
Computer Lab II

Further Introduction to Biogeme Binary Logit Model Estimation

Evanthia Kazagli

`evanthia.kazagli@epfl.ch`

Today

- **Further introduction to BIOGEME**
- Estimation of Binary Logit models

How does BIOGEME work?

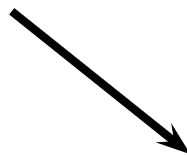
How does BIOGEME work?



BIOGEME

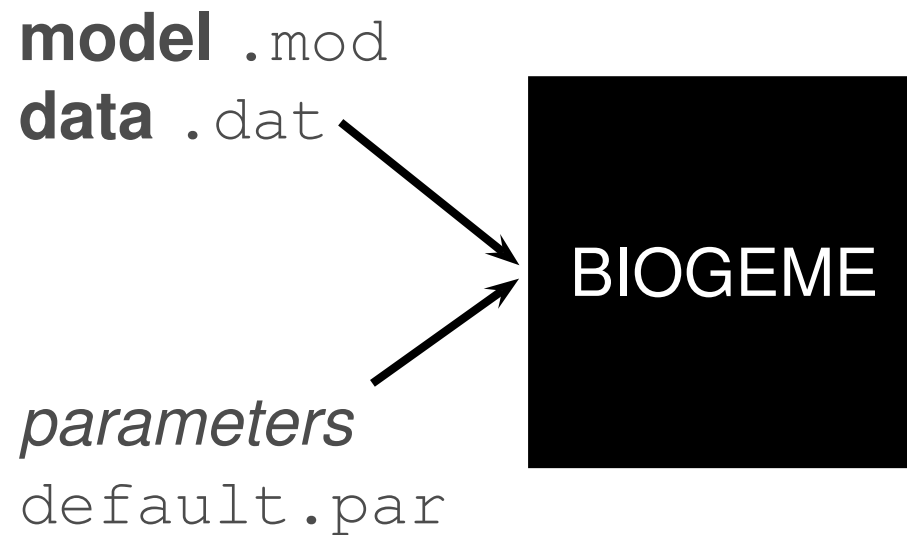
How does BIOGEME work?

