BHAMSLE: A Breakpoint Heuristic Algorithm for Maximum Simulated Likelihood Estimation of Advanced Discrete Choice Models

Tom Haering   Michel Bierlaire

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne

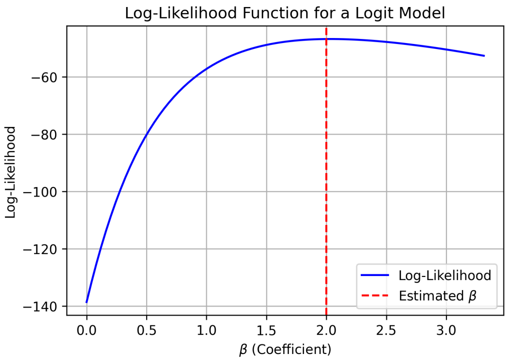
3rd EPFL Symposium on Transportation Research, Barcelona
12-14 February 2025

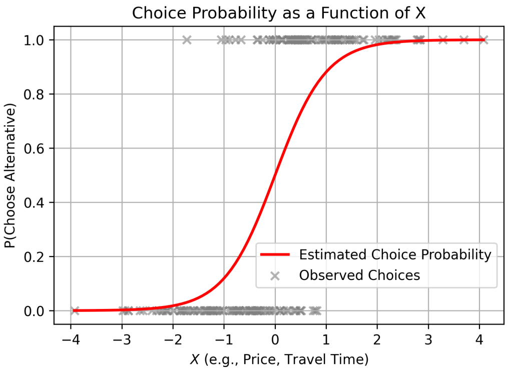
Outline

- Introduction
- Problem formulation & Methodology
- Numerical Results
- Conclusion

Maximum Likelihood Estimation (MLE)

- MLE is widely used for estimating parameters of a distribution based on observed data.
- Applications in physics (Hauschild and Jentschel, 2001), machine learning (Goodfellow et al., 2016), and discrete choice modeling (Bierlaire, 2023).
- Estimation involves maximizing a log-likelihood function.





MLE for Discrete Choice Modeling

- DCM are used for prediction, counterfactual analysis, as well as policy and systems design.
- Parameters are estimated from observations to capture the effect of explanatory variables.

$$
\begin{aligned}
U_{\text{car},n} &= \beta_{\text{ASC\_car}} + \beta_{\text{time}} \cdot \text{time}_{\text{car},n} \\
&\quad + \beta_{\text{cost}} \cdot \text{cost}_{\text{car},n} + \varepsilon_{\text{car},n}, \\
U_{\text{train},n} &= \beta_{\text{time}} \cdot \text{time}_{\text{train},n} \\
&\quad + \beta_{\text{cost}} \cdot \text{cost}_{\text{train},n} + \varepsilon_{\text{train},n}.
\end{aligned}
$$

Parameter	Estimate
$\beta_{\text{ASC\_car}}$	0.30
β_{cost}	-0.75
β_{time}	-1.20
$LL(\beta)$	-1450.39

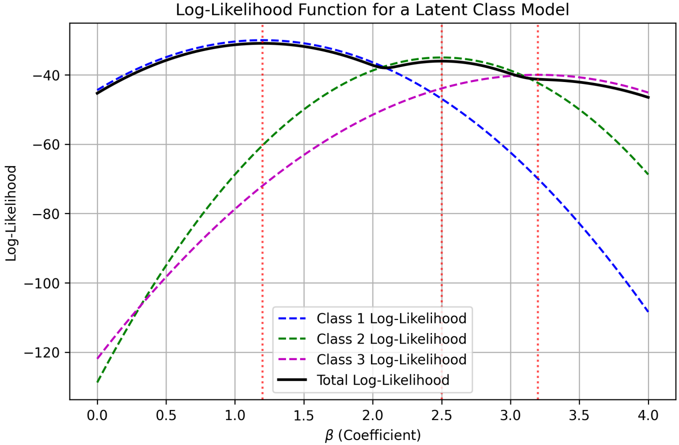
- For logit and nested logit models, concavity guarantees a global optimum.

TRANSP-OR

EPFL

Latent Class Models

- Capture unobserved heterogeneity by segmenting the population into latent groups.
- Prone to many local optima.



Latent Class Models

Why do we care?

- Stuck in suboptimal solutions.
- Class misidentification.
- Affects interpretability and validity of the estimated model.

What can we do?

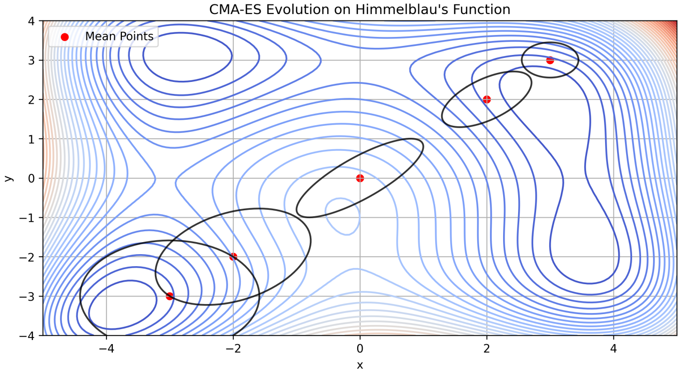
- Perform multiple estimations with diverse initializations (Lubke and Muthén, 2005; Jung and Wickrama, 2008; Peer et al., 2016).
- Invest more effort into finding good starting points (global optimization).

TRANSP-OR

EPFL

Covariance Matrix Adaptation Evolution Strategy (CMA-ES)

- A stochastic, population-based algorithm for black-box optimization.
- Adapts covariance matrix to explore irregular functions (Hansen et al., 2003).
- Effective in nonconvex, noisy, and simulation-based estimation problems.



Outline

- Introduction
- Problem formulation & Methodology
- Numerical Results
- Conclusion

Maximum Simulated Likelihood Estimation (MSLE)

- A set of individuals $\mathcal{N} = \{1, \ldots, N\}$ each selects exactly one option from available choices $\mathcal{I} = \{1, \ldots, I\}$.
- Individual n has access to a subset $C_n \subset \mathcal{I}$.
- The observed choice for individual n is $y_n \in \mathcal{I}$.

For latent class models:

- Population is segmented into C latent classes $\mathcal{C} = \{1, \ldots, C\}$ with class probabilities $\pi_c, c \in \mathcal{C}$, and $\sum_c \pi_c = 1$.
- Each class may have a separate choice model, with different
  - available choices $\rightarrow C_n^c$,
  - attributes and characteristics $\rightarrow \mathcal{K}^c$,
  - parameters to estimate $\rightarrow \beta_k, k \in \mathcal{K}^c$.

Maximum Simulated Likelihood Estimation (MSLE)

Optimization Problem (stochastic):

$$\max_{\beta,\pi} \quad \sum_{n\in\mathcal{N}} \ln\left(\sum_{c\in\mathcal{C}} \pi_c P^c_{y_n n}\right)$$

s.t.

$$\sum_{c\in\mathcal{C}} \pi_c = 1, \qquad \text{(Class probability constraint)} \qquad (1)$$

$$U^c_{in} = \sum_{k\in\mathcal{K}^c} x_{ink}\beta_k + \varepsilon_{in}, \qquad \forall n \in \mathcal{N}, c \in \mathcal{C}, i \in C^c_n, \qquad (2)$$

$$P^c_{in} = \mathbb{P}(U^c_{in} \geq U^c_{jn}, \quad \forall j \in C^c_n), \qquad \forall n \in \mathcal{N}, c \in \mathcal{C}, i \in C^c_n. \qquad (3)$$

$$\beta_k \in \mathbb{R}, \qquad \forall k \in \bigcup_{c\in\mathcal{C}} \mathcal{K}^c,$$

$$\pi_c \geq 0, \qquad \forall c \in \mathcal{C}.$$

Maximum Simulated Likelihood Estimation (MSLE)

Need for Simulation:

- In complex models, P_{in} lacks a closed-form expression.
- Approximated using random draws to simulate error components.

