UNIVERSITÄT BIELEFELD

Fakultät für Wirtschaftswissenschaften

# Estimating Multinomial Probit Models using `Rprobit`

R package developed at Bielefeld University:
Manuel Batram, Sebastian Büscher, Lennart Oelschläger


Dietmar Bauer

Bielefeld University


ZeSt Kolloquium, 4.6. 2024

# Introduction

# Multinomial Probit Models (MNP)

MNPs are used for discrete choice modeling:

- Deciders pick one out of a finite and exhaustive number of distinct alternatives.
- No ordering necessary: categorical data.
- Alternatives could be ordered: for example answer to survey questions (leads to ordered probit/logit).
- Typical setting: conjoint analysis.

# Example data set: Helveston et al. (2015)

Purchase decisions for automobiles within a conjoint study.

**SECTION 3**

**Suppose these 3 vehicles below were the only vehicles available for purchase, which would you choose?**

Each option will look like this:

| **Attribute*** | **Option 1** | **Option 2** | **Option 3** |
|---|---|---|---|
| **Vehicle Type** | Conventional<br>300 mile range on 1 tank | Plug-In Hybrid &<br>300 mile range on 1 tank (first 40 miles electric) | Electric<br>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) |
| | ○ | ○ | ○ |

*To view an attribute description, click on: ⓘ
**The average acceleration for cars in the U.S. is 0 to 60 mph in 7.4 seconds

Source: Helveston et al. (2015).

Each respondent sees 15 of these images and takes 15 decisions (balanced panel structure).

## Other example: mode choice

- For each trip a mode is chosen.
- Data typically collected in household surveys for one week (collecting information for each trip for each member of household).
- rolling panel design: each household is kept in the panel for three consecutive years.
- In Germany: Mopilitätspanel (MOP), collected by KIT Karlsruhe.
- Unbalanced panel data set: respondents have different number of trips.

## Third example: survey

- Within a survey each respondent is asked a number of questions.
- Oftentimes the questions require answers on a Likert scale: discrete choices.
- Contrary to the examples before the answers are *ordered*.
- Answers often are not uncorrelated, but dependent.
- Structural equations models are used in order to define the dependence structure.

## Random utility models (RUM)

MNPs often are interpreted as RUMs:

- Alternatives are characterized via regressors: $X_{i,t,j} \in \mathbb{R}^K$.
- Some regressors may characterize deciders: $S_{i,t} \in \mathbb{R}^d$.
- Each decider assigns a utility to each alternative:

$$U_{i,t,j} = V_{i,t,j} + e_{i,t,j}$$

- systematic part of the utility $V_{i,t,j}$ depends on the characteristics, often linear:

$$V_{i,t,j} = X_{i,t,j}\beta_i$$

- preferences are encoded using the parameter vectors $\beta_i$.
- Error terms and regressors are assumed to be independent of each other.
- Choice $y_{i,t}$: utility maximizing alternative.

In this talk regressors are used as row vectors and parameters are column vectors. Hence $X_{i,t,j}\beta_i$ is a scalar product of two vectors of same dimension.

## Interpretation of random utility model (RUM)

For the mode choice example:

- $X_{i,t,j}$: for example travel time [in min] and costs [in Euro].
- systematic part of the utility $V_{i,t,j}$:

$$\begin{aligned} V_{i,t,j} &= \beta_p \text{price}_{i,t,j} + \beta_{TT} \text{traveltime}_{i,t,j} + \ldots \\ &= \beta_p (\text{price}_{i,t,j} + \frac{\beta_{TT}}{\beta_p} \text{traveltime}_{i,t,j}) + \ldots \end{aligned}$$

- interpretation: tradeoffs: identical systematic part of utility for one minute more travel time, when price drops by $\frac{\beta_{TT}}{\beta_p}$ (VoT)
- The VoT may depend on income, for example, contained in $S_{i,t}$.

## Contents of the talk

Multinomial probit models (MNP) assume that $e_{i,t,:} \in \mathbb{R}^J \sim \mathcal{N}(0, \Sigma)$ and that choices occur independent of each other such that $e_{i,t,j}$ are independent identically distributed (iid) over $i$, $t$.

1. Specification of the model in `Rprobit`
2. Estimation methods: composite marginal likelihood (CML)
3. Modelling of preferences using random effects: mixed models

**REMARK:** In the talk I mark contributions from (some, not all in the area) master theses and dissertations (past and ongoing) in blue.

