**HITS program for Maximum Score Estimation**

Kostas Florios

Department of International and European Economic Studies, Athens University of Economics and Business, cflorios@aueb.gr

# Description

Hyperplanes Intersection Tabu Search (HITS) is a program that calculates the Maximum Score Estimator (MS) of the binary choice model (Manski, 1975; 1985). It is a variant of Tabu Search (TS) method (Glover, 1989) especially modified for MS. Specifically, it uses the gaussian elimination to locate the trial points of hyperplanes intersections (Pinkse, 1993). It uses the concept of “Neighbourhood” which is popular in trajectory-based methods of combinatorial optimization in order to search nearby solutions of the current solution and avoid complete enumeration. It also uses a “Tabu List” as a short-term memory of prohibited neighbouring solutions to which the search is prevented from moving to. The program is coded in Microsoft Visual Studio 2017 with Visual C++.

# Input

The basic input of the HITS program is 4 text files that contain the required information. Assume that we have the following dataset in a file called X.txt:

|  |
| --- |
| 1 11.077 2 0 6 4.1 16.81 1  2 11.835 2 0 6 4.4 19.36 1  3 11.225 2 0 4 4.6 21.16 1  4 11.835 2 0 6 4.4 19.36 1  5 11.835 2 0 6 4.4 19.36 1  …  2335 11.835 2 0 6 5.6 31.36 1  2336 11.352 2 0 4 4.7 22.09 1  2337 10.510 0 0 6 3.0 9.00 1  2338 11.222 0 0 17 5.4 29.16 1  2339 10.964 0 0 17 3.9 15.21 1 |

Also, assume the response variable saved in file y.txt as follows:

|  |
| --- |
| 1 -1  2 -1  3 -1  4 -1  5 -1  …  2335 1  2336 -1  2337 1  2338 1  2339 1 |

Then, the auxiliary text files to complete the estimation problem formulation have the following form:

First the names.txt file looks like:

|  |
| --- |
| OBS HW CHILD YCHILD EDU AGE AGE2 INTCPT |

Second the params.txt file looks like:

|  |
| --- |
| n = 2339  p = 7  b0 = -1  Nreps = 10  d = 1e300  //ntenure = 50  //niterovermax = 50  //niterover = 200 |

The user is welcome to double click the executable application DisplayGrid-MFC. The following screen is displayed with the estimation data already preloaded from the text files:

A screenshot of a computer

Description automatically generated

The observations of the dataset are presented in reverse order, counting from the last one (No. 2339) to the first one (No. 1, truncated in the Figure).

The required parameters for HITS program are five:

1. “**n**”: the number of rows in the loaded dataset or the number of observations in the regression.
2. “**p**”: the number of columns in the loaded dataset or the number of independent variables in the regression.
3. “**b0**”: is set to either 1 or -1 and is the coefficient of the first independent variable. This is done for identification and because the MS is identified up-to-scale. Multiplying the regression coefficients by any positive constant number provides the same objective function value. Prior knowledge can be used to select 1 or -1, or in general both values can be tried, and the regression will be run twice.
4. “**Nreps**”: the number of randomized independent runs of the tabu search algorithm. This is done to avoid getting trapped in local maxima using a single run.
5. “**d**”: the upper bound for the absolute value of each regression coefficient. It can be set to an arbitrary large value to allow unconstrained maximization. Alternatively, it can be set to a stricter value (say 1e5) to put bound constraints on the regression coefficients. The parameter d applied to each regression coefficient.

The selection of parameters depends on the size and nature of the problem. In general, the solution time increases almost proportionally with the Nreps. A rule of thumb is to keep the Nreps between 3 and 20. For an extremely large estimation problem (with N>20,000 and p>10) the user can specify Nreps=1 and see what the program HITS provides as a solution of a single heavy run. The HITS software has currently be compiled with limits: Tmax=50,000, pmax=20. Nevertheless, the user can set different limits and recompile the program for larger estimation problems.

Last, there are 3 secondary parameters, that are hard coded like: ntenure=50, niterover=200, niterovermax=50. (In the code the variable niterovermax is called niterovernotimproved). If the user wishes, they can change these default values and recompile. These values work well for standard problems with N<20,000 and p<10.

