**Code Documentation for the Estimation of a Probit Kernel-based Hybrid Choice Model**

This addendum serves to document the estimation code for a probit kernel-based hybrid choice model, which integrates latent psychological constructs such as attitudes and preferences within traditional choice models. An explanation of the parameters in the model and the different settings available within the code are documented below. Please refer to Bhat, and Dubey (2013) for the notations and the model structure.

Bhat, C.R., and S.K. Dubey (2013) "**A New Estimation Approach to Integrate Latent Psychological Constructs in Choice Modeling**," Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, 2013.

1. **Data Generation Settings**

* The current version of the code allows a user to generate the data within the code as well as specify the external dataset.
* If the user wants to simulate the data inside the code, choose the option “Simulation\_data” (line 25). On the other hand, if the user wants to specify his or her own dataset, choose the option “Own\_data” (line 24) and specify the file name in the section “How to specify your own dataset” in the code. The user can also change the seed for data generation process under the simulation option by changing the value of variable “Run\_No” (line 51).

1. **Dataset Specifications**

The dataset should be a csv file with no header. The dataset should include the columns in the following order.

1. Columns of covariates of structural equation.
2. Columns of covariates of choice model equation.
3. A column filled with the value 1.
4. A column filled with the value 0.
5. Columns of measurement equation variables. The ordering of variables should be as follows: first specify all the continuous variables followed by ordinal variables.
6. A column indicating the chosen alternative number. For example: if the choice model has three alternatives- car, rail and transit, the corresponding row in the column should have a value 1 if the person chooses car, 2 for rail and 3 for transit.
7. **Simulation Data**

The data in the code is generated using the simulation configuration discussed in the paper titled “**A New Estimation Approach to Integrate Latent Psychological Constructs in Choice Modeling”.**  Please refer to the paper for details of simulation configuration.

1. **Code Settings**

The user must specify the value of following variables (lines 30 – 36).

**Table 1: Description of Variables**

|  |  |
| --- | --- |
| **Variables** | **Description** |
| nind | Number of observations |
| nvar\_latent | Number of latent variables in the structural equation |
| nvar | Number of covariates in the structural equation |
| nvar\_mear | Total number of variables in the measurement equation |
| nvar\_mear\_cont | Number of continuous variables in the measurement equation |
| nvar\_ut | Number of exogenous variables in the choice model equation |
| nc | Number of alternatives in the choice model |

1. **Sample Data**

The data consist of 1000 individual records (i.e., the number of rows in the dataset equals 1000). There are six covariates in the structural equation, five variables in the measurement equation and three alternatives with two explanatory variables each. The following table presents the content and the structure of the sample dataset.

**Table 2: Sample Data Description**

|  |  |
| --- | --- |
| **Column No** | **Explanation** |
| 1 to 6 | Structural equation covariates |
| 7 to 12 | Explanatory variables for each of the three alternatives |
| 13 | A column of 1’s |
| 14 | A column of 0’s |
| 15 to 19 | Measurement equation variables |
| 20 | Column indicating the chosen alternative number |

We do not provide labels for the explanatory variables because the labels are generated internally by the program. Please refer to line number 106 for how to define labels for variables. The sample data follows the same structure as discussed in the simulation exercise of the paper.

1. **Estimation Results**

The code has two likelihood expressions as some sort of parameterization is necessary to ensure that diagonal elements of the correlation matrix equal 1 at the end of estimation because the code works with the Cholesky decomposition of various matrices to ensure that the final covariance matrix is positive definite. Thus, the first likelihood expression estimates the parameterized coefficients and then passes the un-parameterized values to the second likelihood expression.

The standard error for the parameterize coefficients is obtained from the Jacobian matrix. The standard error for the un-parameterized coefficients is obtained from the Hessian matrix. Finally, the standard error and the corresponding t-stat value for each of the active un-parameterize coefficients are reported using a sandwich matrix under the section titled “Final Result,” as shown below.

Here the structural equation coefficients are represented by Alpha. The Cholesky elements (lower triangular matrix) of the structural equation correlation matrix are represented by Tild. The measurement equation constants and remaining coefficients are represented by Delta and D\_cap. The variance of the continuous variable of measurement equation is represented by Psi. The choice equation alternative specific constants and exogenous variables are represented by ASC\_alternative name and TT and TC. The elements of the matrix of latent variable loading on the alternatives are represented by Lambda. The elements of Cholesky decomposition (lower triangular matrix) of the error differenced matrix are represented by Lamda and finally, the upper thresholds for the ordinal variables are represented by Th\_Up. Please note that for all the Cholesky decomposition elements, the corresponding position of the elements in the matrix is added at the end of their name. For example: Tild04 indicates that this element is the 4th element of the Cholesky decomposition (lower triangular matrix) of the structural equation correlation matrix.

---------------------------Final Result---------------------------------------------------------------------------

Log-likelihood value: -13870.64

Parameter Estimate ST.BHHH T\_Stat

Alpha01 0.4126 0.1620 2.5470

Alpha03 0.6120 0.2285 2.6786

Alpha08 0.2517 0.1744 1.4434

Alpha10 0.3192 0.1937 1.6482

Alpha13 0.4137 0.1674 2.4712

Alpha20 0.3688 0.2288 1.6116

Alpha23 -0.3146 0.1677 -1.8766

Alpha30 0.5251 0.2335 2.2487

Tild04 0.6808 0.4485 1.5181

Tild08 0.6114 0.5823 1.0500

Tild14 0.6886 0.4035 1.7064

Delta01 1.0281 0.0391 26.2620

Delta02 -1.3026 0.1469 -8.8683

Delta03 -0.9941 0.2484 -4.0018

Delta04 -1.3862 0.1642 -8.4418

Delta05 -1.2782 0.1205 -10.6053

D\_caP05 0.2546 0.0944 2.6980

D\_caP06 0.3592 0.1318 2.7266

D\_caP12 0.5843 0.4611 1.2671

D\_caP18 0.5739 0.3077 1.8652

D\_caP24 0.4876 0.3036 1.6062

Psi01 0.9697 0.0257 37.7879

ASC\_air 1.0650 0.4248 2.5070

ASC\_bus -1.3733 0.6323 -2.1721

TT -1.2379 0.4126 -3.0003

TC -1.0053 0.3307 -3.0403

Lambda06 0.3375 0.1889 1.7864

Lambda08 0.9254 0.3467 2.6694

Lambda10 -0.8519 0.5822 -1.4634

Lambda12 0.3120 0.3752 0.8315

Lambda14 0.5711 0.3234 1.7657

Lambda15 0.5843 0.4376 1.3353

Lamda02 1.1524 0.6761 1.7044

Lamda03 0.5459 0.8206 0.6653

Th\_Up01 1.4768 0.0996 14.8216

Th\_Up02 1.5874 0.3302 4.8065

Th\_Up03 1.5383 0.1512 10.1764

Th\_Up04 1.3335 0.1348 9.8962