model .mod
data .dat

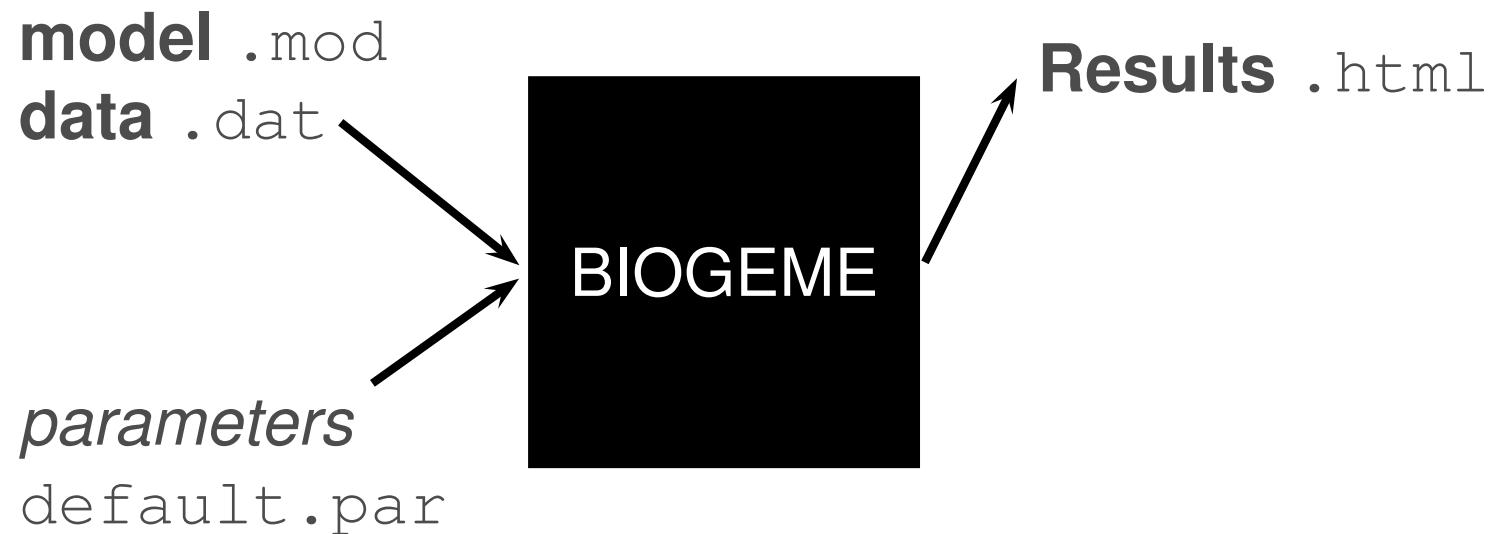


BIOGEME

How does BIOGEME work?



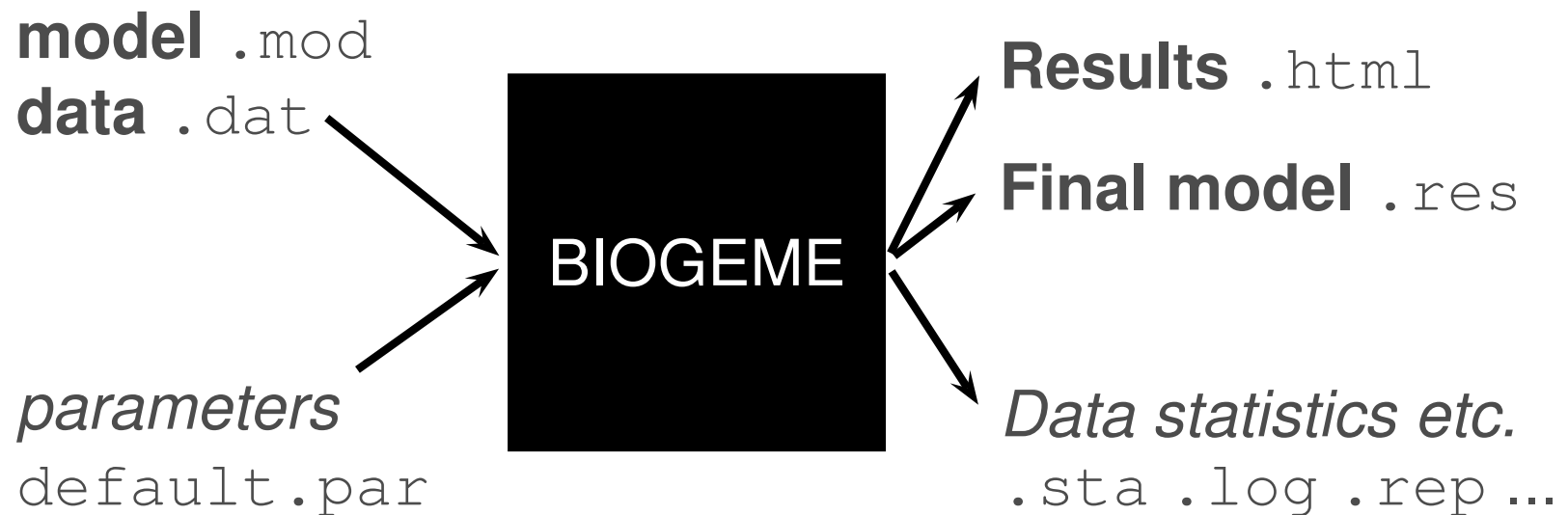
How does BIOGEME work?



How does BIOGEME work?



How does BIOGEME work?



