UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

Modeling unobserved choice behavior heterogeneity

Comparing methods in terms of support recovery and estimation speed

Lennart Oelschläger

Bielefeld University, Department of Empirical Methods

16 February 2024

Outline

Outline

Example: Buying a new car

Consumer has the choice:

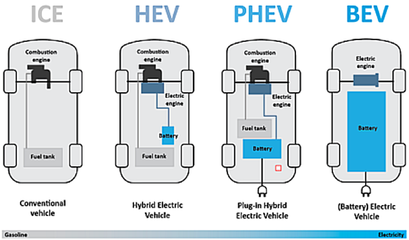
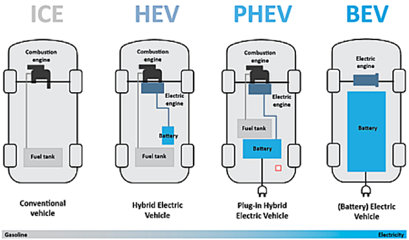


Image source https://www.linkedin.com/pulse/best-choice-ice-vehicles-vs-evs-hybrid-how-shaping-up-kulkarni

Example: Buying a new car

Consumer has the choice:

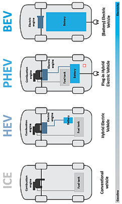


Entities that try to understand the choice process:

- Manufacturers (What to produce?)
- Retailers (How to sell?)
- Politicians (How to change behavior?)

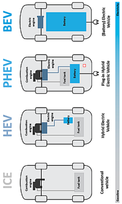
Image source https://www.linkedin.com/pulse/best-choice-ice-vehicles-vs-evs-hybrid-how-shaping-up-kulkarni

Random utility model approach



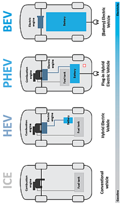
$$\rightarrow \begin{bmatrix} U_{\text{BEV}} \\ U_{\text{PHEV}} \\ U_{\text{HEV}} \\ U_{\text{ICE}} \end{bmatrix}$$

Random utility model approach



$$\rightarrow \begin{bmatrix} U_{\text{BEV}} \\ U_{\text{PHEV}} \\ U_{\text{HEV}} \\ U_{\text{ICE}} \end{bmatrix} = \boldsymbol{X}\beta + \epsilon$$

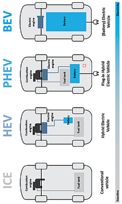
Random utility model approach



$$\rightarrow \begin{bmatrix} U_{\text{BEV}} \\ U_{\text{PHEV}} \\ U_{\text{HEV}} \\ U_{\text{ICE}} \end{bmatrix} = \boldsymbol{X}\beta + \epsilon$$

- $\boldsymbol{X}$ is a matrix of decider and / or alternatives attributes

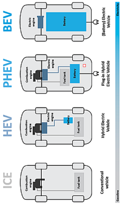
Random utility model approach



$$\rightarrow \begin{bmatrix} U_{\text{BEV}} \\ U_{\text{PHEV}} \\ U_{\text{HEV}} \\ U_{\text{ICE}} \end{bmatrix} = \boldsymbol{X}\boldsymbol{\beta} + \epsilon$$

- $\boldsymbol{X}$ is a matrix of decider and / or alternatives attributes
- $\boldsymbol{\beta}$ is to be estimated, encodes how $\boldsymbol{X}$ influences $\boldsymbol{U}$ and is of primary interest

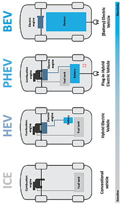
Random utility model approach



$$\rightarrow \quad \begin{bmatrix} U_{\text{BEV}} \\ U_{\text{PHEV}} \\ U_{\text{HEV}} \\ U_{\text{ICE}} \end{bmatrix} = \boldsymbol{X}\beta + \epsilon$$

- $\boldsymbol{X}$ is a matrix of decider and / or alternatives attributes
- β is to be estimated, encodes how $\boldsymbol{X}$ influences $\boldsymbol{U}$ and is of primary interest
- ϵ is the random error, typically $\epsilon \sim \text{MVN}(\mathbf{0}, \Sigma)$ (probit) or $\epsilon \sim$ iid EV (logit)

