Empirical Identification Issues in Semi-Ordered Lexicographic Model

Toshiyuki Yamamoto, Nagoya Univ.

Shinya Kurauchi, Ehime Univ.

Takayuki Morikawa, Nagoya Univ.

Outline

- Background & Objective
- Semi-Ordered lexicographic model
- Estimation methods
- Simulated choice data
- Estimation results
- Conclusions

Background

- Many empirical analyses have criticized conventional discrete choice models
  - consider all the alternatives available
  - correctly recognize each attribute of each alternative
  - evaluate each alternative independently from other alternatives

- Alternative models have been developed

- Lexicographic rule and Elimination-by-Aspect
  - alternatives are processed attribute-by-attribute

- Two-stage decision making process
  - choice set generation stage & choice stage

- Thresholds for perception of attribute values
  - too small change in attribute value is not recognized

Studies suggest that

- sequential decision making rule is realistic
- perception of attributes and the evaluation of alternative may include some threshold

Objective

- Individuals employ more sequentially simplified decision making rules than linear-in-attributes utility function
- However, estimation of such simplified decision rules is not necessarily simpler than linear-in-attributes utility function because of a high non-linearity

- Empirical estimability of semi-ordered lexicographic model is investigated
  - Data mining algorithm as well as maximum likelihood estimation are examined

Semi-ordered lexicographic model

- Evaluation is conducted in the attribute-by-attribute lexicographic order.

- An alternative is rejected if the concerned attribute value of the alternative is inferior beyond the tolerable gap to the “best” alternative.

- If more than one alternative remain after an attribute is evaluated, next most important attribute is evaluated.

Semi-ordered lexicographic model

- The probability of alternative *i* is being rejected with respect to attribute *j* by individual *n*

$$q_{in}(j) = 0, \text{ if } \quad X_{inj} = X_{*nj}$$

$$q_{in}(j) = \Pr\left(\left|X_{inj} - X_{*nj}\right| > \tau_{nj}\right), \text{ if } \quad X_{inj} \neq X_{*nj}$$

Assuming threshold follow normal distribution

$$q_{in}(j) = \Phi\left(\frac{\left|X_{inj} - X_{*nj}\right| - \nu_j}{\sigma_{nj}}\right)$$

Semi-ordered lexicographic model

- Given that there are only 2 alternatives and 2 attributes
- Importance ranking of the attributes is attributes 1 and 2
- Probability that an individual choose alternative 1 is given as

$$P_n(1) = \bar{q}_{1n}(1) \cdot q_{2n}(1)$$

⇦ Choice is determined with the first attribute

$$+ \bar{q}_{1n}(1) \cdot \bar{q}_{1n}(2) \cdot \bar{q}_{2n}(1) \cdot q_{2n}(2)$$

⇦ With the second attribute

$$+ \bar{q}_{1n}(1) \cdot \bar{q}_{1n}(2) \cdot \bar{q}_{2n}(1) \cdot \bar{q}_{2n}(2) \cdot 1/2$$

⇦ 2 alternatives remain after 2 attributes are evaluated, so pure random choice is employed

where $\bar{q}_{in}(j) = 1 - q_{in}(j)$

Semi-ordered lexicographic model

- Given that there are only 3 alternatives and 3 attributes
- Importance ranking of the attributes is attributes 1, 2 and 3
- Probability that an individual choose alternative 1 is given as

