Online Supplement to

“**MaaS in Car-Dominated Cities:**

**Modeling the adoption, frequency, and characteristics of ride-hailing trips in Dallas, TX**”

Patrícia S. Lavieri and Chandra R. Bhat (corresponding author)

**An Overview of the Generalized Heterogeneous Data Model**

Latent Variable Structural Equation Model

Let *l* be an index for latent variables (*l*=1,2,3,4). Consider the latent variable  and write it 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 random error term assumed to be standard normally distributed for identification purposes (See Bhat, 2015). Next, define the matrix , and the vectors  and To accommodate interactions among the unobserved latent variables, we allow a MVN correlation structure for , that is whue  is correlation matrix. A general covariance structure is adopted because in our study there are no conceptual reasons to establish causal relationships between the latent variables. In matrix form, we may write Equation (1) as:

. (2)

Latent Variable Measurement Equation Model Components

As mentioned earlier, we consider a combination of ordinal and nominal outcomes explained by a latent variable vector and, when relevant, a set of other endogeneous and exogenous variables as well.

Consider *N* ordinal outcomes for the individual, and let *n* be the index for the ordinal outcomes , in our application *N*=13 for 906 individuals and *N*=12 for the remainder of the sample. Also, let  be the number of categories for the *nth* ordinal outcome  and let the corresponding index be. Let  be the latent underlying variable whose horizontal partitioning leads to the observed outcome for the *nth* ordinal variable. Assume that the individual under consideration chooses the  ordinal category. Then, in the usual ordered response formulation, for the individual, we may write:

 (3)

where  is a vector of exogenous and possibly endogenous variables as defined earlier,  is a corresponding vector of coefficients to be estimated,  is an vector of latent variable loadings on the *nth* continuous outcome, the  terms represent thresholds, and  is the standard normal random error for the *nth* ordinal outcome. For each ordinal outcome, ; , , and .

For later use, let  and  Stack the *N* underlying continuousvariables  into an vector , and the *N* error terms  into another vector . Define  [ matrix] and  [ matrix], and let  be the identity matrix of dimension *N* representing the correlation matrix of  (so, ; again, this is for identification purposes, given the presence of the unobserved  vector to generate covariance. Finally, stack the lower thresholds for the decision-maker  into an  vector  and the upper thresholds  into another vector  Then, in matrix form, the measurement equation for the ordinal outcomes (indicators) for the decision-maker may be written as:

. (4)

Next, let there be *G* nominal (unordered-response) variables for an individual, and let *g* be the index for the nominal variables, in our application G=2. Also, let *Ig* be the number of alternatives corresponding to the *g*th nominal variable (*Ig*3) and let be the corresponding index. Both nominal outcomes in our application have *Ig*=3. Consider the *g*th nominal variable and assume that the individual under consideration chooses the alternative . Also, assume the usual random utility structure for each alternative .

 (5)

where  is as defined earlier,  is an  column vector of corresponding coefficients, and is a normal error term, and  is an -column vector of coefficients capturing the effects of latent variables. 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  is individual-specific and not alternative-specific, the corresponding element in is set to zero for at least one alternative 

To proceed, define  and . 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  to occupy columns +1 to . Further, define ,   vector), vector), matrix), and  (that is,  is a column vector that includes all elements of the matrices and ). Then, in matrix form, we may write Equation (3) as:

 (6)

where . As earlier, to ensure identification, we specify  as follows:

 (7)

In the general case, this allows the estimation of  terms across all the *G* nominal variables, as originating from  terms embedded in each matrix.

To develop the reduced form equations, replace the right side of Equation (2) for in Equations (4) and (6) to obtain the following system:

, (8)

. (9)

Now, consider the  vector . Define

 and . (10)

Then 

The model estimation is performed using Bhat’s (2011) MACML. We refer the reader to Bhat (2015) for the detailed explanation as well as information on model identification criteria.

**REFERENCES**

Bhat, C.R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part* B, 45(7), 923-939.

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.

**Table 1. Thresholds and constants of indicators and loadings of latent variables on indicators**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Attitudinal and lifestyle indicators** | Threshold 2 | | Threshold 3 | | Threshold 4 | | Constant | | Latent variable loading | |
| Coeff. | t-stat | Coeff. | t-stat | Coeff. | t-stat | Coeff. | t-stat | Coeff. | t-stat |
| **Privacy-sensitivity** |  |  |  |  |  |  |  |  |  |  |
| I don’t mind sharing a ride with strangers if it reduces my costs (inverse scale) | 2.523 | 19.85 | 3.598 | 21.06 | 5.123 | 19.08 | 2.504 | 12.84 | 1.792 | 14.09 |
| Having privacy is important to me when I make a trip | 0.922 | 12.13 | 1.799 | 22.17 | 3.076 | 33.69 | 2.101 | 23.01 | 0.575 | 16.21 |
| I feel uncomfortable sitting close to strangers | 0.954 | 17.55 | 1.737 | 25.04 | 2.777 | 25.44 | 1.409 | 22.24 | 0.427 | 6.19 |
| **Tech-savviness** |  |  |  |  |  |  |  |  |  |  |
| I frequently use online banking services | 1.133 | 8.67 | 2.606 | 18.136 | 4.099 | 28.56 | 2.559 | 12.83 | 1.601 | 55.44 |
| I frequently purchase products online | 0.506 | 6.475 | 1.017 | 11.17 | 1.849 | 19.27 | 1.861 | 14.69 | 0.681 | 26.15 |
| Learning how to use new smartphone apps is easy for me | 1.138 | 9.685 | 1.993 | 16.22 | 2.859 | 23.18 | 2.255 | 15.08 | 0.787 | 30.61 |
| **Variety-seeking lifestyle propensity (VSLP)** |  |  |  |  |  |  |  |  |  |  |
| I think it is important to have all sorts of new experiences and I am always trying new things | 1.159 | 13.78 | 2.374 | 26.41 | 3.676 | 35.80 | 2.631 | 19.33 | 0.930 | 22.40 |
| Looking for adventures and taking risks is important to me | 1.195 | 2.45 | 2.468 | 2.33 | 3.834 | 2.17 | 1.739 | 2.67 | 1.033 | 23.83 |
| I love to try new products before anyone else | 0.910 | 6.67 | 1.859 | 7.37 | 2.934 | 7.38 | 1.908 | 6.88 | 0.704 | 2.69 |
| **Green lifestyle propensity (GLP)** |  |  |  |  |  |  |  |  |  |  |
| When choosing my commute mode, there are many things that are more important than being environmentally friendly (inverse scale) | 1.045 | 15.37 | 1.860 | 16.49 | 2.746 | 15.00 | 0.988 | 12.66 | 0.158 | 1.84 |
| I don’t give much thought to saving energy at home (inverse scale) | 0.708 | 10.87 | 1.182 | 16.44 | 2.203 | 25.18 | 1.910 | 21.34 | 0.132 | 1.80 |