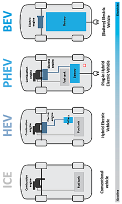
Random utility model approach



$$\rightarrow \begin{bmatrix} U_{\text{BEV}} \\ U_{\text{PHEV}} \\ U_{\text{HEV}} \\ U_{\text{ICE}} \end{bmatrix} = \boldsymbol{X}\beta + \epsilon = \boldsymbol{U}$$

- $\boldsymbol{X}$ is a matrix of decider and / or alternatives attributes
- β is to be estimated, encodes how $\boldsymbol{X}$ influences $\boldsymbol{U}$ and is of primary interest
- ϵ is the random error, typically $\epsilon \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$ (probit) or $\epsilon \sim$ iid EV (logit)

Random utility model approach



$$\rightarrow \begin{bmatrix} U_{\text{BEV}} \\ U_{\text{PHEV}} \\ U_{\text{HEV}} \\ U_{\text{ICE}} \end{bmatrix} = \boldsymbol{X}\beta + \epsilon = \boldsymbol{U} \quad \rightarrow \quad \text{choice} = \arg\max \boldsymbol{U}$$

- $\boldsymbol{X}$ is a matrix of decider and / or alternatives attributes
- β is to be estimated, encodes how $\boldsymbol{X}$ influences $\boldsymbol{U}$ and is of primary interest
- ϵ is the random error, typically $\epsilon \sim \text{MVN}(\mathbf{0}, \Sigma)$ (probit) or $\epsilon \sim$ iid EV (logit)

Typical choice attributes *X*

Attribute*	Option 1	Option 2	Option 3
Vehicle Type	Conventional 300 mile range on 1 tank	Plug-In Hybrid & 300 mile range on 1 tank (first 40 miles electric)	Electric 75 mile range on full charge
Brand	German	American	Japanese
Purchase Price	$18,000	$32,000	$24,000
Fast Charging Capability	--	Not Available	Available
Operating Cost (Equivalent Gasoline Fuel Efficiency)	19 cents per mile (20 MPG equivalent)	12 cents per mile (30 MPG equivalent)	6 cents per mile (60 MPG equivalent)
0 to 60 mph Acceleration Time**	8.5 seconds (Medium-Slow)	8.5 seconds (Medium-Slow)	7 seconds (Medium-Fast)
	○	○	○

Image source Helveston et al. (2015)

Typical choice attributes *X*

Attribute*	Option 1	Option 2	Option 3
Vehicle Type	Conventional 300 mile range on 1 tank	Plug-In Hybrid & 300 mile range on 1 tank (first 40 miles electric)	Electric 75 mile range on full charge
Brand	German	American	Japanese
Purchase Price	$18,000	$32,000	$24,000
Fast Charging Capability	--	Not Available	Available
Operating Cost (Equivalent Gasoline Fuel Efficiency)	19 cents per mile (20 MPG equivalent)	12 cents per mile (30 MPG equivalent)	6 cents per mile (60 MPG equivalent)
0 to 60 mph Acceleration Time**	8.5 seconds (Medium-Slow)	8.5 seconds (Medium-Slow)	7 seconds (Medium-Fast)
	○	○	○

Image source Helveston et al. (2015)

Typical choice attributes *X*

Attribute*	Option 1	Option 2	Option 3
Vehicle Type	Conventional 300 mile range on 1 tank	Plug-In Hybrid & 300 mile range on 1 tank (first 40 miles electric)	Electric 75 mile range on full charge
Brand	German	American	Japanese
Purchase Price	$18,000	$32,000	$24,000
Fast Charging Capability	--	Not Available	Available
Operating Cost (Equivalent Gasoline Fuel Efficiency)	19 cents per mile (20 MPG equivalent)	12 cents per mile (30 MPG equivalent)	6 cents per mile (60 MPG equivalent)
0 to 60 mph Acceleration Time**	8.5 seconds (Medium-Slow)	8.5 seconds (Medium-Slow)	7 seconds (Medium-Fast)
	○	○	○

