Online supplementary note to the paper entitled

"A Comprehensive Dwelling Unit Choice Model Accommodating Psychological Constructs within a Search Strategy for Consideration Set Formation"

Model System Estimation

Let $E=(N+C)$. Define $\vec{\boldsymbol{y}}=\left(\left[\tilde{\boldsymbol{y}}^{*}\right]',\left[\breve{\boldsymbol{y}}^{*}\right]'\right)'$ $[E\times 1$ vector], $\vec{\boldsymbol{\gamma}}=(\tilde{\boldsymbol{\gamma}}',\boldsymbol{0}_{AC})'$ $[E\times A$ matrix], $\vec{\boldsymbol{d}}=(\tilde{\boldsymbol{d}}',\breve{\boldsymbol{d}}')'$ $[E\times L$ matrix], and $\vec{\boldsymbol{\varepsilon}}=(\tilde{\boldsymbol{\varepsilon}}',\breve{\boldsymbol{\varepsilon}}')'$ $(E\times 1$ vector), where $\boldsymbol{0}_{AC}$ is a matrix of zeros of dimension $A\times C$. Let $\boldsymbol{\delta}$ be the collection of parameters to be estimated.

With the definitions above, we obtain the following reduced form system for $\vec{\boldsymbol{y}}$ and $\boldsymbol{U}$:

$$\vec{\boldsymbol{y}}=\vec{\boldsymbol{\gamma}}\boldsymbol{x}+\vec{\boldsymbol{d}}\boldsymbol{z}^{*}+\vec{\boldsymbol{\varepsilon}}=\vec{\boldsymbol{\gamma}}\boldsymbol{x}+\vec{\boldsymbol{d}}(\boldsymbol{\alpha w}+\boldsymbol{\eta})+\vec{\boldsymbol{\varepsilon}}=\vec{\boldsymbol{\gamma}}\boldsymbol{x}+\vec{\boldsymbol{d}}\boldsymbol{\alpha w}+\vec{\boldsymbol{d}}\boldsymbol{\eta}+\vec{\boldsymbol{\varepsilon}},\ \mathrm{Var}\,(\vec{\varepsilon})=\vec{\boldsymbol{\Sigma}}=\begin{bmatrix}\boldsymbol{\Xi} & \boldsymbol{0}\\ \boldsymbol{0} & \mathbf{IDEN}_{C}\end{bmatrix}$$

$$\boldsymbol{U}=\boldsymbol{bx}+\boldsymbol{\varpi}\boldsymbol{z}^{*}+\boldsymbol{\varsigma}=\boldsymbol{bx}+\boldsymbol{\varpi}(\boldsymbol{\alpha w}+\boldsymbol{\eta})+\boldsymbol{\varsigma}=\boldsymbol{bx}+\boldsymbol{\varpi}\boldsymbol{\alpha w}+\boldsymbol{\varpi}\boldsymbol{\eta}+\boldsymbol{\varsigma} \tag{1}$$

Now, consider the $[(E+\vec{G})\times 1)]$ vector $\boldsymbol{yU}=\left[\vec{\boldsymbol{y}}',\boldsymbol{U}'\right]'$. Define

$$\boldsymbol{B}=\begin{bmatrix}\boldsymbol{B}_1\\ \boldsymbol{B}_2\end{bmatrix}=\begin{bmatrix}\vec{\boldsymbol{\gamma}}\boldsymbol{x}+\vec{\boldsymbol{d}}\boldsymbol{\alpha w}\\ \boldsymbol{bx}+\boldsymbol{\varpi}\boldsymbol{\alpha w}\end{bmatrix}\ \text{and}\ \boldsymbol{\Omega}=\begin{bmatrix}\boldsymbol{\Omega}_1 & \boldsymbol{\Omega}_{12}'\\ \boldsymbol{\Omega}_{12} & \boldsymbol{\Omega}_2\end{bmatrix}=\begin{bmatrix}\vec{\boldsymbol{d}}\boldsymbol{\Gamma}\vec{\boldsymbol{d}}'+\vec{\boldsymbol{\Sigma}} & \vec{\boldsymbol{d}}\boldsymbol{\Gamma}\boldsymbol{\varpi}'\\ \boldsymbol{\varpi}\boldsymbol{\Gamma}\vec{\boldsymbol{d}}' & \boldsymbol{\varpi}\boldsymbol{\Gamma}\boldsymbol{\varpi}'+\boldsymbol{\Lambda}\end{bmatrix} \tag{2}$$

