“Mixed logit with a flexible mixing distribution” Kenneth Train

The journal of transport modelling, 2016

Laureen ATTOLOU
BiN paper reading seminar
May 31st 2024

Abstract

This paper focuses on the mixing distribution of the mixed logit, aiming to bring:

- More flexibility
- Easier definition
- Less limitation in terms of properties
- Feasibility from a computational perspective

How: Logit distribution for the mixing distribution

→ possibility to specify any useful function for the distribution shape

→ good properties (summation to 1, positivity, easy sampling)

→ easy to program and fast computationally

- Approximate generalization of previous studies by by Bajari et al. (2007), Fosgerau and Bierlaire (2007), Train (2008), Fox et al. (2011), Burda et al. (2008) and Fosgerau and Mabit (2013)
- **Higher goal:** shift researcher’s focus from distributional constraints to dataset refining

Contents

Introduction: mixed logit model

Standard logit model:

We consider a decision maker n, in a choice situation with J alternatives j:

$$U_{nj} = V_{nj} + \varepsilon_{nj} \; \forall j$$

Observed part of the utility
Known to the researcher

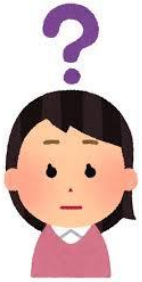
linear in explainable variables

Random part of the utility
Unknown to the researcher

iid extreme value

P(decision maker n chooses option i) $P_{ni} = \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \; \forall j \neq i)$

$$\frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \quad \text{if} \quad V_{nj} = \beta' x_{nj}$$



Unobserved information | Random part of utility

Observed information → Explainable part of utility

Assumptions and limitations:

- Systematic taste variation (if all observed characteristics are the same, the choice will stay the same)
- Proportional substitution across alternatives (ex. price↘ = probability ↗)
- No correlation for unobserved factors over time (ex. for repeated choices)

Introduction: mixed logit model

Standard logit model:

We consider a decision maker n, in a choice situation with J alternatives j:

$$U_{nj} = V_{nj} + \varepsilon_{nj} \, \forall j$$

Observed part of the utility
Known to the researcher

linear in explainable variables

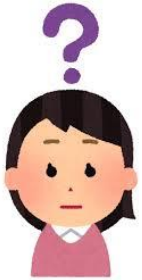
Random part of the utility
Unknown to the researcher

iid extreme value

P(decision maker n chooses option i) = $P_{ni} = \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \, \forall j \neq i)$

$$\frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \quad \text{if} \quad V_{nj} = \beta' x_{nj}$$

Is assumed the same for everyone



Unobserved information

Random part of utility

Observed information → Explainable part of utility

Assumptions and limitations:

- Systematic taste variation (if all observed characteristics are the same, the choice will stay the same)
- Proportional substitution across alternatives (ex. price↘ = probability ↗)
- No correlation for unobserved factors over time (ex. for repeated choices)

Introduction: mixed logit model

~~Standard~~ logit model:

We consider a decision maker n, in a choice situation with J alternatives j:

$$U_{nj} = V_{nj} + \varepsilon_{nj} \; \forall j$$

Observed part of the utility
Known to the researcher

linear in explainable variables

Random part of the utility
Unknown to the researcher

iid extreme value

P(decision maker n chooses option i) = $P_{ni} = \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \; \forall j \neq i)$

$$P_{n_i}(\beta_n) \quad \frac{e^{\beta_n' x_{ni}}}{\sum_j e^{\beta_n' x_{nj}}} \quad \text{if} \quad V_{nj} = \beta_n' x_{nj}$$

Is assumed different for everyone
→ need to define distribution for β: $\beta_n \sim f(\beta|\theta)$

Usually, $\beta_n \sim N(b, w)$ (normal distribution)
(or lognormal, uniform, triangular,..)