Example: Mixed Logit Model

- A parameter β_m follows a normal distribution:

$$\beta_m \sim \mathcal{N}(\beta_m^{\text{mean}}, \beta_m^{\text{std}})$$

- For each simulation scenario $r \in \mathcal{R}$, the deterministic utility is:

$$U_{inr} = \sum_{k \neq m} x_{ink}\beta_k + x_{inm}(\beta_m^{\text{mean}} + \sigma_{nr}\beta_m^{\text{std}}) + \varepsilon_{inr}$$

where $\sigma_{nr} \sim \mathcal{N}(0, 1)$ and $\varepsilon_{inr} \sim \text{Gumbel}(0, 1)$.

Maximum Simulated Likelihood Estimation (MSLE)

Simulating latent class models:

- Let $\gamma_g,\ g \in \mathcal{G} = \{1, \ldots, C-1\}$ represent the parameters that separate the unit interval into C partitions $P_1, \ldots, P_C$. Set $\gamma_0 = 0$ and $\gamma_C = 1$.
- Take draws $u_{nr} \sim \mathcal{U}(0, 1)$.
- If $u_{nr} \in [\gamma_{c-1}, \gamma_c]$ then $U_{inr} = U^c_{inr}\ \forall c \in \mathcal{C}$.

$$\Rightarrow U_{inr} = \sum_{c \in \mathcal{C}} \mathbb{1}_{[u_{nr} \in [\gamma_{c-1}, \gamma_c]]} U^c_{inr}$$

Maximum Simulated Likelihood Estimation (MSLE)

Simulated Choice and Probability Estimation:

- Define binary variables $\forall n \in \mathcal{N}, i \in C_n, r \in \mathcal{R}$:

$$\omega_{inr} = \begin{cases} 1, & \text{if } \exists c \in \mathcal{C} \text{ such that } u_{nr} \in [\gamma_{c-1}, \gamma_c] \\ & \text{and } U^c_{inr} \geq U^c_{jnr} \text{ for all } j \in C^c_n, \\ 0, & \text{otherwise.} \end{cases} = \sum_{c \in \mathcal{C}} \mathbb{1}_{[u_{nr} \in [\gamma_{c-1}, \gamma_c]]} \mathbb{1}_{[U^c_{inr} = \max_j U^c_{jnr}]}$$

- All indicator functions introduced can be linearized.

Maximum Simulated Likelihood Estimation (MSLE)

$\forall n \in \mathcal{N}, r \in \mathcal{R}$: add auxiliary binary variables $\omega^c_{inr} \forall i \in C_n, \delta^c_{nr}, \mu^c_{nr} \in \{0,1\}, \quad \forall c \in \mathcal{C}$.

$$
\begin{aligned}
\text{Interval Membership Constraints:}\quad & \gamma_{c-1} \leq u_{nr} + M_1(1-\delta^c_{nr}), \quad \forall c \in \mathcal{C} \\
& u_{nr} \leq \gamma_c + M_1(1-\delta^c_{nr}), \quad \forall c \in \mathcal{C} \\
\text{Maximum Utility Constraints:}\quad & U^c_{inr} \geq U^c_{jnr} - M_2^{cnr}(1-\mu^c_{inr}), \ \forall c \in \mathcal{C}, i,j \in C^c_n \\
\text{Product Linearization Constraints:}\quad & \tilde{U}^c_{inr} \leq U^c_{inr}, \quad \forall c \in \mathcal{C}, i \in C^c_n \\
& \tilde{U}^c_{inr} \leq M_3^{cnr}\delta^c_{nr}, \quad \forall c \in \mathcal{C}, i \in C^c_n \\
& \tilde{U}^c_{inr} \geq U^c_{inr} - M_3^{cnr}(1-\delta^c_{nr}), \quad \forall c \in \mathcal{C}, i \in C^c_n \\
& \omega^c_{inr} \leq \delta^c_{nr}, \quad \forall c \in \mathcal{C}, i \in C^c_n \\
& \omega^c_{inr} \leq \mu^c_{inr}, \quad \forall c \in \mathcal{C}, i \in C^c_n \\
& \omega^c_{inr} \geq \delta^c_{nr} + \mu^c_{inr} - 1, \quad \forall c \in \mathcal{C}, i \in C^c_n
\end{aligned}
$$

Choose $M_1 = 1$, $M_2^{cnr} = U^{cnr}_{\max} - U^{cnr}_{\min}$, and $M_3^{cnr} = U^{cnr}_{\max}$ where $U^{cnr}_{\max} = \max_i U^c_{inr}$ and $U^{cnr}_{\min} = \min_i U^c_{inr}$.

Maximum Simulated Likelihood Estimation (MSLE)

$\forall n \in \mathcal{N}, r \in \mathcal{R}:$

Interval Membership Constraints:

$$\gamma_{c-1} \leq u_{nr} + (1 - \delta^c_{nr}), \quad \forall c \in \mathcal{C}$$
$$u_{nr} \leq \gamma_c + (1 - \delta^c_{nr}), \quad \forall c \in \mathcal{C}$$

Maximum Utility Constraints:

$$U^c_{inr} \geq U^c_{jnr} - (U^{cnr}_{\max} - U^{cnr}_{\min})(1 - \mu^c_{inr}), \ \forall c \in \mathcal{C}, i, j \in C^c_n$$

Product Linearization Constraints:

$$\tilde{U}^c_{inr} \leq U^c_{inr}, \quad \forall c \in \mathcal{C}, i \in C^c_n$$
$$\tilde{U}^c_{inr} \leq U^{cnr}_{\max}\delta^c_{nr}, \quad \forall c \in \mathcal{C}, i \in C^c_n$$
$$\tilde{U}^c_{inr} \geq U^c_{inr} - U^{cnr}_{\max}(1 - \delta^c_{nr}), \quad \forall c \in \mathcal{C}, i \in C^c_n$$
$$\omega^c_{inr} \leq \delta^c_{nr}, \quad \forall c \in \mathcal{C}, i \in C^c_n$$
$$\omega^c_{inr} \leq \mu^c_{inr}, \quad \forall c \in \mathcal{C}, i \in C^c_n$$
$$\omega^c_{inr} \geq \delta^c_{nr} + \mu^c_{inr} - 1, \quad \forall c \in \mathcal{C}, i \in C^c_n$$

Finally:

$$\omega_{inr} = \sum_{c \in \mathcal{C}} \omega^c_{inr} \quad \forall i \in C_n.$$

Maximum Simulated Likelihood Estimation (MSLE)

Simulated Choice and Probability Estimation:

Approximate choice probability:

$$P_{in} \approx \frac{1}{R} \sum_{r \in \mathcal{R}} \omega_{inr} \quad \forall n \in \mathcal{N}, i \in C_n.$$

Objective function:

$$\max_{\beta, \pi_c} \quad \sum_{n \in \mathcal{N}} \ln \left(P_{y_n n} \right) \approx \sum_{n \in \mathcal{N}} \ln \left(\frac{1}{R} \sum_{r} \omega_{y_n nr} \right) = -NR + \sum_{n \in \mathcal{N}} \ln \left(\sum_{r} \omega_{y_n nr} \right),$$

Maximum Simulated Likelihood Estimation (MSLE)

Linearizing the log (Fernandez Antolin, 2018):

Add auxiliary variables z_{in} $\forall n \in \mathcal{N}, i \in C_n$, as well as constants $L_r = (1+r)\ln(r) - r\ln(1+r), K_r = \ln(r) - \ln(1+r)$ $\forall r \in \mathcal{R}$. Then:

$$z_{in} \leq L_r - K_r \sum_{r \in \mathcal{R}} \omega_{inr}, \quad \forall n \in \mathcal{N}, i \in C_n.$$

Final Objective:

The simulated log-likelihood (sLL) becomes:

$$sLL(\pi, \beta) = \sum_{n \in \mathcal{N}} z_{y_n n}.$$

TRANSP-OR

EPFL

Maximum Simulated Likelihood Estimation (MSLE)

MILP formulation:

$$\max_{\beta,\pi} \sum_{n\in\mathcal{N}} z_{y_n n} \quad \text{s.t.}$$

$$U^c_{inr} = \textstyle\sum_{k\in\mathcal{K}^c\setminus\{m\}} x_{ink}\beta_k + x_{inm}(\beta^{\text{mean}}_m + \sigma_{nr}\beta^{\text{std}}_m) + \varepsilon_{inr}, \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, i\in C^c_n, r\in\mathcal{R}, \tag{4}$$

$$U_{inr} = \textstyle\sum_{c\in\mathcal{C}} \tilde{U}^c_{inr}, \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, i\in C_n, r\in\mathcal{R}, \tag{5}$$

$$\omega_{inr} = \textstyle\sum_{c\in\mathcal{C}} \omega^c_{inr}, \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, i\in C_n, r\in\mathcal{R}, \tag{6}$$

$$\gamma_{c-1} \le u_{nr} + (1-\delta^c_{nr}), \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, r\in\mathcal{R}, \tag{7}$$

$$u_{nr} \le \gamma_c + (1-\delta^c_{nr}), \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, r\in\mathcal{R}, \tag{8}$$

$$U^c_{inr} \ge U^c_{jnr} - (U^{cnr}_{\max} - U^{cnr}_{\min})(1-\mu^c_{inr}), \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, i,j\in C^c_n, r\in\mathcal{R} \tag{9}$$

$$\tilde{U}^c_{inr} \le U^c_{inr}, \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, i\in C^c_n, r\in\mathcal{R} \tag{10}$$

$$\tilde{U}^c_{inr} \le U^{cnr}_{\max}\delta^c_{nr}, \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, i\in C^c_n, r\in\mathcal{R} \tag{11}$$

$$\tilde{U}^c_{inr} \ge U^c_{inr} - U^{cnr}_{\max}(1-\delta^c_{nr}), \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, i\in C^c_n, r\in\mathcal{R} \tag{12}$$

$$\omega^c_{inr} \le \delta^c_{nr}, \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, i\in C^c_n, r\in\mathcal{R}, \tag{13}$$

$$\omega^c_{inr} \le \mu^c_{inr}, \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, i\in C^c_n, r\in\mathcal{R}, \tag{14}$$

$$\omega^c_{inr} \ge \delta^c_{nr} + \mu^c_{inr} - 1, \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, i\in C^c_n, r\in\mathcal{R}, \tag{15}$$

$$z_{in} \le L_r - K_r \textstyle\sum_{r\in\mathcal{R}} \omega_{inr}, \quad \forall n\in\mathcal{N}, i\in C_n, \tag{16}$$

$$\beta_k \in \mathbb{R}, \quad \forall k\in \bigcup_{c\in\mathcal{C}} \mathcal{K}^c,$$

$$U_{inr}, \tilde{U}^c_{inr}, z_{in} \in \mathbb{R}, \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, i\in C_n, r\in\mathcal{R},$$

$$\gamma_c \in [0,1], \quad \forall c\in\mathcal{C},$$

$$\omega_{inr}, \omega^c_{inr}, \delta^c_{nr}, \mu^c_{nr} \in \{0,1\}, \quad \forall c\in\mathcal{C}, n\in\mathcal{N}, i\in C_n, r\in\mathcal{R}.$$

Maximum Simulated Likelihood Estimation (MSLE)

- Solving the MILP gives a globally optimal solution.
- Issue: only very small instances are solvable.
- Breakpoint Heuristic Algorithm (BHA) for choice-based pricing (Haering et al., 2024) solves a similar MILP with high accuracy in short time.
  - Systematically explores local optima using decision-making breakpoints.
  - Can be categorized as a coordinate descent algorithm.
- Idea: adapt BHA to MSLE and hope for good results.

BHAMSLE: Breakpoint Heuristic Algorithm for MSLE

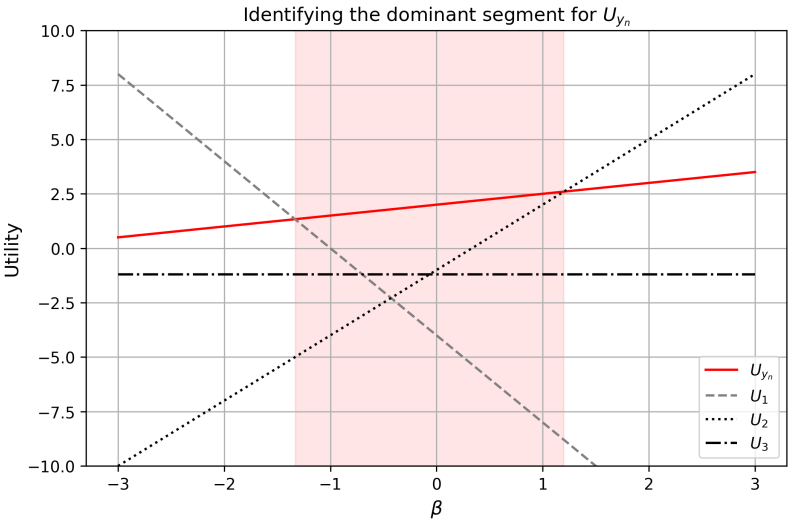
1. Choose a starting point for the estimation, usually, $\beta_k^* = 0,\ k \in \mathcal{K}$, $\gamma_g^* = \frac{g}{C},\ g \in \mathcal{G}$, and compute its objective value $o^* = sLL(\pi^*, \beta^*)$.
2. Set $j = 1$.
3. Fix all other parameters $\beta_k = \beta_k^*$, $k \neq j$ and $\gamma_g = \gamma_g^*$, $g \neq j - K$.

BHAMSLE: Breakpoint Heuristic Algorithm for MSLE

4. Compute the set of breakpoints, initialized as $\mathcal{B} = \{\}$:

for $n \in \mathcal{N}, r \in \mathcal{R}$ **:**
  if $j \leq K$ **:**
    for $c \in \mathcal{C}$ **:**
      if $u_{nr} \in P_c$ **:**
        Compute the segment $[s_1, s_2]$ where $U^c_{y_n nr} \geq U^c_{inr}\ \forall i \in C^c_n$.
        Add (s_1, n) as an entry breakpoint and (s_2, n) as an exit breakpoint to $\mathcal{B}$.
      end
    end

BHAMSLE: Breakpoint Heuristic Algorithm for MSLE



BHAMSLE: Breakpoint Heuristic Algorithm for MSLE

4. Compute the set of breakpoints, initialized as $\mathcal{B} = \{\}$:

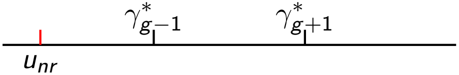
for $n \in \mathcal{N}, r \in \mathcal{R}$ **:**
    if $j \leq K$ **:**
        for $c \in \mathcal{C}$ **:**
            if $\sigma_{nr} \in P_c$ **:**
                Compute the segment $[s_1, s_2]$ where $U^c_{y_n nr} \geq U^c_{inr}$ $\forall i \in C^c_n$.
                Add (s_1, n) as an entry breakpoint and (s_2, n) as an exit breakpoint to $\mathcal{B}$.
            end
        end

BHAMSLE: Breakpoint Heuristic Algorithm for MSLE

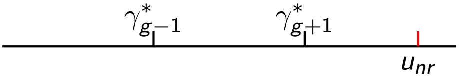
4. Compute the set of breakpoints, initialized as $\mathcal{B} = \{\}$:

```
    else
        Let g ← j − K.
        if u_nr ∈ (γ*_{g−1}, γ*_{g+1}] :
            Let W ← {c ∈ {g, g+1} | U^c_{y_n nr} ≥ U^c_{inr}, ∀i ∈ I}.
            if W = {g, g+1} :
                Add (−∞, n) as an entry breakpoint to B.
            elseif W = {g} :
                Add (u_nr, n) as an entry breakpoint to B.
            elseif W = {g+1} :
                Add (u_nr, n) as an exit breakpoint to B.
            end
        end
    end
end
```

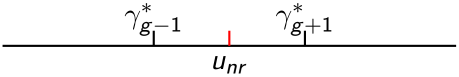
Classification Cases



Case 1: $u_{nr} \leq \gamma^*_{g-1}$
(n, r) guaranteed to be in class $c \leq g - 1$.