Then $\boldsymbol{yU}\sim \boldsymbol{MVN}_{E+\vec{G}}(\boldsymbol{B},\boldsymbol{\Omega})$. (3)

Bhat (2015) has identified sufficient identification conditions for the GHDM model, which are summarized here (details are in Bhat, 2015): (1) there are at least two latent variables, with each latent variable correlated with at least one other latent variable in the correlation matrix $\boldsymbol{\Gamma}$, (2) diagonality is maintained across the elements of the error term vector $\vec{\boldsymbol{\varepsilon}}$ (that is, $\vec{\boldsymbol{\Sigma}}$ is diagonal), (3) block-diagonality is maintained for the matrix $\boldsymbol{\Lambda}$ as in Equation (11) of the paper, (4) The error term vectors $\vec{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\varsigma}$ are independent of each other, (5) for each latent variable, there are at least two outcome variables that load only on that latent variable and no other latent variable (that is, there is at least two factor complexity one outcome variables for each latent variable), (6) if a specific variable in the vector $\boldsymbol{x}$ loads onto an element of the non-nominal outcome vector $\vec{\boldsymbol{y}}$, then that element does not depend on any latent variable that contains the specific variable as a covariate in the structural equation system, (7) if a specific variable in the vector $\boldsymbol{x}$ affects the utility of alternative i_g of a nominal variable, then the utility of alternative i_g does not depend on any latent variable that contains that specific variable as a covariate in the structural equation system, and (8) endogenous variable effects can be specified only in a single direction (as discussed in the footnote in Section 2.2 of the paper); in addition, when a continuous observed variable (say variable A) appears as a right side variable in the regression for another continuous observed variable, or as a right side variable in the latent regression underlying another count or ordinal variable, each latent variable appearing in the regression/latent regression for the other continuous/count/ordinal variable (say variable B) should have two factor complexity one outcome variables after excluding the equation for variable B. This latter condition is not needed when a non-continuous observed variable appears as a right side variable in the regression of any other observed variable because of the non-linear nature of the relationship between the latent regressions and the observed non-continuous variables.

To estimate the model, note that, under the utility maximization paradigm, $U_{gi_g} - U_{gm_g}$ must be less than zero for all $i_g \neq m_g$ corresponding to the *gth* 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 $\boldsymbol{u_g} = \left[\left(u_{g1m_g}, u_{g2m_g}, ..., u_{gI_g m_g} \right)'; i_g \neq m_g \right]$. Also, define $\boldsymbol{u} = \left([\boldsymbol{u_1}]', [\boldsymbol{u_2}]', ..., [\boldsymbol{u_G}]' \right)'$ and develop the distribution of the vector $\boldsymbol{yu} = (\ddot{\boldsymbol{y}}', \boldsymbol{u}')'$ from that of $\boldsymbol{yU} = [\ddot{\boldsymbol{y}}', \boldsymbol{U}']'$. To do so, define a matrix $\mathbf{M}$ of size $[E + \tilde{G}] \times [E + \vec{G}]$. Fill this matrix up with values of zero. Then, insert an identity matrix of size *E* into the first *E* rows and *E* columns of the matrix **M**. Next, consider the rows from $E+1$ to $E + I_1 - 1$, and columns from $E+1$ to $E + I_1$. These rows and columns correspond to the first nominal 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 $E + I_1$ through $E + I_1 + I_2 - 2$ and columns $E + I_1 + 1$ through $E + I_1 + I_2$ 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. Continue this procedure for all *G* nominal variables. With the matrix $\mathbf{M}$ as defined, we can write $\boldsymbol{yu} \sim \boldsymbol{MVN}_{\boldsymbol{E}+\tilde{G}}(\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Omega}})$, where $\tilde{\boldsymbol{B}} = \mathbf{M}\boldsymbol{B}$ and $\tilde{\boldsymbol{\Omega}} = \mathbf{M\Omega M'}$.