統計学では、複数の関数における重み付き平均は混合関数(mixing function)と呼ぶ

$$P_{n_i} = \int P_{ni}(\beta_n) f(\beta)\, d\beta$$

"Mixed function" means weighted average of several functions
f(β) is the mixing distribution, which gives the weights

Introduction: mixed logit model

Mixed logit model: (random coefficient specification -ランダム係数ミックスドロジット)

We consider a decision maker n, in a choice situation with J alternatives j:

$$U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj}$$

Observed part of the utility
Known to the researcher

Random part of the utility
Unknown to the researcher

β :Random with distribution f
ε : iid extreme value

x_{nj} は観測変数のベクトルである
β_n は個人nにおける係数のベクトルである
ε_{nj} はi.i.d. 極値分布に従う誤差項である

P(decision maker n chooses option i) = $P_{ni} = \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \ \forall j \neq i)$

$$P_{n_i}(\beta_n) \quad \frac{e^{\beta_n' x_{ni}}}{\sum_j e^{\beta_n' x_{nj}}} \quad \text{if} \quad V_{nj} = \beta_n' x_{nj}$$

Is assumed different for everyone
→ need to define distribution for β: $\beta_n \sim f(\beta|\theta)$

$$P_{n_i} = \int P_{ni}(\beta_n) f(\beta)\, d\beta$$

Usually, $\beta_n \sim N(b, w)$ (normal distribution)
(or lognormal, uniform, triangular,..)

Introduction: mixed logit model

Mixed logit model: (random coefficient specification)

We consider a decision maker n, in a choice situation with J alternatives j:

$$U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj}$$

Observed part of the utility
Known to the researcher

Random part of the utility
Unknown to the researcher

β :Random with distribution f
ε : iid extreme value

P(decision maker n chooses option i) = $P_{ni} = \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \; \forall j \neq i)$

$$\boldsymbol{P_{n_i}(\beta_n)} = \frac{e^{\beta_n' x_{ni}}}{\sum_j e^{\beta_n' x_{nj}}} \quad \text{if} \quad V_{nj} = \beta_n' x_{nj}$$

Is assumed different for everyone
→ need to define distribution for β:

Usually, $\beta_n \sim N(b, w)$ (normal distribution)
(or lognormal, uniform, triangular,..)

$$\boldsymbol{P_{n_i} = \int P_{ni}(\beta_n) f(\beta)\, d\beta}$$

Simulation:

1. Draw a value of β, labelled β^r (sampling)
2. Calculate $P_{n_i}(\beta^r)$
3. Repeat 1 and 2 R times
4. Average the results to get $\check{P}_{ni} = \frac{1}{R}\sum_{r=1}^{R} L_{ni}(\beta^r)$, then maximize the simulated log-likelihood: $\text{SLL} = \sum_{n=1}^{N}\sum_{j=1}^{J} d_{nj} \ln \check{P}_{nj}$

Introduction: mixed logit model

Mixed logit model: (random coefficient specification)

We consider a decision maker n, in a choice situation with J alternatives j:

$$U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj}$$

Observed part of the utility
Known to the researcher

Random part of the utility
Unknown to the researcher

β :Random with distribution f
ε : iid extreme value

P(decision maker n chooses option i) = $P_{ni} = \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \ \forall j \neq i)$

$$P_{n_i}(\boldsymbol{\beta_n}) = \frac{e^{\beta_n' x_{ni}}}{\sum_j e^{\beta_n' x_{nj}}} \quad \text{if} \quad V_{nj} = \beta_n' x_{nj}$$

Is assumed different for everyone
→ need to define distribution for β:

Usually, $\beta_n \sim N(b, w)$ (normal distribution)
(or lognormal, uniform, triangular,..)

$$\boldsymbol{P_{n_i}} = \int \boldsymbol{P_{ni}(\beta_n) f(\beta)\, d\beta}$$

Flexibility and advantages

- Random taste variation,
- Unrestricted substitution patterns
- Correlation in unobserved factors over time
- Can approximate any random utility model (see McFadden and Train (2000), or Train textbook)