Case 2: $u_{nr} > \gamma^*_{g+1}$
(n, r) guaranteed to be in class $g + 2$.

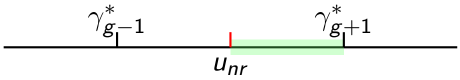


Case 3: $u_{nr} \in (\gamma^*_{g-1}, \gamma^*_{g+1}]$
(n, r) either in class g or $g + 1$, depending on parameter γ_g.

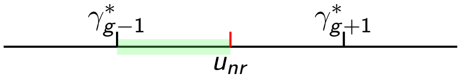
TRANSP-OR EPFL

Dominant Segments and Breakpoints

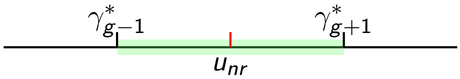
If we match observed only in class g:
$\Rightarrow$ Dominant segment: $[u_{nr}, \gamma^*_{g+1}]$
$\Rightarrow$ **Entry** breakpoint $= u_{nr}$



If we match observed only in class $g+1$:
$\Rightarrow$ Dominant segment: $[\gamma^*_{g-1}, u_{nr}]$
$\Rightarrow$ **Exit** breakpoint $= u_{nr}$



If both in class g and $g+1$ match observed:
$\Rightarrow$ Dominant segment: $[\gamma^*_{g-1}, \gamma^*_{g+1}]$ (guaranteed capture)
$\Rightarrow$ **Entry** breakpoint $= -\infty$



BHAMSLE: Breakpoint Heuristic Algorithm for MSLE

4. Compute the set of breakpoints, initialized as $\mathcal{B} = \{\}$:

```
    else
        Let g ← j − K.
        if u_nr ∈ (γ*_{g−1}, γ*_{g+1}] :
            Let W ← {c ∈ {g, g+1} | U^c_{y_n nr} ≥ U^c_{inr}, ∀i ∈ I}.
            if W = {g, g+1} :
                Add (−∞, n) as an entry breakpoint to B.
            elseif W = {g} :
                Add (u_nr, n) as an entry breakpoint to B.
            elseif W = {g+1} :
                Add (u_nr, n) as an exit breakpoint to B.
            end
        end
    end
end
```

else
- Let $g \leftarrow j - K$.
- **if** $u_{nr} \in (\gamma^*_{g-1}, \gamma^*_{g+1}]$ **:**
  - Let $W \leftarrow \{c \in \{g, g+1\} \mid U^c_{y_n nr} \geq U^c_{inr},\ \forall i \in \mathcal{I}\}$.
  - **if** $W = \{g, g+1\}$ **:**
    - Add $(-\infty, n)$ as an entry breakpoint to $\mathcal{B}$.
  - **elseif** $W = \{g\}$ **:**
    - Add (u_{nr}, n) as an entry breakpoint to $\mathcal{B}$.
  - **elseif** $W = \{g+1\}$ **:**
    - Add (u_{nr}, n) as an exit breakpoint to $\mathcal{B}$.
  - **end**
- **end**

end

end

BHAMSLE: Breakpoint Heuristic Algorithm for MSLE

5. Sort $\mathcal{B}$ in ascending order. Define $\Sigma_n = |\{\text{entry point } (x, y) \in \mathcal{B} : x = -\infty, y = n\}|$, $n \in \mathcal{N}$, $o = -N \ln(R) + \sum_n \ln(\Sigma_n)$ and $\mathcal{B} \leftarrow \{(x, y) \in \mathcal{B} : x \neq -\infty\}$. Then evaluate all $b \in \mathcal{B}$:

for $b \in \mathcal{B}$ **:**
  if b is an entry point **:**
    $o \mathrel{+}= \ln(\Sigma_n + 1) - \ln(\Sigma_n)$.
  else
    $o \mathrel{+}= \ln(\Sigma_n - 1) - \ln(\Sigma_n)$.
  end
  if $o > o^*$ **:**
    $o^* = o$, if $j \leq K$ set $\beta_j^* = b$, else set $\gamma_{j-K}^* = b$.
  end
end

BHAMSLE: Breakpoint Heuristic Algorithm for MSLE

6. Set $j = j + 1$ (if now $j = K + C$, set $j = 1$) and repeat from step 3.
7. Terminate when no improvement is found over $K + C - 1$ iterations.

- Differentiating "entry" / "exit" breakpoints is important:
  - Enables processing them in ascending order.
  - Allows for the incremental computation of changes in the sLL in $\mathcal{O}(1)$ time per breakpoint.
  - Evaluating at each breakpoint would take $\mathcal{O}(NR)$ operations.

Outline

- Introduction
- Problem formulation & Methodology
- Numerical Results
- Conclusion

Numerical Results

Model Variants:

- Latent class logit vs. latent class mixed logit.
- Observed choices vs. synthetically generated choices.

Optimization Methods:

- CMA-ES (`Evolutionary.jl` in Julia).
- BHAMSLE.
- Biogeme 3.2.14 (`PandasBiogeme` in Python).

Simulation and Computation:

- 100 samples per (N, R), averaged results.
- Final log-likelihood for latent class mixed logit evaluated with $R = 10,000$.

Numerical Results

Test 1: Stated preference data on hypothetical mode choice for SwissMetro (SM), collected in Switzerland (Bierlaire et al., 2001).

$$
\begin{aligned}
V^{(1)}_{\text{car}} &= \mathsf{ASC}_{\mathsf{car}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{car}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{car}}, \\
V^{(1)}_{\text{rail}} &= \mathsf{ASC}_{\mathsf{rail}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{rail}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{rail}} + \beta_{\text{headway}} \cdot \text{headway}_{\text{rail}}, \\
V^{(1)}_{\text{SM}} &= \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{SM}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{SM}} + \beta_{\text{headway}} \cdot \text{headway}_{\text{SM}},
\end{aligned}
$$

$$
\begin{aligned}
V^{(2)}_{\text{car}} &= \mathsf{ASC}'_{\mathsf{car}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{car}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{car}}, \\
V^{(2)}_{\text{rail}} &= \mathsf{ASC}'_{\mathsf{rail}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{rail}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{rail}} + \beta_{\text{headway}} \cdot \text{headway}_{\text{rail}}, \\
V^{(2)}_{\text{SM}} &= \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{SM}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{SM}} + \beta_{\text{headway}} \cdot \text{headway}_{\text{SM}}.
\end{aligned}
$$

[1] Latent Class Logit: Log-likelihood & Estimation time

N	R	LL-Bio	LL-Bio-C	Gap (%)	LL-Bio-B	Gap (%)	T-Bio	T-C	T-Bio-C	T-B	T-Bio-B
500	1	-390.267	-397.371	-1.82	-390.267	0.00	3	9	3	0	3
500	5	-390.267	-412.224	-5.63	-390.267	0.00	3	9	3	0	3
500	10	-390.267	-397.582	-1.87	-382.090	2.10	3	9	3	0	4
500	20	-390.267	-404.122	-3.55	-377.073	3.38	3	9	3	1	4
500	50	-390.267	-407.086	-4.31	-374.006	4.17	3	9	3	9	3
500	100	-390.267	-409.461	-4.92	-380.120	2.60	3	9	3	8	3
500	500	-390.267	-404.289	-3.59	-374.499	4.04	3	9	3	63	3
500	1,000	-390.267	-391.796	-0.39	-376.737	3.47	3	9	3	238	3
1,000	1	-779.195	-803.627	-3.14	-779.195	0.00	4	18	4	0	4
1,000	5	-779.195	-827.014	-6.14	-779.195	0.00	3	18	3	0	3
1,000	10	-779.195	-813.654	-4.42	-758.929	2.60	3	18	4	1	4
1,000	20	-779.195	-819.721	-5.20	-760.612	2.38	3	18	3	2	5
1,000	50	-779.195	-808.425	-3.75	-761.902	2.22	3	17	4	8	4
1,000	100	-779.195	-820.438	-5.29	-759.129	2.58	3	17	4	15	4
1,000	500	-779.195	-797.136	-2.30	-758.910	2.60	3	16	3	148	4
1,000	1,000	-779.195	-815.383	-4.64	-756.742	2.88	3	18	3	1,035	4

