where  denotes the ordinal category assumed by  and  denotes the lower boundary of the *k*th discrete interval of the continous measure associated with the *j*th outcome.  for all *j* and all *k*. Since  may take any value in , we fix the value of and  for all *j*. Since the location of the thresholds on the real-line is not uniquely identifiable, we also set .  is expressed as a function of its explanatory variables as,

,

where is a vector of size of explanatory variables for the continuous measure .  is a column vector of the coefficients associated with  and  is the vector of coefficients of the latent variables for outcome *j*.  is a stochastic error term that captures the effect of unobserved variables on .  is assumed to follow a standard normal distribution. Jointly, the continuous measures of the *J* outcome variables may be expressed as,

,

where  and  are the vectors formed by vertically stacking and , respectively, of the *J* dependent variables.  is a matrix formed by vertically stacking the vectors  and  is a matrix formed by vertically stacking .  follows a multivariate normal distribution centered at the origin with an identity matrix as the covariance matrix (independent error terms). . We assume the terms in  to be independent because it is not possible to uniquely identify all the correlations between the elements in and all the correlations between the elements in . Further, because of the ordinal nature of the outcome variables, the scale of  cannot be uniquely identified. Therefore, the variances of all elements in  is fixed to one. 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),  can be expressed in the reduced form as

,

.

In the right side of Equation (7),  and  are random vectors that follow the multivariate normal distribution and the other variables are constants. Therefore,  also follows the multivariate normal distribution with a mean of  (all the elements of  and  have a mean of zero) and a covariance matrix of .

.

The parameters that are to be estimated are the elements of , strictly upper triangular elements of **Γ**, elements of ***β***, elements of ***d*** and  for all *j* and . Let ***θ*** 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*th observation will be,

,

where,  denotes the probability density of a *J* dimensional multivariate normal distribution centered at the origin with a covariance matrix **Σ** at the point 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., 2018. New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function. *Transportation Research Part B*, 109, 238-256.