# MNP structure

## Structure of utilities

$$U_{i,t,j} = V_{i,t,j} + e_{i,t,j}, i = 1, ..., N, t = 1, ..., T_i, j = 1, ..., J_{i,t}.$$

- utilities are not measured, thus only differences are relevant: $\tilde{U}_{i,t,j} := U_{i,t,j} - U_{i,t,1}$ leads to same decisions $\Rightarrow$ normalization of the level required for identification.
- scale of utility does not matter for decisions: $\tilde{U}_{i,t,j} = cU_{i,t,j}, c > 0$, leads to same decisions $\Rightarrow$ normalization of scale necessary.
- `Rprobit` needs to provide flexibility for these normalisations.

## Regressors and model formulae

- Some regressors are constant in the sample: sex typically does not change. Others vary across alternatives $\Rightarrow$ different types of regressors.
- `Rprobit` uses three types of regressors: (taken from R package `mlogit`)

| | |
|---|---|
| type I | vary across alt., coefficients constant across alt.: e.g. price. |
| type II | regressors do not vary, coefficients vary across alt.: e.g. sex. |
| type III | regressors and coefficients vary across alt.: e.g. travel time. |

- Special case: $S_i = 1$: *alternative specific constants (ASCs)*.

Formula:

```
choice ~ type I | type II  | type III
```

## Regressors and model formulae

For now: $\beta_{i,t} = \beta$!

$$
\begin{aligned}
U_{i,t,j} &= X_{i,t,j}\beta_I + S_i\beta_{j,II} + Z_{i,t,j}\beta_{j,III} + e_{i,t,j}, \\
U_{i,t,j} - U_{i,t,1} &= X_{i,t,j}\beta_I + S_i\beta_{j,II} + Z_{i,t,j}\beta_{j,III} + e_{i,t,j} - X_{i,t,1}\beta_I + S_i\beta_{1,II} + Z_{i,t,1}\beta_{1,III} + e_{i,t,1} \\
&= (X_{i,t,j} - X_{i,t,1})\beta_I + S_i(\beta_{j,II} - \beta_{1,II}) + (Z_{i,t,j}\beta_{j,III} - Z_{i,t,1}\beta_{1,III}) + e_{i,t,j} - e_{i,t,1}.
\end{aligned}
$$

- Thus we have to fix one coefficient for one alternative for the type II variables to zero.
- This is then the reference alternative. Coefficients need to be interpreted with respect to this alternative.

# 2. MNP structure

One type II variable coded as many type I variables:

$$S_i \begin{pmatrix} 0 \\ \beta_{2,II} \\ \vdots \\ \vdots \\ \beta_{J,II} \end{pmatrix} = \begin{pmatrix} 0 \\ S_i\beta_{2,II} \\ \vdots \\ \vdots \\ S_i\beta_{J,II} \end{pmatrix} = \begin{pmatrix} 0 & \dots & \dots & 0 \\ S_i & \ddots & & \vdots \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & S_i \end{pmatrix} \begin{pmatrix} \beta_{2,II} \\ \vdots \\ \beta_{J,II} \end{pmatrix}.$$

- Analogously we can recode each type III variable into a number of type I variables.
- This makes the representation of the model easier, as one only needs to work with one type of regressors:

$$U_{i,t,j} = X_{i,t,j}\beta + e_{i,t,j}.$$

**Parameters to estimate:** $\beta, \Sigma$.

## Representation of Constraints

Different kind of restrictions for $\beta$ are possible / useful:

- setting entries of $\beta$ equal to zero: regressor selection.
- setting entries of $\beta$ to $\pm 1$: normalisation of scale: e.g. price -> utility in monetare units.
- setting differences of coefficients to zero: VoT identical for different travel time components.