Limitations of mixed logit

- For the researcher, using the mixed logit is a two-part process
  - Specification of the logit (parameters used in the utility function)
  - Specification of the mixing distribution f(β)
    - Usually normal or lognormal
    - Johnson's Sb
    - Gamma
    - Triangular

  - Most distributions are limiting: "most researchers will probably agree that: whatever parametric distribution the researcher specifies, he/she quickly becomes dissatisfied with its properties"

A Logit-Mixed-Logit model（LML）

- **Situation**: 1 decision-maker n, faced with one choice

Utility: $U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj}$

Prob(n chooses i |$\boldsymbol{\beta_n}$) :

$$Q_{ni}(\beta_n) = \frac{e^{\beta_n' x_{ni}}}{\sum_{j \in J} e^{\beta_n' x_{nj}}}$$

- x: Vector of observed attributes
- β: Vector of utility coefficient, varies randomly over people
- ε: Random term representing the unobserved component of utility

x_{nj} は観測変数のベクトルである

β_n は個人nにおける係数のベクトルである

ε_{nj} はi.i.d. 極値分布に従う誤差項である

- Mixing distribution F(β):
  - Discrete with a finite support set S (WLOG as long as S is dense enough)
  - F is a logit distribution, i.e:

$$\text{Prob}(\beta_n = \beta_r) \equiv W(\beta_r|\alpha) = \frac{e^{\alpha' z(\beta_r)}}{\sum_{s \in S} e^{\alpha' z(\beta_s)}}$$

- $z(\beta_r)$: vector function of β, chosen to fit a certain shape
- α: vector of coefficients

- Then the choice probability is:

$$\text{Prob}(n \text{ chooses } i) = \sum_{r \in S} W(\beta_r|\alpha) \cdot Q_{ni}(\beta_r) = \sum_{r \in S} \left(\frac{e^{\alpha' z(\beta_r)}}{\sum_{s \in S} e^{\alpha' z(\beta_s)}} \right) \cdot \left(\frac{e^{\beta_r' x_{ni}}}{\sum_{j \in J} e^{\beta_r' x_{nj}}} \right)$$

To be defined by the researcher

A Logit-Mixed-Logit model (LML)

- **Situation**: 1 decision-maker n, faced with one choice

Utility: $U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj}$

Prob(n chooses i $|\boldsymbol{\beta_n}$) $Q_{ni}(\beta_n) = \frac{e^{\beta_n' x_{ni}}}{\sum_{j \in J} e^{\beta_n' x_{nj}}}$

- x: Vector of observed attributes
- β: Vector of utility coefficient, varies randomly over people
- ε: Random term representing the unobserved component of utility

- Mixing distribution F(β):
  - Discrete with a finite support set S (WLOG as long as S is dense enough)
  - F is a logit distribution, i.e:

$$\text{Prob}(\beta_n = \beta_r) \equiv W(\beta_r|\alpha) = \frac{e^{\alpha' z(\beta_r)}}{\sum_{s \in S} e^{\alpha' z(\beta_s)}}$$

- $z(\beta_r)$: vector function of β, chosen to fit a certain shape
- α: vector of coefficients

- Then the choice probability is:

$$\text{Prob}(n \text{ chooses } i) = \sum_{r \in S} W(\beta_r|\alpha) \cdot Q_{ni}(\beta_r) = \sum_{r \in S} \left(\frac{e^{\alpha' z(\beta_r)}}{\sum_{s \in S} e^{\alpha' z(\beta_s)}} \right) \cdot \left(\frac{e^{\beta_r' x_{ni}}}{\sum_{j \in J} e^{\beta_r' x_{nj}}} \right)$$

Properties:

- Easy and flexible specification of probabilities
- Only need to describe the shape of the distribution as summation to one, positivity are already assured
- Can also approximate any choice model

Estimating LML model

- Situation: Multiple choices by each decision maker
  - t: choice situation
  - j,i: alternative
  - n: choice maker
- We consider T choice situations, and the probability of choice sequence $(i_1, i_2, \dots, i_T)$

