Latent variables

Michel Bierlaire

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

Outline

Motivation

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

Subscription to The Economist

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

Example: The Economist

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

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

Latent concepts

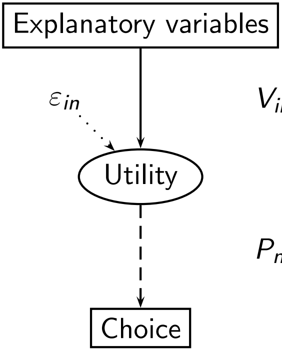
Latent

- **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

Measuring 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.

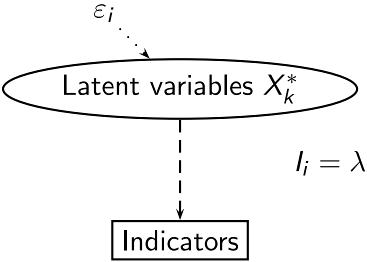
Indicators

Model specification

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

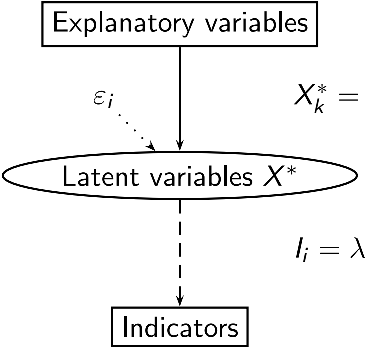
1. Measurement errors
   - Scale is arbitrary and discrete
   - People may overreact
   - Justification bias may produce exaggerated responses
2. No forecasting possibility
   - No way to predict the indicators in the future

Factor analysis



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

Measurement equation



$$X_k^* = \sum_j \beta_j x_j$$

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

Measurement equation

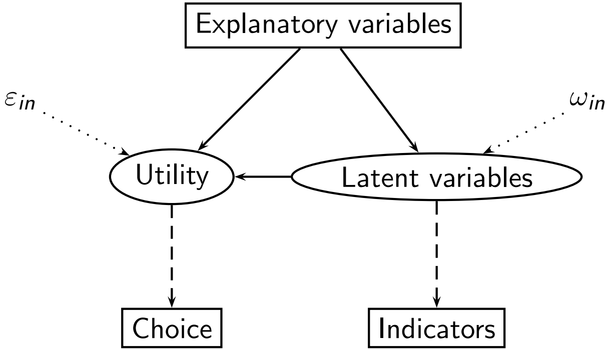
Continuous model: regression

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

Discrete model: thresholds

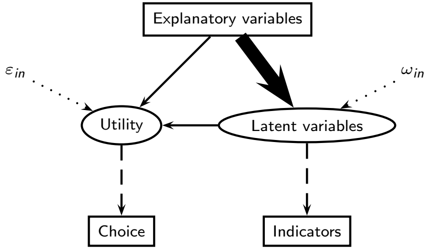
$$I = \begin{cases} 1 & \text{if } -\infty < X^* \leq \tau_1 \\ 2 & \text{if } \tau_1 < X^* \leq \tau_2 \\ 3 & \text{if } \tau_2 < X^* \leq \tau_3 \\ 4 & \text{if } \tau_3 < X^* \leq \tau_4 \\ 5 & \text{if } \tau_4 < X^* \leq +\infty \end{cases}$$

Choice model



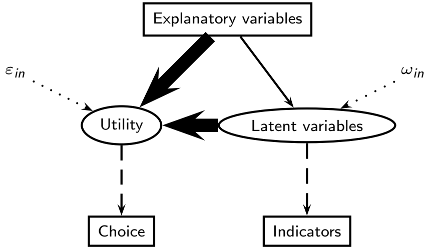
Outline

Structural equations



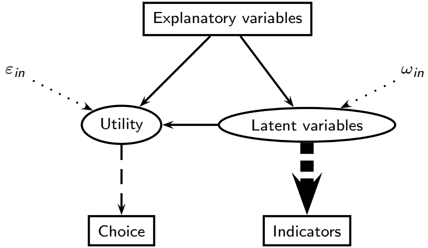
$$X_n^* = h(X_n; \lambda) + \omega_n, \quad \omega_n \sim N(0, \Sigma_\omega).$$

Structural equations



$$U_n = V(X_n, X_n^*; \beta) + \varepsilon_n, \quad \varepsilon_n \sim \text{EV}(0, \mu).$$

