Mixtures and latent variables in discrete choice models: an introduction

Michel Bierlaire

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

July 16, 2013

Outline

Discrete choice

- Decision maker: n, with characteristics s_n
- Choice set: $\mathcal{C}_n$
- Attributes: $z_n = (z_{1n}, \ldots, z_{J_n n})$,
- Choice model:

$$P(i|s_n, z_n, \mathcal{C}_n)$$

- Exogenous variables: $x_n = (s_n, z_n)$
  - both continuous and discrete
- Endogenous variable: i
  - discrete

Utility

- Utility functions:

$$U_n = U_n(s_n, z_n, \varepsilon_n)$$

- $U_n \in \mathbb{R}^{J_n} : (U_{1n}, \ldots, U_{J_n n})$
- Assumption: i is chosen if

$$U_{in} \geq U_{jn},\ \forall j \in \mathcal{C}_n.$$

Random utility

- Issue: ε_n is unobserved.
- Random vector.

$$P(i|\mathcal{C}_n) = \Pr(U_{in} \geq U_{jn}, \ \forall j \in \mathcal{C}_n)$$

- Assumptions must be made on ε_n.

Additive utility

- Utility function:

$$U_{in} = V_{in} + \varepsilon_{in}$$

- Deterministic part:

$$V_{in} = V_{in}(s_n, z_{in})$$

- Error term: ε_{in}
  - Expectation: alternative specific constant.
  - Scale: unidentified, must be normalized.
  - Distribution: extreme value, normal, ...

Logit

Assumption: error terms ε_{in} are

- independent
- identically distributed
- across i and across n

$$P(i|s_n, z_n, \mathcal{C}_n) = \frac{e^{V_{in}(s_n,z_{in})}}{\sum_{j\in\mathcal{C}_n} e^{V_{jn}(s_n,z_{jn})}}$$

Outline

Continuous mixtures

In statistics, a mixture probability distribution function is a convex combination of other probability distribution functions.

If $f(\varepsilon, \theta)$ is a distribution function, and if $w(\theta)$ is a non negative function such that

$$\int_\theta w(\theta) d\theta = 1$$

then

$$g(\varepsilon) = \int_\theta w(\theta) f(\varepsilon, \theta) d\theta$$

is also a distribution function. We say that g is a w-mixture of f.

If f is a logit model, g is a continuous w-mixture of logit

Discrete mixtures

Discrete mixtures are also possible. If w_i, $i = 1, \ldots, n$ are non negative weights such that

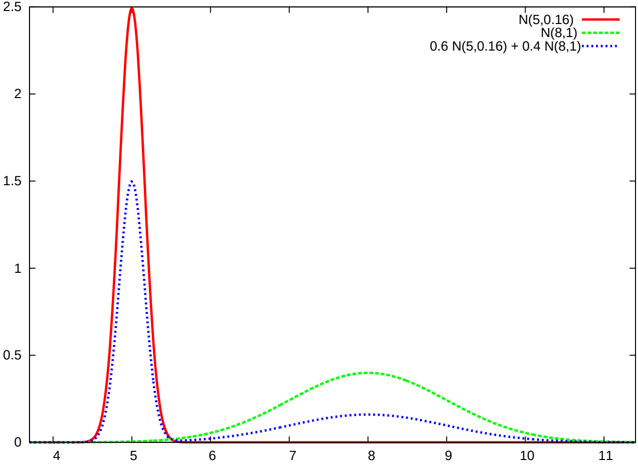
$$\sum_{i=1}^{n} w_i = 1$$

then

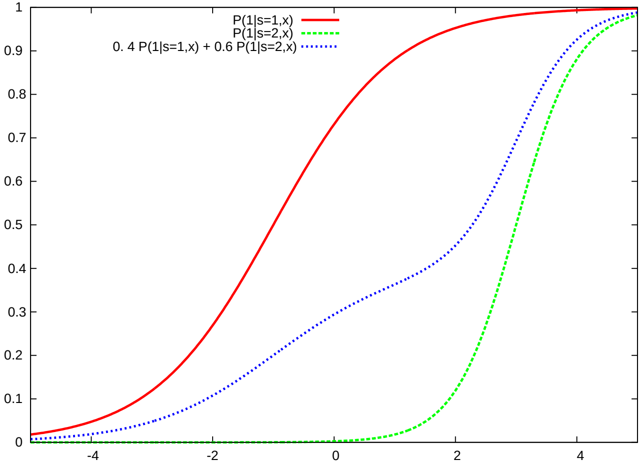
$$g(\varepsilon) = \sum_{i=1}^{n} w_i f(\varepsilon, \theta_i)$$

is also a distribution function where θ_i, $i = 1, \ldots, n$ are parameters. We say that g is a discrete w-mixture of f.

Example: discrete mixture of normal distributions



Example: discrete mixture of binary logit models



Motivation

- General motivation: generate flexible distributional forms
- For discrete choice:
  - correlation across alternatives
  - alternative specific variances
  - taste heterogeneity
  - . . .

