**Code Documentation for the GHDM**

This addendum serves to document the estimation code for a probit kernel-based generalized heterogeneous data model (an extension of the traditional hybrid choice model). An explanation of the parameters in the model and the different settings available within the code are documented below. Please refer to Bhat (2015) for the notations and the model structure.

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

1. **Dataset Specifications**

The dataset should be a csv file with no header. The name of the columns is generated inside the code. Alternatively, one can use atog command available in Gauss to create a data file with column names. The dataset provided with the code contains variables described in the simulation framework of the main paper. The ordering of columns in the dataset is provided below in Table 1.

**Table 1: Description of dataset**

|  |  |
| --- | --- |
| **Column Number** | **Variable Name** |
| 1 | Bachelor’s degree or higher or not |
| 2 | Male or not |
| 3 | High income or not |
| 4 | # of children <11 years |
| 5 | # of young adults |
| 6 | Immigrant household |
| 7 | Own house |
| 8 | Urban dwelling |
| 9 | Auto ownership |
| 10 | Commute distance |
| 11 | Observed outcome for non-commute (NC) propensity by non-motorized mode |
| 12 | Observed outcome for non-commute (NC) propensity by public transport |
| 13 | Observed outcome for non-commute (NC) propensity by motorized mode |
| 14 | Observed outcome for residential location (1: rural, 2: urban, and 3: suburban) |
| 15 | Observed outcome for commute mode (1: motorized-mode, 2: public transport, 3: non-motorized mode) |
| 16 | Column of 1’s |
| 17 | Column of 0’s |

1. **Code Settings**

The user must specify the value of following variables (lines 15-16, 30-52, and 118-120).

**Table 2: Description of Variables**

|  |  |
| --- | --- |
| **Variables** | **Description** |
| MACML or GHK | Set MACML=1 for evaluation of multivariate normal cumulative distribution function using Joe-Solow approach, otherwise set GHK=1 for using GHK simulator |
| nind | Number of observations (3000 in the current example provided in the code) |
| nvar\_latent | Number of latent variables in the structural equation |
| nvar\_mear\_cont | Number of continuous variables in the measurement equation |
| nvar\_mear\_ordl | Number of ordinal variables in the measurement equation |
| nvar\_mear\_count | Number of count variables in the measurement equation |
| All\_Nominal | Number of alternatives for each nominal variable. Please refer to the code for additional details on how to specify the number of alternatives. |
| gradient\_limit | Tolerance used for convergence |
| iteration\_limit | Maximum number of iterations for deciding the convergence |
| Force\_convergence | Set this to 1, if you want convergence based on maximum number of iterations |

The code is commented with appropriate descriptions and examples to help users change the settings to suit their data needs and requirements. We hope that this should help users run the program without any difficulties. However, in the event of any issues, please contact Subodh Dubey at [subbits@gmail.com](mailto:subbits@gmail.com)

Further, the current version of the code uses six threads (i.e., it is a multithreaded version). We recommended using at least six threads for estimation given the high estimation time of complex models such as GHDM. However, users can change the number of threads by setting the variable “Num\_Threads” in the section Thread Settings in the code and then appropriately change the number of receiving functions inside the functions “lpr1, lgd1, and User\_Hess”.

1. **Estimation Results**

The code has two likelihood expressions. The first likelihood expression estimates the parameterized version of parameters (for required parameters as discussed in the paper). Once estimated, parameters are un-parametrized and sent to second likelihood expression to obtain meaningful parameter values and their standard errors. Finally, the standard error and the corresponding t-stat value for each of the active un-parameterize coefficients are reported using the sandwich matrix under the section titled “Snapshot of final result,” as shown below. All the functions used in the likelihood, gradient and Hessian expressions are provided at the end of the code. Further, the code uses analytic first and second order conditions for the calculation of Jacobian and Hessian matrices.

Here the structural equation coefficients are represented by Alpha. The non-diagonal elements (lower triangular matrix) of the structural equation correlation matrix are represented by Tild. The measurement equation exogenous variable coefficients of non-nominal variables are represented by Gamma followed by loading of latent variables on non-nominal variables by D\_cap. The variance of the continuous variable of measurement equation is represented by Psi. The exogenous variable coefficients for nominal variables of measurement equation are represented by Beta. The elements of the matrix of latent variable loading on the alternatives are represented by Lambda. The elements of (lower triangular matrix) of the error differenced matrix are represented by Err1 and finally, the upper thresholds for the ordinal variables, flexibility parameters, and dispersion parameters for the count variables are represented by Th, Phi, and Theta respectively.

Please note that we have tried to provide the name of coefficients in the output (from the code) similar to the notations used in the paper for easy understanding of readers.

**Snapshot of final result**

-----------------Parameters and T-Statistics-------------------------------------------------------------------

Log-likelihood value : -97482.9

Parameter Estimate ST.Err T-Stat

Alpha01 0.774 0.084 9.240

Alpha02 -0.286 0.058 -4.892

Alpha03 0.297 0.346 0.859

Alpha04 0.577 0.334 1.730

Tild02 -0.503 0.321 -1.565

Gamma01 0.999 0.057 17.500

Gamma02 0.395 0.052 7.629

Gamma03 0.084 0.051 1.631

Gamma04 1.243 0.082 15.124

Gamma05 -0.165 0.033 -4.997

Gamma06 0.338 0.090 3.766

Gamma07 1.036 0.041 25.409

Gamma08 0.173 0.071 2.435

Gamma09 0.928 0.087 10.723

Gamma10 0.423 0.033 12.922

Gamma11 -0.274 0.030 -8.999

Gamma12 0.787 0.089 8.873

Gamma13 -0.560 0.081 -6.932

D\_caP01 0.194 0.060 3.234

D\_caP02 0.725 0.061 11.808

D\_caP03 0.126 0.042 3.038

D\_caP04 0.204 0.069 2.948

D\_caP05 -0.499 0.092 -5.425

D\_caP06 0.549 0.185 2.965

Psi01 0.770 0.030 26.049

Beta01 0.298 0.054 5.546

Beta02 0.369 0.016 22.403

Beta03 -0.527 0.025 -20.932

Beta04 0.346 0.031 11.239

Beta05 0.198 0.016 12.341

Beta06 0.251 0.021 11.938

Beta07 -0.426 0.026 -16.253

Beta08 0.245 0.020 12.534

Beta09 0.336 0.038 8.894

Beta10 -0.956 0.069 -13.813

Beta11 -0.122 0.008 -15.078

Beta12 -0.089 0.026 -3.430

Lambda02 0.233 0.051 4.582

Lambda04 0.146 0.085 1.714

Lambda05 0.350 0.038 9.119

Lambda06 0.203 0.061 3.345

Err1\_02 0.625 0.065 9.588

Err1\_03 1.512 0.208 7.282

Err1\_05 0.327 0.136 2.399

Err1\_06 1.325 0.207 6.395

Th\_01 1.610 0.066 24.276

Th\_02 1.494 0.050 30.173

Th\_03 1.479 0.058 25.659

Phi01 0.757 0.087 8.701

Theta01 1.964 0.265 7.412

--------------------------------------------------------------------------------------------------------------