BIOGEME - Data file

- File extension `.dat`
- First row contains column / variable names
- One observation per row
- Each row must contain a choice indicator
- Example with the Netherlands transportation mode choice data:
choice between car and train

BIOGEME - Data file (cont.)

netherlands.dat

id	choice	rail_cost	rail_time	car_cost	car_time
1	0	40	2.5	5	1.167
2	0	35	2.016	9	1.517
3	0	24	2.017	11.5	1.966
4	0	7.8	1.75	8.333	2
5	0	28	2.034	5	1.267
...					
219	1	35	2.416	6.4	1.283
220	1	30	2.334	2.083	1.667
221	1	35.7	1.834	16.667	2.017
222	1	47	1.833	72	1.533
223	1	30	1.967	30	1.267

BIOGEME - Data file (cont.)

netherlands.dat

id	choice	rail_cost	rail_time	car_cost	car_time
1	0	40	2.5	5	1.167
2	0	35	2.016	9	1.517
3	0	24	2.017	11.5	1.966
4	0	7.8	1.75	8.333	2
5	0	28	2.034	5	1.267
...					
219	1	35	2.416	6.4	1.283
220	1	30	2.334	2.083	1.667
221	1	35.7	1.834	16.667	2.017
222	1	47	1.833	72	1.533
223	1	30	1.967	30	1.267

Unique identifier of observations

BIOGEME - Data file (cont.)

netherlands.dat

id	choice	rail_cost	rail_time	car_cost	car_time
1	0	40	2.5	5	1.167
2	0	35	2.016	9	1.517
3	0	24	2.017	11.5	1.966
4	0	7.8	1.75	8.333	2
5	0	28	2.034	5	1.267
...					
219	1	35	2.416	6.4	1.283
220	1	30	2.334	2.083	1.667
221	1	35.7	1.834	16.667	2.017
222	1	47	1.833	72	1.533
223	1	30	1.967	30	1.267

Choice indicator, 0: car and 1: train

BIOGEME - Model file

- File extension `.mod`
- Must be consistent with data file
- Contains deterministic utility specifications, model type etc.
- The model file contains different *sections* describing different elements of the model specification

BIOGEME - Model file (cont.)

- How can we write the following deterministic utility functions for BIOGEME?

$$V_{\text{car}} = \text{ASC}_{\text{car}} + \beta_{\text{time}} \text{time}_{\text{car}} + \beta_{\text{cost}} \text{cost}_{\text{car}}$$

$$V_{\text{rail}} = \beta_{\text{time}} \text{time}_{\text{rail}} + \beta_{\text{cost}} \text{cost}_{\text{rail}}$$

BIOGEME - Model file (cont.)

[Choice]

choice

[Beta]

// Name	DefaultValue	LowerBound	UpperBound	status
ASC_CAR	0.0	-100.0	100.0	0
ASC_RAIL	0.0	-100.0	100.0	1
BETA_COST	0.0	-100.0	100.0	0
BETA_TIME	0.0	-100.0	100.0	0

[Utilities]

//Id	Name	Avail	linear-in-parameter expression
0	Car	one	ASC_CAR * one + BETA_COST * car_cost + BETA_TIME * car_time
1	Rail	one	ASC_RAIL * one + BETA_COST * rail_cost + BETA_TIME * rail_time

BIOGEME - Model file (cont.)

[Choice]

choice

[Beta]

// Name	DefaultValue	LowerBound	UpperBound	status
ASC_CAR	0.0	-100.0	100.0	0
ASC_RAIL	0.0	-100.0	100.0	1
BETA_COST	0.0	-100.0	100.0	0
BETA_TIME	0.0	-100.0	100.0	0

[Utilities]

//Id	Name	Avail	linear-in-parameter expression
0	Car	one	ASC_CAR * one + BETA_COST * car_cost + BETA_TIME * car_time
1	Rail	one	ASC_RAIL * one + BETA_COST * rail_cost + BETA_TIME * rail_time

BIOGEME - Model file (cont.)

[Choice]

choice

[Beta]

// Name	DefaultValue	LowerBound	UpperBound	status
ASC_CAR	0.0	-100.0	100.0	0
ASC_RAIL	0.0	-100.0	100.0	1
BETA_COST	0.0	-100.0	100.0	0
BETA_TIME	0.0	-100.0	100.0	0

[Utilities]

//Id	Name	Avail	linear-in-parameter expression
0	Car	one	ASC_CAR * one + BETA_COST * car_cost + BETA_TIME * car_time
1	Rail	one	ASC_RAIL * one + BETA_COST * rail_cost + BETA_TIME * rail_time

BIOGEME - Model file (cont.)

```
[Choice]
```

```
choice
```

What is one?