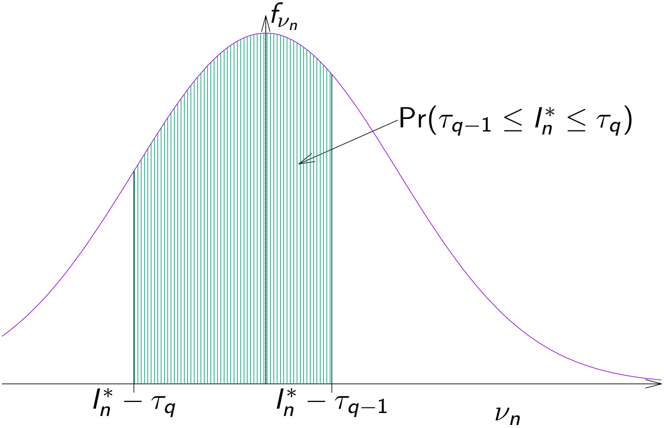
Measurement equations



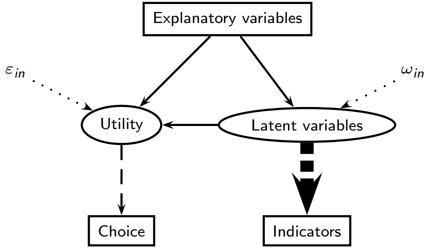
Ordinal discrete variable: ordered probit model

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

Ordered probit

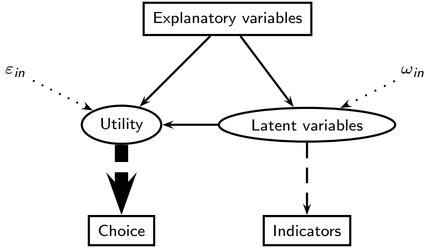


Measurement equations



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

Measurement equations



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

Estimation: likelihood

Assuming ω_n, ε_n and ν_n are independent, we have

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

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

Maximum likelihood estimation:

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

Source: Walker (2001)

Outline

Case studies

Walker (2001)

- Mode choice
- Latent variables:
  - Ride comfort
  - Convenience
- Indicators: (from "very poor" to "very good")
  - Relaxation during the trip
  - Reliability of the arrival time
  - Flexibility of choosing departure time
  - Ease of traveling with children
  - Safety during the trip
  - Overall rating of the mode

Case studies

Walker (2001)

- Employees' adoption of telecommuting
- Latent variables:
  - Perceived costs
    - Impact on your expenditures on home utilities
    - Impact on your expenditures on child care
    - Impact on your expenditures on elder care
    - Impact on your expenditures on overall working costs
  - Benefits
    - Impact on your schedule flexibility
    - Impact on your productivity
    - Impact on your autonomy in your job
    - Impact on the productivity of the group you work with
    - Impact on your family life
    - Impact on your social life
    - etc.

Case study: Optima

Effect of attitude on mode choice

- Switzerland, 2009–2010
- 1124 completed surveys
- 1906 trip chains from home to home

Attitudinal questions

Statements

- **Envir01** Fuel price should be increased to reduce congestion and air pollution.
- **Envir02** More public transportation is needed, even if taxes are set to pay the additional costs.
- **Envir03** Ecology disadvantages minorities and small businesses.
- **Mobil11** It is difficult to take the public transport when I carry bags or luggage.
- **Mobil14** When I take the car I know I will be on time.
- **Mobil16** I do not like changing the mean of transport when I am traveling.
- **Mobil17** If I use public transportation I have to cancel certain activities I would have done if I had taken the car.

Factor analysis

	Factor1	Factor2	Factor3
Envir01	−0.565		
Envir02	−0.407		
Envir03	0.414		
Mobil11	0.484		
Mobil14	0.473		
Mobil16	0.462		
Mobil17	0.434		
Mobil26			0.408
ResidCh01		0.577	
ResidCh04		0.406	
ResidCh05		0.635	
ResidCh06		0.451	
ResidCh07		−0.418	
LifSty07		0.430	

Car lovers

- Latent variable: car loving attitude
- Structural equation:

$$z = \beta_0^s + \sum_{k=1}^{K_s-1} \beta_k^s x_k + \sigma_s \varepsilon^s$$

Explanatory variables

- age_65_more: the respondent is 65 or older;
- moreThanOneCar: the number of cars in the household is strictly greater than 1;
- moreThanOneBike: the number of bikes in the household is strictly greater than 1;
- individualHouse: the type of house is individual or terraced;
- male: the respondent is a male;

Car lovers

Explanatory variables (ctd)

- haveChildren: the family is a couple or a single with children;
- haveGA: the respondent owns a season ticket;
- highEducation: the respondent has obtained a degree strictly higher than high school.
- ScaledIncome: income, in 1000 CHF;
- ContIncome_0_4000: min(ScaledIncome,4)
- ContIncome_4000_6000: max(0,min(ScaledIncome-4,2))
- ContIncome_6000_8000: max(0,min(ScaledIncome-6,2))
- ContIncome_8000_10000: max(0,min(ScaledIncome-8,2))
- ContIncome_10000_more: max(0,ScaledIncome-10)