Image source Helveston et al. (2015)

Typical choice attributes *X*

Attribute*	Option 1	Option 2	Option 3
Vehicle Type	Conventional 300 mile range on 1 tank	Plug-In Hybrid & 300 mile range on 1 tank (first 40 miles electric)	Electric 75 mile range on full charge
Brand	German	American	Japanese
Purchase Price	$18,000	$32,000	$24,000
Fast Charging Capability	--	Not Available	Available
Operating Cost (Equivalent Gasoline Fuel Efficiency)	19 cents per mile (20 MPG equivalent)	12 cents per mile (30 MPG equivalent)	6 cents per mile (60 MPG equivalent)
0 to 60 mph Acceleration Time**	8.5 seconds (Medium-Slow)	8.5 seconds (Medium-Slow)	7 seconds (Medium-Fast)

Image source Helveston et al. (2015)

How to interpret β

- Sign: $\beta_{\text{Acceleration time}} < 0$

How to interpret β

- Sign: $\beta_{\text{Acceleration time}} < 0$
- Magnitude: $\beta_{\text{Operating cost}} \lesseqgtr \beta_{\text{Purchase price}}$ (after standardization)

How to interpret β

- Sign: $\beta_{\text{Acceleration time}} < 0$
- Magnitude: $\beta_{\text{Operating cost}} \lessgtr \beta_{\text{Purchase price}}$ (after standardization)
- Substitution: $\frac{\beta_{\text{BEV fast-charge}}}{\beta_{\text{Purchase price}}}$ (willingness to pay)

How to interpret $\boldsymbol{\beta}$

- Sign: $\beta_{\text{Acceleration time}} < 0$
- Magnitude: $\beta_{\text{Operating cost}} \lesseqgtr \beta_{\text{Purchase price}}$ (after standardization)
- Substitution: $\dfrac{\beta_{\text{BEV fast-charge}}}{\beta_{\text{Purchase price}}}$ (willingness to pay)

But is a constant $\boldsymbol{\beta}$ across deciders realistic?

UNIVERSITÄT BIELEFELD
Faculty of Business Administration and Economics

How to interpret $\boldsymbol{\beta}$

- Sign: $\beta_{\text{Acceleration time}} < 0$
- Magnitude: $\beta_{\text{Operating cost}} \lesseqgtr \beta_{\text{Purchase price}}$ (after standardization)
- Substitution: $\frac{\beta_{\text{BEV fast-charge}}}{\beta_{\text{Purchase price}}}$ (willingness to pay)

But is a constant $\boldsymbol{\beta}$ across deciders realistic?

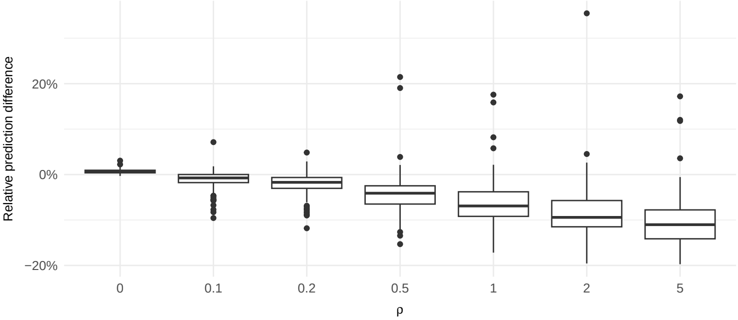
In general, **no**, we should assume that people are heterogeneous in their choice behavior.