Conditional probability:

$$P_n(\beta_n) = \prod_{t=1,\dots,T} Q_{n i_t t}(\beta_n)$$

Unconditional probability:

$$P_n = \sum_{r \epsilon S} P_{nr}(\beta_r) W(\beta_r|\alpha)$$

to be estimated

Does not need to be recalculated at each iteration

Log-likelihood $$LL = \sum_{n=1,\dots,N} \ln(P_n) = \sum_{n=1,\dots,N} \ln(\sum_{r \epsilon S} P_{nr}(\beta_r) W(\beta_r|\alpha))$$

Simulated log-likelihood Using a subset S_n

$$SLL = \sum_{n=1,\dots,N} \ln(\sum_{r \epsilon S_n} P_{nr}(\beta_r) w_n(\beta_r|\alpha))$$

Variables for the mixing distribution

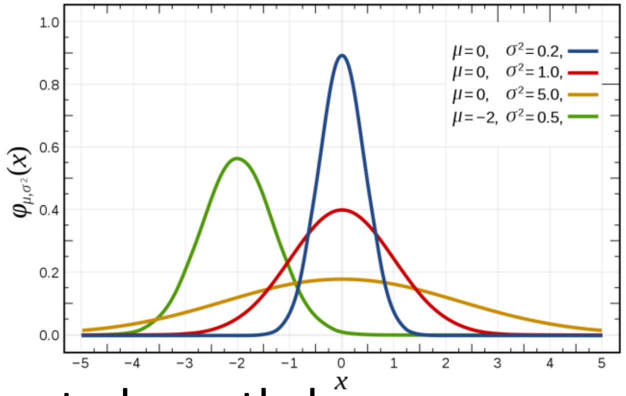
how to specify the z variables ?

$$\text{Prob}(\beta_n = \beta_r) \equiv W(\beta_r|\alpha) = \frac{e^{\alpha' z(\beta_r)}}{\sum_{s\in S} e^{\alpha' z(\beta_s)}}$$

Normal

- **Why**: To try the LML on a simple example? To eliminate long tails of the usual normal and lognormal
- **What**: Normal distribution (正規分布) of mean and variance V
- **How**:

$$f(\beta) = m(V)\exp\left(-\frac{1}{V}\left(\frac{\beta'\beta}{2} - b'\beta + \frac{b'b}{2}\right)\right)$$



Z is a second order polynomial (多項式) in β and this density can be represented exactly by

$$\text{Prob}(\beta_n = \beta_r) \equiv W(\beta_r|\alpha) = \frac{e^{\alpha' z(\beta_r)}}{\sum_{s\in S} e^{\alpha' z(\beta_s)}}$$

Higher order polynomials

- **Why**: Greater flexibility, reduce collinearity, describe a wide variety of shapes
- **What**: Specify z to be a higher order polynomial (多項式) in β
- **How**: (case of 1-dimensional β)

Legendre polynomials: family of polynomials $L_1, \ldots, L_n$ such that $\int_1^1 L_m(x)L_n(x)\,\mathrm{d}x = 0 \quad \text{if } n \neq m$ and $L_n(1) = 1$

The Legendre polynomials are only defined on [-1,..,1], so we use the transformation $\tilde{\beta} = -1 + 2\frac{\beta - \min_S(\beta)}{\max_S(\beta) - \min_S(\beta)}$

We specify the z variables as $z_k(\beta) = L_k(\tilde{\beta})$ for k = 1,…,K (K is the highest degree specified by the researcher)

Therefore,

$$\text{Prob}(\beta_n = \beta_r) \equiv W(\beta_r|\alpha) = \frac{e^{\alpha' z(\beta_r)}}{\sum_{s \in S} e^{\alpha' z(\beta_s)}}$$

With $e^{\alpha' z(\beta_r)} = e^{\alpha'(L_1(\tilde{\beta}r) + L_2(\tilde{\beta}_r) + \ldots + L_k(\tilde{\beta}_r))}$

For multi-dimensional β, dependence among the elements of is captured though cross-products of the terms of each element's polynomial.