Continuous mixtures of logit

- Combining probit and logit
- Error decomposed into two parts

$$U_{in} = V_{in} + \underbrace{\xi_{in}}_{\text{Normal distribution (probit): flexibility}} + \underbrace{\nu_{in}}_{\text{i.i.d EV (logit): tractability}}$$

Choice model

$$U_{in} = V_{in} + \xi_{in} + \nu_{in}$$

- Assumptions:
  - ν_{in} i.i.d. extreme value,
  - $\xi_{in} \sim N(0, \Sigma)$
- If ξ_{in} were observed, we would have a logit model

$$P(i|\xi_n, \mathcal{C}_n) = \frac{e^{V_{in}+\xi_{in}}}{\sum_{j\in\mathcal{C}_n} e^{V_{jn}+\xi_{in}}}$$

Choice model

- To obtain the model, we must integrate over ξ_n

$$P(i|\mathcal{C}_n) = \int_\xi P(i|\xi,\mathcal{C}_n) f(\xi) d\xi = \int_\xi \frac{e^{V_{in}+\xi_{in}}}{\sum_{j\in\mathcal{C}_n} e^{V_{jn}+\xi_{in}}} f(\xi) d\xi$$

- $f(\xi)$ is the pdf of the normal distribution.
- Complex integral, requires Monte-Carlo simulation

Simulation

- In order to approximate

$$P(i|\mathcal{C}_n) = \int_\xi P(i|\xi, \mathcal{C}_n) f(\xi) d\xi = \int_\xi \frac{e^{V_{in}+\xi_{in}}}{\sum_{j\in\mathcal{C}_n} e^{V_{jn}+\xi_{in}}} f(\xi) d\xi$$

- Draw from $f(\xi)$ to obtain $r_1, \ldots, r_R$
- Compute

$$\begin{aligned} P(i|\mathcal{C}_n) \approx \tilde{P}(i|\mathcal{C}_n) &= \frac{1}{R}\sum_{k=1}^{R} P(i|\mathcal{C}_n, r_k) \\ &= \frac{1}{R}\sum_{k=1}^{R} \frac{e^{V_{in}+r_{ki}}}{\sum_{j\in\mathcal{C}_n} e^{V_{jn}+r_{kj}}} \end{aligned}$$

Application: relaxing the independence assumption

- Utility:

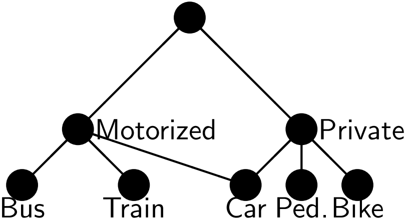
$$\begin{array}{rcllll} U_{\text{auto}} & = & \beta X_{\text{auto}} & & & + \; \nu_{\text{auto}} \\ U_{\text{bus}} & = & \beta X_{\text{bus}} & + & \sigma_{\text{transit}} \xi_{\text{transit}} & + \; \nu_{\text{bus}} \\ U_{\text{subway}} & = & \beta X_{\text{subway}} & + & \sigma_{\text{transit}} \xi_{\text{transit}} & + \; \nu_{\text{subway}} \end{array}$$

- ν i.i.d. extreme value, $\xi_{\text{transit}} \sim N(0,1)$, $\sigma^2_{\text{transit}}$ =cov(bus,subway)
- Probability:

$$\Pr(\text{auto}|X, \xi_{\text{transit}}) = \frac{e^{\beta X_{\text{auto}}}}{e^{\beta X_{\text{auto}}} + e^{\beta X_{\text{bus}} + \sigma_{\text{transit}} \xi_{\text{transit}}} + e^{\beta X_{\text{subway}} + \sigma_{\text{transit}} \xi_{\text{transit}}}}$$

$$P(\text{auto}|X) = \int_{\xi} \Pr(\text{auto}|X, \xi) f(\xi) d\xi$$

Cross nesting



$$\begin{array}{rclcccl}
U_{\text{bus}} & = & V_{\text{bus}} & +\xi_1 & & +\nu_{\text{bus}} \\
U_{\text{train}} & = & V_{\text{train}} & +\xi_1 & & +\nu_{\text{train}} \\
U_{\text{car}} & = & V_{\text{car}} & +\xi_1 & +\xi_2 & +\nu_{\text{car}} \\
U_{\text{ped}} & = & V_{\text{ped}} & & +\xi_2 & +\nu_{\text{ped}} \\
U_{\text{bike}} & = & V_{\text{bike}} & & +\xi_2 & +\nu_{\text{bike}}
\end{array}$$

$$P(\text{car}) = \int_{\xi_1}\int_{\xi_2} P(\text{car}|\xi_1,\xi_2)f(\xi_1)f(\xi_2)d\xi_2 d\xi_1$$

Application: relaxing the identical distribution assumption

- Error terms in logit are identically distributed and, in particular, have the same variance

$$U_{in} = \beta^T x_{in} + \text{ASC}_i + \varepsilon_{in}$$

- ε_{in} i.i.d. extreme value $\Rightarrow \text{Var}(\varepsilon_{in}) = \pi^2/6\mu^2$
- In order allow for different variances, we use mixtures

$$U_{in} = \beta^T x_{in} + \text{ASC}_i + \sigma_i\xi_i + \nu_{in}$$

where $\xi_i \sim N(0,1)$ and ν_{in} are i.i.d extreme value.