Ignored variance in β

1. Simulated 200 data sets from probit model with $\beta \sim N(\boldsymbol{b}, \boldsymbol{\Omega})$ where $\boldsymbol{\Omega} = \rho \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$
2. Estimated probit model with $\boldsymbol{\Omega} = \mathbf{0}$ fixed as well as $\boldsymbol{\Omega}$ flexible
3. Computed relative loss when $\boldsymbol{\Omega} = \mathbf{0}$ in average out-of-sample predicted choice probabilities

Ignored variance in β

1. Simulated 200 data sets from probit model with $\beta \sim N(\boldsymbol{b}, \boldsymbol{\Omega})$ where $\boldsymbol{\Omega} = \rho \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$
2. Estimated probit model with $\boldsymbol{\Omega} = \mathbf{0}$ fixed as well as $\boldsymbol{\Omega}$ flexible
3. Computed relative loss when $\boldsymbol{\Omega} = \mathbf{0}$ in average out-of-sample predicted choice probabilities



Outline

How to model variation in $\boldsymbol{\beta}$?

Three options:

1. Control for heterogeneity via exogenous regressors (in many cases infeasible)
2. Fixed effects: $\boldsymbol{\beta}_n$ for each decider n (would need many choice occasions per decider)
3. Random effects: $\boldsymbol{\beta}_n \sim F$ (instead of $\boldsymbol{\beta}_n$, estimate F)

How to model variation in $\boldsymbol{\beta}$?

Three options:

1. Control for heterogeneity via exogenous regressors (in many cases infeasible)
2. Fixed effects: $\boldsymbol{\beta}_n$ for each decider n (would need many choice occasions per decider)
3. Random effects: $\boldsymbol{\beta}_n \sim F$ (instead of $\boldsymbol{\beta}_n$, estimate F)

If we opt for option 3 (most people in practice do), we need to decide

- how much structure we want to impose on F,
- how to estimate F.

Parametric F

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Will subsidies drive electric vehicle adoption? Measuring consumer preferences in the U.S. and China

CrossMark

John Paul Helveston [a,1], Yimin Liu [b,2], Elea McDonnell Feit [c,3], Erica Fuchs [a,4], Erica Klampfl [b,5], Jeremy J. Michalek [a,d,*]

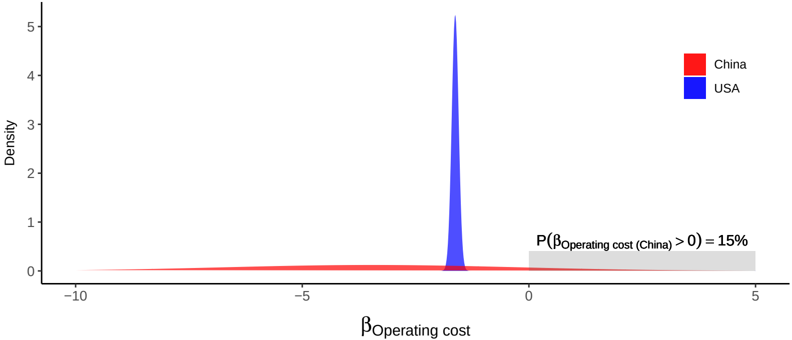
[a] Department of Engineering and Public Policy, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, United States
[b] Ford Motor Company, Dearborn, MI, United States
[c] Department of Marketing, Drexel University, 828 Gerri C. LeBow Hall, 3141 Chestnut St., Philadelphia, PA 19104, United States
[d] Department of Mechanical Engineering, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, United States

Parametric F

In order to relax some limiting assumptions from the basic logit model (e.g. the independence from irrelevant alternatives (IIA) property (Train, 2009)), we also apply a random coefficients mixed logit model (McFadden and Train, 2000) in the WTP space, which allows for heterogeneity of preferences across the population and more general substitution patterns. While the basic logit model effectively assumes $\boldsymbol{\gamma}_i = \boldsymbol{\gamma} \forall i$ and captures variation in WTP across individuals only in the error term ε_{ijt}, the mixed logit model instead assumes that the $\boldsymbol{\gamma}_i$ coefficients are drawn from a parametric distribution.[11] Following convention, we assume each element γ_{ij} of the vector $\boldsymbol{\gamma}_i$ is drawn from an independent normal distribution, where $\gamma_{ij} \sim N(\mu_i, \sigma_i^2)$. We assume a fixed (non-random) α_i coefficient for all mixed logit models. While WTP could also be computed from a preference space mixed logit model post hoc, Train and Weeks (2005) show that such estimates have unreasonably large variance in comparison to those from a WTP space model.