Measurement equations

Indicators

- Likert scale (5 levels)
- 1 — strongly approve $\cdots$ 5 — strongly disapprove

Thresholds

$$I_i^* = \beta_{0i}^m + \beta_i^m z + \sigma_i^* \varepsilon_i^*$$

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

Symmetry

$$\begin{aligned} \tau_1 &= -\delta_1 - \delta_2 \\ \tau_2 &= -\delta_1 \\ \tau_3 &= \delta_1 \\ \tau_4 &= \delta_1 + \delta_2 \end{aligned}$$

Measurement equations: ordered probit

Contribution to the likelihood

$$
\begin{aligned}
\Pr(I_i = j_i) &= \Pr(\tau_{i-1} \leq I_i^* \leq \tau_i) \\
&= \Pr(\tau_{i-1} \leq \beta_{0i}^m + \beta_i^m z + \sigma_i^* \varepsilon_i^* \leq \tau_i) \\
&= \Pr\left(\frac{\tau_{i-1} - \beta_{0i}^m - \beta_i^m z}{\sigma_i^*} < \varepsilon_i^* \leq \frac{\tau_i - \beta_{0i}^m - \beta_i^m z}{\sigma_i^*}\right) \\
&= \Phi\left(\frac{\tau_i - \beta_{0i}^m - \beta_i^m z}{\sigma_i^*}\right) - \Phi\left(\frac{\tau_{i-1} - \beta_{0i}^m - \beta_i^m z}{\sigma_i^*}\right).
\end{aligned}
$$

Choice model

Specification table

	Public transp.	Car	Slow modes
β_1	0	1	0
β_2	0	0	1
β_3'	Travel time	0	0
β_5'		Travel time	0
β_7	Waiting time	0	0
β_8	Cost if HWH	Cost if HWH	0
β_9	Cost if not HWH	Cost if not HWH	0
β_{10}	0	0	Distance

Travel time coefficients

$$\beta_3' = \beta_3 e^{\beta_4 \text{CarLovers}}$$

$$\beta_5' = \beta_5 e^{\beta_6 \text{CarLovers}}$$

Value of time

Public transportation — HWH

$$\text{VOT} = \frac{\beta_3 e^{\beta_4 \text{CarLovers}}}{\beta_8}$$

Car — HWH

$$\text{VOT} = \frac{\beta_5 e^{\beta_6 \text{CarLovers}}}{\beta_8}$$

Model estimation

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

Measurement equations

Envir01 Fuel price should be increased to reduce congestion and air pollution.

$$I_1^* = -z$$

Envir02 More public transportation is needed, even if taxes are set to pay the additional costs.

$$I_2^* = 0.460 - 0.459z + 0.918\varepsilon_2^*$$

Envir03 Ecology disadvantages minorities and small businesses.

$$I_3^* = -0.367 + 0.484z + 0.857\varepsilon_3^*$$

Measurement equations

Mobil11 It is difficult to take the public transport when I carry bags or luggage.

$$I_{11}^{*} = 0.418 + 0.572z + 0.895\varepsilon_{11}^{*}$$

Mobil14 When I take the car I know I will be on time.

$$I_{14}^{*} = -0.173 + 0.575z + 0.760\varepsilon_{14}^{*}$$

Mobil16 I do not like changing the mean of transport when I am traveling.

$$I_{16}^{*} = 0.147 + 0.525z + 0.873\varepsilon_{16}^{*}$$

Mobil17 If I use public transportation I have to cancel certain activities I would have done if I had taken the car.

$$I_{17}^{*} = 0.140 + 0.514z + 0.877\varepsilon_{17}^{*}$$

Structural equation

age_65_more	0.0411
moreThanOneCar	0.710
moreThanOneBike	-0.366
individualHouse	-0.116
male	0.0773
haveChildren	-0.0253
haveGA	-0.743
highEducation	-0.267
ContIncome_0_4000	0.147
ContIncome_4000_6000	-0.281
ContIncome_6000_8000	0.322
ContIncome_8000_10000	-0.666
ContIncome_10000_more	0.119

Choice model

Specification table

	Public transp.	Car	Slow modes	
β_1	0	1	0	0.703
β_2	0	0	1	0.261
β_3	Travel time (ref)	0	0	-3.22
β_4	Travel time (att)	0	0	-0.454
β_5		Travel time (ref)	0	-9.50
β_6		Travel time (att)	0	-0.953
β_7	Waiting time	0	0	-0.0204
β_8	Cost if HWH	Cost if HWH	0	-1.43
β_9	Cost if not HWH	Cost if not HWH	0	-0.525
β_{10}	0	0	Distance	-1.41

Outline

Conclusion

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