# Results from the simulations using Stepmixr

Félix Laliberté
Charles-Édouard Giguère
Éric Lacourse
Roxane de la Sablonnière
Sacha Morin
Robin Legault
Zsusza Bakk

2024-05-05

The following article shows results from the simulations presented in the article StepMix: a Python Package for Pseudo-Likelihood Estimation of Generalized Mixture Models with External Variables. In the original article, the results are presented in table 3, 5 and 7 for respectively the Distal outcome, Covariate, and Complete simulations.

# 1 Distal outcome

| Mn Bias sep level | n | method | 1 (NA) | 2 (NA) | 3 (NA) | 3 (BCH) | 3 (ML) |
|---|---|---|---|---|---|---|---|
| 0.7 | 500 | | 0.01 | -0.19 | -0.64 | -0.38 | -0.14 |
| | 1000 | | 0.01 | -0.07 | -0.61 | -0.23 | -0.05 |
| | 2000 | | 0.00 | -0.02 | -0.57 | -0.11 | -0.01 |
| 0.8 | 500 | | 0.00 | -0.03 | -0.31 | -0.04 | -0.02 |
| | 1000 | | 0.00 | -0.02 | -0.29 | -0.01 | -0.01 |
| | 2000 | | 0.00 | -0.01 | -0.28 | 0.00 | 0.00 |
| 0.9 | 500 | | 0.00 | -0.01 | -0.08 | 0.00 | 0.00 |
| | 1000 | | 0.00 | 0.00 | -0.08 | 0.00 | 0.00 |
| | 2000 | | 0.00 | 0.00 | -0.08 | 0.00 | 0.00 |
| RMSE sep level | n | method | 1 (NA) | 2 (NA) | 3 (NA) | 3 (BCH) | 3 (ML) |
| 0.7 | 500 | | 0.16 | 0.33 | 0.66 | 0.48 | 0.31 |
| | 1000 | | 0.11 | 0.22 | 0.63 | 0.36 | 0.22 |
| | 2000 | | 0.08 | 0.17 | 0.58 | 0.24 | 0.17 |
| 0.8 | 500 | | 0.11 | 0.15 | 0.33 | 0.18 | 0.15 |
| | 1000 | | 0.08 | 0.09 | 0.30 | 0.12 | 0.09 |
| | 2000 | | 0.06 | 0.07 | 0.29 | 0.09 | 0.07 |
| 0.9 | 500 | | 0.09 | 0.09 | 0.12 | 0.09 | 0.09 |
| | 1000 | | 0.07 | 0.06 | 0.10 | 0.06 | 0.06 |
| | 2000 | | 0.05 | 0.04 | 0.09 | 0.04 | 0.04 |

Table 1: Simulation results for StepMix point estimates of one parameter in an outcome SM (with true value $\mu_2 = 1$). Three degrees of class separation in the MM (lower is harder) and three sample sizes are tested. Each configuration is simulated 500 times and the average parameter bias and RMSE are reported for one-, two- and three-step estimation, as well as the BCH and ML bias correction methods in the three-step case. NA refers to no correction.