Which is the type of model?

```
[Beta]
```

// Name	DefaultValue	LowerBound	UpperBound	status
ASC_CAR	0.0	-100.0	100.0	0
ASC_RAIL	0.0	-100.0	100.0	1
BETA_COST	0.0	-100.0	100.0	0
BETA_TIME	0.0	-100.0	100.0	0

```
[Utilities]
```

//Id	Name	Avail	linear-in-parameter expression
0	Car	one	ASC_CAR * one + BETA_COST * car_cost + BETA_TIME * car_time
1	Rail	one	ASC_RAIL * one + BETA_COST * rail_cost + BETA_TIME * rail_time

BIOGEME - Model file (cont.)

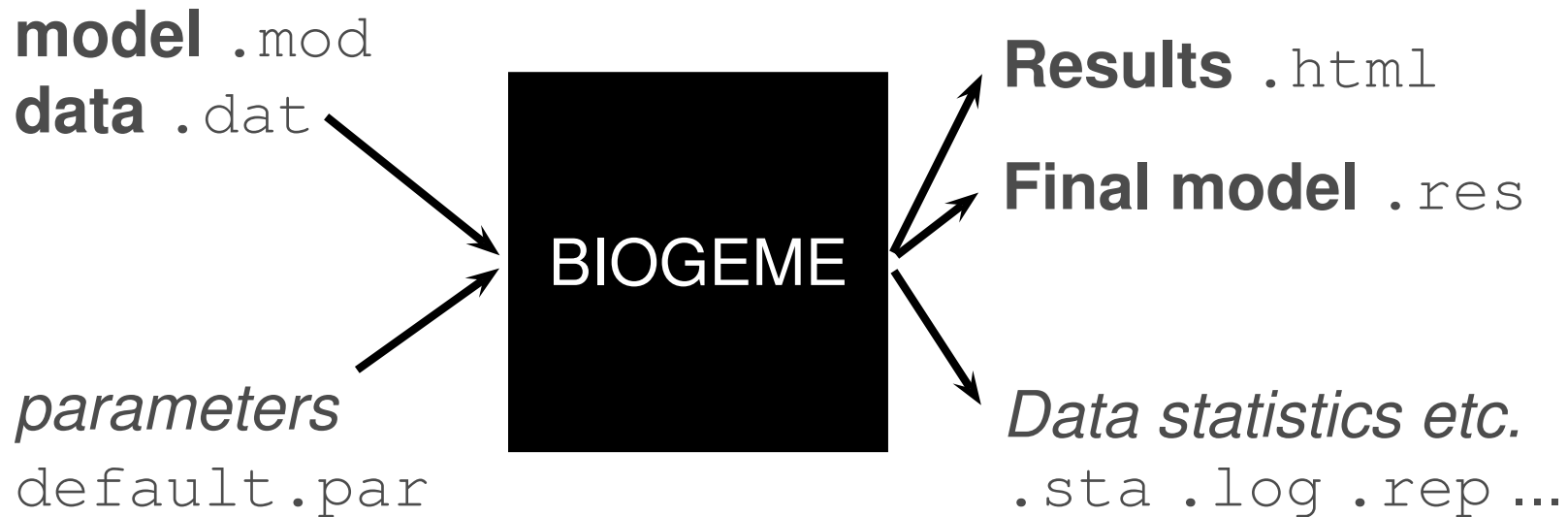
```
[Expressions]
```

```
// Define here arithmetic expressions for name that are not directly  
// available from the data  
one = 1
```

```
[Model]
```

```
// Currently, only $MNL (multinomial logit), $NL (nested logit), $CNL  
// (cross-nested logit) and $NGEV (Network GEV model) are valid keywords  
//  
$MNL
```

How does BIOGEME work?



Model and Data Files

- How to read and modify model files?
How to read data files?
 - GNU Emacs, vi, TextEdit (Mac) or Wordpad (Windows)
 - **Notepad (Windows) should not be used!**

BIOGEME Results: Netherlands dataset

binary_generic_netherlands.html - Report from BIOGEME Version 1.8 [Sat Mar 7 14:36:56 CEST 2009] - Moz...