Parametric F

In their paper, each $\beta_p \sim$ iid $N(\mu_p, \sigma_p^2)$, for example:



Estimates from Helveston et al. (2015)

F as a mixture

$$F = \sum_{c=1}^{C} a_c F_c, \text{ for example } F_c = \text{MVN}(\boldsymbol{b}_c, \boldsymbol{\Omega}_c):$$

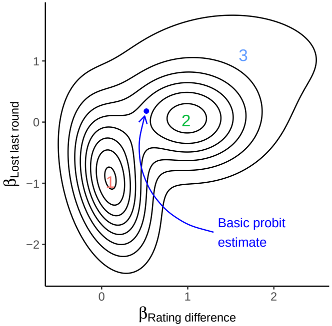
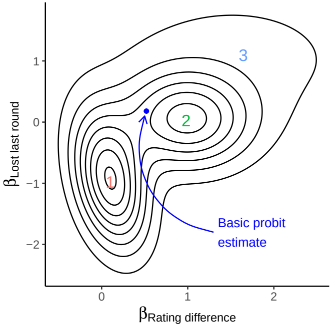


Image source Oelschläger and Bauer (2023)

Another option: Non-parametric F

F as a mixture

$$F = \sum_{c=1}^{C} a_c F_c, \text{ for example } F_c = \text{MVN}(\boldsymbol{b}_c, \boldsymbol{\Omega}_c):$$



- more flexible
- allows for classification
- harder to estimate
- class number C ?

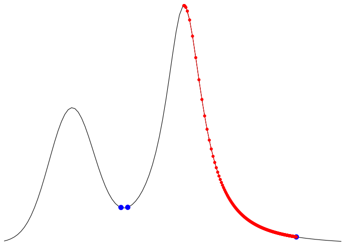
Image source Oelschläger and Bauer (2023)

