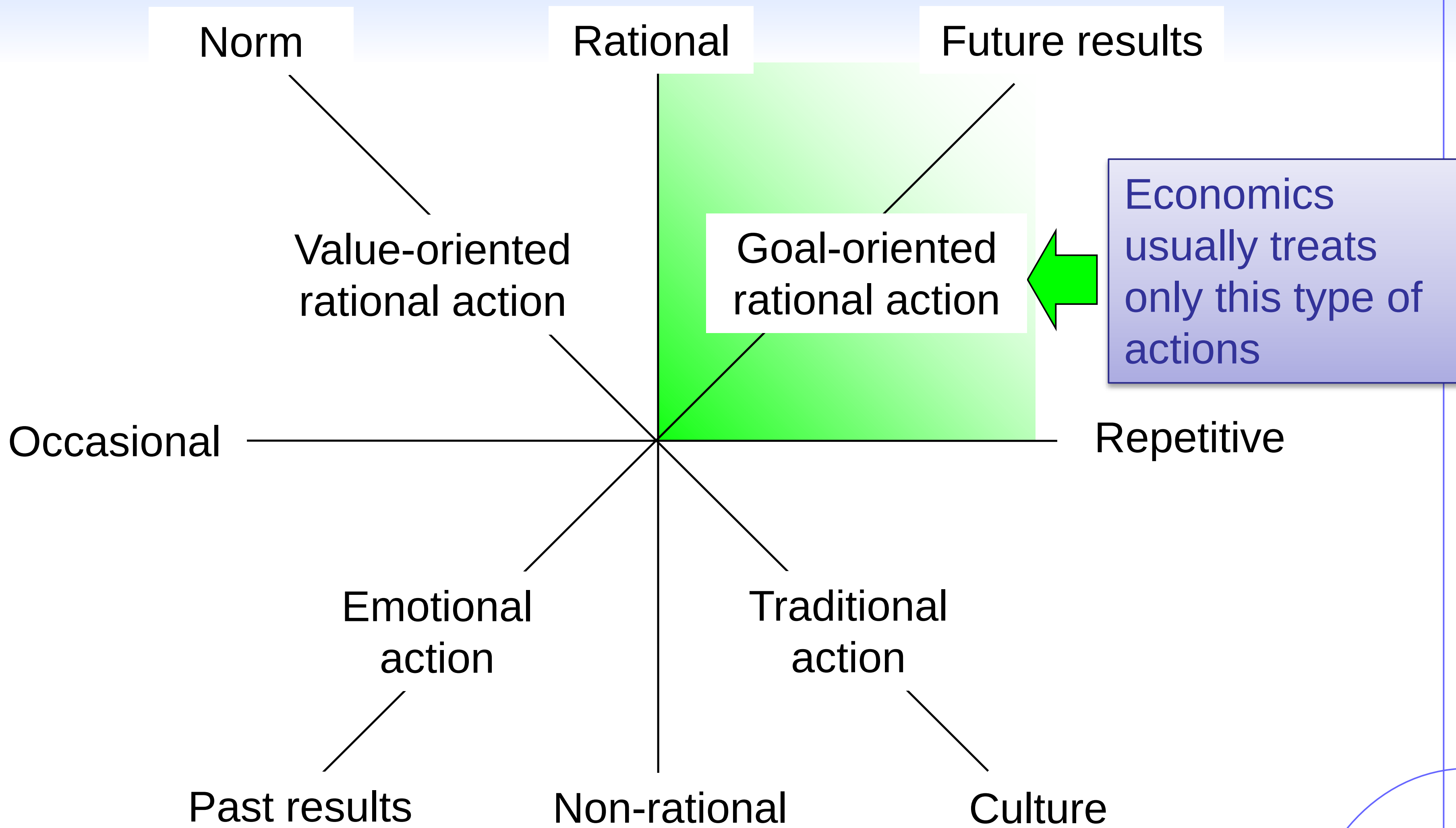
Incorporating Psychological Factors into Discrete Choice Models

September 15, 2018

Takayuki Morikawa
Institutes of Innovation for Future Society
Nagoya University

Max Weber's Four Types of Social Action



Goal-oriented Rational Action

Consumer's Behavior

Tries to reach the goal (become happiest) by consuming various goods under some constraints (typically "income").