- Variance:

$$\text{Var}(\sigma_i\xi_i + \nu_{in}) = \sigma_i^2 + \frac{\pi^2}{6\mu^2}$$

Alternative specific variance

Example with Swissmetro

	ASC_CAR	ASC_SBB	ASC_SM	B_COST	B_FR	B_TIME
Car	1	0	0	cost	0	time
Train	0	0	0	cost	freq.	time
Swissmetro	0	0	1	cost	freq.	time

$+$ alternative specific variance

	Logit		ASV		ASV norm.	
$\mathcal{L}$	-5315.39		-5241.01		-5242.10	
	Value	Scaled	Value	Scaled	Value	Scaled
ASC_CAR	0.189	1.000	0.248	1.000	0.241	1.000
ASC_SM	0.451	2.384	0.903	3.637	0.882	3.657
B_COST	-0.011	-0.057	-0.018	-0.072	-0.018	-0.073
B_FR	-0.005	-0.028	-0.008	-0.031	-0.008	-0.032
B_TIME	-0.013	-0.067	-0.017	-0.069	-0.017	-0.071
SIGMA_CAR			0.020			
SIGMA_TRAIN			0.039		0.061	
SIGMA_SM			3.224		3.180	

Taste heterogeneity

- Population is heterogeneous
- Taste heterogeneity is captured by segmentation
- Deterministic segmentation is desirable but not always possible
- Distribution of a parameter in the population

Disributed time coefficient

$$\begin{array}{rcl} U_i & = & \beta_t T_i + \beta_c C_i + \varepsilon_i \\ U_j & = & \beta_t T_j + \beta_c C_j + \varepsilon_j \end{array}$$

Let $\beta_t \sim N(\bar{\beta}_t, \sigma_t^2)$, or, equivalently,

$$\beta_t = \bar{\beta}_t + \sigma_t \xi, \text{ with } \xi \sim N(0,1).$$

$$\begin{array}{rcl} U_i & = & \bar{\beta}_t T_i + \sigma_t \xi T_i + \beta_c C_i + \varepsilon_i \\ U_j & = & \bar{\beta}_t T_j + \sigma_t \xi T_j + \beta_c C_j + \varepsilon_j \end{array}$$

If ε_i and ε_j are i.i.d. EV and ξ is given, we have

$$P(i|\xi) = \frac{e^{\bar{\beta}_t T_i + \sigma_t \xi T_i + \beta_c C_i}}{e^{\bar{\beta}_t T_i + \sigma_t \xi T_i + \beta_c C_i} + e^{\bar{\beta}_t T_j + \sigma_t \xi T_j + \beta_c C_j}}, \text{ and}$$

$$P(i) = \int_\xi P(i|\xi) f(\xi) d\xi.$$

Example with Swissmetro

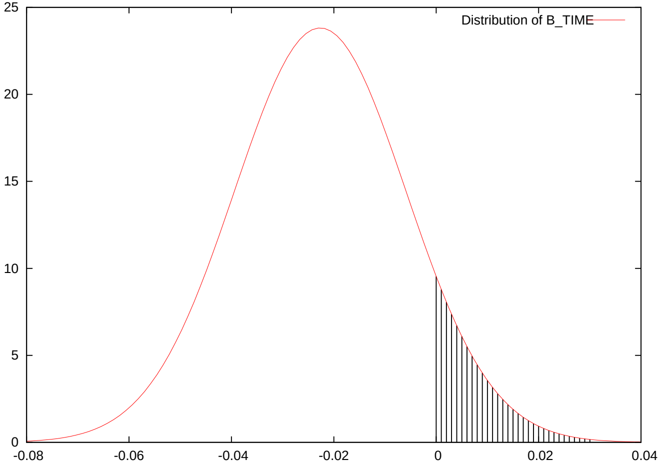
	ASC_CAR	ASC_SBB	ASC_SM	B_COST	B_FR	B_TIME
Car	1	0	0	cost	0	time
Train	0	0	0	cost	freq.	time
Swissmetro	0	0	1	cost	freq.	time

B_TIME randomly distributed across the population, normal distribution

Estimated parameters

	Logit	RC
$\mathcal{L}$	-5315.4	-5198.0
ASC_CAR_SP	0.189	0.118
ASC_SM_SP	0.451	0.107
B_COST	-0.011	-0.013
B_FR	-0.005	-0.006
B_TIME	-0.013	-0.023
S_TIME		0.017
Prob(B_TIME $\geq$ 0)		8.8%
χ^2		234.84