Next, define threshold vectors as follows: $\ddot{\boldsymbol{\psi}}_{\boldsymbol{low}} = \left[\tilde{\boldsymbol{\psi}}'_{\boldsymbol{low}}, \breve{\boldsymbol{\psi}}'_{\boldsymbol{low}}, \left(-\infty_{\tilde{G}}\right)', \right]'$ ($[(N + C + \tilde{G}) \times 1]$ vector) and $\ddot{\boldsymbol{\psi}}_{\boldsymbol{up}} = \left[\tilde{\boldsymbol{\psi}}'_{up}, \breve{\boldsymbol{\psi}}'_{up}, \left(\boldsymbol{0}_{\tilde{G}}\right)'\right]'$ ($[(N + C + \tilde{G}) \times 1]$ vector), where $-\infty_{\tilde{\boldsymbol{G}}}$ is a $\tilde{G} \times 1$-column vector of negative infinities, and $\boldsymbol{0}_{\tilde{G}}$ is another $\tilde{G} \times 1$-column vector of zeros. Then the likelihood function may be written as:

$$L(\boldsymbol{\delta}) = \Pr\left[\ddot{\boldsymbol{\psi}}_{low} \leq \boldsymbol{yu} \leq \ddot{\boldsymbol{\psi}}_{up}\right] = \int\limits_{\boldsymbol{D_r}} f_{N+C+\tilde{G}}(\boldsymbol{r} \mid \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Omega}})dr, \tag{4}$$

where the integration domain $\boldsymbol{D_r} = \{\boldsymbol{r} : \ddot{\boldsymbol{\psi}}_{low} \leq \boldsymbol{r} \leq \ddot{\boldsymbol{\psi}}_{up}\}$ is simply the multivariate region implied by the observed non-nominal indicator outcomes, and the range $(-\infty_{\tilde{\boldsymbol{G}}}, \boldsymbol{0}_{\tilde{G}})$ for the utility differences taken with respect to the utility of the observed choice alternative for the nominal outcome. $f_{N+C+\tilde{G}}(\boldsymbol{r} \mid \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Omega}})$ is the multivariate normal density function of dimension $N + C + \tilde{G}$ with a mean of $\tilde{\boldsymbol{B}}$ and a covariance of $\tilde{\boldsymbol{\Omega}}$, and evaluated at $\boldsymbol{r}$. The likelihood function for a sample of Q households is obtained as the product of the household-level likelihood functions.

The above likelihood function involves the evaluation of a $N + C + \tilde{G}$-dimensional rectangular integral for each decision-maker, which can be computationally expensive. An alternative estimation technique is Bhat’s (2011) Maximum Approximate Composite Marginal Likelihood (MACML) approach.

The Joint Mixed Model System and the MACML Estimation Approach

In the MACML procedure, we develop the following (pairwise) composite marginal likelihood function formed by taking the products (across the *N* grouped variables, the *C* count variables, and *G* nominal variables) of the joint pairwise probability of the chosen alternatives for a household:

$$
\begin{aligned}
L_{CML}(\boldsymbol{\delta}) = & \left(\prod_{n=1}^{N-1} \prod_{n'=n+1}^{N} \Pr(j_n = a_n, j_{n'} = a_{n'}) \right) \times \left(\prod_{c=1}^{C-1} \prod_{c'=c+1}^{C} \Pr(k_c = r_c, k_{c'} = r_{c'}) \right) \times \\
& \times \left(\prod_{n=1}^{N} \prod_{c=1}^{C} \Pr(j_n = a_n, k_c = r_c) \right) \times \left(\prod_{g=1}^{G} \prod_{n=1}^{N} \Pr(j_n = a_n, i_g = m_g) \right) \\
& \times \left(\prod_{g=1}^{G} \prod_{c=1}^{C} \Pr(k_c = r_c, i_g = m_g) \right) \times \left(\prod_{g=1}^{G-1} \prod_{g'=g+1}^{G} \Pr(i_g = m_g, i_{g'} = m_{g'}) \right).
\end{aligned} \tag{5}
$$

