Computer Lab II

Further Introduction to Biogeme
Binary Logit Model Estimation

Anna Fernández Antolín

anna.fernandezantolin@epfl.ch

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Today

- **Further introduction to BIOGEME**

- Estimation of Binary Logit models

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

How does BIOGEME work?

How does BIOGEME work?

How does BIOGEME work?

model `.mod`
data `.dat`

BIOGEME

How does BIOGEME work?
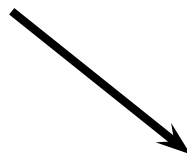
model `.mod`
data `.dat`

BIOGEME

parameters
`default.par`

How does BIOGEME work?

model `.mod`
data `.dat`

BIOGEME

Results `.html`

parameters
`default.par`

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

How does BIOGEME work?

model `.mod`
data `.dat`

parameters
`default.par`

BIOGEME

Results `.html`

Final model `.res`

How does BIOGEME work?

model `.mod`
data `.dat`

parameters
`default.par`

BIOGEME

Results `.html`

Final model `.res`

Data statistics etc.
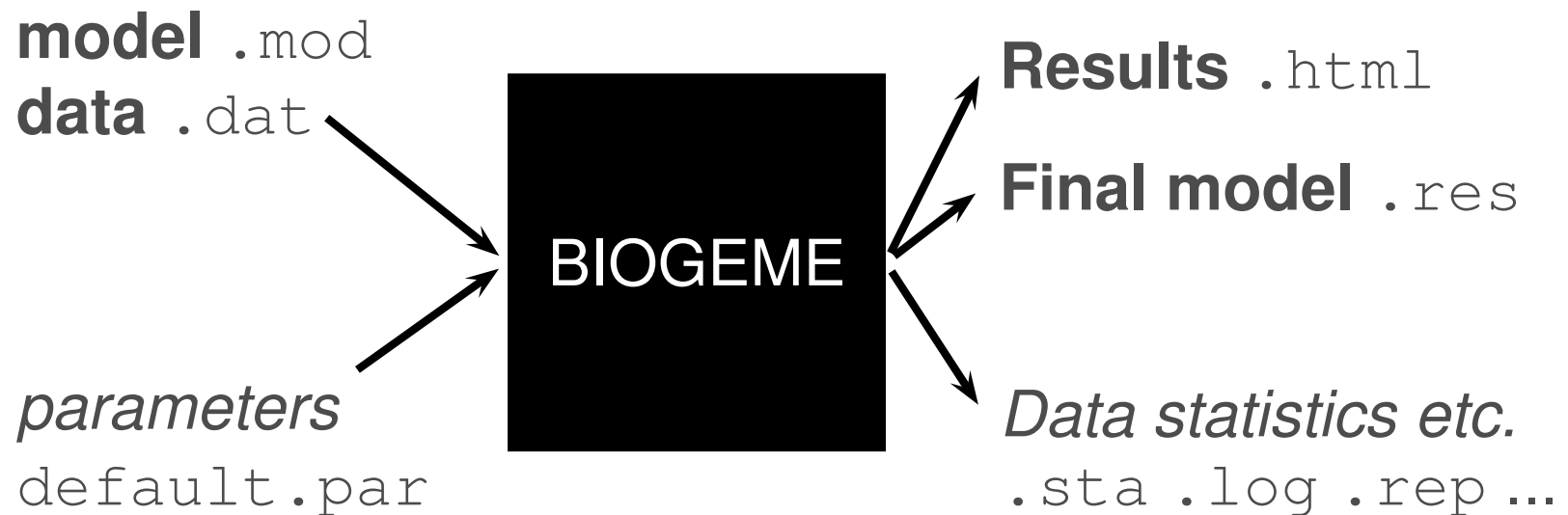`.sta .log .rep ...`
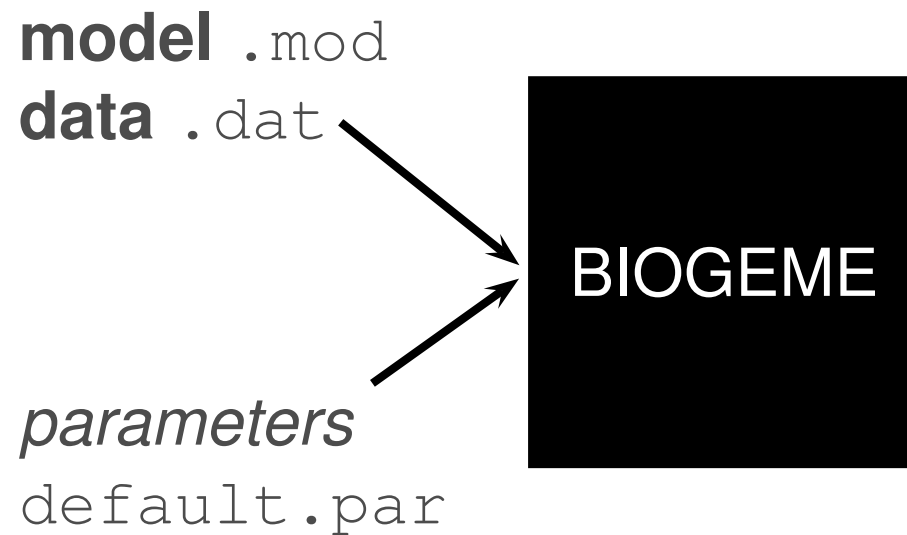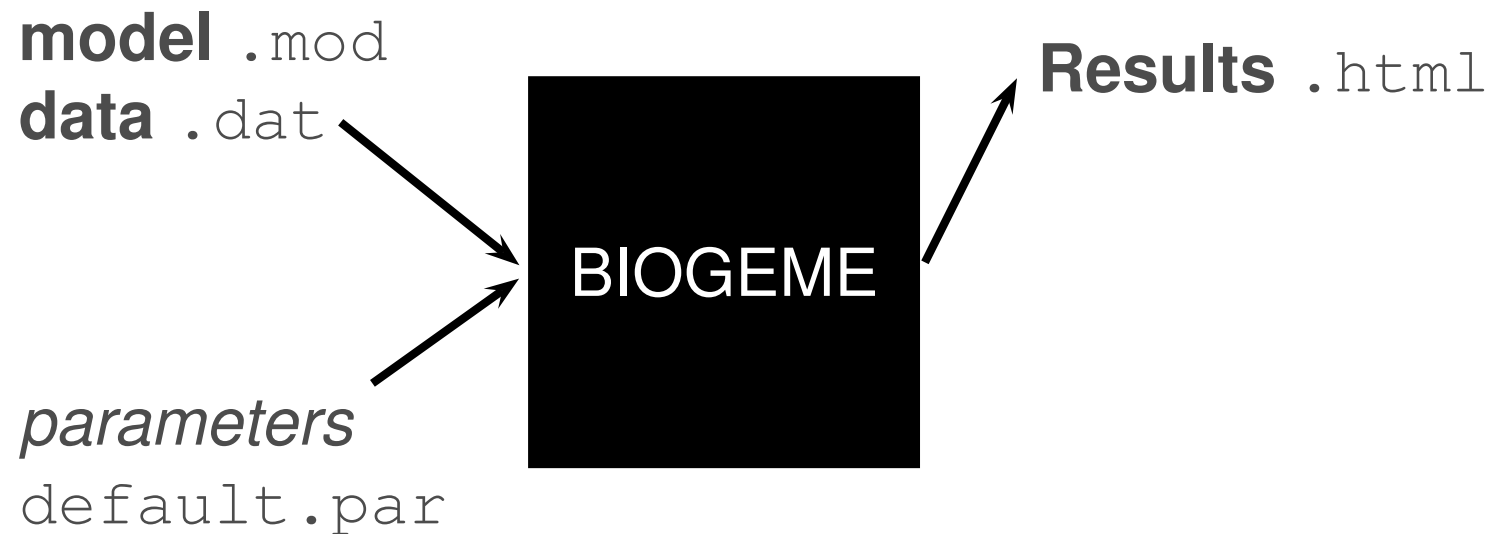
BIOGEME - Data file