$$
\begin{aligned}
P_n(1) = {} & \bar{q}_{1n}(1)\cdot q_{2n}(1)\cdot q_{3n}(1) \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{2n}(1)\cdot q_{2n}(2)\cdot q_{3n}(1) \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot q_{2n}(1)\cdot \bar{q}_{3n}(1)\cdot q_{3n}(2) \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{2n}(1)\cdot q_{2n}(2)\cdot \bar{q}_{3n}(1)\cdot q_{3n}(2) \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{1n}(3)\cdot \bar{q}_{2n}(1)\cdot \bar{q}_{2n}(2)\cdot q_{2n}(3)\cdot q_{3n}(1) \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{1n}(3)\cdot q_{2n}(1)\cdot \bar{q}_{3n}(1)\cdot \bar{q}_{3n}(2)\cdot q_{3n}(3) \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{1n}(3)\cdot \bar{q}_{2n}(1)\cdot \bar{q}_{2n}(2)\cdot q_{2n}(3)\cdot \bar{q}_{3n}(1)\cdot q_{3n}(2) \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{1n}(3)\cdot \bar{q}_{2n}(1)\cdot q_{2n}(2)\cdot \bar{q}_{3n}(1)\cdot \bar{q}_{3n}(2)\cdot q_{3n}(3) \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{1n}(3)\cdot \bar{q}_{2n}(1)\cdot \bar{q}_{2n}(2)\cdot q_{2n}(3)\cdot \bar{q}_{3n}(1)\cdot \bar{q}_{3n}(2)\cdot q_{3n}(3) \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{1n}(3)\cdot \bar{q}_{2n}(1)\cdot \bar{q}_{2n}(2)\cdot \bar{q}_{2n}(3)\cdot q_{3n}(1)\cdot 1/2 \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{1n}(3)\cdot q_{2n}(1)\cdot \bar{q}_{3n}(1)\cdot \bar{q}_{3n}(2)\cdot \bar{q}_{3n}(3)\cdot 1/2 \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{1n}(3)\cdot \bar{q}_{2n}(1)\cdot \bar{q}_{2n}(2)\cdot \bar{q}_{2n}(3)\cdot \bar{q}_{3n}(1)\cdot q_{3n}(2)\cdot 1/2 \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{1n}(3)\cdot \bar{q}_{2n}(1)\cdot q_{2n}(2)\cdot \bar{q}_{3n}(1)\cdot \bar{q}_{3n}(2)\cdot \bar{q}_{3n}(3)\cdot 1/2 \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{1n}(3)\cdot \bar{q}_{2n}(1)\cdot \bar{q}_{2n}(2)\cdot \bar{q}_{2n}(3)\cdot \bar{q}_{3n}(1)\cdot \bar{q}_{3n}(2)\cdot q_{3n}(3)\cdot 1/2 \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{1n}(3)\cdot \bar{q}_{2n}(1)\cdot \bar{q}_{2n}(2)\cdot q_{2n}(3)\cdot \bar{q}_{3n}(1)\cdot \bar{q}_{3n}(2)\cdot \bar{q}_{3n}(3)\cdot 1/2 \\
& + \bar{q}_{1n}(1)\cdot \bar{q}_{1n}(2)\cdot \bar{q}_{1n}(3)\cdot \bar{q}_{2n}(1)\cdot \bar{q}_{2n}(2)\cdot \bar{q}_{2n}(3)\cdot \bar{q}_{3n}(1)\cdot \bar{q}_{3n}(2)\cdot \bar{q}_{3n}(3)\cdot 1/3
\end{aligned}
$$

Estimation methods

1. Maximum likelihood estimation

- importance ranking of the attributes should be predetermined
  - Search for correct ranking becomes extremely time-consuming with large numbers of attributes

Motivation of alternative estimation techniques shown below

2. C4.5 algorithm

- one of supervised learning algorithms
- algorithm subdivides the sample cases recursively into segments based on attributes
- the decision tree is generated by top down sequence

importance ranking of the attributes is determined in the algorithm

- recursive subdivision is based on the concept of entropy

3. Hybrid estimation procedure

1) C4.5 algorithm is applied to identify

- the order of the importance ranking
- mean of the threshold
- (standard deviation of the threshold is not obtained by C4.5 algorithm)

2) Maximum likelihood estimation is executed

- using mean of the threshold estimated by C4.5 as the starting values

Maximum likelihood estimation is executed just once, so the computational burden is greatly reduced