Distribution of the parameter



Another distribution

Example with Swissmetro

	ASC_CAR	ASC_SBB	ASC_SM	B_COST	B_FR	B_TIME
Car	1	0	0	cost	0	time
Train	0	0	0	cost	freq.	time
Swissmetro	0	0	1	cost	freq.	time

B_TIME randomly distributed across the population, log normal distribution

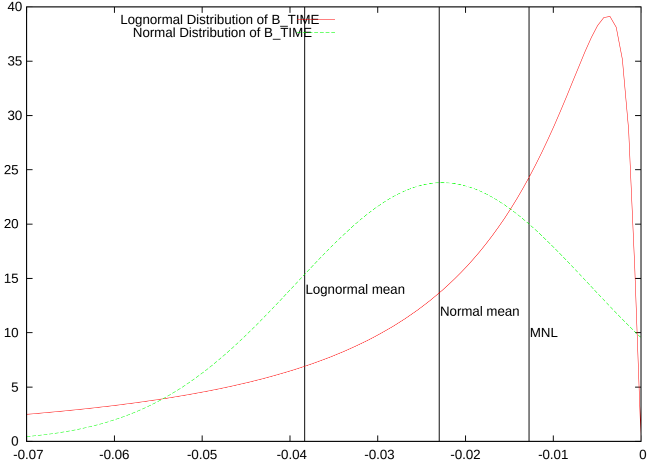
Syntax for Biogeme

```
[Utilities]
11 SBB_SP TRAIN_AV_SP ASC_SBB_SP * one        +
                      B_COST     * TRAIN_COST +
                      B_FR       * TRAIN_FR
21 SM_SP SM_AV        ASC_SM_SP  * one        +
                      B_COST     * SM_COST    +
                      B_FR * SM_FR
31 Car_SP CAR_AV_SP   ASC_CAR_SP * one        +
                      B_COST     * CAR_CO
[GeneralizedUtilities]
11 - exp( B_TIME [ S_TIME ] ) * TRAIN_TT
21 - exp( B_TIME [ S_TIME ] ) * SM_TT
31 - exp( B_TIME [ S_TIME ] ) * CAR_TT
```

Estimation results

	Logit	RC-norm.	RC-logn.	
	-5315.4	-5198.0	-5215.81	
ASC_CAR_SP	0.189	0.118	0.122	
ASC_SM_SP	0.451	0.107	0.069	
B_COST	-0.011	-0.013	-0.014	
B_FR	-0.005	-0.006	-0.006	
B_TIME	-0.013	-0.023	-4.033	-0.038
S_TIME		0.017	1.242	0.073
Prob($\beta > 0$)		8.8%	0.0%	
χ^2		234.84	199.16	

Distribution of the parameter



Another distribution

Example with Swissmetro

	ASC_CAR	ASC_SBB	ASC_SM	B_COST	B_FR	B_TIME
Car	1	0	0	cost	0	time
Train	0	0	0	cost	freq.	time
Swissmetro	0	0	1	cost	freq.	time

B_TIME randomly distributed across the population, discrete distribution

$$P(\beta_{\text{time}} = \hat{\beta}) = \omega_1 \quad P(\beta_{\text{time}} = 0) = \omega_2 = 1 - \omega_1$$

Syntax for Biogeme

```
[DiscreteDistributions]
B_TIME < B_TIME_1 ( W1 ) B_TIME_2 ( W2 ) >

[LinearConstraints]
W1 + W2 = 1.0
```

Estimation results

	Logit	RC-norm.	RC-logn.		RC-disc.
	-5315.4	-5198.0	-5215.8		-5191.1
ASC_CAR_SP	0.189	0.118	0.122		0.111
ASC_SM_SP	0.451	0.107	0.069		0.108
B_COST	-0.011	-0.013	-0.014		-0.013
B_FR	-0.005	-0.006	-0.006		-0.006
B_TIME	-0.013	-0.023	-4.033	-0.038	-0.028
					0.000
S_TIME		0.017	1.242	0.073	
W1					0.749
W2					0.251
Prob($\beta > 0$)		8.8%	0.0%		0.0%
χ^2		234.84	199.16		248.6

Summary

- Logit mixtures models
  - Computationally more complex than logit
  - Allow for more flexibility than logit
- Continuous mixtures: alternative specific variance, nesting structures, random parameters

$$P(i) = \int_{\xi} \Pr(i|\xi) f(\xi) d\xi$$

- Discrete mixtures:

$$P(i) = \sum_{s=1}^{S} w_s \Pr(i|s).$$

Tips for applications

- Be careful: simulation can mask specification and identification issues
- Do not forget about the systematic portion

Outline

1. Discrete choice models
   - Introduction
   - Random utility
   - Logit
2. Mixtures
   - Introduction
   - Error component
   - Random parameter
   - Discrete Mixtures
   - Summary
3. Beyond rationality
   - Examples
4. Latent concepts
   - Utility
   - Indicators
   - Measurement equation
   - Hybrid choice model
   - Case study

Beyond rationality

- Standard random utility assumptions are often violated.
- Factors such as attitudes, perceptions, knowledge are not reflected.