All restrictions can be embedded into:

$$\beta = H_b\theta_b + f_b : \beta = \begin{pmatrix} 1 \\ \theta_1 \\ \theta_1 \\ 0 \\ \theta_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}}_{H_b} \underbrace{\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}}_{\theta_b} + \underbrace{\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}}_{f_b}.$$

## Constraints for variance matrices

- harder for variances: $\Sigma \geq 0, \Sigma = \Sigma'$.
- Use Cholesky factors: $\Sigma = L_S L_S'$, square root $L_S$ is lower triangular (automatically symmetric).
- Then we can parameterize the vectorization using affine mappings are above.

$$L_S = \begin{pmatrix} l_1 & 0 & 0 \\ l_2 & l_4 & 0 \\ l_3 & l_5 & l_6 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & l_4 & 0 \\ 0 & l_5 & l_6 \end{pmatrix}, \begin{pmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \\ l_6 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}}_{H_L} \underbrace{\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}}_{\theta_L} + \underbrace{\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}}_{f_L} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ \theta_1 \\ \theta_2 \end{pmatrix}$$

$$L_S = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \theta_1 & \theta_2 \end{pmatrix} \Sigma = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & \theta_1 \\ 0 & \theta_1 & \theta_1^2 + \theta_2^2 \end{pmatrix}.$$

## Random Effects Models

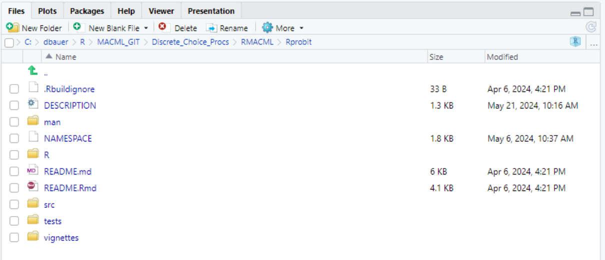
- Preferences $\beta_i$ may depend on characteristics $S_i$ : $\beta_i = S_i\delta$ leads to interactions effects
$$X_{i,t,j}\beta_i = X_{i,t,j}S_i\delta$$
- Often we did not measure the relevant variables.
- Then we may use concepts from panel data: *random effects / coefficients*.
- Assume that coefficient vectors $\beta_i$ are drawn randomly independent of everything observed:
$$\beta_i = \beta + \gamma_i \sim \mathcal{N}(\beta; \Omega)$$
- MNP with normally distributed random effects: Mixed MNP (MMNP).
- Leads to
$$U_{i,t,:} = X_{i,t,:}\beta + \underbrace{X_{i,t,:}\gamma_i + e_{i,t,:}}_{\check{e}_{i,t,:}|X_{i,t,:}\sim\mathcal{N}(0,\Sigma+X_{i,t,:}\Omega X'_{i,t,:})}$$
- Specification of $\Omega$: $\Omega = L_O L_O', \mathrm{vech}(L_O) = H_O\theta_O + f_O$.

**Three parts of parameter vector estimated:** $\theta_b, \theta_O, \theta_L$.

# Rprobit

# Availability

- `Rprobit` is available at https://github.com/dbauer72/Rprobit
- There exists a restricted version using Bayesian estimation for some of the models: `RprobitB` (on CRAN) Oelschläger and Bauer (2020).
- Much of it is still experimental and constantly under development.



# 3. Rprobit

Generate the model in `Rprobit` :

```
re = "opCost"
P_re <- length(re)

mod_hel <- mod_cl$new(
  alt = 3,
  Hb = make_Hb(P),
  fb = make_fb(P),
  HO = make_HO(P_re),
  fO = make_fO(P_re),
  HL   = matrix(0,6,0),
  fL   = matrix(0,6,1),
  ordered = FALSE
)

mod_hel$fL[1] = 1.264911 # sqrt(1.6)
mod_hel$fL[4] = 1.264911
mod_hel$fL[6] = 1.264911
```

# 3. Rprobit

`Rprobit_cl` object stores all information necessary for estimation:

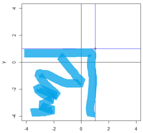
```
hel_obj <- setup_Rprobit(
  form = form,
  data_raw = data_wide,
  re = re,
  id = "id",
  mod = mod_hel
)
```

# Estimation

## Choice probabilities

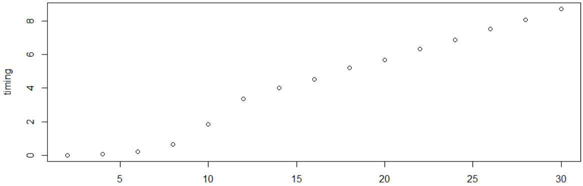
$$\begin{aligned}\mathbb{P}(y_{i,t} = j) &= \mathbb{P}(U_{i,t,k} \leq U_{i,t,j}, k \neq j) \\ &= \mathbb{P}(V_{i,t,k} + e_{i,t,k} \leq V_{i,t,j} + e_{i,t,j}, k \neq j) \\ &= \mathbb{P}(e_{i,t,k} - e_{i,t,j} \leq V_{i,t,j} - V_{i,t,k}, k \neq j)\end{aligned}$$