[1] Latent Class Logit: Class Probabilities

N	R	(p_1, p_2)-Bio	(p_1, p_2)-C	(p_1, p_2)-Bio-C	(p_1, p_2)-B	(p_1, p_2)-Bio-B
500	1	(0.50, 0.50)	(0.42, 0.58)	(0.40, 0.60)	(0.50, 0.50)	(0.50, 0.50)
500	5	(0.50, 0.50)	(0.39, 0.61)	(0.43, 0.57)	(0.50, 0.50)	(0.50, 0.50)
500	10	(0.50, 0.50)	(0.39, 0.61)	(0.39, 0.61)	(0.58, 0.42)	(0.72, 0.28)
500	20	(0.50, 0.50)	(0.49, 0.51)	(0.49, 0.51)	(0.53, 0.47)	(0.69, 0.31)
500	50	(0.50, 0.50)	(0.25, 0.75)	(0.00, 1.00)	(0.61, 0.39)	(0.53, 0.47)
500	100	(0.50, 0.50)	(0.56, 0.44)	(0.60, 0.40)	(0.55, 0.45)	(0.72, 0.28)
500	500	(0.50, 0.50)	(0.47, 0.53)	(0.48, 0.52)	(0.67, 0.33)	(0.66, 0.34)
500	1,000	(0.50, 0.50)	(0.36, 0.64)	(0.01, 0.99)	(0.62, 0.38)	(0.61, 0.39)
1,000	1	(0.50, 0.50)	(0.51, 0.49)	(0.51, 0.49)	(0.50, 0.50)	(0.50, 0.50)
1,000	5	(0.50, 0.50)	(0.44, 0.56)	(0.45, 0.55)	(0.50, 0.50)	(0.50, 0.50)
1,000	10	(0.50, 0.50)	(0.45, 0.55)	(0.43, 0.57)	(0.54, 0.46)	(0.77, 0.23)
1,000	20	(0.50, 0.50)	(0.45, 0.55)	(0.45, 0.55)	(0.59, 0.41)	(0.78, 0.22)
1,000	50	(0.50, 0.50)	(0.61, 0.39)	(0.60, 0.40)	(0.63, 0.37)	(0.74, 0.26)
1,000	100	(0.50, 0.50)	(0.42, 0.58)	(0.40, 0.60)	(0.68, 0.32)	(0.62, 0.38)
1,000	500	(0.50, 0.50)	(0.36, 0.64)	(0.38, 0.62)	(0.62, 0.38)	(0.53, 0.47)
1,000	1,000	(0.50, 0.50)	(0.33, 0.67)	(0.33, 0.67)	(0.66, 0.34)	(0.63, 0.37)

[1] Latent Class Logit: Estimation Results

Parameter	Biogeme	CMA-ES	Biogeme-C	BHAMSLE	Biogeme-B
ASC_{car}	-0.452	70.802	17.579	-11.987	-12.076
ASC'_{car}	-0.657	38.071	10.754	-2.887	-2.360
ASC_{train}	-1.066	1.764	1.761	-1.449	-1.571
ASC'_{train}	-5.299	-11.665	-11.421	-9.955	-5.622
β_{cost}	-1.098	-2.192	-2.195	-0.642	-1.811
β_{HE}	-0.452	-21.005	-5.335	-0.055	0.399
β_{time}	-0.657	0.407	0.401	-0.284	0.261
p_1	0.50	0.33	0.33	0.66	0.63
p_2	0.50	0.67	0.67	0.34	0.37
$LL(\beta)$	-779.195	-821.813	-815.383	-783.631	-756.742

Numerical Results

Test 2: For class 1 we consider $\beta^{\text{mixed}}_{\text{time}} = \beta^{\text{mean}}_{\text{time}} + \beta^{\text{std}}_{\text{time}} \cdot U_n$, where $U_n \sim \mathcal{N}(0,1)$.

$$
\begin{aligned}
V^{(1)}_{\text{car}} &= \text{ASC}_{\text{car}} + \beta^{\text{mixed}}_{\text{traveltime}} \cdot \text{traveltime}_{\text{car}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{car}}, \\
V^{(1)}_{\text{rail}} &= \text{ASC}_{\text{rail}} + \beta^{\text{mixed}}_{\text{traveltime}} \cdot \text{traveltime}_{\text{rail}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{rail}} + \beta_{\text{headway}} \cdot \text{headway}_{\text{rail}}, \\
V^{(1)}_{\text{SM}} &= \beta^{\text{mixed}}_{\text{traveltime}} \cdot \text{traveltime}_{\text{SM}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{SM}} + \beta_{\text{headway}} \cdot \text{headway}_{\text{SM}}, \\
\\
V^{(2)}_{\text{car}} &= \text{ASC}'_{\text{car}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{car}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{car}}, \\
V^{(2)}_{\text{rail}} &= \text{ASC}'_{\text{rail}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{rail}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{rail}} + \beta_{\text{headway}} \cdot \text{headway}_{\text{rail}}, \\
V^{(2)}_{\text{SM}} &= \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{SM}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{SM}} + \beta_{\text{headway}} \cdot \text{headway}_{\text{SM}}.
\end{aligned}
$$

TRANSP-OR

EPFL

[2] Latent Class Mixed Logit: Log-likelihood & Estimation time

N	R	LL-Bio	LL-Bio-C	Gap (%)	LL-Bio-B	Gap (%)	T-Bio	T-C	T-Bio-C	T-B	T-Bio-B
500	1	-438.812	-453.410	-2.88	-438.760	0.01	17	25,774	4	0	20
500	5	-431.788	-444.526	-2.41	-428.005	0.88	14	25,025	5	0	13
500	10	-427.414	-442.143	-2.87	-428.099	-0.16	20	25,665	13	0	19
500	20	-426.447	-443.612	-4.30	-426.925	-0.11	23	27,526	19	1	20
500	50	-439.559	-437.951	-3.04	-435.483	0.93	24	26,428	33	3	17
500	100	-431.124	-445.574	-4.82	-433.809	-0.62	23	27,690	37	6	23
500	500	-490.745	-435.685	-2.28	-436.676	11.02	38	37,066	245	46	48
500	1,000	-488.010	-436.262	-2.42	-435.165	10.83	90	48,358	380	107	55
500	3,000	-474.640	-	-	-433.381	8.69	312	>20h	-	347	135
1,000	1	-877.418	-900.399	-2.62	-875.202	0.25	11	26,643	4	0	11
1,000	5	-868.597	-883.455	-2.82	-867.473	0.13	15	26,253	8	0	15
1,000	10	-855.605	-885.834	-3.91	-855.563	0.00	19	28,093	11	1	21
1,000	20	-856.742	-893.263	-3.77	-853.567	0.37	28	27,758	35	2	30
1,000	50	-869.742	-870.438	-0.56	-866.792	0.34	23	28,612	85	7	23
1,000	100	-888.778	-892.162	-3.82	-870.692	2.04	26	30,897	130	14	38
1,000	500	-867.117	-854.552	-0.70	-845.290	2.52	88	48,886	135	96	83
1,000	1,000	-869.915	-	-	-845.012	2.87	166	>20h	-	219	169
1,000	3,000	-868.542	-	-	-843.699	2.86	477	>20h	-	619	493

