Chapter 1

BIOGEME

BIerlaire Optimization toolbox for GEv Model Estimation (BIOGEME) is a freeware designed for the estimation of binary, Multinomial and Nested Logit models and more complex models in the GEV family as well as mixture of these models (eg. mixed logit). All information (an indeed documentation with tutorial and examples) relative to BIOGEME is maintained at:

<div align="center">roso.epfl.ch/biogeme</div>

1.1 Invoke BIOGEME under Windows

BIOGEME is invoked in a DOS command window or a Cygwin command window under Windows using the following statement structure:

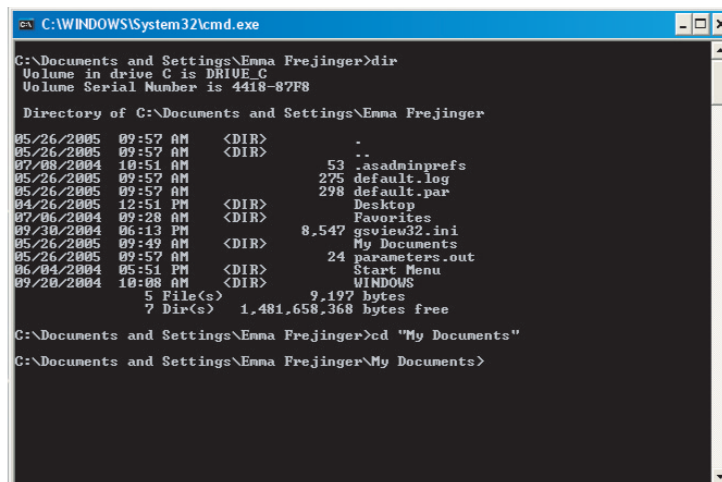<div align="center">biogeme model_file sample_file.dat</div>

Note that the model file is given without the file extension while the sample file does have the extension. When typing this command, the files are assumed to be located in the current directory.

Some useful DOS commands are listed below:

- To select a disk (e.g. C), just type C: in the prompt.

- To connect to a directory (e.g. \biogeme), just type cd \biogeme.

- To see the content of a directory, use Windows Explorer, or type dir.

An example of DOS commands is given in Figure 1.1. The current directory in the example is first C:\Documents and Settings\Emma Frejinger. When typing the command dir the content of this directory is displayed. In order to move to the directory My Documents, the command cd ''My Documents'' is used (note that the quotations marks are optional). Finally the current directory is:
C:\Documents and Settings\Emma Frejinger\My Documents.
In order to return to the previous (top) directory, type cd ...

Figure 1.1: DOS example of commands

1.2   Input Files

BIOGEME reads the following files:

- a file containing the model specifications: `model_file.mod`;

- a file containing the data: `sample_file.dat`;

- a file containing the parameters controlling the behavior of BIOGEME and of its optimization algorithms: `default.par`.

Parameters File

This file is divided into different sections associated with different types of parameters. Each section contains a list of parameters and their corresponding values. The most useful parameters for standard users are defined in the section [GEV], in particular the following ones:

- `gevAlgo` which allows selection of the optimization algorithm to be used for the maximum likelihood estimation;

- `gevTtestThreshold` which sets the threshold for the t-test hypothesis tests on explanatory variables in the model.

This is an example of a parameter file:

```
[GEV]
gevAlgo=''CFSQP''
//gevAlgo=''SOLVOPT''
//gevAlgo=''DONLP2''
//gevAlgo=''BIO''
gevTtestThreshold=1.96
```

The remaining sections are designed for advanced users to allow flexibility to change parameters' default values in the different optimization algorithms.

Note that if you do not specify a parameter file, BIOGEME will create a default one called `default.par` where the "BIO" algorithm is selected.

Model Specification File

You should have a look at the examples given during the labs (available on the course web site) and read the tutorial to understand the details about this file. We list here some important facts for the labs.

- Attribute names are case-sensitive and should be typed exactly as they appear in the list of variable names in the corresponding data file;

- Every string in the file must be ended with a blank space (even if it is followed by a parenthesis);

- Starting values, lower bounds and upper bounds for all model parameters to be estimated should be in float format (including decimal point);

- If there is an Alternative Specific Constant (ASC) defined for each utility function, at least one of these must be fixed (typically set to zero), or absent from the model;

- 0.0 is a reasonable starting value for ASC's and other parameters β in the utility functions;

Data File

All data files needed for the labs are provided on the course web site. Their structure is as follows:

- The first row contains the list of the attributes in the file (the case is important);

- Each subsequent row contains the associated data, one row for each observation;

- No missing value is allowed and all rows must have the exact same number of entries. If a value is missing, a meaningless value must be written (e.g. 99999.9) ;

- Typical information for a given observation is:

  - the observed choice,

  - the description of the choice set through attributes describing the availability of each alternative;

  - the attributes of each alternatives;

  - the socio-economic characteristics of the decision-maker.

1.3 Output Files

BIOGEME automatically generates the following output files:

- A file containing the results of the maximum likelihood estimation: `mymodel.rep`;

- The same file in HTML format: `mymodel.html`

- A file containing the specification of the estimated model, in the same format as the model specification file `mymodel.mod`: `mymodel.res`;

- A file containing some descriptive statistics on the sample such as the number of excluded observations, the total number of observations, details of group membership, . . . : `mymodel.sta`;

as well as the following files to help understanding possible problems:

- A file containing messages produced by BIOGEME during the run: `mymodel.log`;

- A file containing the specification of the model, as it has actually been understood by BIOGEME: `__speFile.debug`;

- A file containing the data stored in BIOGEME to represent the model: `model.debug`;

- A file containing the values of the parameters which have been actually used by BIOGEME: parameters.out;

These file names may be modified according to the following rules:

1. If an input file mymodel.xxx does not exist, BIOGEME attempts to open the file default.xxx. If this file does not exist, BIOGEME exits with an error. Typically, the parameter file is not model dependent. Therefore, it is recommended to call it default.par to avoid copying it for each different model to be estimated.

2. If an output file mymodel.xxx already exists, BIOGEME does not overwrite it. Instead, it creates the file mymodel~1.xxx. If the file mymodel~1.xxx exists, BIOGEME creates the file mymodel~2.xxx, and so on.

Therefore, to avoid any ambiguity, BIOGEME displays filenames it has actually used for a specific run.

If you want more detailed information on the output files generated by BIOGEME, see the tutorial which is available on the BIOGEME web site (roso.epfl.ch/biogeme).