$$\text{Max. } U=f(x_1, x_2, x_3,\ldots, x_n)$$

s.t.

$$p_1x_1+p_2x_2+\ldots+p_nx_n = y$$

U : Utility   x : Goods   p : Price   y : Income

Goal-oriented Rational Action

$$\text{Max. } U=f(x_1, x_2, x_3,\ldots, x_n) \qquad \text{Eq. (1)}$$

s.t.

$$p_1x_1+p_2x_2+\ldots+p_nx_n = y \qquad \text{Eq. (2)}$$

Plug Eq.(2) in Eq. (1) and solve the maximization problem to obtain the following indirect utility function:

$$U'=f(p_1, p_2,\ldots,p_n; y)$$

Discrete Choice Model

Instead of solving the big model that determines the amount of all the goods x_1, x_2, x_3,..., x_n, we can deal with one or a few goods by looking at only those goods and ignoring the effects to the other goods.

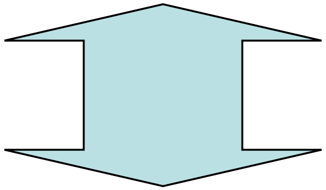
In this kind of small decisions, one alternative is chosen among discrete ones in many situations.

<Examples>

- What mode should I use among rail, bus, and car.
- Should I go to downtown or a suburban mall for shopping?
- What brand of beer should I buy?

Utility Maximization in Discrete Choice

Individual n chooses an alternative i from a choice set $\{1,2,\ldots,J_n\}$



$$U_n(i) > U_n(j) \qquad j=1,2,\ldots,J_n \quad i \neq j$$

$U_n(i)$: (Indirect) Utility function of individual n
choosing an alternative i

Random Utility Function

Unobservable factors in a model

→ Random variable

Random utility function:

$$U_n(i) = V_n(i) + \varepsilon_n(i)$$

$V_n(i)$: utility part by observable factors

$\varepsilon_n(i)$: utility part by unobservable factors (random variable)

$V_n(i)$ is often represented by the linearly additive form of observable factors and their weights

$$V_n(i) = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}$$

Logit Model

If we assume the Gumbel distribution for the random part of utility, then the choice probability is given by:

$$P_n(i) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}}$$

The unknown weights, $\beta_1, \beta_2, \ldots, \beta_k$, are estimated so that the observed choice results are best fitted.

→ Maximum Likelihood Estimation

Daniel McFadden
(UC Berkeley)

Nobel laureate in Economics in 2000

Data Used for Model Estimation

Choice Results

- RP (Revealed Preference) data

Choice results in real the market

- SP (Stated Preference) data

→ **Psychometric data**

Preference information in hypothetical situations

"If MAGLEV trains connect Tokyo and Osaka by one hour and 20,000 yen, which will you choose, the MAGLEV or the existing Shinkansen?"

Data Used for Model Estimation

Attributes of alternatives

- Objective attributes
- Subjective perception → **Psychometric data**

  e.g., comfort, beauty

Attributes of decision-makers

- Objective socio-economic attributes
- Subjective attitude → **Psychometric data**

  e.g., attitude toward reliability, price consciousness

Why Combine Economic and Psychometric Data?

Economic data alone can be used to predict market response to new products or marketing programs if historical natural experiments are rich enough, or field experiments on the new product are feasible.

→ *But often historical natural experiments are inadequate and field experiments are impractical.*

Why Combine Economic and Psychometric Data?

Psychometric data alone can provide useful direct information on perceptions, values and preferences.

→ *But psychometric tasks may elicit different cognitive protocols than market decisions, and psychometric scaling criteria do not necessarily maximize the market behavior predictive power of psychometric data.*

Why Combine Economic and Psychometric Data?

Synthesis of the two types of data:

→ provides a framework for translating psychometric data into bottom line forecasts of market sales and profitability.

→ yields statistical methods for testing hypotheses and assessing the precision of forecasts.