[2] Latent Class Mixed Logit: Class Probabilities

N	R	(p_1, p_2)-Bio	(p_1, p_2)-C	(p_1, p_2)-Bio-C	(p_1, p_2)-B	(p_1, p_2)-Bio-B
500	1	(0.45, 0.55)	(0.47, 0.53)	(0.35, 0.65)	(0.50, 0.50)	(0.44, 0.56)
500	5	(0.44, 0.56)	(0.67, 0.33)	(0.67, 0.33)	(0.50, 0.50)	(0.41, 0.59)
500	10	(0.45, 0.55)	(0.99, 0.01)	(0.98, 0.02)	(0.52, 0.48)	(0.43, 0.57)
500	20	(0.46, 0.54)	(1.00, 0.00)	(0.99, 0.01)	(0.49, 0.51)	(0.49, 0.51)
500	50	(0.44, 0.56)	(1.00, 0.00)	(0.98, 0.02)	(0.57, 0.43)	(0.42, 0.58)
500	100	(0.40, 0.60)	(1.00, 0.00)	(1.00, 0.00)	(0.54, 0.46)	(0.41, 0.59)
500	500	(0.33, 0.67)	(0.61, 0.39)	(0.62, 0.38)	(0.43, 0.57)	(0.39, 0.61)
500	1,000	(0.34, 0.66)	(1.00, 0.00)	(1.00, 0.00)	(0.42, 0.58)	(0.38, 0.62)
500	3,000	(0.34, 0.66)	-	-	(0.39, 0.61)	(0.39, 0.61)
1,000	1	(0.39, 0.61)	(0.27, 0.73)	(0.05, 0.95)	(0.50, 0.50)	(0.48, 0.52)
1,000	5	(0.41, 0.59)	(1.00, 0.00)	(0.99, 0.01)	(0.50, 0.50)	(0.45, 0.55)
1,000	10	(0.46, 0.54)	(0.71, 0.29)	(0.72, 0.28)	(0.51, 0.49)	(0.46, 0.54)
1,000	20	(0.45, 0.55)	(1.00, 0.00)	(1.00, 0.00)	(0.41, 0.59)	(0.41, 0.59)
1,000	50	(0.44, 0.56)	(0.51, 0.49)	(0.44, 0.56)	(0.54, 0.46)	(0.41, 0.59)
1,000	100	(0.44, 0.56)	(1.00, 0.00)	(1.00, 0.00)	(0.37, 0.53)	(0.42, 0.58)
1,000	500	(0.44, 0.56)	(1.00, 0.00)	(0.44, 0.56)	(0.41, 0.59)	(0.39, 0.61)
1,000	1,000	(0.43, 0.57)	-	-	(0.40, 0.60)	(0.39, 0.61)
1,000	3,000	(0.43, 0.57)	-	-	(0.41, 0.59)	(0.40, 0.60)

[2] Latent Class Mixed Logit: Estimation Results

Parameter	Biogeme	CMA-ES	Biogeme-C	BHAMSLE	Biogeme-B
ASC_{car}	-4.498	12.427	0.983	-0.257	3.437
ASC'_{car}	-0.108	-31.996	-9.131	-3.736	2.068
ASC_{train}	-1.548	0.187	-1.042	-1.28	-3.163
ASC'_{train}	-6.313	-7.328	-7.389	-6.998	-10.272
β_{cost}	-1.286	-1.510	-1.284	-1.301	-2.714
β_{HE}	-0.037	-45.085	0.046	0.731	-6.767
$\beta_{time,\ mean}$	0.654	-28.847	-16.483	-16.825	-11.937
$\beta_{time,\ std.}$	-5.124	-0.098	-0.166	0.224	-0.432
p_1	0.43	1.00	0.44	0.41	0.40
p_2	0.57	0.00	0.56	0.59	0.60
$LL(\beta)$	-868.542	-863.920	-854.552	-863.905	-843.699

Numerical Results

Test 3: Revealed preference data on mode choice collected in London (Hillel et al., 2018). Choices are synthetically generated:

- Class 2 has a traveltime parameter that is 5x smaller than class 1.
- 70% of the population belongs to class 1, 30% to class 2.

$$\begin{aligned}
V^{(1)}_{\text{walking}} &= \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{walking}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{walking}}, \\
V^{(1)}_{\text{cycling}} &= \mathsf{ASC}_{\mathsf{cycling}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{cycling}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{cycling}}, \\
V^{(1)}_{\text{pt}} &= \mathsf{ASC}_{\mathsf{pt}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{pt}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{pt}}, \\
V^{(1)}_{\text{driving}} &= \mathsf{ASC}_{\mathsf{driving}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{driving}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{driving}}, \\
V^{(2)}_{\text{walking}} &= \beta'_{\text{traveltime}} \cdot \text{traveltime}_{\text{walking}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{walking}}, \\
V^{(2)}_{\text{cycling}} &= \mathsf{ASC}_{\mathsf{cycling}} + \beta'_{\text{traveltime}} \cdot \text{traveltime}_{\text{cycling}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{cycling}}, \\
V^{(2)}_{\text{pt}} &= \mathsf{ASC}_{\mathsf{pt}} + \beta'_{\text{traveltime}} \cdot \text{traveltime}_{\text{pt}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{pt}}, \\
V^{(2)}_{\text{driving}} &= \mathsf{ASC}_{\mathsf{driving}} + \beta'_{\text{traveltime}} \cdot \text{traveltime}_{\text{driving}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{driving}}.
\end{aligned}$$

[3] Latent Class Logit (synthetic): Log-likelihood & Estimation time

N	R	LL-Bio	LL-Bio-C	Gap (%)	LL-Bio-B	Gap (%)	T-Bio	T-C	T-Bio-C	T-B	T-Bio-B
500	1	-524.054	-524.054	0.00	-524.054	0.00	2	8	1	0	3
500	5	-525.484	-525.638	-0.03	-525.484	0.00	2	9	1	0	2
500	10	-525.483	-524.955	0.10	-525.483	0.00	4	9	1	0	1
500	20	-524.371	-535.487	-2.12	-517.502	1.31	2	9	1	2	2
500	50	-523.352	-521.152	0.42	-493.735	5.66	2	8	1	9	2
500	100	-523.352	-527.380	-0.77	-493.267	5.75	3	9	1	24	2
500	500	-525.485	-524.797	0.13	-494.588	5.88	1	9	1	211	1
500	1000	-525.489	-521.119	0.83	-495.370	5.73	1	10	1	445	1
1000	1	-1051.925	-1039.297	1.20	-1051.921	0.00	2	16	1	0	2
1000	5	-1050.030	-1054.650	-0.44	-1050.034	0.00	2	19	2	0	4
1000	10	-1051.926	-1084.740	-3.12	-1051.929	0.00	2	17	1	0	2
1000	20	-1051.928	-1042.348	0.91	-1039.099	1.22	2	15	1	4	3
1000	50	-1051.923	-1051.394	0.05	-988.700	6.01	2	18	2	24	2
1000	100	-1051.929	-1074.221	-2.12	-987.545	6.12	2	19	2	51	2
1000	500	-1051.434	-1056.687	-0.50	-989.194	5.92	1	19	1	506	1
1000	1000	-1051.926	-1042.242	0.92	-988.494	6.03	2	16	1	1121	1