$J = 3$



- Since the vector $e_{i,t,:} - \iota_J e_{i,t,j} \sim \mathcal{N}(0, \tilde{\Sigma}_j)$ this requires the evaluation of a $J-1$ dimensional multivariate Gaussian CDF.
- In panels with random effects the MNP likelihood involves all choices of one individual: dimension $(J-1)T_i$

Evaluation of CDF at 100 points for different dimensions (in sec.): R fct. `pmvnorm`

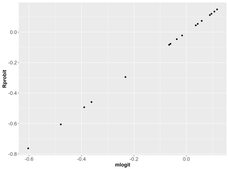


## Estimation using MaCML

Estimation of (M)MNP model for Helveston et al. (2015) US data set using the MaCML approach of Bhat (2011):

- $N = 384$ deciders, 3 alternatives, $T = 15$ choice occasions
- 5760 choices on total
- Estimation using `mlogit` (MSL with $R = 200$) without panel structure versus `Rprobit` without random coefficients and `Rprobit` with random coefficient for `opCost`.

| | without panel | | with panel | |
|---|---|---|---|---|
| | mlogit | MLE | FP | AP |
| opCost/price | 1.66 | 1.65 | 1.84 | 1.84 |
| accelTime/price | 1.80 | 1.80 | 1.78 | 1.78 |
| timing [sec] | 624 | 18 | +317 | +46 |
| in sample log-like | -4633.0 | -4613.1 | -4607.8 | -4607.1 |




How do we achieve this? Are there costs?

`mlogit` and `Rprobit` use slightly different model formulations. For the computation times this does not matter, for the likelihood it does.

## Maximum approximate Composite Marginal Likelihood (MaCML)

The approach is based on two ideas:

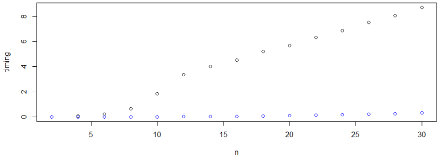
1. analytic approximation of the CDF
2. composite marginal likelihood replacing the full likelihood

**REMARK:**

- The standard approach to estimation is maximum simulated likelihood, wherein numerical approximations are used. Bhat and Sidhartan (2011) demonstrate in one simulation example, that MaCML is numerically faster and has comparable accuracy.
- Pohle (2016) provides additional evidence in favor of MaCML versus MSL for large number of alternatives.

## Approximation

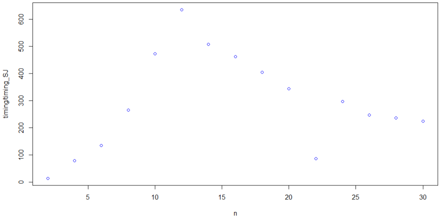
- **Common feature:** analytic approximation, not numerical. No 'bandwidth' parameter that trades off accuracy and numerical speed.
- Two interpretations:
  1. Approximation changes the criterion function
  2. Approximation changes the mapping from parameters to conditional choice probabilities (Batram and Bauer, 2019).
- Different approximation concepts: Solow-Joe (Joe, 1995), Mendel-Elston (1974), TVBS (Bhat, 2018). Rodenburger (2015)




Ratio time for 30-dim versus 4-dim: roughly 50 times.

## Approximation

- **Common feature:** analytic approximation, not numerical. No 'bandwidth' parameter that trades off accuracy and numerical speed.
- Two interpretations:
  1. Approximation changes the criterion function
  2. Approximation changes the mapping from parameters to conditional choice probabilities (Batram and Bauer, 2019).
- Different approximation concepts: Solow-Joe (Joe, 1995), Mendel-Elston (1974), TVBS (Bhat, 2018). Rodenburger (2015)




Ratio time for 30-dim versus 4-dim: roughly 50 times.

## Origin of CML ideas

- Concept aims at numerical speed in situation of large datasets.
- Due to Varin and Vidoni (2005) (see Varin (2008), Varin, Reid, Firth (2011)).
- Introduced to MMNP model estimation by Bhat (2011).