Example: pain lovers

Kahneman, D., Fredrickson, B., Schreiber, C.M., and Redelmeier, D., When More Pain Is Preferred to Less: Adding a Better End, *Psychological Science*, Vol. 4, No. 6, pp. 401-405, 1993.

- Short trial: immerse one hand in water at 14° for 60 sec.
- Long trial: immerse the other hand at 14° for 60 sec, then keep the hand in the water 30 sec. longer as the temperature of the water is gradually raised to 15°.
- Outcome: most people prefer the long trial.
- Explanation:
    - duration plays a small role
    - the peak and the final moments matter

Example: *The Economist*

Example: subscription to *The Economist*

Web only	@ \$59
Print only	@ \$125
Print and web	@ \$125

Example: *The Economist*

Example: subscription to *The Economist*

Experiment 1	Experiment 2
Web only @ $59	Web only @ $59
Print only @ $125	
Print and web @ $125	Print and web @ $125

Example: *The Economist*

Example: subscription to *The Economist*

	Experiment 1	Experiment 2	
16	Web only @ $59	Web only @ $59	68
0	Print only @ $125		
84	Print and web @ $125	Print and web @ $125	32

Source: Ariely (2008)

- Dominated alternative
- According to utility maximization, should not affect the choice
- But it affects the perception, which affects the choice.

Example: good or bad wine?

Choose a bottle of wine...

	Experiment 1	Experiment 2
1	McFadden red at $10	McFadden red at $10
2	Nappa red at $12	Nappa red at $12
3		McFadden special reserve pinot noir at $60
	Most would choose 2	Most would choose 1

- Context plays a role on perceptions

Example: live and let die

Population of 600 is threatened by a disease. Two alternative treatments to combat the disease have been proposed.

Experiment 1 # resp. = 152	Experiment 2 # resp. = 155
Treatment A: 200 people saved	Treatment C: 400 people die
Treatment B: 600 people saved with prob. 1/3 0 people saved with prob. 2/3	Treatment D: 0 people die with prob. 1/3 600 people die with prob. 2/3

Example: live and let die

Population of 600 is threatened by a disease. Two alternative treatments to combat the disease have been proposed.

	Experiment 1 # resp. = 152	Experiment 2 # resp. = 155	
72%	Treatment A: 200 people saved	Treatment C: 400 people die	22%
28%	Treatment B: 600 people saved with prob. 1/3 0 people saved with prob. 2/3	Treatment D: 0 people die with prob. 1/3 600 people die with prob. 2/3	78%

Source: Tversky & Kahneman (1986)

Example: to be free

Choice between a fine and a regular chocolate

	Experiment 1	Experiment 2
Lindt	\$0.15	\$0.14
Hershey	\$0.01	\$0.00
Lindt chosen	73%	31%
Hershey chosen	27%	69%

Source: Ariely (2008) *Predictably irrational*, Harper Collins.

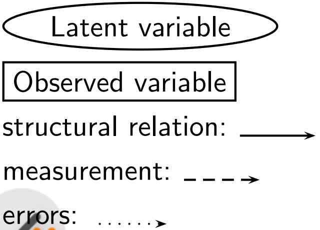
Outline

1. Discrete choice models
   - Introduction
   - Random utility
   - Logit
2. Mixtures
   - Introduction
   - Error component
   - Random parameter
   - Discrete Mixtures
   - Summary
3. Beyond rationality
   - Examples
4. Latent concepts
   - Utility
   - Indicators
   - Measurement equation
   - Hybrid choice model
   - Case study

Latent concepts

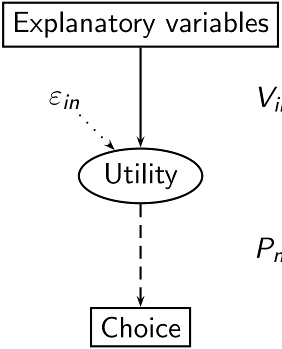
- **latent**: potentially existing but not presently evident or realized (from Latin: *lateo* = lie hidden)
- Here: not directly observed
- Standard models are already based on a latent concept: utility

Drawing convention:



- Latent variable
- Observed variable
- structural relation: ⟶
- measurement: - - - →
- errors: ······→