Another option: Non-parametric F

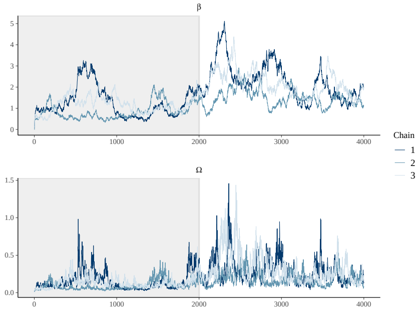
Outline

Two estimation methods

Likelihood optimization



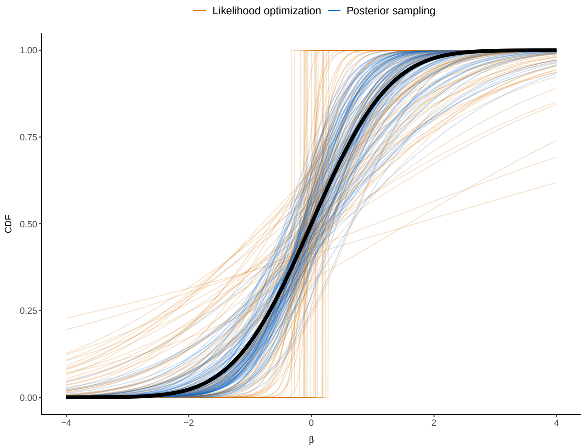
Posterior sampling



Convergence of Gibbs sampler

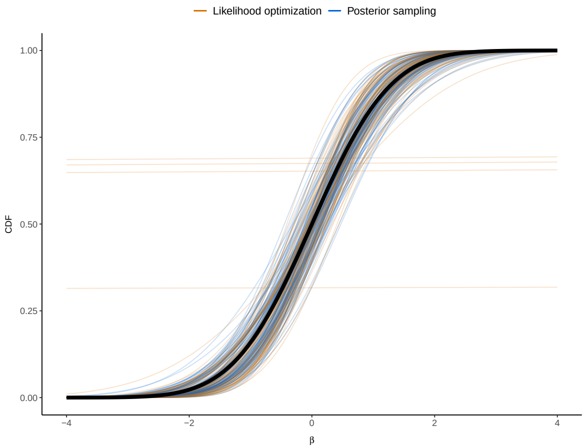
Comparison of approximation

$\beta \sim N(0, 1)$, number of choice occasions per decider: $T = 1$



Comparison of approximation

$\beta \sim N(0,1)$, number of choice occasions per decider: $T = \mathbf{10}$

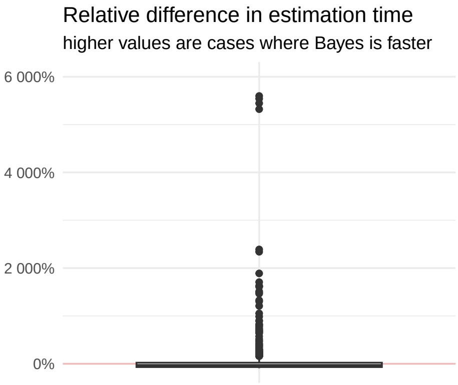


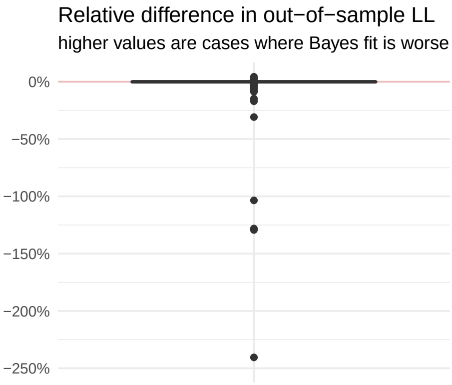
Comparison in different data scenarios

80 simulated data sets from each combination of:

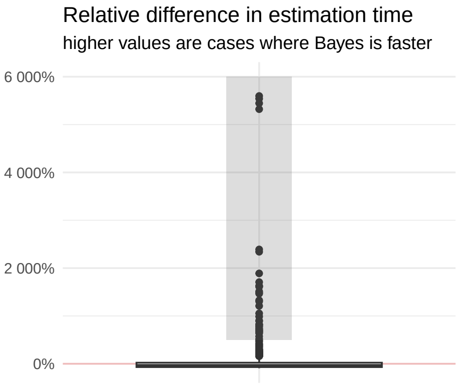
- Number of deciders $N \in \{50, 100\}$
- Number of choices per decider $T \in \{10, 20\}$
- Number of choice alternatives $J \in \{2, 3\}$
- Number of random effects $P \in \{1, 2\}$