**Main idea**

Evaluation of likelihood is much faster, if we use only two choices of each decider.

- Considering only two choices per decider leads to a likelihood function that provides consistent estimates.
- **BUT:** This neglects information: sum over all pairs of choices includes all observations.
- We may also add (logs of) choice probabilities for each choice alone.

## Definition

- Change the criterion function from log-likelihood to pairwise log-CML:

$$
\begin{aligned}
ll(\theta; y, X, S) &= \sum_{i=1}^{N} lcml(\theta; y_i, X_i, S_i),\\
lcml(\theta; y_i, X_i, S_i) &= \sum_{1 \leq a < b \leq T_i} w_{t_a,t_b,i} \log \mathbb{P}(y_{i,t_a}, y_{i,t_b} | X_{i,t_a}, X_{i,t_b}, S_i; \theta)\\
&\quad + \sum_{1 \leq a \leq T_i} v_{t_a,i} \log \mathbb{P}(y_{i,t_a} | X_{i,t_a} S_i; \theta).
\end{aligned}
$$

$w_{t_a,t_b,i} \geq 0$, $v_{t_a,i} \geq 0$... power weights.

## Numerical Optimization

- We started in R, moved to `MATLAB` (compiling code is much faster). Then migrated back to R.
- R provides an interface to C++ called `Rcpp` that speeds up calculations a lot.
- Parallelization can help in addition.
- Estimation uses numerical optimization. This requires initialisation of the search algorithm: Different approximation concepts show different sensitivity to initialisation, so it is important (Wüllner, 2019).
- Clever ideas can provide speed up and increase the chance to get to the global optimum compared to random initialisation (Oelschläger, 2024, Diss).

```
hel_obj$control$probit <- FALSE
theta_0 <- c(hel_fit$theta, rep(0.05, P_re))
hel_obj$set_theta(theta_0)

hel_fit0 <- fit_Rprobit(hel_obj, cml_pair_type = 0)
```

## Power weights choice

- Power weights can be used to tune the numerical load and have impact on accuracy.
- $w_{t_a,t_b,i} = 0$ or $v_{t_a,i} = 0$ implies: Do not need to calculate this term.
- Often used:
  - full pairwise $w_{t_a,t_b,i} = 1, v_{t_a,i} = 0$ (FP ).
  - adjacent pairwise: $w_{t_a,t_b,i} = \mathbb{I}(b = a + 1), v_{t_a,i} = 0$ (AP ).
  - adjacent pairwise plus marginals: $v_{t_a,i} = 1$ (APU )
- While speedup is clear, impact on accuracy is less clear a-priori:
  - Under which assumptions on the power weights do we get consistency?
  - How to get an estimate for the accuracy?
  - How large is the difference to MLE?
  - Which situations are bad for CML? Which are good?
  - What about the various variants?
  - How shall we choose the power weights?

```
hel_fit1 <- fit_Rprobit(hel_obj, cml_pair_type = 1)
```

## Results

- Identifiability holds under mild assumptions on regressors (Batram, 2015).
- Consistency and asymptotic normality for the MaCML estimators can be shown under standard conditions (Batram and Bauer, 2019)
- The CML criterion function leads to *sandwich form* for asymptotic variance. Various consistent estimators thereof can be derived, that use the structure of the problem cleverly making the calculation of the Hessian unnecessary (which is good for numerical speed; see, e.g., Büscher and Bauer, 2024).
- FP shows only a moderate loss of information in some situations, while AP and APU are considerably less accurate. Large $T$ makes the difference to full probit substantial in all cases (Bauer, 2022, ICMC).
- Power weights can be used to tune performance for unbalanced panels in a two stage approach similar to FGLS estimation (Büscher and Bauer, 2024)

# Latent Class (LC) Models

## Random Effects Models

MNP with random effects: Mixed MNP (MMNP).

$$\beta_i \sim F(.;\theta)$$

- *parametric*: multivariate Gaussian, often independent over parameters (diagonal variance matrix).
- extension: *latent class* models using, e.g., Gaussian mixtures for the mixing.
- *non-parametric* models, estimated using, e.g., penalization, see e.g. Heiss, Hetzenecker, Osterhaus (2021) or Train (2016).