Random utility



$$V_{in} = \sum_k \beta_{ik} x_{ikn}$$

$$P_n(i) = e^{V_{in}} / \sum_j e^{V_{jn}}$$

Attitudes

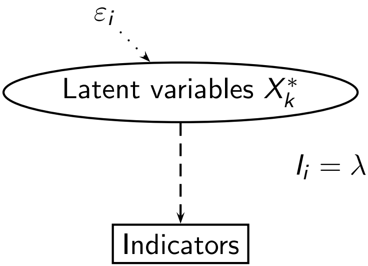
- Psychometric indicators
- Example: attitude towards the environment.
- For each question, response on a scale: strongly agree, agree, neutral, disagree, strongly disagree, no idea.
  - The price of oil should be increased to reduce congestion and pollution
  - More public transportation is necessary, even if it means additional taxes
  - Ecology is a threat to minorities and small companies.
  - People and employment are more important than the environment.
  - I feel concerned by the global warming.
  - Decisions must be taken to reduce the greenhouse gas emission.

Indicators

Indicators cannot be used as explanatory variables. Mainly two reasons:

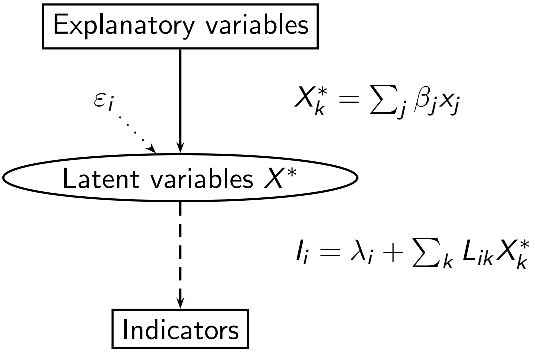
1. Measurement errors
    - Scale is arbitrary and discrete
    - People may overreact
    - Justification bias may produce exaggerated responses
2. No forecasting possibility
    - No way to predict the indicators in the future

Factor analysis



$$I_i = \lambda_i + \sum_k L_{ik} X_k^*$$

Measurement equation



Explanatory variables

ε_i

$X_k^* = \sum_j \beta_j x_j$

Latent variables X^*

$I_i = \lambda_i + \sum_k L_{ik} X_k^*$

Indicators

Measurement equation

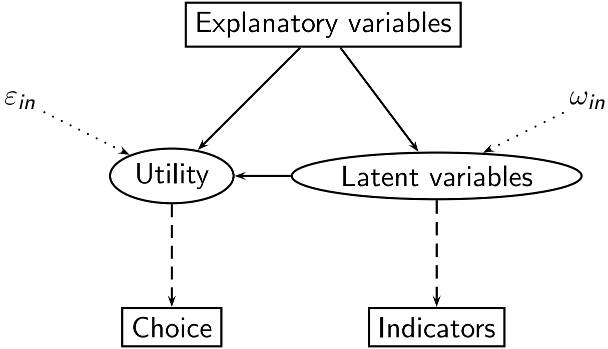
Continuous model: regression

$$I = f(X^*; \beta) + \varepsilon$$

Discrete model: thresholds

$$I = \begin{cases} 1 & \text{if } -\infty < X^* \leq \tau_1 \\ 2 & \text{if } \tau_1 < X^* \leq \tau_2 \\ 3 & \text{if } \tau_2 < X^* \leq \tau_3 \\ 4 & \text{if } \tau_3 < X^* \leq \tau_4 \\ 5 & \text{if } \tau_4 < X^* \leq +\infty \end{cases}$$

Choice model



Estimation: likelihood

Structural equations:

1. Distribution of the latent variables:

$$f_1(X^*_n | X_n; \lambda, \Sigma_\omega)$$

For instance $X^*_n = h(X_n; \lambda) + \omega_n, \quad \omega_n \sim N(0, \Sigma_\omega).$

2. Distribution of the utilities:

$$f_2(U_n | X_n, X^*_n; \beta, \Sigma_\varepsilon)$$

For instance $U_n = V(X_n, X^*_n; \beta) + \varepsilon_n, \quad \varepsilon_n \sim N(0, \Sigma_\omega).$

Estimation: likelihood

Measurement equations:

1. Distribution of the indicators:

$$f_3(I_n | X_n, X_n^*; \alpha, \Sigma_\nu)$$

   For instance:

$$I_n = m(X_n, X_n^*; \alpha) + \nu_n, \quad \nu_n \sim N(0, \Sigma_\nu).$$

2. Distribution of the observed choice:

$$P(y_{in} = 1) = \Pr(U_{in} \geq U_{jn}, \forall j).$$

Indicators: continuous output

$$f_3(I_n|X_n, X_n^*; \alpha, \Sigma_\nu)$$

For instance:

$$I_n = m(X_n, X_n^*; \alpha) + \nu_n, \quad \nu_n \sim N(0, \sigma_{\nu_n}^2)$$

So,

$$f_3(I_n|\cdot) = \frac{1}{\sigma_{\nu_n}\sqrt{2\pi}} \exp\left(-\frac{(I_n - m(\cdot))^2}{2\sigma_{\nu_n}^2}\right)$$