[3] Latent Class Logit (synthetic): traveltime parameter ratios

N	R	Ratio-Bio	Ratio-C	Ratio-Bio-C	Ratio-B	Ratio-Bio-B
500	1	1.00	0.94	0.79	1.00	1.00
500	5	1.00	-3.47	-11.77	1.00	1.00
500	10	1.00	0.59	0.30	1.00	1.00
500	20	1.00	1.06	1.29	0.68	-0.24
500	50	1.00	9.35	-54.83	4.36	6.11
500	100	1.00	0.78	0.57	3.29	4.07
500	500	1.00	-1.58	-34.68	4.15	5.70
500	1,000	1.00	0.57	0.23	4.25	5.70
1,000	1	1.00	7.94	-12.20	1.00	1.00
1,000	5	1.00	1.11	-21.86	1.00	1.00
1,000	10	1.00	0.97	0.69	1.00	1.00
1,000	20	1.00	0.57	0.69	0.65	-0.67
1,000	50	1.00	1.03	1.21	3.95	5.98
1,000	100	1.00	-2.56	-0.17	3.11	3.67
1,000	500	1.00	0.89	-0.21	4.24	5.17
1,000	1,000	1.00	1.30	1.28	4.17	5.20

[3] Latent Class Logit (synthetic): Class Probabilities

N	R	(p_1, p_2)-Bio	(p_1, p_2)-C	(p_1, p_2)-Bio-C	(p_1, p_2)-B	(p_1, p_2)-Bio-B
500	1	(0.50, 0.50)	(0.62, 0.38)	(0.49, 0.51)	(0.50, 0.50)	(0.50, 0.50)
500	5	(0.50, 0.50)	(0.50, 0.50)	(0.13, 0.87)	(0.50, 0.50)	(0.50, 0.50)
500	10	(0.50, 0.50)	(0.57, 0.43)	(0.43, 0.57)	(0.50, 0.50)	(0.50, 0.50)
500	20	(0.50, 0.50)	(0.57, 0.43)	(0.45, 0.55)	(0.40, 0.60)	(0.06, 0.94)
500	50	(0.50, 0.50)	(0.55, 0.45)	(0.53, 0.47)	(0.50, 0.50)	(0.70, 0.30)
500	100	(0.50, 0.50)	(0.61, 0.39)	(0.39, 0.61)	(0.63, 0.37)	(0.67, 0.33)
500	500	(0.50, 0.50)	(0.59, 0.41)	(0.36, 0.64)	(0.66, 0.34)	(0.71, 0.29)
500	1,000	(0.50, 0.50)	(0.55, 0.45)	(0.31, 0.69)	(0.66, 0.34)	(0.71, 0.29)
1,000	1	(0.50, 0.50)	(0.55, 0.45)	(0.21, 0.79)	(0.50, 0.50)	(0.50, 0.50)
1,000	5	(0.50, 0.50)	(0.63, 0.37)	(0.40, 0.60)	(0.50, 0.50)	(0.50, 0.50)
1,000	10	(0.50, 0.50)	(0.63, 0.37)	(0.46, 0.54)	(0.50, 0.50)	(0.50, 0.50)
1,000	20	(0.50, 0.50)	(0.50, 0.50)	(0.43, 0.57)	(0.46, 0.54)	(0.21, 0.79)
1,000	50	(0.50, 0.50)	(0.61, 0.39)	(0.54, 0.46)	(0.68, 0.32)	(0.79, 0.21)
1,000	100	(0.50, 0.50)	(0.57, 0.43)	(0.50, 0.50)	(0.64, 0.36)	(0.66, 0.34)
1,000	500	(0.50, 0.50)	(0.60, 0.40)	(0.39, 0.61)	(0.65, 0.35)	(0.69, 0.31)
1,000	1,000	(0.50, 0.50)	(0.58, 0.42)	(0.53, 0.47)	(0.69, 0.31)	(0.69, 0.31)

[3] Latent Class Logit (synthetic): Estimation Results

Parameter	Biogeme	CMA-ES	Biogeme-C	BHAMSLE	Biogeme-B
ASC_{bike}	-3.420	-4.271	-3.892	-3.794	-3.961
ASC_{car}	-0.685	-0.592	-0.912	-0.927	-1.465
ASC_{pb}	-0.356	-0.215	-0.123	-0.374	-0.585
β_{cost}	-0.158	-0.251	-1.983	-0.159	-0.145
β_{time}	-2.496	-2.472	-3.245	-6.296	-6.503
β'_{time}	-2.496	-1.901	-2.535	-1.509	-1.251
p_1	0.50	0.58	0.53	0.69	0.69
p_2	0.50	0.42	0.47	0.31	0.31
$LL(\beta)$	-1,051.926	-1,046.672	-1,042.242	-1,014.289	-988.494

Numerical Results

Test 4: For class 1 we consider $\beta_{\text{cost}}^{\text{mixed}} = \beta_{\text{cost}}^{\text{mean}} + \beta_{\text{cost}}^{\text{std}} \cdot U_n$, where $U_n \sim \mathcal{N}(0,1)$. We add a third class, with "lazy" people (don't consider walking or cycling). Split: 50%, 30%, 20%.

$$
\begin{aligned}
V_{\text{walking}}^{(1)} &= \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{walking}} + \beta_{\text{cost}}^{\text{mixed}} \cdot \text{cost}_{\text{walking}}, \\
V_{\text{cycling}}^{(1)} &= \text{ASC}_{\text{cycling}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{cycling}} + \beta_{\text{cost}}^{\text{mixed}} \cdot \text{cost}_{\text{cycling}}, \\
V_{\text{pt}}^{(1)} &= \text{ASC}_{\text{pt}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{pt}} + \beta_{\text{cost}}^{\text{mixed}} \cdot \text{cost}_{\text{pt}}, \\
V_{\text{driving}}^{(1)} &= \text{ASC}_{\text{driving}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{driving}} + \beta_{\text{cost}}^{\text{mixed}} \cdot \text{cost}_{\text{driving}}, \\
\\
V_{\text{pt}}^{(3)} &= \text{ASC}_{\text{pt}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{pt}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{pt}}, \\
V_{\text{driving}}^{(3)} &= \text{ASC}_{\text{driving}} + \beta_{\text{traveltime}} \cdot \text{traveltime}_{\text{driving}} + \beta_{\text{cost}} \cdot \text{cost}_{\text{driving}}.
\end{aligned}
$$

[4] Latent Class Mixed Logit (synthetic): Log-likelihood & Estimation time

N	R	LL-Bio	LL-Bio-C	Gap C (%)	LL-Bio-B	Gap B (%)	T-Bio	T-C	T-Bio-C	T-B	T-Bio-B
500	1	-528.517	-529.759	-0.23	-546.381	-3.38	1	27,195	1	0	1
500	5	-529.650	-527.686	0.37	-576.577	-8.86	6	27,031	6	0	6
500	10	-528.341	-533.415	-0.96	-553.279	-4.72	12	27,007	12	0	9
500	20	-531.867	-530.720	0.22	-534.740	-0.54	23	26,880	23	78	23
500	50	-527.150	-529.982	-0.54	-525.410	0.33	58	27,097	58	131	59
500	100	-530.017	-528.852	0.22	-528.374	0.31	108	27,162	108	271	122
500	500	-529.292	-527.867	0.27	-512.725	3.13	719	32,881	719	1,196	563
500	1,000	-525.036	-528.851	-0.73	-509.862	2.89	1,260	33,086	1,260	2,084	1,216
500	3,000	-525.564	-526.246	-0.13	-507.590	3.42	1,359	37,063	1,359	5,543	1,188
1,000	1	-1,051.300	-1,050.300	0.10	-1,053.298	-0.19	3	33,615	3	0	3
1,000	5	-1,051.880	-1,065.680	-1.31	-1,052.301	-0.04	12	33,763	12	0	11
1,000	10	-1,051.470	-1,048.940	0.24	-1,051.155	0.03	26	32,964	26	0	26
1,000	20	-1,049.410	-1,048.650	0.07	-1,051.824	-0.23	45	33,770	45	179	49
1,000	50	-1,050.840	-1,051.740	-0.09	-1,049.790	0.10	137	33,939	137	213	113
1,000	100	-1,054.330	-1,050.760	0.34	-1,030.502	2.26	296	32,651	296	551	240
1,000	500	-1,051.370	-1,053.790	-0.23	-1,016.570	3.31	1,659	38,360	1,659	2,803	1,171
1,000	1,000	-1,049.230	-1,059.890	-1.02	-1,014.081	3.35	2,805	37,943	2,805	5,248	2,423
1,000	3,000	-1,059.610	-1,066.090	-0.61	-1,022.137	3.53	2,459	40,720	2,459	23,423	2,642