We could also use another polynomial family such as Chebyshev, Bernstein,…

Step functions

- **Why**: Estimate the model over different parts of the set S
- **What**: step functions
- **How**: (example with 2 dimensions)

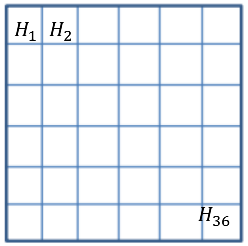
  We partition (分割) S into G subsets (部分集合) (possibly overlapping) $H_1, \dots, H_G$ and define the mixing distribution for each subset

  The z variables are the G indicators (指示関数) of which subset contains β_r: $z(\beta_r) = \mathbf{1}_{H_g}(\beta_r) = \begin{cases} 1 \; if \, \beta_r \in H_g \\ 0 \; if \, \beta_r \notin H_g \end{cases}$

  Therefore, $\text{Prob}(\beta_n = \beta_r) \equiv W(\beta_r|\alpha) = \frac{e^{\alpha' z(\beta_r)}}{\sum_{s \in S} e^{\alpha' z(\beta_s)}}$ With $e^{\alpha' z(\beta_r)} = e^{\alpha'(\mathbf{1}_{H_1}(\beta_r), \dots, \mathbf{1}_{H_G}(\beta_r))}$

❖ Choosing the number of subsets (which is also related to the number of parameters):

Saturated specification:



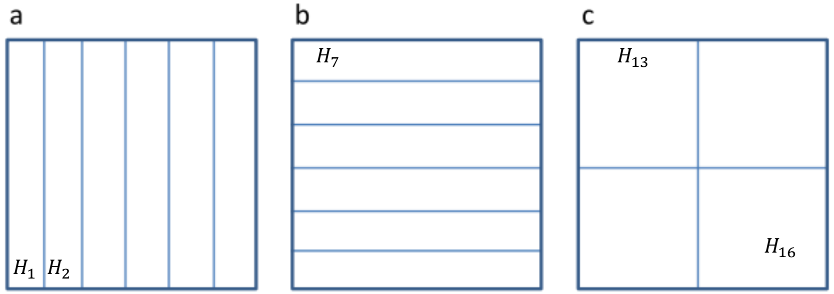


Fig. 1. Overlapping step functions.

Splines

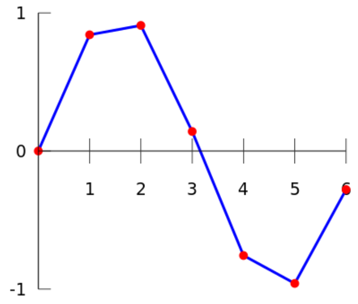
- **Why**: Interpolation of a set of points (線形補間)
- **What**: Specify z so that the mixing distribution has certain values at certain points
- **How**: Splines – function defined piecewise by polynomials

  (example with 4 points and 1-dimensional β)

Let's say you know the values that f(β) takes at points $\overline{\beta_1}, \overline{\beta_2}, \overline{\beta_3}, \bar{\beta}_4$, ie $f\left(\bar{\beta}_1\right) = \alpha_1, \ldots, f\left(\bar{\beta}_4\right) = \alpha_4$

Therefore,

$$f(\beta) = \left\{ \begin{array}{ll} \alpha_1 + \frac{\alpha_2 - \alpha_1}{\bar{\beta}_2 - \bar{\beta}_1}(\beta - \bar{\beta}_1) & \text{if } \beta \leq \bar{\beta}_2 \\ \alpha_2 + \frac{\alpha_3 - \alpha_2}{\bar{\beta}_3 - \bar{\beta}_2}(\beta - \bar{\beta}_2) & \text{if } \bar{\beta}_2 < \beta \leq \bar{\beta}_3 \\ \alpha_3 + \frac{\alpha_4 - \alpha_3}{\bar{\beta}_4 - \bar{\beta}_3}(\beta - \bar{\beta}_3) & \text{if } \bar{\beta}_3 < \beta \end{array} \right\} = \alpha' z(\beta)$$