To explicitly write out the CML function in terms of the standard and bivariate standard normal density and cumulative distribution function, define $\boldsymbol{\omega}_{\Delta}$ as the diagonal matrix of standard deviations of matrix $\boldsymbol{\Delta}$, $\phi_R(.;\boldsymbol{\Delta}^{**})$ for the multivariate standard normal density function of dimension R and correlation matrix $\boldsymbol{\Delta}^*$ ($\boldsymbol{\Delta}^* = \boldsymbol{\omega}_{\Delta}^{-1}\boldsymbol{\Delta}\boldsymbol{\omega}_{\Delta}^{-1}$), and $\Phi_E(.;\boldsymbol{\Delta}^*)$ for the multivariate standard normal cumulative distribution function of dimension E and correlation matrix $\boldsymbol{\Delta}^*$. Consider two selection matrices as follows: (1) $\boldsymbol{D}_{vg}$, an $I_g \times (N + C + \tilde{G})$ selection matrix, with an entry of '1' in the first row and the v^{th} column, an identity matrix of size $I_g - 1$ occupying the last $I_g - 1$ rows and the $N + C + \left[\sum_{j=1}^{g-1}(I_j - 1) + 1\right]^{th}$ through $N + C + \left[\sum_{j=1}^{g}(I_j - 1)\right]^{th}$ columns (with the convention that $\sum_{j=1}^{0}(I_j - 1) = 0$), and entries of '0' everywhere else, (2) $\boldsymbol{R}_{gg'}$, an $(I_g + I_{g'} - 2) \times (N + C + \tilde{G})$ selection matrix, with an identity matrix of size ($I_g - 1$) occupying the first ($I_g - 1$) rows and the $N + C + \left[\sum_{j=1}^{g-1}(I_j - 1) + 1\right]^{th}$ through $N + C + \left[\sum_{j=1}^{g}(I_j - 1)\right]^{th}$ columns (with the convention that $\sum_{j=1}^{0}(I_j - 1) = 0$), and another identity matrix of size ($I_{g'} - 1$) occupying the last ($I_{g'} - 1$) rows and the $N + C + \left[\sum_{j=1}^{g'-1}(I_j - 1) + 1\right]^{th}$ through $N + C + \left[\sum_{j=1}^{g'}(I_j - 1)\right]^{th}$ columns; all other elements of $\boldsymbol{R}_{gg'}$ take a value of zero. Also, let $\hat{\boldsymbol{\Omega}}_{vg} = \boldsymbol{D}_{vg}\tilde{\boldsymbol{\Omega}}\boldsymbol{D}'_{vg}$, $\vec{\boldsymbol{\Omega}}_{gg'} = \boldsymbol{R}_{gg'}\tilde{\boldsymbol{\Omega}}\boldsymbol{R}'_{gg'}$, $\mu_{v,up} = \dfrac{[\vec{\tilde{\boldsymbol{\psi}}}_{\boldsymbol{up}}]_v - [\tilde{\boldsymbol{B}}]_v}{\sqrt{[\tilde{\boldsymbol{\Omega}}]_{vv}}}$, $\mu_{v,low} = \dfrac{[\vec{\tilde{\boldsymbol{\psi}}}_{\boldsymbol{low}}]_v - [\tilde{\boldsymbol{B}}]_v}{\sqrt{[\tilde{\boldsymbol{\Omega}}]_{vv}}}$, $\rho_{vv'} = \dfrac{[\tilde{\boldsymbol{\Omega}}]_{vv'}}{\sqrt{[\tilde{\boldsymbol{\Omega}}]_{vv}[\tilde{\boldsymbol{\Omega}}]_{v'v'}}}$, where $[\vec{\tilde{\boldsymbol{\psi}}}_{\boldsymbol{up}}]_v$

represents the v^{th} element of $\breve{\boldsymbol{\psi}}_{\boldsymbol{up}}$ (and similarly for other vectors), and $[\tilde{\boldsymbol{\Omega}}]_{vv'}$ represents the vv'^{th} element of the matrix $\tilde{\boldsymbol{\Omega}}$. Then,

$$L_{CML}(\boldsymbol{\delta}) = \left(\prod_{v=1}^{N+C-1} \prod_{v'=v+1}^{N+C} \left[\begin{array}{l} \Phi_2(\mu_{v,up}, \mu_{v',up}, \rho_{vv'}) \; - \Phi_2(\mu_{v,up}, \mu_{v',low}, \rho_{vv'}) \\ -\Phi_2(\mu_{v,low}, \mu_{v',up}, \rho_{vv'}) + \Phi_2(\mu_{v,low}, \mu_{v',low}, \rho_{vv'}) \end{array} \right] \right) \times$$