- 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

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

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
```

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

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
```

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

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
```

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

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
```

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

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 `.mod`
data `.dat`

parameters
`default.par`

BIOGEME

Results `.html`

Final model `.res`

Data statistics etc.
`.sta .log .rep` ...

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
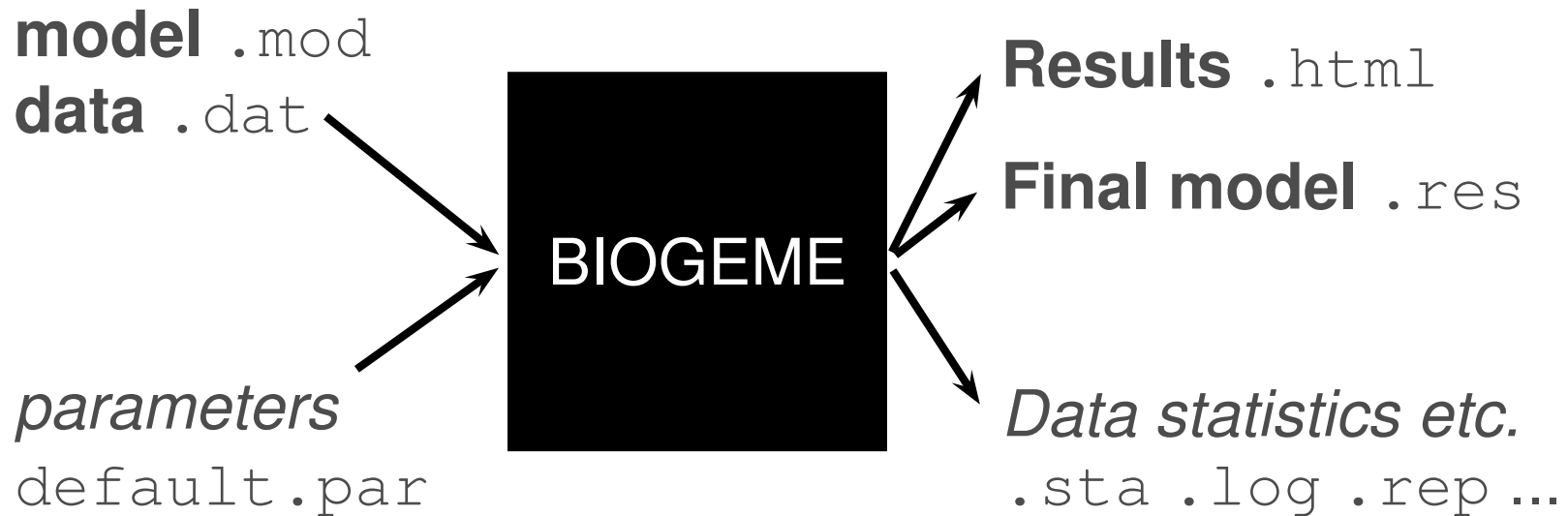
Model and Data Files

- 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!

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

BIOGEME Results: Netherlands dataset

BIOGEME Results

BIOGEME Results (cont.)

Coefficient estimates

Today

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Binary Logit Case Study

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

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.

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Course website (under laboratories)

- `http://transp-or.epfl.ch/courses/decisionAid2014/labs.php`

- BIOGEME software
  (including documentation and utilities)

- For each Case Study:
  - Data files for available datasets;
  - Model specification files;
  - Possible interpretation of results.

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

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 in a 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;

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

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.

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

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 $\rightarrow$ 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.

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

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).

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE