**Code Documentation for the Spatial GHDM**

This addendum serves to document the estimation code for a probit kernel-based spatial 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 and Pinjari (2015) for the notations and the model structure.

Bhat, C. R. and Pinjari, A. R. (2015). On Accommodating Spatial Interactions in a Generalized Heterogeneous Data Model (GHDM) of Mixed Types of Dependent Variables. *Technical paper, University of Texas at Austin.*

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 18-19, 33-54, and 149-152).

**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 of both the aspatial and spatial models |
| iteration\_limit1 | Maximum number of iterations for deciding the convergence of aspatial model |
| iteration\_limit2 | Maximum number of iterations for deciding the convergence of spatial model |
| Force\_convergence | Set this to 1, if you want convergence based on maximum number of iterations for both the aspatial and spatial models |

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, lpr, lgd, and User\_Hess”.

1. **Estimation Results**

The code has three likelihood expressions. The first likelihood expression estimates the parameterized version of parameters (for required parameters as discussed in the paper) assuming no spatial dependency. This is done to obtain the good starting value for the spatial GHDM model as complex models, especially spatial models require good starting value for optimization. Once estimated, parameters are sent to second likelihood expression (parameterize spatial model likelihood expression) to obtain parametrized parameters. Once the estimation is finished, parameters from the second likelihood are un-parametrized and their standard errors are obtained using the sandwich estimator. The code then displays the final parameter estimate, standard error and T-stat as shown below in the section titled “Snapshot of final result,” as shown below. During optimization, the Gauss software will display an error indicating that the cross-product calculation failed. Users don’t need to be alarmed on display of this error as calculation of the Jacobian matrix in the spatial models are different from aspatial model and need to be done separately. Since the code doesn’t collect any Jacobian or Hessian matrix values from the Gauss software, we simply assign the covariance matrix function to provide Jacobian matrix to save on time. The code uses analytic first and second order conditions for the calculation of Jacobian and Hessian matrices. All the functions used in the likelihood, gradient and Hessian expressions are provided at the end of the code.

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, dispersion parameters, and spatial auto-correlation parameters for the count variables and latent variables are represented by Th, Phi, Theta, and Delta 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 : -832879.00

Parameter Estimate ST.Err T-Stat

Alpha01 0.737 0.174 4.230

Alpha02 -0.232 0.090 -2.576

Alpha03 0.382 0.515 0.742

Alpha04 0.589 0.498 1.182

Tild02 -0.677 0.483 -1.402

Gamma01 0.949 0.121 7.864

Gamma02 0.333 0.116 2.864

Gamma03 0.205 0.068 3.005

Gamma04 1.210 0.193 6.278

Gamma05 -0.171 0.027 -6.302

Gamma06 0.364 0.074 4.952

Gamma07 1.039 0.043 24.335

Gamma08 0.166 0.058 2.887

Gamma09 0.914 0.090 10.133

Gamma10 0.425 0.026 16.118

Gamma11 -0.273 0.025 -10.910

Gamma12 0.740 0.256 2.887

Gamma13 -0.591 0.051 -11.586

D\_caP01 0.196 0.118 1.660

D\_caP02 0.729 0.079 9.288

D\_caP03 0.117 0.045 2.624

D\_caP04 0.194 0.080 2.420

D\_caP05 -0.497 0.179 -2.777

D\_caP06 0.532 0.266 1.998

Psi01 0.895 0.090 9.915

Beta01 0.217 0.054 3.982

Beta02 0.191 0.030 6.348

Beta03 -0.210 0.011 -18.435

Beta04 0.654 0.039 16.751

Beta05 0.081 0.008 9.978

Beta06 0.130 0.017 7.703

Beta07 -0.303 0.054 -5.586

Beta08 0.114 0.025 4.567

Beta09 0.129 0.029 4.378

Beta10 -1.093 0.085 -12.839

Beta11 -0.046 0.004 -12.845

Beta12 -0.049 0.024 -2.046

Lambda02 0.144 0.060 2.404

Lambda04 0.092 0.054 1.690

Lambda05 0.155 0.027 5.736

Lambda06 0.081 0.080 1.022

Err1\_02 0.516 0.065 7.888

Err1\_03 1.562 0.208 7.513

Err1\_05 0.026 0.207 0.124

Err1\_06 1.140 0.195 5.848

Th\_01 1.596 0.056 28.549

Th\_02 1.492 0.041 36.538

Th\_03 1.476 0.047 31.213

Phi01 0.770 0.150 5.118

Theta01 2.015 0.845 2.385

Delta01 0.286 0.363 0.789

Delta02 0.280 0.427 0.655