$$\left(\prod_{v=1}^{N+C} \prod_{g=1}^{G} \Phi_{I_g} \left[\boldsymbol{\omega}_{\hat{\Omega}_{vg}}^{-1} \boldsymbol{D}_{vg} \left\{ \ddot{\boldsymbol{\psi}}_{\boldsymbol{up}} - \tilde{\boldsymbol{B}} \right\}; \hat{\boldsymbol{\Omega}}_{vg}^{*} \right] - \Phi_{I_g} \left[\boldsymbol{\omega}_{\hat{\Omega}_{vg}}^{-1} \boldsymbol{D}_{vg} \left\{ \ddot{\boldsymbol{\psi}}_{\boldsymbol{low}} - \tilde{\boldsymbol{B}} \right\}; \hat{\boldsymbol{\Omega}}_{vg}^{*} \right] \right) \times \tag{6}$$

$$\left(\prod_{g=1}^{G-1} \prod_{g'=1}^{G} \Phi_{I_g + I_{g'} - 2} \left[\boldsymbol{\omega}_{\hat{\Omega}_{gg'}}^{-1} \boldsymbol{R}_{\boldsymbol{gg'}} \left\{ - \tilde{\boldsymbol{B}} \right\}; \vec{\boldsymbol{\Omega}}_{gg'}^{*} \right] \right),$$

where $\ddot{\boldsymbol{\psi}}_{\boldsymbol{low}} = \left[\tilde{\boldsymbol{\psi}}'_{\boldsymbol{low}}, \breve{\boldsymbol{\psi}}'_{\boldsymbol{low}}, \left(0_{\tilde{G}}\right)', \right]'$.

In the MACML approach, all MVNVD function evaluation greater than two dimensions in the expression above are evaluated using an *analytic approximation* method rather than a simulation method (see Bhat, 2011). As has been demonstrated by Bhat and Sidharthan (2011), the MACML method has the virtue of computational robustness in that the approximate CML surface is smoother and easier to maximize than traditional simulation-based likelihood surfaces. Write the resulting equivalent of Equation (6) computed using the analytic approximation for the MVNCD function as $L_{MACML,q}(\boldsymbol{\delta})$, after introducing the index q for households. The MACML estimator is then obtained by maximizing the following function:

$$\log L_{MACML}(\boldsymbol{\delta}) = \sum_{q=1}^{Q} \log L_{MACML,q}(\boldsymbol{\delta}). \tag{7}$$

In the actual empirical analysis in the paper, lot size is not defined for dwelling units in apartment complexes. Thus, the likelihood function in Equation (4) and the CML function in Equation (5) have to be modified in a minor way. Specifically, in Equation (4), the likelihood function becomes the product of two components: (1) one component for dwelling units in apartment complexes, where the likelihood corresponds to the probability of the multidimensional set of chosen attributes but sans the lot size dimension (this reduces the dimensionality of the integral), and (2) a second component for non-apartment complexes that takes the exact form as in the current Equation (4). In the CML function of Equation (5), there are two multiplicative components once again: (1) one component for dwelling units in apartment complexes where pairwise combinations of other dimensions with the lot size dimension do not appear, and (2) a second component for non-apartment complexes where the CML function is exactly as is in the current Equation (5).

The covariance matrix of the parameters $\boldsymbol{\delta}$ may be estimated by the inverse of Godambe's (1960) sandwich information matrix (see Zhao and Joe, 2005).

$$V_{MACML}(\boldsymbol{\delta}) = \frac{\left[\hat{\boldsymbol{G}}(\boldsymbol{\delta})\right]^{-1}}{Q} = \frac{\left[\hat{\boldsymbol{H}}^{-1}\right]\left[\hat{\boldsymbol{J}}\right]\left[\hat{\boldsymbol{H}}^{-1}\right]'}{Q}, \tag{8}$$

with $\hat{\boldsymbol{H}} = -\frac{1}{Q}\left[\sum_{q=1}^{Q} \frac{\partial^2 \log L_{MACML,q}(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\right]_{\hat{\boldsymbol{\delta}}_{MACML}}$

$$\hat{\boldsymbol{J}} = \frac{1}{Q}\sum_{q=1}^{Q} \left[\left(\frac{\partial \log L_{MACML,q}(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}}\right)\left(\frac{\partial \log L_{MACML,q}(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}'}\right)\right]_{\hat{\boldsymbol{\delta}}_{MACML}} \tag{9}$$