## 2 Covariate

| Mn bias | | method | 1 (NA) | 2 (NA) | 3 (NA) | 3 (BCH) | 3 (ML) |
|---|---|---|---|---|---|---|---|
| sep level | n | | | | | | |
| 0.7 | 500 | | 0.03 | -0.27 | -0.63 | -0.29 | -0.31 |
| | 1000 | | 0.01 | -0.15 | -0.61 | -0.19 | -0.18 |
| | 2000 | | 0.00 | -0.09 | -0.62 | -0.11 | -0.10 |
| 0.8 | 500 | | 0.02 | -0.03 | -0.39 | -0.03 | -0.05 |
| | 1000 | | 0.01 | -0.02 | -0.37 | 0.01 | -0.02 |
| | 2000 | | 0.00 | -0.01 | -0.37 | 0.01 | -0.01 |
| 0.9 | 500 | | 0.02 | 0.01 | -0.12 | 0.01 | 0.01 |
| | 1000 | | 0.01 | 0.00 | -0.12 | 0.01 | 0.01 |
| | 2000 | | 0.00 | 0.00 | -0.13 | 0.00 | 0.00 |
| RMSE | | method | 1 (NA) | 2 (NA) | 3 (NA) | 3 (BCH) | 3 (ML) |
| sep level | n | | | | | | |
| 0.7 | 500 | | 0.27 | 0.41 | 0.65 | 0.57 | 0.45 |
| | 1000 | | 0.17 | 0.29 | 0.62 | 0.42 | 0.33 |
| | 2000 | | 0.12 | 0.20 | 0.62 | 0.33 | 0.23 |
| 0.8 | 500 | | 0.16 | 0.18 | 0.40 | 0.29 | 0.21 |
| | 1000 | | 0.12 | 0.13 | 0.38 | 0.21 | 0.14 |
| | 2000 | | 0.08 | 0.09 | 0.38 | 0.14 | 0.10 |
| 0.9 | 500 | | 0.13 | 0.13 | 0.16 | 0.15 | 0.14 |
| | 1000 | | 0.09 | 0.09 | 0.15 | 0.11 | 0.10 |
| | 2000 | | 0.07 | 0.07 | 0.14 | 0.07 | 0.07 |

Table 2: Simulation results for StepMix point estimates of one parameter in a covariate SM (with true value $\beta_2 = 1$) over 500 simulated data sets. NA stands for no correction.

# 3 Complete

| Mn bias<br>nan ratio | n | method | 1 (NA) | 2 (NA) | 3 (NA) | 3 (BCH) | 3 (ML) |
|---|---|---|---|---|---|---|---|
| 0 | 500 | | 0.01 | 0.00 | -0.30 | -0.03 | 0.00 |
| | 1000 | | 0.00 | 0.00 | -0.29 | 0.00 | 0.00 |
| | 2000 | | 0.00 | 0.00 | -0.28 | 0.01 | 0.00 |
| 0.25 | 500 | | 0.00 | -0.04 | -0.45 | -0.13 | 0.00 |
| | 1000 | | 0.00 | -0.01 | -0.43 | -0.08 | 0.00 |
| | 2000 | | 0.00 | 0.00 | -0.41 | -0.03 | 0.00 |
| 0.5 | 500 | | 0.00 | -0.09 | -0.57 | -0.22 | 0.00 |
| | 1000 | | 0.00 | -0.04 | -0.57 | -0.16 | 0.00 |
| | 2000 | | 0.00 | 0.00 | -0.54 | -0.08 | 0.01 |
| RMSE<br>nan ratio | n | method | 1 (NA) | 2 (NA) | 3 (NA) | 3 (BCH) | 3 (ML) |
| 0 | 500 | | 0.09 | 0.11 | 0.33 | 0.18 | 0.11 |
| | 1000 | | 0.06 | 0.07 | 0.30 | 0.13 | 0.08 |
| | 2000 | | 0.05 | 0.05 | 0.29 | 0.09 | 0.05 |
| 0.25 | 500 | | 0.11 | 0.17 | 0.47 | 0.28 | 0.16 |
| | 1000 | | 0.08 | 0.10 | 0.45 | 0.21 | 0.10 |
| | 2000 | | 0.05 | 0.07 | 0.42 | 0.14 | 0.07 |
| 0.5 | 500 | | 0.14 | 0.24 | 0.60 | 0.40 | 0.20 |
| | 1000 | | 0.10 | 0.17 | 0.59 | 0.31 | 0.15 |
| | 2000 | | 0.07 | 0.12 | 0.56 | 0.23 | 0.11 |

Table 3: Simulation results for StepMix point estimates of a SM's parameter in a model with six indicators, a covariate, and an outcome variable. We study the estimates of a single parameter in the outcome model (with true value $\mu_2 = 1$) over 500 simulated data sets. For this simulation, we fix the degree of class separation in the MM ($\gamma = 0.8$) and instead vary the ratio of data missing completely at random in the indicators and the outcome.