Simulated choice data

- 1,000 hypothetical individuals with the semi-ordered lexicographic rule
- Travel mode choice: car, train and bus
- 3 attributes: travel time, travel cost and access time (in order of importance)
- Person trip survey data is used to generate the synthetic data

	Mean	s.d.	Max	Min
Travel time (m)	20.4	8.3	69	9
Travel cost (JPY)	117.1	163.9	1938	13
Access time (m)	12.9	8.8	24	1

- Heterogeneity among individual in threshold

Synthetic threshold

	Travel time (m)		Travel cost (JPY)		Access time (m)	
Heterogeneity	Mean	s.d.	Mean	s.d.	Mean	s.d.
Small	5	0.5	20	2	1	0.1
Medium	5	2.5	20	10	1	0.5
Large	5	10.0	20	40	1	2.0

- Correlation between travel time and travel cost

Distribution of the number of compared attributes by correlation coefficient

	Correlation coefficient		
	+0.7	0.0	-0.7
1 (Travel time)	45 %	45 %	45 %
2 (Travel time and travel cost)	35 %	45 %	35 %
3 (Travel time, travel cost and access time)	20 %	10 %	20 %
Total	100 %	100 %	100 %

Estimation results

1. Maximum likelihood estimation
   – Convergence & identification of importance ranking
   – Empirical efficiency of parameter estimates

2. C4.5 algorithm
   – Convergence & identification of importance ranking
   – Empirical efficiency of parameter estimates

3. Hybrid estimation procedure
   – Convergence & identification of importance ranking

4. Conventional MNL model

5. Comparison of the goodness-of-fit statistic

1. Maximum likelihood estimation

- 2 alternatives and 2 attributes case

Convergence of the estimation

Correlation	+0.7			0.0			-0.7		
Heterogeneity	Small	Med.	Large	Small	Med.	Large	Small	Med.	Large
Correct order	80%	100%	90%	100%	90%	100%	100%	70%	70%
Wrong order	80%	90%	60%	10%	20%	20%	10%	30%	50%

Adjusted ρ^2

Correlation	+0.7			0.0			-0.7		
Heterogeneity	Small	Med.	Large	Small	Med.	Large	Small	Med.	Large
Correct order	0.87	0.82	0.77	0.85	0.78	0.68	0.72	0.61	0.43
Wrong order	0.70	0.69	0.66	0.46	0.44	0.40	0.10	0.07	0.05

- Convergence rate is satisfactory
- Convergence rate and adjusted ρ^2 are higher with correct order than with wrong order

1. Maximum likelihood estimation

- 3 alternatives and 3 attributes case

Convergence of the estimation

Correlation	+0.7			0.0			-0.7		
Heterogeneity	Small	Med.	Large	Small	Med.	Large	Small	Med.	Large
Correct order	20%	40%	0%	0%	20%	20%	30%	10%	0%
Wrong order	0%	10%	10%	0%	20%	0%	0%	0%	0%

Adjusted ρ^2

Correlation	+0.7			0.0			-0.7		
Heterogeneity	Small	Med.	Large	Small	Med.	Large	Small	Med.	Large
Correct order	0.81	0.77	-	-	0.76	0.66	0.73	0.60	-
Wrong order	-	0.58	0.57	-	0.47	-	-	-	-

- Convergence rate is insufficient
- Convergence rate and adjusted ρ^2 are higher with correct order than with wrong order

1. Maximum likelihood estimation

Parameter estimates (corr. = -0.7, heterogeneity = med.)