[4] Latent Class Mixed Logit (synthetic): Class Probabilities

N	R	(p_1, p_2, p_3)-Bio	$(p1, p2, p3)$-C	$(p1, p2, p3)$-Bio-C	(p_1, p_2, p_3)-B	(p_1, p_2, p_3)-Bio-B
500	1	(0.95, 0.05, 0.01)	(1.00, 0.00, 0.00)	(0.97, 0.01, 0.02)	(0.33, 0.33, 0.33)	(0.83, 0.05, 0.12)
500	5	(0.94, 0.05, 0.01)	(1.00, 0.00, 0.00)	(0.91, 0.01, 0.08)	(0.33, 0.33, 0.33)	(0.94, 0.06, 0.00)
500	10	(0.93, 0.05, 0.02)	(0.00, 1.00, 0.00)	(0.98, 0.01, 0.02)	(0.33, 0.33, 0.33)	(0.95, 0.00, 0.05)
500	20	(0.93, 0.07, 0.00)	(1.00, 1.00, 0.00)	(0.93, 0.01, 0.06)	(0.34, 0.34, 0.32)	(0.82, 0.04, 0.14)
500	50	(0.97, 0.02, 0.01)	(0.00, 1.00, 0.00)	(0.91, 0.00, 0.08)	(0.31, 0.39, 0.30)	(0.85, 0.01, 0.14)
500	100	(0.97, 0.03, 0.00)	(0.00, 1.00, 0.00)	(0.99, 0.01, 0.00)	(0.36, 0.36, 0.28)	(0.81, 0.02, 0.17)
500	500	(0.98, 0.02, 0.00)	(1.00, 0.00, 0.00)	(0.91, 0.04, 0.06)	(0.40, 0.42, 0.18)	(0.45, 0.35, 0.20)
500	1,000	(0.96, 0.04, 0.00)	(0.00, 0.00, 1.00)	(1.00, 0.00, 0.00)	(0.36, 0.38, 0.14)	(0.44, 0.31, 0.25)
500	3,000	(0.89, 0.11, 0.00)	(0.00, 1.00, 0.00)	(0.93, 0.07, 0.00)	(0.43, 0.32, 0.25)	(0.48, 0.29, 0.23)
1,000	1	(0.98, 0.02, 0.00)	(0.00, 1.00, 0.00)	(0.93, 0.03, 0.04)	(0.33, 0.33, 0.33)	(0.82, 0.14, 0.04)
1,000	5	(0.87, 0.12, 0.00)	(0.00, 1.00, 0.00)	(0.98, 0.01, 0.01)	(0.33, 0.33, 0.33)	(0.81, 0.18, 0.01)
1,000	10	(0.94, 0.06, 0.00)	(1.00, 0.00, 0.00)	(1.00, 0.00, 0.00)	(0.33, 0.33, 0.33)	(0.97, 0.03, 0.00)
1,000	20	(0.95, 0.05, 0.00)	(1.00, 0.00, 0.00)	(1.00, 0.00, 0.00)	(0.35, 0.35, 0.30)	(0.90, 0.09, 0.00)
1,000	50	(0.93, 0.07, 0.00)	(0.00, 1.00, 0.00)	(0.98, 0.01, 0.01)	(0.32, 0.42, 0.26)	(0.90, 0.10, 0.00)
1,000	100	(0.89, 0.11, 0.00)	(1.00, 0.00, 0.00)	(0.99, 0.01, 0.00)	(0.43, 0.32, 0.25)	(0.81, 0.06, 0.13)
1,000	500	(0.92, 0.08, 0.00)	(1.00, 0.00, 0.00)	(0.96, 0.04, 0.00)	(0.46, 0.28, 0.26)	(0.47, 0.26, 0.27)
1,000	1,000	(0.72, 0.09, 0.19)	(0.00, 0.00, 1.00)	(0.94, 0.04, 0.02)	(0.43, 0.32, 0.25)	(0.49, 0.31, 0.20)
1,000	3,000	(0.91, 0.01, 0.09)	(0.00, 1.00, 0.00)	(0.96, 0.01, 0.03)	(0.46, 0.31, 0.23)	(0.50, 0.31, 0.19)

[4] Latent Class Mixed Logit (synthetic): Estimation Results

Parameter	Biogeme	CMA-ES	Biogeme-C	BHAMSLE	Biogeme-B
ASC_{bike}	-4.041	11.740	-3.386	-3.936	-3.890
ASC_{car}	-18.563	-20.829	-0.524	-1.439	-1.138
ASC_{pb}	-16.195	-6.060	-0.035	-1.281	-0.245
$\beta_{cost,\ mean}$	-0.156	-18.890	-0.172	-0.134	-0.175
$\beta_{cost,\ std.}$	-2.426	23.218	-3.156	-2.546	-1.277
β_{time}	-1.314	-4.257	-2.285	-5.283	-4.144
β'_{time}	-8.157	-4.738	-4.210	-2.120	-1.843
p_1	0.91	0.00	0.96	0.46	0.50
p_2	0.01	1.00	0.01	0.31	0.31
p_3	0.09	0.00	0.03	0.23	0.19
$LL(\beta)$	-1,059.610	-1,063.112	-1,066.090	-1,055.974	-1,022.137

TRANSP-OR

EPFL

Outline

- Introduction
- Problem formulation & Methodology
- Numerical Results
- Conclusion

Conclusion

Key Contributions:

- Systematic exploration of local optima via decision-making breakpoints.
- BHAMSLE provides robust initialization for complex estimation problems.
- Achieves improvements in class identification as well as log-likelihood, ranging from 2% to 10%.

Comparison with CMA-ES:

- CMA-ES underperforms in log-likelihood gains and computational efficiency.
- BHAMSLE reduces reliance on costly random re-initialization.

Future Research:

- Extend BHAMSLE to more complex DCMs.
- Optimize performance with parallelization techniques.

Bierlaire, M. (2023). A short introduction to biogeme. Technical Report TRANSP-OR 230620, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

Bierlaire, M., Axhausen, K., and Abay, G. (2001). The acceptance of modal innovation: The case of swissmetro. In *Swiss transport research conference*.

Fernandez Antolin, A. (2018). *Dealing with correlations in discrete choice models*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Machine learning basics. *Deep learning*, 1(7):98–164.

Haering, T., Torres, Legault, R., Fabian, and Bierlaire, M. (2024). Heuristics and exact algorithms for choice-based capacitated and uncapacitated continuous pricing. Technical Report TRANSP-OR 240918, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

Hansen, N., Müller, S. D., and Koumoutsakos, P. (2003). Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (cma-es). *Evolutionary computation*, 11(1):1–18.

Hauschild, T. and Jentschel, M. (2001). Comparison of maximum likelihood estimation and chi-square statistics applied to counting experiments. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 457(1-2):384–401.

Hillel, T., Elshafie, M. Z., and Jin, Y. (2018). Recreating passenger mode choice-sets for transport simulation: A case study of london, uk. *Proceedings of the Institution of Civil Engineers-Smart Infrastructure and Construction*, 171(1):29–42.

Jung, T. and Wickrama, K. A. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and personality psychology compass*, 2(1):302–317.

Lubke, G. H. and Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, 10(1):21–39.

Peer, S., Knockaert, J., and Verhoef, E. T. (2016). Train commuters' scheduling preferences: Evidence from a large-scale peak avoidance experiment. *Transportation Research Part B: Methodological*, 83:314–333.

TRANSP-OR

EPFL