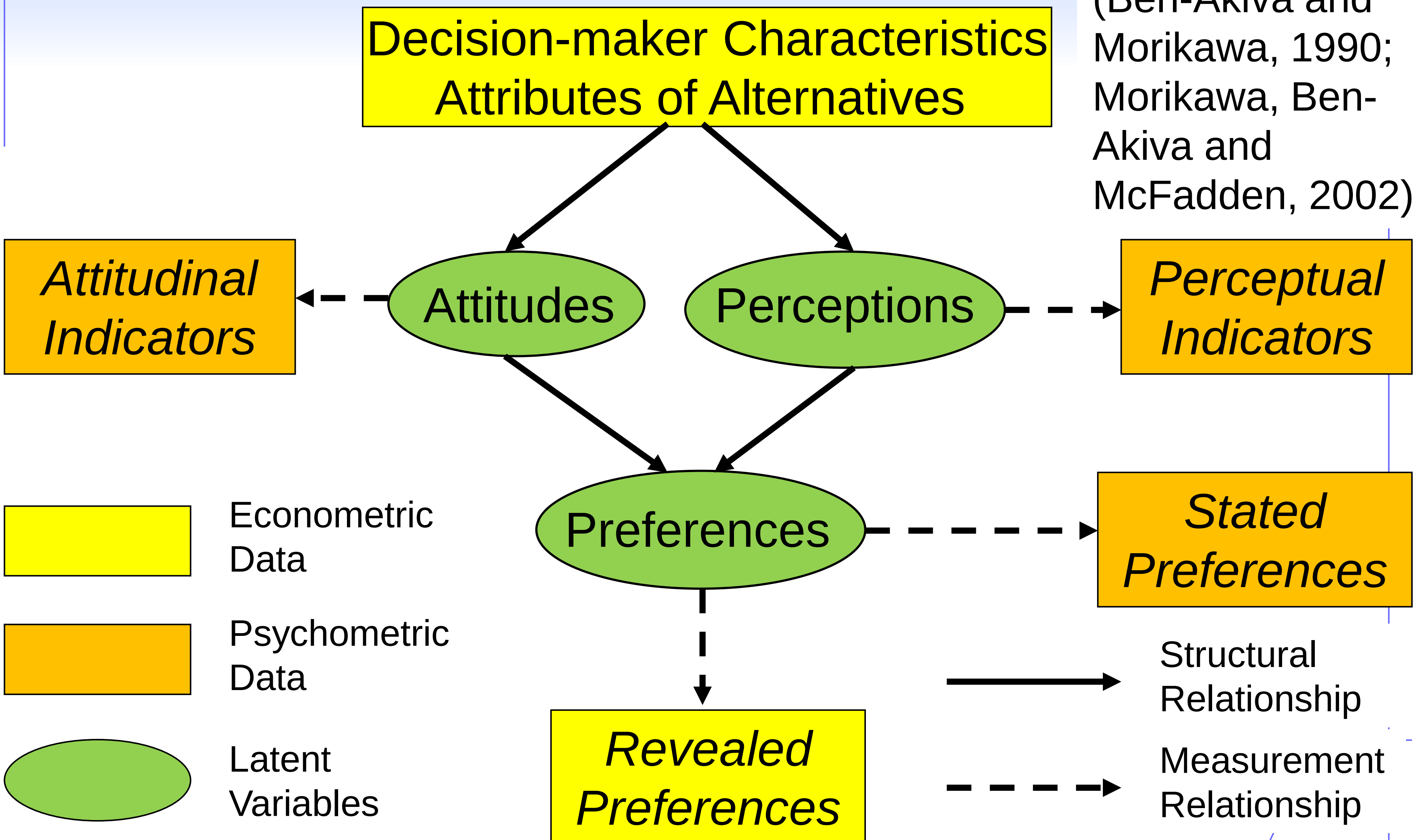
→ extends economic market forecasting methodology so that it can handle new products and marketing programs.

Comparison of RP and SP Data

	RP Data	SP Data
Preference	Choice behavior in actual market Cognitively congruent with actual behavior	Preference statement for hypothetical scenarios May be cognitively incongruent with actual behavior
Alternatives	Actual alternatives Responses to non-existing alternatives are not observable	Generated alternatives Can elicit preference for new (non-existing) alternatives
Attributes	May include measurement errors Correlated attributes Ranges are limited	No measurement errors Multicollinearity can be avoided Ranges can be extended
Choice set	Ambiguous in many cases	Prespecified
Number of responses	Difficult to obtain multiple responses by questionnaire	Repetitive questioning is easily implemented
Response format	Choice	Choice, ranking, rating, matching

Integrated Framework for Demand Analysis

(Ben-Akiva and Morikawa, 1990; Morikawa, Ben-Akiva and McFadden, 2002)