# Running

During the run the user can see:

1. The progress bar filling in (proportionally to: run i over Nreps, the total number of runs).

A screenshot of a computer

Description automatically generated

1. The log file: DemoExample\_Cpp\_tabusearch\_opt\_log\_iter200.txt with the trajectory of the search.

A screenshot of a computer

Description automatically generated

1. The info file: DemoExample\_Cpp\_tabusearch\_opt\_info\_iter200 .txt with advanced debug information.

A screenshot of a computer

Description automatically generated

# Output

The basic output of HITS is a text file with name DemoExample\_Cpp\_tabusearch\_opt\_log\_iter200.txt that contains all the information:

this is Discrete tabu search algorithm

for maximum score estimation

==========1==========

Initial score is 0.000000

Initial attributes: 452 819 1122 1318 1368 1891

MaxScore = 0.557931 dzmax = 0.557931 Max Score Ever = 0.557931

MaxScore = 0.667379 dzmax = 0.109448 Max Score Ever = 0.667379

MaxScore = 0.667379 dzmax = 0.000000 Max Score Ever = 0.667379

MaxScore = 0.682343 dzmax = 0.014964 Max Score Ever = 0.682343

MaxScore = 0.687046 dzmax = 0.004703 Max Score Ever = 0.687046

…

MaxScore = 0.692604 dzmax = 0.000000 Max Score Ever = 0.693031

MaxScore = 0.692176 dzmax = -0.000428 Max Score Ever = 0.693031

MaxScore = 0.692604 dzmax = 0.000428 Max Score Ever = 0.693031

MaxScore = 0.692604 dzmax = 0.000000 Max Score Ever = 0.693031

MaxScore = 0.691749 dzmax = -0.000855 Max Score Ever = 0.693031

Argmax is -1.443236 -0.043792 0.598771 11.926891 -1.653683 -9.622923

Max Score Ever is 0.693031

MaxScore at Argmax is 0.693031

coefficient of variable 0 b(0) is -1.000000

attributes are 71 317 1291 1651 892 2016

ntenure= 50 niterover = 200 niternotimproved = 50

Time of computation was 76.170000 seconds

==========2==========

Then moving on to independent run no. 2 in the same file

this is Discrete tabu search algorithm

for maximum score estimation

…

==========2==========

Initial score is 0.000000

Initial attributes: 407 1201 1661 1746 1924 2009

MaxScore = 0.647713 dzmax = 0.647713 Max Score Ever = 0.647713

MaxScore = 0.669517 dzmax = 0.021804 Max Score Ever = 0.669517

MaxScore = 0.669089 dzmax = -0.000428 Max Score Ever = 0.669517

MaxScore = 0.681915 dzmax = 0.012826 Max Score Ever = 0.681915

MaxScore = 0.684481 dzmax = 0.002565 Max Score Ever = 0.684481

…

MaxScore = 0.693886 dzmax = -0.000428 Max Score Ever = 0.694314

MaxScore = 0.693886 dzmax = 0.000000 Max Score Ever = 0.694314

MaxScore = 0.694314 dzmax = 0.000428 Max Score Ever = 0.694314

MaxScore = 0.694314 dzmax = 0.000000 Max Score Ever = 0.694314

MaxScore = 0.694314 dzmax = 0.000000 Max Score Ever = 0.694314

Argmax is -0.764674 -1.187283 0.670576 7.147826 -1.077536 -1.618442

Max Score Ever is **0.694314**

MaxScore at Argmax is 0.694314

coefficient of variable 0 b(0) is -1.000000

attributes are 237 918 1094 1328 1360 2016

ntenure= 50 niterover = 200 niternotimproved = 50

Time of computation was 76.980000 seconds

==========3==========

Up to finally the independent run no. 10 (Nreps was set to 10):