Then the mixing distribution is defined using the interpolation parameters

Variables for the mixing distribution

Combination

To take advantage of the properties of each previous described variable setting, it is possible to combine different functions:

- Step-function or spline for each single coefficient,
- Second-order polynomial to capture correlation over coefficients
- Fewer parameters for correlations than creating multi-dimensional step-functions or splines.

Equivalence to the method of sieves:

In the method of sieves, the estimation is performed by dividing the range of the function into a sequence of intervals, called "sieves" or "bins." Within each sieve, a simpler parametric model is assumed to approximate the true function.

Here, the use of polynomials and slives in the LML can be seen as a sieve method, with the number of α parameters (i.e., the order of the polynomial and/or the number of steps/nodes) rising with sample size, providing more flexibility in fitting the true distribution.

WTP (Willingness-to-pay) space where price variables should be estimated

- **Willingness to pay (支払意思額)**: amount of money or resources that an individual or group of individuals is willing to increase the quantity of an attribute, usually given by

$$\frac{attribute\ coefficient}{price\ coefficient}$$

- WTP is usually overestimated in the preference space ; the better the fit, the less reasonable wtp
- Utility in WTP space and preference space:

Preference space:

(we divide the utility by the scale parameter and separate price and non-price attributes)

Scale parameter, fixed

$$U_{njt} = -(\alpha_n/k_n)p_{njt} + (\beta_n/k_n)' x_{njt} + \varepsilon_{njt}$$

$$U_{njt} = -\lambda_n p_{njt} + c_n' x_{njt} + \varepsilon_{njt}$$

$$w_n = c_n/\lambda_n$$

WTP space:

$$U_{njt} = -\lambda_n p_{njt} + (\lambda_n w_n)' x_{njt} + \varepsilon_{njt}$$

See Train and Weeks (2004)

LML in WTP space

Preference space:

Utility: $U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj}$

$$\text{Prob}(\beta_n = \beta_r) \equiv W(\beta_r|\alpha) = \frac{e^{\alpha' z(\beta_r)}}{\sum_{s \in S} e^{\alpha' z(\beta_s)}}$$

$w_n = c_n / \lambda_n$

WTP space:

$U_{nj} = -\lambda_n p_{nj} + (\lambda_n w_n)' x_{nj} + \varepsilon_{nj}$

Let X, B as:
$B_n' = [-\lambda_n, \lambda_n \omega_n]$ and $x_{nj} = [p_{nj}, x_{nj}]$

Then, $\boldsymbol{U_{n_j} = B_n' X_{n_j} + \varepsilon_{n_j}}$
and the LML can be used

LML with unequal probability sampling

- The log-likelihood was previously defined with equal probability of a sample β to be selected
- In the case of unequal probability sampling,

$$\mathrm{LL} = \sum_{n=1,\ldots N} \ln\left(\sum_{r \in S} (L_n(\beta_r)/q(\beta_r))W(\beta_r|\alpha)q(\beta_r) \right)$$

q(β): a probability mass function

$$\mathrm{SLL} = \sum_{n} \ln\left(\sum_{r \in S_n} (L_n(\beta_r)/q(\beta_r))w_n(\beta_r|\alpha) \right)$$

Application

Application

- **Dataset**: experiment about consumer's choice among video streaming services
  - Price and non-price attributes
  - 4 alternative video services + no service alternative
  - 11 choice situations
  - 260 respondents

Table 1
Non-price attributes.