**Identification:** holds, if usual restrictions to identify level and scale are used and if regressors are rich enough. *Details:*

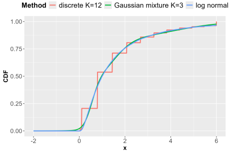
# 5. 1. Latent class models

## Definition

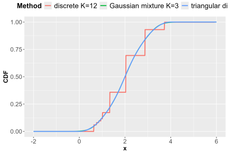
$$\beta_i \sim \sum_{c=1}^{C} \pi_c F(., \theta_c) \quad \text{(CDF)}$$

- We assume that deciders can be grouped into a finite number $C$ of 'classes'
- Within each class the same model applies
  - point masses $F(., \theta_c)$ corresponds to deterministic value $\beta_c$
  - mixing density, for example Gaussian: $F(., \theta_c) = \mathcal{N}(\beta_c, \Omega_c)$
- The model, thus, may be using a point mass $\beta_c$ or allow for mixing also within each class
- class membership is latent (not observed, not modelled as depending on observables)

Approximation of Log-normal by sum of point masses and Gaussian mixture.




Same for triangular distribution




This model is fully flexible, if a large number of classes are used: approximation of general CDF using point masses or via mixture of Gaussians, e.g.

# 5. 1. Latent class models

Specification of models in `Rprobit` :

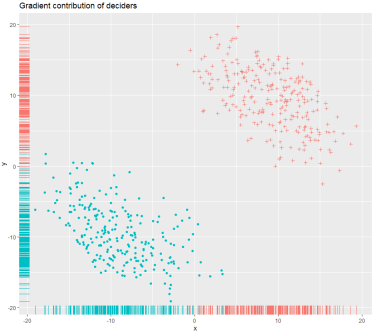
```
mode_LC <- mod_latclass_cl$new(
  Hb    = matrix(0,16,1),
  fb    = matrix(0,16,1),
  HO    = matrix(1,1,0),
  fO    = matrix(1,1,1)*0.01, # fix std of random effects
  HL    = matrix(0,6,0),
  fL    = matrix(0,6,1),
  ordered = FALSE
)

[...]
mode_LC$Hb[1,1] <- 1 # H_b contains only one column, only first
     entry is estimated.
mode_LC$fb <- as.matrix(hel_fit0$theta[1:16],ncol=1) # f_b fixes the
      values of all other coefficients.
mode_LC$fb[1] <- 0

mode_LC$num_class <- 3 # this is the essential line: 3 classes.
```

# Frequentist estimation: initialization

- Estimation would be simple, if class membership was known.
- Gradient of restricted model with one class provides information on class membership (Lampe (2023), Büscher (2024, Diss)):

**Extreme example:**

- Two class MNP model, 5 choices from 3 alternatives for 250 deciders each.
- Two standard normal regressors with coefficients $\beta = \pm(2, 2)$.
- Gradient contributions of each decider for a joint one class MNP model estimated using MaCML



# Current topics

## Non-parametric mixing distribution

If one wants to avoid to impose much structure:

$$\beta_i \sim \sum_{g=1}^{G} \pi_g(\theta) F(., \beta_g) \quad \text{(CDF)}$$

for fixed $\beta_g, g = 1, ..., G$.

- approximation as point masses on grid points (not estimated but prespecified)
- only operational for small dimensions. Already $d = 3$ is a challenge.

Middle ground: splines in order to smooth the PMF (Train, 2016):

- reduces the number of parameters;
- works in slightly higher dimensions;
- again initialization is an issue, many tuning parameters.

Bauer and Oelschläger, 2024, ICMC

## Temporal Dependence

- Most often the error term $e_{i,t,j}$ is always assumed to be independent over choice occasions.
- This may not always hold: habit formation/feedback versus variety seeking.
- Lagrange-multiplier type tests can be used to detect deviations from uncorrelatedness (Büscher, 2024, Diss)
- Conceptionally including autoregressive models for time dependency is straightforward, but number of parameters grows rapidly -> future research.

## Ordered responses

- For ordered MNP the answers can be ordered such for survey questions on a Likert scale.
- Then there is only one utility, the choice is provided by intervals.
- Again treating a number of questions jointly high dimensional cdfs may occur:
  - one question: interval: 2 points
  - two questions: rectangle: 4 points
  - three questions: box: 8 points.
  - $K$ questions: $2^K$ points.