Framework: A Binary Choice Model with Latent Attributes

- **Structural Equations:**

$$u^{*RP} = \mathbf{a}'\mathbf{x}^{RP} + \mathbf{b}'\mathbf{w}^{RP} + \mathbf{c}'\mathbf{w}^{*RP} + \nu^{RP}$$

$$u^{*SP} = \mathbf{a}'\mathbf{x}^{SP} + \mathbf{e}'\mathbf{z}^{SP} + \nu^{SP}$$

$$\mathbf{w}^{*RP} = \mathbf{B}\mathbf{s}^{RP} + \zeta^{RP}$$

where

u^* = latent utility;

$\mathbf{x}$, $\mathbf{w}$, $\mathbf{z}$ = vectors of observable explanatory variables;

$\mathbf{w}^*$ = vector of latent explanatory variables;

$\mathbf{s}$ = vector of observable variables that influence $\mathbf{w}^*$;

$\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{e}$, $\mathbf{B}$ = arrays of unknown parameters;

ν = random component of utility; and

ζ = vector of normally distributed disturbances.

- **Measurement Equations:**

$$d^{RP} = \begin{cases} 1, & \text{if } u^{*RP} \geq 0 \\ -1, & \text{if } u^{*RP} < 0 \end{cases}$$

$$d^{SP} = \begin{cases} 1, & \text{if } u^{*SP} \geq 0 \\ -1, & \text{if } u^{*SP} < 0 \end{cases}$$

$$\mathbf{y}^{RP} = \mathbf{\Lambda} \mathbf{w}^{*RP} + \boldsymbol{\varepsilon}^{RP}$$