Attribute	Levels
Commercials between content	Yes ("commercials") No (baseline category)
Speed of content availability	TV episodes next day, movies in 3 months ("fast content") TV episodes in 3 months, movies in 6 months (baseline)
Catalog	5000 movies and 2500 TV episodes (baseline) 10,000 movies and 5000 TV episodes ("more content") 2000 movies and 13,000 TV episodes ("more TV, fewer movies")
Data-sharing policies	Information is collected but not shared (baseline) Usage info is shared with third parties ("share usage") Usage and personal info shared ("share usage and personal")

Application: normal distribution

- **Model**: model in WTP space with normal WTP and lognormal price/scale coefficient
  - Estimated first with Hierarchical Bayes (HB) to get the initial values for Maximum Simulated Likelihood method
  - Computation time: 4h in Stata
  - Higher log-likelihood with the maximum likelihood estimates than the HB method
- **Results**:
  - People are willing to pay $1.56 per month on average to avoid commercials
  - Fast availability is valued highly, with an average WTP of $3.94 per month in order to see TV shows and movies soon after their original showing
  - People are willing to pay $2.70 per month to avoid their data to be shared

Application: polynomials and splines

Polynomials

- Model: definition of S with 10^{24} points, 2000 draws of β for each person
  - Z variables: 6th order polynomials on each utility parameter and a second utility parameter on each WTP pair
- Estimation using Maximum Simulated Likelihood method in Matlab
  - Computation time: 16 minutes (optimized setup)
  - SLL at convergence of 3864.85, compared to 3903.47 for the model with a normal distribution

Splines

- Model:
  - 83 parameters
- Estimation in Matlab
  - Computation time: 16 minutes (optimized setup)
  - SLL at convergence of 3886.70
- Similar utility parameters shapes for polynomials and splines, but different than normal

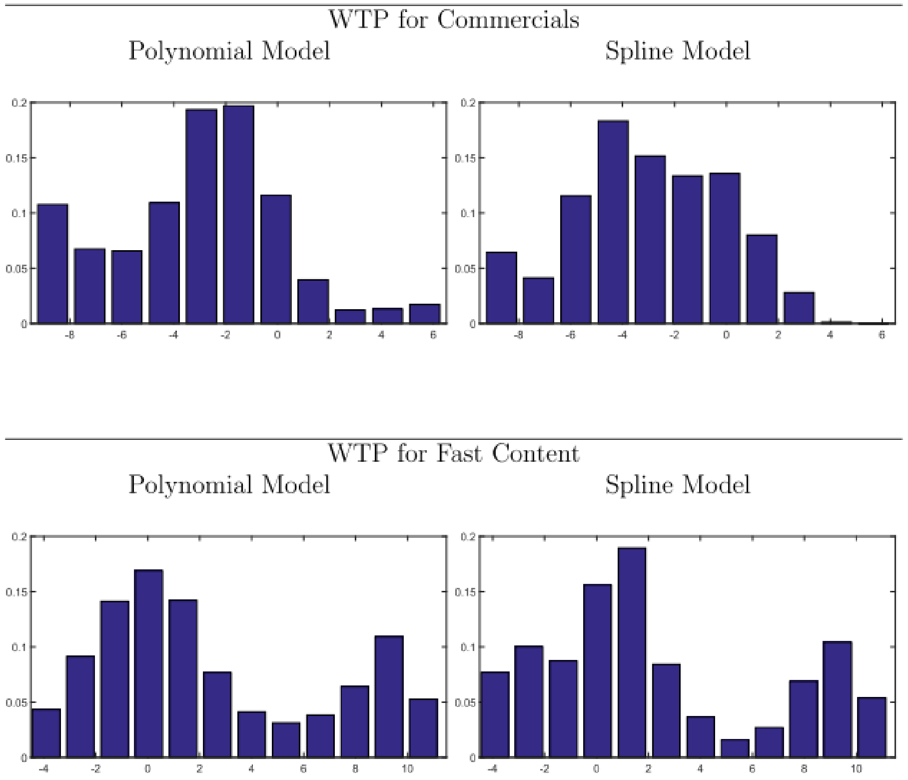


Fig. 2. Distribution of utility parameters.