Threshold	True value	Converged (s.e.)	Not converged
Travel time: mean	5.00	4.89 (0.09)	4.95
Travel time: log(s.d.)	0.92	0.44 (0.06)	0.53
Travel cost: mean	20.00	16.67 (1.26)	18.77
Travel cost: log(s.d.)	2.30	1.68 (0.24)	1.63
Access time: mean	1.00	4.46 (445.9)	9.29
Access time: log(s.d.)	-0.69	17.94 (896.9)	17.02
Log-likelihood at convergence		-428.60	-463.36
Adjusted ρ^2		0.60	0.57

- Parameters on the higher order of the importance ranking cannot be estimated well
- Estimation procedure could not find the optimum values on such parameters

2. C4.5 algorithm

- Convergence: No problem. Always converged
- Identification of importance ranking: No problem with correlation coefficient of 0.0 and -0.7, but misidentify order with correlation coefficient of +0.7

Parameter estimates (Standard error in parenthesis)

Correlation	+0.7			0.0			-0.7		
Heterogeneity	Small	Med.	Large	Small	Med.	Large	Small	Med.	Large
Travel time (True = 5.0)	4.5 (0.00)	4.1 (0.84)	3.5 (1.58)	5.4 (0.32)	5.4 (0.74)	5.2 (1.25)	5.5 (0.00)	6.0 (0.53)	7.1 (0.84)
Travel cost (True = 20.0)	19.7 (0.97)	18.4 (1.73)	19.1 (1.65)	20.8 (0.48)	21.0 (1.72)	19.5 (3.27)	21.0 (2.12)	20.9 (2.24)	20.5 (5.07)

- Thresholds for travel time and travel cost are well estimated
- Parameters on the higher order of the importance ranking cannot be estimated well

3. Hybrid estimation procedure

Convergence of the estimation by starting value

Correlation	0.0		-0.7	
Heterogeneity	Small	Med.	Small	Med.
Starting values are set as 0 for all the parameters	0%	20%	30%	10%
Starting values are set at the estimates by C4.5 algorithm	30%	0%	50%	20%
Starting values are set at the true values	30%	70%	60%	70%

- By using estimated values of C4.5 algorithm as starting values, the convergence rate improves when the heterogeneity is small
- Even if the true values are used as the starting values, the convergence is not perfect.

4. Conventional MNL model

Parameter estimates (heterogeneity = med.)

Correlation	+0.7	0.0	-0.7
Travel time	-27.20	-24.89	-19.50
(s.e.)	(-12.89)	(-18.51)	(-17.30)
Travel cost	-52.48	-9.85	-5.58
(s.e.)	(-12.03)	(-12.10)	(-8.19)
Access time	-4.28	-1.28	-0.89
(s.e.)	(-2.73)	(-1.16)	(-0.99)
Train constant	-0.67	-0.34	-0.19
(s.e.)	(-2.04)	(-1.45)	(-0.98)
Bus constant	-0.78	-0.39	-0.28
(s.e.)	(-2.29)	(-1.61)	(-1.43)
Log-likelihood at convergence	-250.59	-461.94	-728.35
Adjusted ρ^2	0.77	0.58	0.34

- Coefficient estimates look very natural
- Wrong assumption on the decision rule can not be identified

5. Comparison of the goodness-of-fit

Adjusted ρ^2 (heterogeneity = med.)

Correlation	+0.7	0.0	-0.7
Semi-ordered lexicographic model	0.78	0.76	0.59
C4.5 algorithm	0.81	0.71	0.60
Multinomial logit model	0.77	0.58	0.34

- Maximum likelihood estimation and C4.5 algorithm have almost the same goodness-of-fit
- Accuracy of the approximation by MNL model is low for the case with negative correlation among attributes

Conclusions

- MNL model may not approximate the non-compensatory decision making rule, especially when negative correlation among attributes
- Maximum likelihood estimation can avoid the misspecification of the order by lower goodness-of-fit
- However, even if the correct order is predetermined, the convergence rate is low especially for the case with negative correlation of the attributes.
- C4.5 algorithm has no problem in convergence, but generates wrong order of compared attributes for the case with positive correlation.
- Hybrid method improves the practical estimability only for the case with small heterogeneity in the thresholds