Define

$$Z = \frac{I_n - m(\cdot)}{\sigma_{\nu_n}} \sim N(0,1), \quad \phi(Z) = \frac{1}{\sqrt{2\pi}} e^{-Z^2/2}$$

and

$$f_3(I_n|\cdot) = \frac{1}{\sigma_{\nu_n}} \phi(Z)$$

Indicators: discrete output

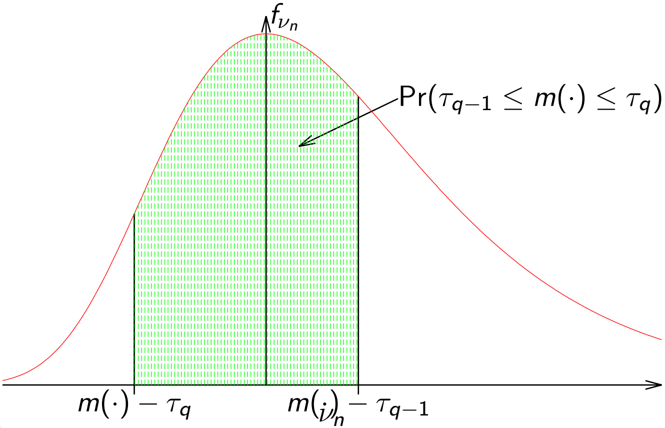
$$f_3(I_n | X_n, X_n^*; \alpha, \Sigma_\nu)$$

For instance:

$$I_n = m(X_n, X_n^*; \alpha) + \nu_n, \quad \nu_n \sim \text{Logistic(0,1)}$$

$$\begin{aligned}
P(I_n = 1) &= & \Pr(m(\cdot) \leq \tau_1) &= \frac{1}{1 + e^{-\tau_1 + m(\cdot)}} \\
P(I_n = 2) &= & \Pr(m(\cdot) \leq \tau_2) - \Pr(m(\cdot) \leq \tau_1) &= \frac{1}{1 + e^{-\tau_2 + m(\cdot)}} - \frac{1}{1 + e^{-\tau_1 + m(\cdot)}} \\
&\vdots & &\vdots \\
P(I_n = 5) &= & 1 - \Pr(m(\cdot) \leq \tau_4) &= 1 - \frac{1}{1 + e^{-\tau_4 + m(\cdot)}}
\end{aligned}$$

Indicators: discrete output



Estimation: likelihood

Assuming ω_n, ε_n and ν_n are independent, we have

$$\mathcal{L}_n(y_n, I_n | X_n; \alpha, \beta, \lambda, \Sigma_\varepsilon, \Sigma_\nu, \Sigma_\omega) =$$

$$\int_{X^*} P(y_n | X_n, X^*; \beta, \Sigma_\varepsilon) f_3(I_n | X_n, X^*; \alpha, \Sigma_\nu) f_1(X^* | X_n; \lambda, \Sigma_\omega) dX^*.$$

Maximum likelihood estimation:

$$\max_{\alpha,\beta,\lambda,\Sigma_\varepsilon,\Sigma_\nu,\Sigma_\omega} \sum_n \log \left(\mathcal{L}_n(y_n, I_n | X_n; \alpha, \beta, \lambda, \Sigma_\varepsilon, \Sigma_\nu, \Sigma_\omega) \right)$$

Source: Walker (2001)

Case study: value of time

- Effect of attitude on value of time
- SP survey, Stockholm, Sweden, 2005
- 2400 households surveyed
- Married couples with both husband and wife working or studying
- Choice between car alternatives
- Data used: 554 respondents, 2216 SP responses
- Attributes:
  - travel time
  - travel cost
  - number of speed cameras

Attitudinal questions

- It feels safe to go by car.
- It is comfortable to go by car to work.
- It is very important that traffic speed limits are not violated.
- Increase the motorway speed limit to 140 km/h.

Likert scale:

- 1: do not agree at all
- 5: do fully agree

Structural models

Attitude model, capturing the positive attitude towards car

$$\begin{aligned}
\text{Attitude} = \quad & \theta_0 \cdot 1 && \text{(intercept)} \\
& + \theta_f \cdot \text{female} && \\
& + \theta_{\text{inc}} \cdot \text{income} && \text{(monthly, in Kronas)} \\
& + \theta_{\text{age1}} \cdot (\text{Age} < 55) && \\
& + \theta_{\text{age2}} \cdot (\text{Age } 55\text{–}65) && \\
& + \theta_{\text{age3}} \cdot (\text{Age} > 65) && \\
& + \theta_{\text{edu1}} \cdot (\text{basic/pre high school}) && \\
& + \theta_{\text{edu2}} \cdot (\text{university}) && \\
& + \theta_{\text{edu3}} \cdot (\text{other}) && \\
& \sigma \cdot \omega && \text{(normal error term)}
\end{aligned}$$

Structural models

Choice model: 3 alternatives

- Car on route 1
- Car on route 2
- Indifferent (utility = 0)

$$\begin{aligned}\text{Utility}_i = \quad & \beta_i & \text{(ASC)}\\ & + \beta_{\text{t}} \cdot \text{travel time}_i & \\ & + \beta_{\text{c}} \cdot \text{cost}_i \;/\; \text{Income} & \\ & + \gamma \cdot \text{cost}_i \cdot \text{Attitude} \;/\; \text{Income} & \\ & + \beta_{\text{cam}} \cdot \# \text{ cameras}_i & \\ & + \varepsilon_i & \text{(EV error term)}\end{aligned}$$

Note: standard model obtained with $\gamma = 0$.