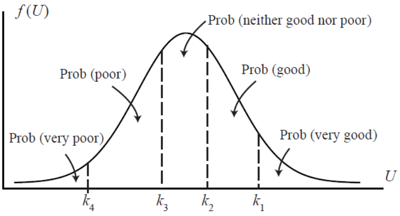



Figure 7.1: Distribution of opinion about president's job.

# Summary

# 7. Summary

- Writing a package requires many choices with respect to features and defaults.
- These choices should be taken based on experience from theory and simulations.
- Typically these involve 'case studies' examination of examples.
- Using these we gain knowledge, experience.
- Each step leads to additional functionalities, new challenges. Every door we open leads to several new doors.
- The work never ends ...

# 7. Summary

Thank you for your attention!

Questions? Dietmar.Bauer@uni-bielefeld.de

# 8. Literature


- Batram, M. and Bauer, D. (2019). On consistency of the MaCML approach to discrete choice modelling, *Journal of Choice Modelling* **30** 1-16.
- Bhat, C. R. (2011). The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models, *Transportation Research Part B* **45**(7): 923-939
- Bhat, C. R. (2018). New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function, *Transportation Research Part B: Methodological* **109**, 238-256
- Bhat, C. R., Sidharthan, R. (2011). A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models. *Transportation Research Part B: Methodological*, **45(7)**, 940-953
- Büscher, S., Bauer, D. (2024). Weighting strategies for pairwise composite marginal likelihood estimation in case of unbalanced panels and unaccounted autoregressive structure of the errors. *Transportation Research Part B: Methodological* **181**, 102890.
- Fox, J. (2017) A Note on Nonparametric Identification of Distributions of Random Coefficients in Multinomial Choice Models. Technical report, National Bureau of Economic Research, Cambridge, MA.
- Fox, J., il Kim, K., Ryan, St., and Bajari, P. (2012). The random coefficients logit model is identified. *Journal of Econometrics*, **166**(2):204-212.
- Grün, B. and Leisch, F. (2008). Identifiability of finite mixtures of multinomial logit models with varying and fixed effects. *Journal of Classification*, **25**(2):225-247.
- Heiss, F., Hetzenecker, S., Osterhaus, M. (2022). Nonparametric estimation of the random coefficients model: An elastic net approach. *Journal of Econometrics*, **229**(2), 299-321.
- Helveston, J. P., Liu, Y., Feit, E. M., Fuchs, E., Klampfl, E., Michalek, J. J. (2015). Will subsidies drive electric vehicle adoption? Measuring consumer preferences in the US and China. *Transportation Research Part A: Policy and Practice*, **73**, 96-112
- Joe, H. (1995). Approximations to multivariate normal rectangle probabilities based on conditional expectations, *Journal of the American Statistical Association* **90**(431): 957-964
- Mendell, N. and Elston, R. (1974). Multifactorial qualitative traits: Genetic analysis and prediction of recurrence risks, *Biometrics* **30**, 41-57
- Oelschläger L., Bauer D. (2021). Bayes Estimation of Latent Class Mixed Multinomial Probit Models. Online presentation of poster at the TRB Annual Meeting Jan. 2021, Washington.
- Train, K. (2009) Discrete choice methods with simulation. Cambridge university press.
- Train, K. (2016) Mixed logit with a flexible mixing distribution. *Journal of Choice Modelling*, 19:40-53.
- Varin, C. (2008). On composite marginal likelihoods, *Advances in Statistical Analysis* **92**, 1-28.
- Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods, *Statistica Sinica* **21**, 5-42.
- Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection, *Biometrika* **92**(3): 519-528.

## Identification

- Level: only utility differences are identified, hence the level needs to be specified. $U_{i,t,1} = 0$ works.
- Scale: $cU_{i,t,j}$ leads to the same choices for $c > 0$. Often the variance is fixed to identify scale.

**Mixing**:

- Parametric mixing for MNL models is identified for different mixing distributions (Train, 2009) and also for non-parametric mixing (Train, 2016).
- Fox (2017) in the general case and Fox et al. (2012) for MNL models provide assumptions such that the mixing distribution is (non-parametrically) identified from cross-sectional data: at least one regressor needs to be supported in $\mathbb{R}$ where the coefficient is supported in subset of $\mathbb{R}_+$ (as would be typical for costs).
- Grün and Leisch (2008) show that in order to identify $S$ latent classes in mixed MNL models the regressors need to take on at least $2S - 1$ different values.