Echier Édition Affichage Historique Marque-pages Outils 2

file:///C:/BinaryNetherlands/binary_generic_netherlands.html

Google

Les plus visités Débuter avec Firefox À la une

binary_generic_netherlands...

BIOGEME Version 1.8 [Sat Mar 7 14:36:56 CEST 2009]

Michel Bierlaire, EPFL

This file has automatically been generated.

09/23/10 12:08:00

Model: Multinomial Logit

Number of estimated parameters: 3

Number of observations: 228

Number of individuals: 228

Null log-likelihood: -158.038

Cte log-likelihood: -148.347

Init log-likelihood: -158.038

Final log-likelihood: -123.133

Likelihood ratio test: 69.809

Rho-square: 0.221

Adjusted rho-square: 0.202

Final gradient norm: +6.045e-004

Diagnostic: Convergence reached...

Iterations: 7

Run time: 00:00

Variance-covariance: from analytical hessian

Sample file: netherlands05.dat

Utility parameters

Name	Value	Std err	t-test	p-value	Robust Std err	Robust t-test	p-value
ASC_CAR	-0.798	0.270	-2.95	0.00	0.275	-2.90	0.00
ASC_RAIL	0.00	fixed					
BETA_COST	-0.0499	0.0103	-4.85	0.00	0.0107	-4.67	0.00
BETA_TIME	-1.33	0.344	-3.86	0.00	0.354	-3.75	0.00

Terminé

BIOGEME Results

```
binary_generic_netherlands.html - Report from BIOGEME Version 1.8 [Sat Mar 7 14:36:56 CEST 2009] - Moz...
Echier  Édition  Affichage  Historique  Marque-pages  Outils  2
file:///C:/BinaryNetherlands/binary_generic_netherlands.html
Les plus visités  Débuter avec Firefox  À la une
binary_generic_netherlands...
BIOGEME Version 1.8 [Sat Mar 7 14:36:56 CEST 2009]
Michel Bierlaire, EPFL
This file has automatically been generated.
09/23/10 12:08:00

Model: Multinomial Logit
Number of estimated parameters: 3
Number of observations: 228
Number of individuals: 228
Null log-likelihood: -158.038
Cte log-likelihood: -148.347
Init log-likelihood: -158.038
Final log-likelihood: -123.133
Likelihood ratio test: 69.809
Rho-square: 0.221
Adjusted rho-square: 0.202
Final gradient norm: +6.045e-004
Diagnostic: Convergence reached...
Iterations: 7
Run time: 00:00
Variance-covariance: from analytical hessian
Sample file: netherlands05.dat

Utility parameters


| Name      | Value   | Std err | t-test | p-value | Robust Std err | Robust t-test | p-value |
|-----------|---------|---------|--------|---------|----------------|---------------|---------|
| ASC_CAR   | -0.798  | 0.270   | -2.95  | 0.00    | 0.275          | -2.90         | 0.00    |
| ASC_RAIL  | 0.00    | fixed   |        |         |                |               |         |
| BETA_COST | -0.0499 | 0.0103  | -4.85  | 0.00    | 0.0107         | -4.67         | 0.00    |
| BETA_TIME | -1.33   | 0.344   | -3.86  | 0.00    | 0.354          | -3.75         | 0.00    |


Terminé
```

General model information

BIOGEME Results (cont.)

binary_generic_netherlands.html - Report from BIOGEME Version 1.8 [Sat Mar 7 14:36:56 CEST 2009] - Moz...

Echier Édition Affichage Historique Marque-pages Outils 2

file:///C:/BinaryNetherlands/binary_generic_netherlands.html

Google

Les plus visités Débuter avec Firefox À la une

binary_generic_netherlands...

BIOGEME Version 1.8 [Sat Mar 7 14:36:56 CEST 2009]

Michel Bierlaire, EPFL

This file has automatically been generated.