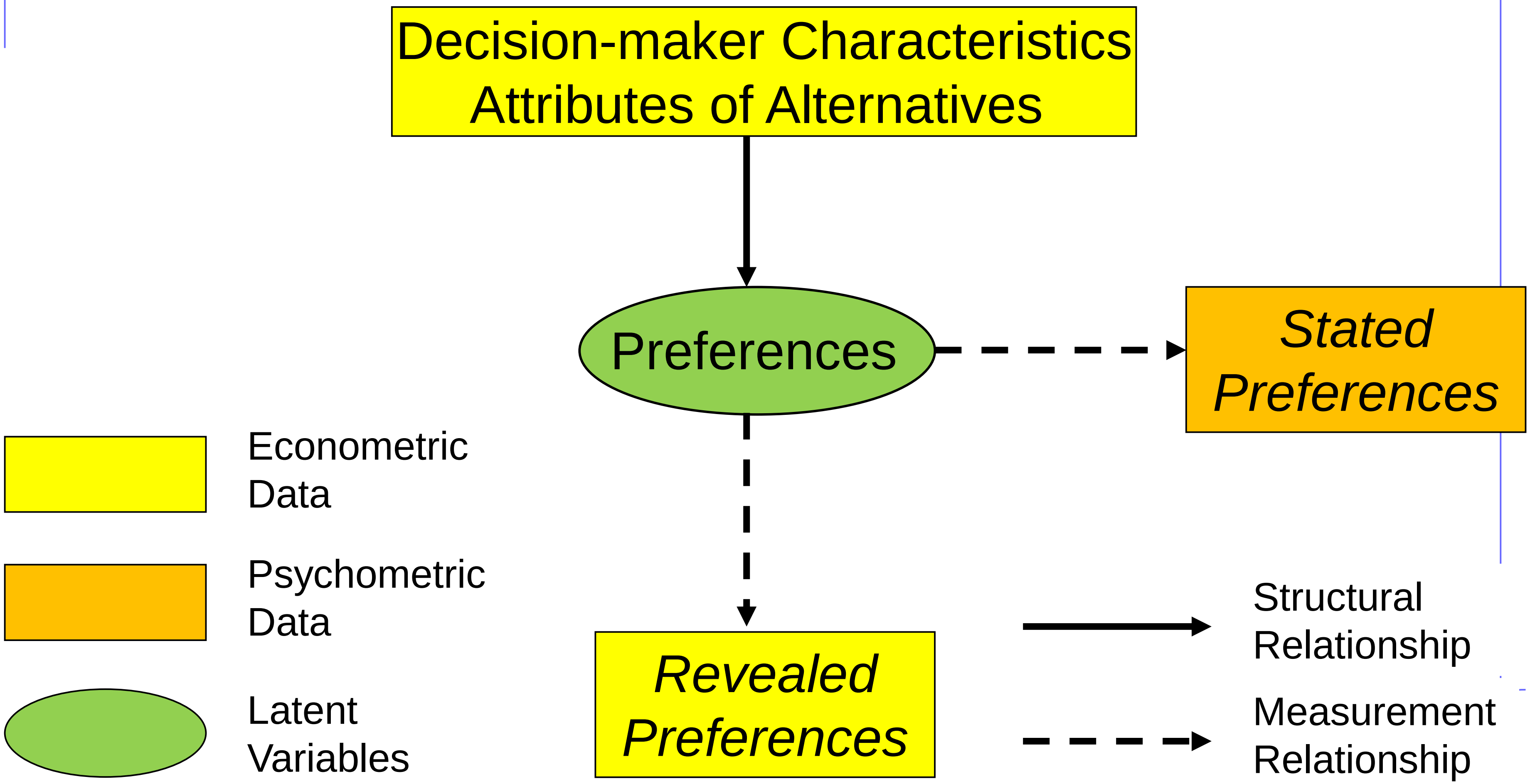
where

$\mathbf{y}$ = vector of observed indicators of $\mathbf{w}^*$;

Λ = matrix of unknown parameters; and

ε = vector of normally distributed disturbances.

Submodel 1: Combined Estimation with RP and SP Data



Submodel 1: Combined Estimation with RP and SP Data

Key Features of the Methodology

1. **Bias correction** through explicitly specifying the SP model
2. **Improving the efficiency** by jointly estimating the parameters from all the available data
3. **Identifying** the effect of the new services using SP data

Submodel 1: Combined Estimation with RP and SP Data

RP model:

$$u^{*\mathrm{RP}} = \mathbf{a}'\mathbf{x}^{\mathrm{RP}} + \mathbf{b}'\mathbf{w}^{\mathrm{RP}} + \nu^{\mathrm{RP}}$$

$$d^{\mathrm{RP}} = \begin{cases} 1, & \text{if } u^{*\mathrm{RP}} \geq 0 \\ -1, & \text{if } u^{*\mathrm{RP}} < 0 \end{cases}$$

SP model:

$$u^{*\mathrm{SP}} = \mathbf{a}'\mathbf{x}^{\mathrm{SP}} + \mathbf{e}'\mathbf{z}^{\mathrm{SP}} + fd^{\mathrm{RP}} + \nu^{\mathrm{SP}}$$

$$d^{\mathrm{SP}} = \begin{cases} 1, & \text{if } u^{\mathrm{SP}} \geq 0 \\ -1, & \text{if } u^{\mathrm{SP}} < 0 \end{cases}$$

Correction of the scale:

$$Var(v^{RP}) = \mu^2 Var(v^{SP})$$

Systematic utility used for prediction:

$$v = \mathrm{a'x} + \mathrm{b'w} + \mathrm{e'z}$$

Joint Log-Likelihood to be maximized:

$$\mathrm{L}(\mathbf{a},\mathbf{b},\mathbf{e},f,\mu)=\sum_{n=1}^{N^{\mathrm{RP}}}\log\left\{\Phi\left[d_n^{\mathrm{RP}}\left(\mathbf{a}'\mathbf{x}_n^{\mathrm{RP}}+\mathbf{b}'\mathbf{w}_n^{\mathrm{RP}}\right)\right]\right\}$$

$$+\sum_{n=1}^{N^{\mathrm{SP}}}\log\left\{\Phi\left[d_n^{\mathrm{SP}}\mu\left(\mathbf{a}'\mathbf{x}_n^{\mathrm{SP}}+\mathbf{e}'\mathbf{z}_n^{\mathrm{SP}}\right)\right]\right\}$$

Due to the introduction of scale correction parameter μ, non-linearity in parameters occurs.