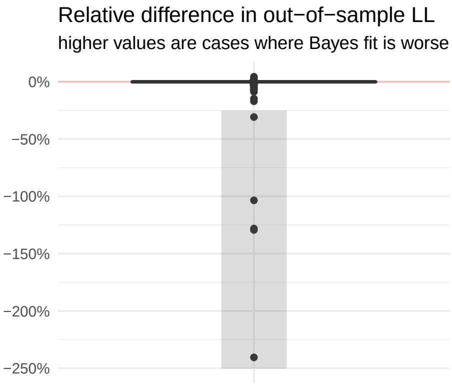
Comparison in different data scenarios



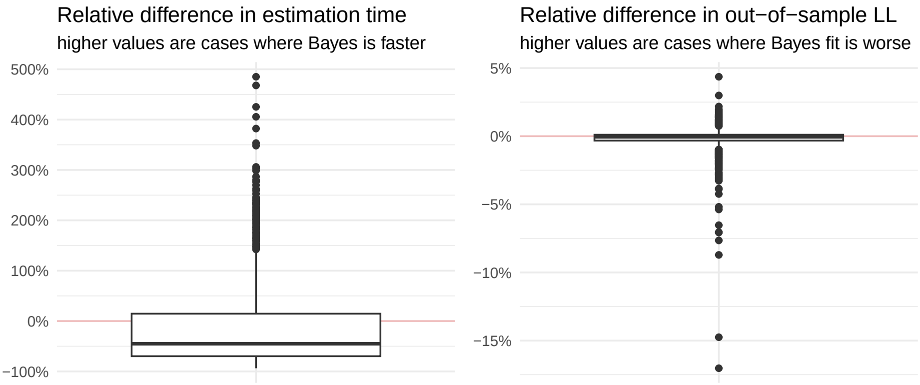


Comparison in different data scenarios

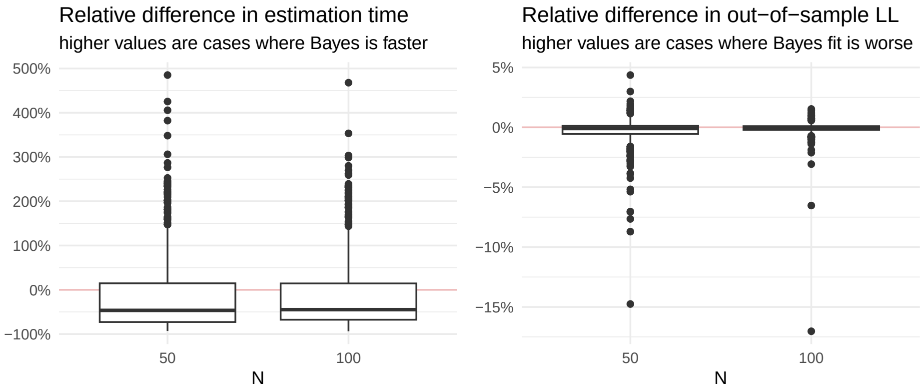




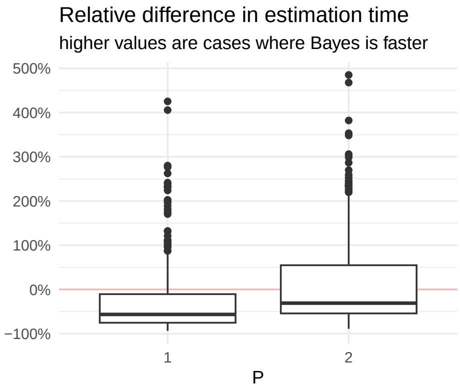
Comparison in different data scenarios

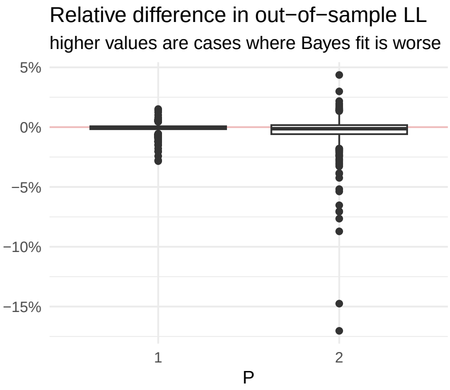


Comparison in different data scenarios

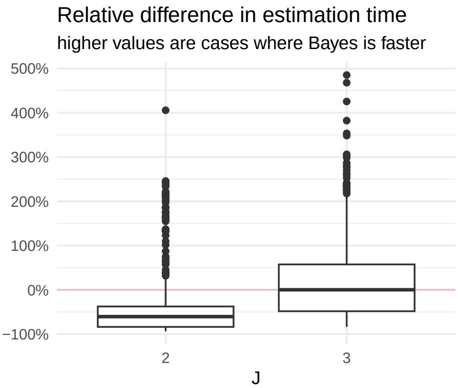


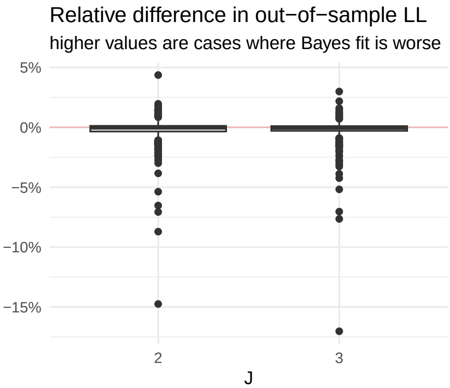
Comparison in different data scenarios



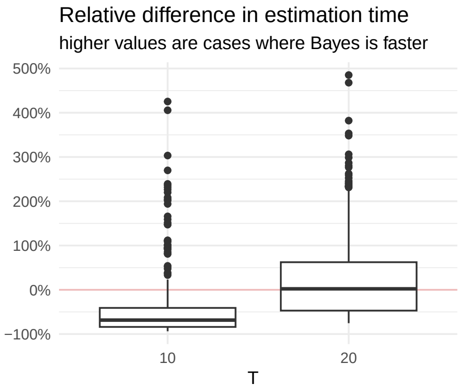


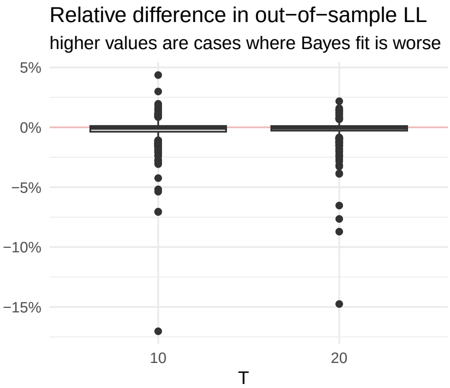
Comparison in different data scenarios





Comparison in different data scenarios





Outline

Takeaways and outlook

- Ignoring unobserved choice behavior heterogeneity decreases prediction power
- Modeling methods available with trade-off between flexibility and numerical feasibility
- In the parametric case, frequentist and Bayes estimation have similar out-of-sample prediction power but Bayes estimation becomes faster with rising J and T
- Next steps (ICMC in April):
  - also compare mixture models and non-parametric methods
  - apply the methods to the car purchase data set from the beginning

Takeaways and outlook

- Ignoring unobserved choice behavior heterogeneity decreases prediction power
- Modeling methods available with trade-off between flexibility and numerical feasibility
