**Code Documentation for “On the use of Probit based models for ranking data analysis”**

**Gopindra S. Nair, Chandra R. Bhat, Ram M. Pendyala, Becky P.Y. Loo, William H.K. Lam**

This addendum serves to document the GAUSS matrix programming codes for estimating the heteroskedastic and non-heteroskedastic rank ordered probit and logit models. If you use any part of the codes, either within the GAUSS environment or after translating the codes to a different environment, please do cite,

Nair, G.S., C.R. Bhat, R.M. Pendyala, B.P.Y. Loo and W.H.K. Lam, **"On the Use of Probit Based Models for Ranking Data Analysis,"** Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, October 2018

And,

Bhat, C.R., 2018. **“New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function.”** Transportation Research Part B: Methodological 109, 238–256.

1. **Setting up**

To run estimations using the ranking models, the entry point to the code is the file *Ranking Model Setup.gss*. Before running this code, it must be ensured that the libraries *maxlik*, *cdfmvnaT* and *rankingmodels* are available in your version of GAUSS. *maxlik* is a proprietary package that needs to be purchased from Aptech Systems, Inc. The code for *cdfmvnaT* and *rankingmodels* libraries is provided with this document. Only the file *cdfmvnaT.src* needs to be included in the *cdfmvnaT* library. The files *ranking\_config.sdf*, *rankingmodels.dec*, *rankingmodels.ext* and *rankingmodels.src* are to be included in the *rankingmodels* library. You may refer the link,

<https://www.aptech.com/gauss13-whats-new/library-tool/>

for more information on how to add libraries in GAUSS.

1. **Dataset**

The dataset to be used for estimation may be provided in the .dat format used by GAUSS or .csv format (in which case it will automatically be converted to the .dat format). When the .csv format is used, the first row must have the headers and the subsequent rows must have observations. The dataset should have columns for all independent variables, alternative ranks, a column called ‘UNO’ which contains only ‘1’s in all rows and a column called ‘SERO’ which contains only ‘0’ on all rows. For reliable performance of the code, please ensure that all independent variables used for estimations have relatively the same magnitude and variance. If this is not the case, the independent variable may be normalized and recentered into a new variable before estimation. There should be the same number of rank columns (dependent variables) as the number of alternatives. The dependent variable column for alternative *k* would contain the ranks given to alternative *k* by the different individuals. (Another common way of storing ranking data is when a column represents a rank and the column entries would be the indices of alternatives that is given that particular rank by different individuals. Please ensure that this is not the format followed in the dataset used with this code)

A sample dataset has been provided in the file *game.csv*. (van Dijk et al., 2007)

1. **Specification**

For most use cases, all the fields that needs to be edited before running the code can be found in the configuration section of *Ranking Model Setup.gss*. Use this section to specify the input dataset, utility functions and model to be used.

1. **Results**

The output from successfully running the code on the sample dataset will be as follows.

===============================================================================

MAXLIK Version 4.0.26 10/21/2018 1:58 am

===============================================================================

Data Set: C:/Users/gs27556/Box Sync/IATBR/Datasets/Game/game.dat

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

return code = 0

normal convergence

Mean log-likelihood -5.56493

Number of cases 91

Covariance matrix of the parameters computed by the following method:

Inverse of computed Hessian

Parameters Estimates Std. err. Est./s.e. Prob. Gradient

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

BASC\_PS 0.8318 0.3284 2.533 0.0057 0.0000

BASC\_PSP 0.5339 0.3567 1.497 0.0672 0.0000

BASC\_GC 0.0435 0.3600 0.121 0.4519 0.0000

BASC\_GB -0.3025 0.3951 -0.766 0.2219 0.0000

BASC\_XBO 1.2650 0.3510 3.604 0.0002 0.0000

BOWN 1.2316 0.2874 4.285 0.0000 0.0000

BHRS\_PS -0.1132 0.0509 -2.224 0.0131 0.0000

BHRS\_PSP -0.2138 0.0609 -3.511 0.0002 0.0000

BHRS\_GC -0.2156 0.0692 -3.117 0.0009 0.0000

BHRS\_GB -0.2430 0.0723 -3.362 0.0004 0.0000

BHRS\_XBO -0.1526 0.0532 -2.866 0.0021 0.0000

RS02 -0.0958 0.2414 -0.397 0.3457 0.0000

RS03 0.1478 0.2364 0.625 0.2660 0.0000

RS04 -0.2873 0.5061 -0.568 0.2851 0.0000

RS05 -0.3923 0.7051 -0.556 0.2890 0.0000

Correlation matrix of the parameters

1.000 0.621 0.570 0.537 0.656 0.359 -0.599 -0.277 -0.206 -0.183 -0.348 0.039 -0.064 0.081 0.103

0.621 1.000 0.552 0.607 0.646 0.330 -0.363 -0.554 -0.177 -0.180 -0.380 0.005 0.046 0.247 0.199

0.570 0.552 1.000 0.532 0.588 0.396 -0.290 -0.230 -0.507 -0.309 -0.282 0.019 0.002 -0.163 -0.078

0.537 0.607 0.532 1.000 0.545 0.091 -0.339 -0.250 -0.105 -0.291 -0.365 0.269 0.341 0.222 0.409

0.656 0.646 0.588 0.545 1.000 0.476 -0.356 -0.308 -0.247 -0.224 -0.552 -0.064 -0.117 0.038 0.126

0.359 0.330 0.396 0.091 0.476 1.000 -0.012 -0.076 -0.243 -0.207 0.018 -0.440 -0.565 -0.322 -0.277

-0.599 -0.363 -0.290 -0.339 -0.356 -0.012 1.000 0.511 0.375 0.354 0.605 -0.058 -0.030 -0.135 -0.152

-0.277 -0.554 -0.230 -0.250 -0.308 -0.076 0.511 1.000 0.509 0.518 0.586 0.260 0.104 -0.106 -0.084

-0.206 -0.177 -0.507 -0.105 -0.247 -0.243 0.375 0.509 1.000 0.690 0.417 0.405 0.269 0.469 0.190

-0.183 -0.180 -0.309 -0.291 -0.224 -0.207 0.354 0.518 0.690 1.000 0.403 0.409 0.193 0.438 0.175

-0.348 -0.380 -0.282 -0.365 -0.552 0.018 0.605 0.586 0.417 0.403 1.000 -0.014 -0.089 -0.149 -0.210

0.039 0.005 0.019 0.269 -0.064 -0.440 -0.058 0.260 0.405 0.409 -0.014 1.000 0.498 0.411 0.323

-0.064 0.046 0.002 0.341 -0.117 -0.565 -0.030 0.104 0.269 0.193 -0.089 0.498 1.000 0.366 0.347

0.081 0.247 -0.163 0.222 0.038 -0.322 -0.135 -0.106 0.469 0.438 -0.149 0.411 0.366 1.000 0.394

0.103 0.199 -0.078 0.409 0.126 -0.277 -0.152 -0.084 0.190 0.175 -0.210 0.323 0.347 0.394 1.000

Number of iterations 74

Minutes to convergence 0.18378

Rank Depth: 5.0000000

error\_unknown: 0.00000000

rank\_scaling: 1.0000000

model\_type: ROP

**References**

van Dijk, B., Fok, D., Paap, R., 2007. A rank-ordered logit model with unobserved heterogeneity in ranking capabilities (Econometric Institute Research Paper No. EI 2007-07). Erasmus University Rotterdam, Erasmus School of Economics (ESE), Econometric Institute.