09/23/10 12:08:00

Model: Multinomial Logit

Number of estimated parameters: 3

Number of observations: 228

Number of individuals: 228

Null log-likelihood: -158.038

Cte log-likelihood: -148.347

Init log-likelihood: -158.038

Final log-likelihood: -123.133

Likelihood ratio test: 69.809

Rho-square: 0.221

Adjusted rho-square: 0.202

Final gradient norm: +6.045e-004

Diagnostic: Convergence reached...

Iterations: 7

Run time: 00:00

Variance-covariance: from analytical hessian

Sample file: netherlands05.dat

Utility parameters

Name	Value	Std err	t-test	p-value	Robust Std err	Robust t-test	p-value
ASC_CAR	-0.798	0.270	-2.95	0.00	0.275	-2.90	0.00
ASC_RAIL	0.00	fixed					
BETA_COST	-0.0499	0.0103	-4.85	0.00	0.0107	-4.67	0.00
BETA_TIME	-1.33	0.344	-3.86	0.00	0.354	-3.75	0.00

Terminé

Coefficient estimates

Today

- Further introduction to BIOGEME
- **Estimation of Binary Logit models**

Binary Logit Case Study

- Available datasets:
 - Mode choice in Netherlands
- Descriptions available on the course web site

How to go through the Case Studies

- Download the files related to the Netherlands dataset and case study from the course website;
- Study the `.mod` files with the help of the descriptions;
- Run the `.mod` files with BIOGEME;
- Interpret the results and compare your interpretation with the one we have provided;
- Develop other model specifications.

Course website (under laboratories)

- <http://transp-or.epfl.ch/courses/decisionAid2013/labs.php>
- BIOGEME software
(including documentation and utilities)
- For each Case Study:
 - Data files for available datasets;
 - Model specification files;
 - Possible interpretation of results.

Today's plan

Group work

- Listen to the description of dataset;
- Gather in groups;
- Generate .mod file (base);
- Test an idea/ hypothesis.

Lab assignment

- Work with your group on your own specification of a Binary Logit on the Netherlands mode choice data;
- Examine the data & and the variables' description;
- Write a .mod file;
- Formulate your own hypothesis;
- Test your hypothesis;
- Send your results by e-mail on Friday, March 8, by noon:
 - the .html file
 - .mod file
 - 5-one-sided-pages report.

Specifying models: Recommended steps

- Formulate a-priori hypothesis:
 - Expectations and intuition regarding the explanatory variables that appear to be significant for mode choice.
- Specify a minimal model:
 - Start simple;
 - Include the main factors affecting the mode choice of (rational) travelers;
 - This will be your starting point.
- Continue adding and testing variables that improve the initial model in terms of *causality*, and *efficiency* to predict what actually happened in the sample.

Evaluating models

The main indicators used to evaluate and compare the various models are summarised here:

- Informal tests:
 - *signs* and *relative magnitudes* of the parameters β values (under our a-priori expectations);
 - *trade-offs* among some attributes and ratios of pairs of parameters (e.g. reasonable value of time).
- Overall goodness of fit measure:
 - *adjusted rho-square* (likelihood ratio index): takes into account the different number of explanatory variables used in the models and normalizes for their effect → suitable to compare models with different number of independent variables. We check this value to have a first idea about which model might be better (among models of the same type), but it is not a statistical test.

Evaluating models (cont.)

- Statistical tests:
 - *t-test values*: statistically significant explanatory variables are denoted by t-statistic values remarkably higher/ lower than ± 2 (for a 95% level of confidence);
 - *final log-likelihood* for the full set of parameters: should be remarkably different from the ones in the naive approach (null log-likelihood and log-likelihood at constants); we ask for high values of likelihood ratio test $[-2(LL(0) - LL(\beta))]$ in order to have a model significantly different than the naive one.
- Test of entire models:
 - *likelihood ratio test* $[-2(LL(\hat{\beta}_R) - LL(\hat{\beta}_U))]$: used to test the null hypothesis that two models are equivalent, under the requirement that the one is the restricted version of the other. The likelihood ratio test is X^2 distributed, with degrees of freedom equal to $K_U - K_R$ (where K the number of parameters of the unrestricted and restricted model, respectively).