- In the parametric case, frequentist and Bayes estimation have similar out-of-sample prediction power but Bayes estimation becomes faster with rising J and T
- Next steps (ICMC in April):
  - also compare mixture models and non-parametric methods
  - apply the methods to the car purchase data set from the beginning

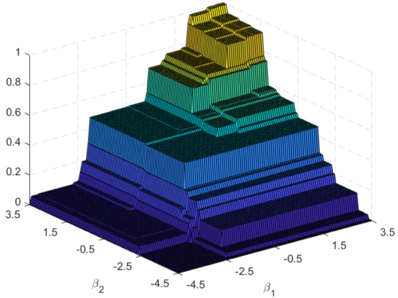
Thanks for your attention! Do you have any comments or questions for me? 🙂

References

- Florian Heiss, Stephan Hetzenecker, and Maximilian Osterhaus, *Nonparametric estimation of the random coefficients model: An elastic net approach*, Journal of Econometrics **229** (2022), no. 2, 299–321.
- John Paul Helveston, Yimin Liu, Elea McDonnell Feit, Erica Fuchs, Erica Klampfl, and Jeremy J. Michalek, *Will subsidies drive electric vehicle adoption? measuring consumer preferences in the u.s. and china*, Transportation Research Part A: Policy and Practice **73** (2015), 96–112.
- Lennart Oelschläger and Dietmar Bauer, *Bayesian probit models for preference classification: an analysis of chess players' propensity for risk-taking*, Proceedings of the 37th International Workshop on Statistical Modelling, TU Dortmund University, 2023.

Appendix: Non-parametric F

$$\widehat{F}(\beta) = \sum_{r=1}^{R} \hat{\theta}_r \, 1(\beta_r \leq \beta),$$ for example:



- 👍 / 👎 even more flexible
- 👎 even harder to estimate
- 👎 grid size R ?

Image source Heiss et al. (2022)

Appendix: Convergence of Gibbs sampler