this is Discrete tabu search algorithm

for maximum score estimation

…

==========10==========

Initial score is 0.617785

Initial attributes: 20 638 645 1375 1616 2149

MaxScore = 0.640017 dzmax = 0.022232 Max Score Ever = 0.640017

MaxScore = 0.649423 dzmax = 0.009406 Max Score Ever = 0.649423

MaxScore = 0.657973 dzmax = 0.008551 Max Score Ever = 0.657973

MaxScore = 0.662676 dzmax = 0.004703 Max Score Ever = 0.662676

MaxScore = 0.667379 dzmax = 0.004703 Max Score Ever = 0.667379

…

MaxScore = 0.690894 dzmax = 0.000000 Max Score Ever = 0.690894

MaxScore = 0.690894 dzmax = 0.000000 Max Score Ever = 0.690894

MaxScore = 0.690894 dzmax = 0.000000 Max Score Ever = 0.690894

MaxScore = 0.689611 dzmax = -0.001283 Max Score Ever = 0.690894

MaxScore = 0.689611 dzmax = 0.000000 Max Score Ever = 0.690894

Argmax is -1.158853 0.147361 0.786379 8.194124 -1.107319 -5.667072

Max Score Ever is 0.690894

MaxScore at Argmax is 0.690894

coefficient of variable 0 b(0) is -1.000000

attributes are 9 337 616 1217 1846 2206

ntenure= 50 niterover = 200 niternotimproved = 50

Time of computation was 89.513000 seconds

The software at the time of inspection of results presents the following screen with the progress bar filled in 100%:

A screenshot of a computer

Description automatically generated

Then, the user can complete the analysis hitting CLOSE button and inspect the solution output files already discussed above.

The final result is taken by inspecting the lines

Max Score Ever is **0.694314**

MaxScore at Argmax is 0.694314

For each of the runs and manually selecting the run with the highest Max Score Ever / MaxScore at Argmax.

In our case, this is run No. 2 (and run No. 8 in a tie, not shown in this document) with objective function value 0.694314.

The coefficients are taken from the run No. 2 as

Argmax is -0.764674 -1.187283 0.670576 7.147826 -1.077536 -1.618442

Max Score Ever is **0.694314**

MaxScore at Argmax is 0.694314

coefficient of variable 0 b(0) is -1.000000

The first coefficient is -1.0 (we set it like that in params.txt for this run). Then the remaining coefficients are -0.764674 -1.187283 0.670576 7.147826 -1.077536 -1.618442 as they come in the columns of X.txt. Note that in X.txt the constant term is included in the last column.

The following information is used for debugging:

attributes are 237 918 1094 1328 1360 2016

ntenure= 50 niterover = 200 niternotimproved = 50

Time of computation was 76.980000 seconds

The computation time was 76.98 seconds for run no.2 only and the attributes 237 918 1094 1328 1360 2016 are the observations id’s (numbers from 1 to 2339) that define the linear system p×p that is equivalent to the final solution of HITS.

# Acknowledgements

I would like to thank Dr. **Alexandros Louka** and Professor **Yannis Bilias** for their valuable help and the shared ideas for the development of HITS. The development of the HITS software was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the '2nd Call for H.F.R.I. Research Projects to support Post-Doctoral Researchers' (Project Number: 902).

# References

C. F. Manski, Maximum score estimation of the stochastic utility model of choice, Journal of Econometrics 3 (1975) 205–228.

C. F. Manski, Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator, Journal of econometrics 27 (3) (1985) 313–333.

C. A. Pinkse, On the computation of semiparametric estimates in limited dependent variable models, Journal of Econometrics 58 (1993) 185–205.

Glover, F., 1989. Tabu search—part i. ORSA Journal on Computing 1, 190–206.

**Contents of the zip file**

**Programs**

**DisplayGrid-MFC.exe:** GUI program for HITS method for maximum score estimation.

## Input files

**X.txt:** input file regarding the design matrix X of the regression

**y.txt:** input file regarding the target y of the regression (values -1 or 1)

**names.txt** input file regarding the headers of names of columns of X

**params.txt:** input file regarding the options of the estimation problem

**3kp250.dat:** input file including a MOMDKP with 3 objective functions, 3 constraints and 250 items

**3kp500.dat:** input file including a MOMDKP with 3 objective functions, 3 constraints and 500 items

**4kp100.dat:** input file including a MOMDKP with 4 objective functions, 4 constraints and 100 items

## Output files

**DemoExample\_Cpp\_tabusearch\_opt\_log\_iter200.txt:** main output file of Martins regression (N=2339, p=7)

**DemoExample\_Cpp\_tabusearch\_opt\_info\_iter200.txt:** extra output file of Martins regression (N=2339, p=7)

**DemoExample\_Cpp\_tabusearch\_opt\_res\_iter200.txt:** same as the \_info\_ file above (a simple backup file).

## Reference files

**X\_Martins.txt:** simply a backup of X.txt

**y\_Martins.txt:** simply a backup of y.txt

**X\_Martins\_Gams .txt:** simply a backup of X.txt with first two lines suitable for GAMS input.

**y\_Martins\_Gams .txt:** simply a backup of y.txt with first line suitable for GAMS input.

## Information file

**manual\_HITS.pdf**: the manual of the program