Value of time

- Model without attitude variable ($\gamma = 0$)

$$\mathrm{VOT} = \frac{\beta_{\mathrm{t}}}{\beta_{\mathrm{c}}} * \mathit{Income}$$

- Model with attitude variable

$$\mathrm{VOT} = \frac{\beta_{\mathrm{t}}}{\beta_{\mathrm{c}} + \gamma \cdot \mathrm{Attitude}} * \mathit{Income}$$

Note: distributed

Measurement equations

- Choice:

$$y_i = \begin{cases} 1 & \text{if } U_i \geq U_j, j \neq i \\ 0 & \text{otherwise} \end{cases}$$

- Attitude questions: $k = 1, \ldots, 4$

$$I_k = \alpha_k + \lambda_k \text{Attitude} + \mu_k$$

where I_k is the response to question k.

Model estimation

- Simultaneous estimation of all parameters
- with Biogeme 2.0
- Important: both the choice and the indicators reveal something about the attitude.

Measurement equations

- It feels safe to go by car.

$$I_1 = \text{Attitude} + 0.5666\ \nu_1$$

- It is comfortable to go by car to work.

$$I_2 = 1.13 + 0.764\ \text{Attitude} + 0.909\ \nu_2$$

- It is very important that traffic speed limits are not violated.

$$I_3 = 3.53 - 0.0716\ \text{Attitude} + 1.25\ \nu_3$$

- Increase the motorway speed limit to 140 km/h.

$$I_4 = 1.94 + 0.481\ \text{Attitude} + 1.37\ \nu_4$$

Structural model

Attitude towards car:

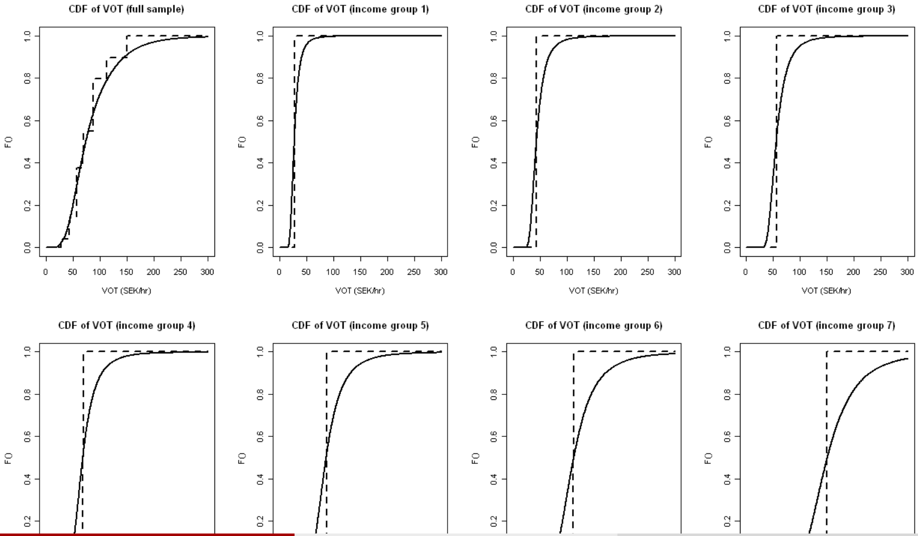
Param.	Estim.	t-stat.
θ_0	5.25	8.99
θ_f	-0.0185	-0.34
θ_{inc}	0.0347	1.99
θ_{age1}	-0.0217	-1.85
θ_{age2}	0.00797	0.88
θ_{age3}	0.0231	2.35
θ_{edu1}	-0.147	-0.94
θ_{edu2}	-0.252	-5.22
θ_{edu3}	-0.157	-0.85
σ	0.934	16.18

Structural model

Utility:

Param.	Estim.	*t*-stat.
β_1	4.01	15.58
β_2	2.84	10.57
Time	-0.0388	-8.10
Cost/Income	-2.02	-3.63
Cost $\cdot$ Attitude/Income	0.265	2.11
Speed camera	-0.109	-2.75

Value of time



Conclusion

- Flexible models with more structure
- Translate more assumptions into equations
- More complicated to estimate
- Currently very active field for research and applications.