Discussion: more flexibility, or better data?

- LML = Fast and easy specification of flexible mixing distribution
- **But**, additional burden on the researcher: What shape for the distribution ? What range for each utility parameter ?
  - Different distributions can provide similar log-likelihood, so which one is the best and how to choose ?
  - The range defined for the parameter (ex. Negative, positive,…) might lead to worst results than a model without range restrictions
- The issue might be that the data does not contain enough information to exclude theoretically implausible behavior

→ need for richer data that would lead to bigger differences in distributions and meet expected results

Sampling considerations in 2), 3), 4) might relate to *Kim and Bansal (2023)* and 門坂さん's presentation ?

Discussion: Limitations of the LML model:

Limitations of the LML model:

1. Delineating the relation of LML models to nonparametric estimation would enhance specification and interpretation of models
2. Unequal probability sampling should improve the performance of the model and should be studied
3. The relationship between the statistics (mean, SD) of the estimated mixing distribution and the range of coefficients used to define S should be studied, to define the best ways to define S for accuracy
4. The tails of distributions (not frequently observed behaviour) are important for policy and marketing purposes, but the use of flexible methods increase the need for those behaviours to be actually observed in the data for them to be simulated.

Sampling considerations in 2), 3), 4) might relate to *Kim and Bansal (2023)* and 門坂さん's presentation ?

Impressions

- This paper has a lot of mathematical results, but it's difficult to grasp all of them and imagine the dimensions, functions etc., as there are few examples, and the explanation is succinct
- There's a big difference between understanding something and explaining it to someone else
- Since most researchers already only use a normal distribution for β, I wonder how many would switch to an even more complicated setting, even if it allows for more flexibility

Related articles

- About discrete choice models: K. Train, Discrete Choice Methods with Simulation (2nd ed.), Cambridge University Press, New York (2009)
- About the ability of the mixed logit to approximate any discrete choice model: McFadden, K. Train, Mixed MNL models of discrete response, J. Appl. Econ., 15 (2000), pp. 447-470
- About reasons to use mixed logit: D. Revelt, K. Train, Mixed logit with repeated choices, Rev. Econ. Stat., 80 (1998), pp. 647-657
- About willingness to pay: K. Train, M. Weeks, Discrete choice models in preference space and willingness-to-pay space, Applications of Simulation Methods in Environmental and Resource Economics, (2005), pp. 1-17
- About the method of sieves: X. Chen, Large sample sieve estimation of semi-nonparametric models, Handbook of Econometrics, vol. 6A, (2007) (Chapter 76)
- About the introduction of fixed parameters: P. Bansal, R .A. Daziano, M. Achtnicht, Extending the logit-mixed logit model for a combination of random and fixed parameters, Journal of Choice Modelling, Volume 27, (2018), Pages 88-96,

Introduction: mixed logit model

Mixed logit model: (error components specification)

We consider a decision maker n, in a choice situation with J alternatives j:

$$U_{nj} = \alpha' x_{nj} + \mu_n' z_{nj} + \varepsilon_{nj}$$

Random part of the utility
Unknown to the researcher

Observed part of the utility
Known to the researcher

x_{nj} is a vector of observed variables relating to alternative j

z_{nj} is a vector of observed variables that define the error correlation among alternatives

α is a vector of fixed coefficients

μ is a vector of random coefficients with zero mean

ε_{nj} is an i.i.d. extreme value error term

Here, the goal is to create and estimate correlation among the utilities for different alternatives:

$$\text{Cov}(\eta_{ni}, \eta_{nj}) = E(\mu_n' z_{ni} + \varepsilon_{ni})(\mu_n' z_{nj} + \varepsilon_{nj}) = z_{ni}' W z_{nj}$$ where W is the covariance of μ_n

How to choose the mixed logit specification:

- What is the goal: estimate the pattern of taste (random parameters) or choice prediction (error components)
- Number of parameters (usually less parameters used for random parameters specification so that the joint distribution can be estimated)