An alternative estimator for $\hat{\boldsymbol{H}}$ may be obtained by computing the quantity below for each household, and averaging across households:

$$\hat{\boldsymbol{H}} \text{ for each } q = \left(\begin{array}{l}
\sum_{n=1}^{N-1}\sum_{n'=n+1}^{N}\left[\frac{\partial \log\left[\Pr(j_n = a_n, j_{n'} = a_{n'})\right]}{\partial\boldsymbol{\delta}}\right]\left[\frac{\partial \log\left[\Pr(j_n = a_n, j_{n'} = a_{n'})\right]}{\partial\boldsymbol{\delta}'}\right] + \\
\sum_{c=1}^{C-1}\sum_{c'=c+1}^{C}\left[\frac{\partial \log[\Pr(k_c = r_c, k_{c'} = r_{c'})]}{\partial\boldsymbol{\delta}}\right]\left[\frac{\partial \log\left[\Pr(k_c = r_c, k_{c'} = r_{c'})\right]}{\partial\boldsymbol{\delta}'}\right] + \\
\sum_{n=1}^{N}\sum_{c=1}^{C}\left[\frac{\partial \log[\Pr(j_n = a_n, k_c = r_c)]}{\partial\boldsymbol{\delta}}\right]\left[\frac{\partial \log\left[\Pr(j_n = a_n, k_c = r_c)\right]}{\partial\boldsymbol{\delta}'}\right] + \\
\sum_{g=1}^{G}\sum_{n=1}^{N}\left[\frac{\partial \log\left[\Pr(j_n = a_n, i_g = m_g)\right]}{\partial\boldsymbol{\delta}}\right]\left[\frac{\partial \log\left[\Pr(j_n = a_n, i_g = m_g)\right]}{\partial\boldsymbol{\delta}'}\right] + \\
\sum_{g=1}^{G}\sum_{c=1}^{C}\left[\frac{\partial \log\left[\Pr(k_c = r_c, i_g = m_g)\right]}{\partial\boldsymbol{\delta}}\right]\left[\frac{\partial \log[\Pr(k_c = r_c, i_g = m_g)]}{\partial\boldsymbol{\delta}'}\right] \\
\sum_{g=1}^{G-1}\sum_{g'=g+1}^{G}\left[\frac{\partial \log\left[\Pr(i_g = m_g, i_{g'} = m_{g'})\right]}{\partial\boldsymbol{\delta}}\right]\left[\frac{\partial \log\left[\Pr(i_g = m_g, i_{g'} = m_{g'})\right]}{\partial\boldsymbol{\delta}'}\right] +
\end{array}\right)$$

Positive Definiteness

The matrix $\tilde{\boldsymbol{\Omega}}$ for each household has to be positive definite. The simplest way to guarantee this in our mixed model system is to ensure that the $(L \times L)$ correlation matrix $\boldsymbol{\Gamma}$ is positive definite, and each matrix $\breve{\mathbf{\Lambda}}_{\mathbf{g}}$ $(g=1,2,...,G)$ is also positive definite. An easy way to ensure the positive-definiteness of these matrices is to use a Cholesky-decomposition and parameterize the CML function in terms of the Cholesky parameters. Further, because the matrix $\mathbf{\Gamma}$ is a correlation matrix, we write each diagonal element (say the aa^{th} element) of the lower triangular Cholesky matrix of $\mathbf{\Gamma}$ as $\sqrt{1-\sum_{j=1}^{a-1} p_{aj}^2}$, where the p_{aj} elements are the Cholesky factors that are to be estimated. In addition, note that the top diagonal element of each $\breve{\mathbf{\Lambda}}_{\mathbf{g}}$ matrix has to be normalized to one (as discussed earlier in Section 2.2 of the paper), which implies that the first element of the Cholesky matrix of each $\breve{\mathbf{\Lambda}}_{\mathbf{g}}$ is fixed to the value of one.

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*, forthcoming.

Bhat, C.R., Sidharthan, R., 2011. A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models. *Transportation Research Part B*, 45(7), 940-953.

Godambe, V.P., 1960. An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31(4), 1208-1211.

Zhao, Y., Joe, H., 2005 . Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, 33(3), 335-356.