Sequential estimation method:

Step 1: Estimate SP model and calculate the fitted utility value

Step 2: Estimate RP model with RP specific variables and the above fitted values multiplied by the unknown scale parameter

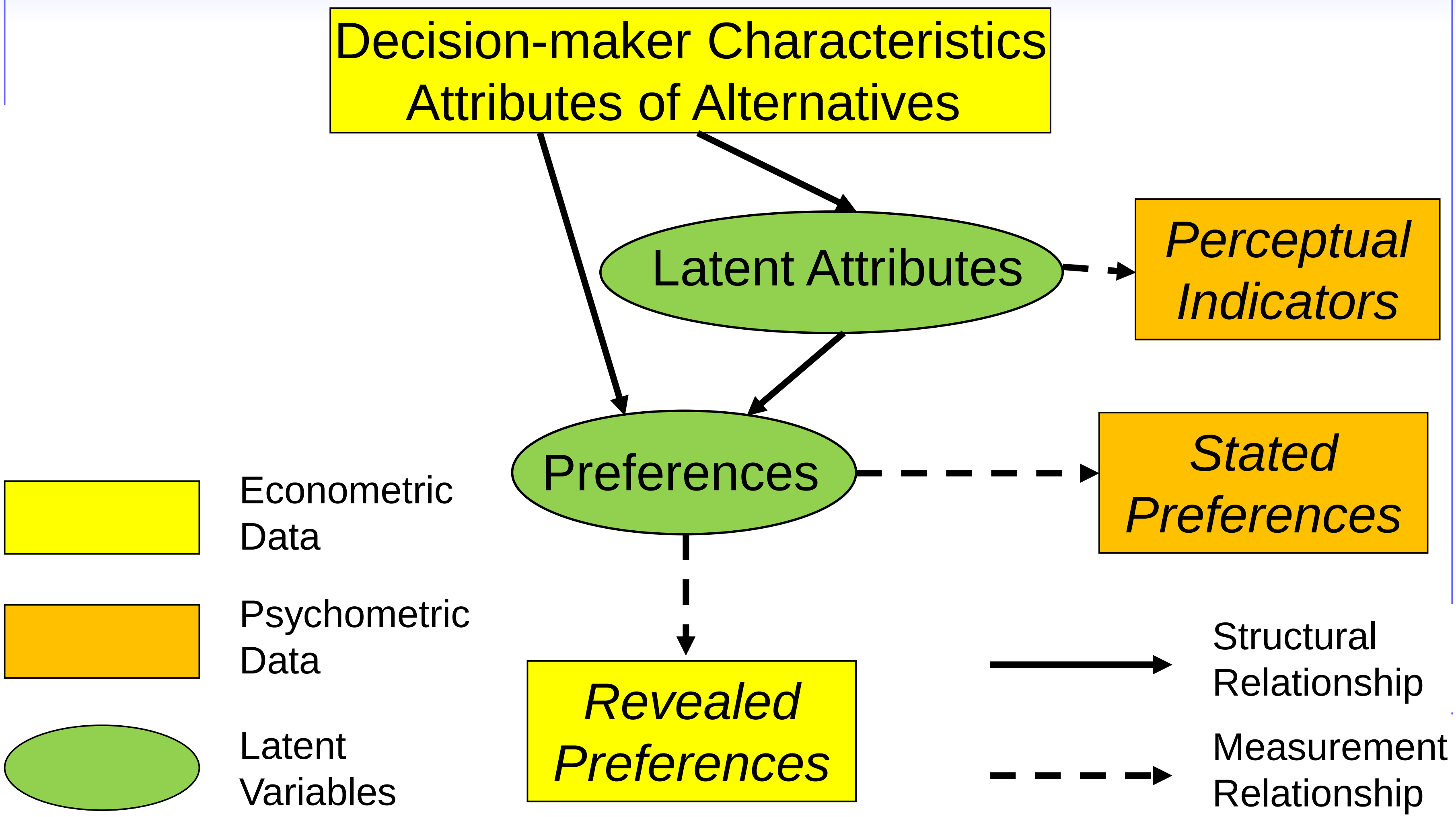
Can use an estimation package software

Simultaneous estimation method:

Maximize the joint log-likelihood

Requires programming

Submodel 2: Discrete Choice Models with Latent Attributes



Submodel 2: Discrete Choice Models with Latent Attributes

Key Features of the Methodology

1. Include latent attributes such as reliability and comfort as explanatory variables

2. Use psychometric data such as subjective rating of attributes only as indicators of the latent attributes

3. Predicted values of the latent attributes can be obtained without future values of psychometric data

Submodel 2: Discrete Choice Models with Latent Attributes

Structural equations:

$$u^* = \mathbf{a}'\mathbf{x} + \mathbf{c}'\mathbf{w}^* + \nu$$

$$\mathbf{w}^* = \mathbf{B}\mathbf{s} + \boldsymbol{\zeta}$$

Measurement equations:

$$d = \begin{cases} 1, \text{ if } u^* \geq 0 \\ -1, \text{ if } u^* < 0 \end{cases}$$

$$\mathbf{y} = \boldsymbol{\Lambda}\mathbf{w}^* + \boldsymbol{\varepsilon}$$

Sequential Estimation Method

Step 1: Use a LISREL type software and calculate the fitted values:

$$\widehat{\mathbf{w}^*} = \hat{\mathbf{B}}\mathbf{s} + \hat{\Psi}\hat{\Lambda}'[\hat{\Lambda}\hat{\Psi}\hat{\Lambda}' + \hat{\Theta}]^{-1}(\mathbf{y} - \hat{\Lambda}\hat{\mathbf{B}}\mathbf{s})$$

$$\hat{\omega} = \hat{\Psi} - \hat{\Psi}\hat{\Lambda}'[\hat{\Lambda}\hat{\Psi}\hat{\Lambda}' + \hat{\Theta}]^{-1}\hat{\Lambda}\hat{\Psi}$$

Step 2: Estimate the probit model using the above fitted values, i.e., estimate *a*, b and c using the following choice probability:

$$P(d \mid \mathbf{y}, \mathbf{x}, \mathbf{s}) = \Phi\left(d \, \frac{\mathbf{a}'\mathbf{x} + \mathbf{c}'\widehat{\mathbf{w}^*}}{\sqrt{1 + \mathbf{c}'\hat{\omega}\mathbf{c}}} \right).$$

Findings

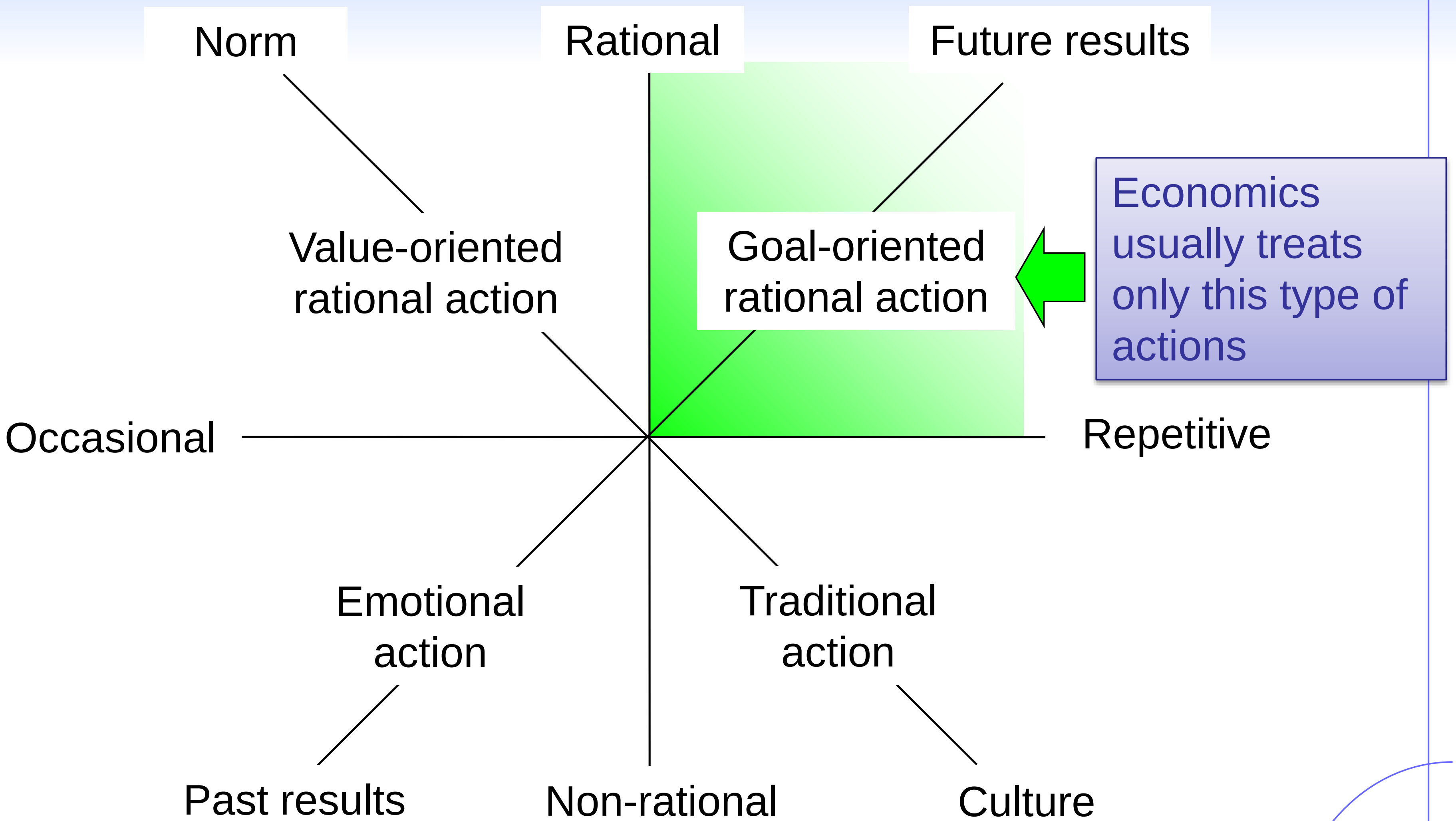
1. The RP/SP combined modeling is found to be a strong method to analyze the demand for new services and obtain robust trade-off parameters such as the value-of-time by many empirical researches.

2. It has been also utilized to assess the quantitative value of the intangible, i.e., monetary value of environmental quality.

3. Discrete choice models with latent attributes are effective incorporate intangible factors, i.e., comfort and privacy.

4. Latent attitudes can be incorporated by modeling the latent heterogeneity of preference parameters.

Max Weber's Four Types of Social Action



Treatment of Non-"Goal-oriented Rational Action"

- Bounded Rationality
  - People only can behave rationally in a bounded way because their ability of information processing is limited.
  - Can we really evaluate the trade-offs among many attributes of alternatives?
    - Compensatory rule:

$$V = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots$$

  - We often use easier (or brain-resource-saving) rule in decision-making.
    - Non-compensatory rule:
      - Conjunctive → minimum standards in some attributes
      - Disjunctive → one outstanding attribute covers the others

Two Stage Choice Models with Compensatory and Non-compensatory Rules

- Two Stage Probabilistic Choice Set (PCS) Model
  1) First Stage: Choice Set Formation
  2) Second Stage: Discrete Choice Given the Choice Set

(Manski, 1977)

$$P_n(i) = \sum_{C \in G} P_n(i|C) \cdot Q_n(C|G)$$

where

$P_n(i|C)$: choice probability of alternative i given choice set C

$Q_n(C|G)$: probability of n's choice set being C

G : set of all non-empty subsets of M (master set)

Practical Problem of PCS Model

$$P_n(i) = \sum_{C \in G} P_n(i|C) \cdot Q_n(C|G)$$

G is the set of all non-empty subsets of the master set.

Three alternative case:

$$G = \{\{1\},\{2\},\{3\},\{1,2\},\{2,3\},\{1,3\},\{1,2,3\}\}$$

In general, when the master set has J alternatives, G has 2^J-1 elements.

→ 1,023 elements for 10 alternative case!

Choice Set Formation Model

<Assumption>

Individuals check whether all the constraints are satisfied.

→ Non-compensatory model

Assume K independent constraints:

$$q_n(i) = \prod_{k=1}^{K} q_{kn}(i)$$ (Swait and Ben-Akiva, 1987)

where

$q_n(i)$: probability of i included in n's choice set

$q_{kn}(i)$: probability of i satisfying the k-th constraint

Latent variable > Threshold value

Choice Set Formation Model

Assume that each constraint is satisfied iff a latent variable $E_{kn}(i)$ is greater than a threshold value μ_k.

Assuming $E_{kn}(i)$ consists of a systematic part and a random part,

$$E_{kn}(i) = \boldsymbol{\alpha}_k^{\prime}\boldsymbol{w}_{kin} - \vartheta_{kin}$$

the probability that the k-th constraint is satisfied is given by,

$$\begin{aligned} q_{kn}(i) &= Prob[E_{kn}(i) \geq \mu_k] \\ &= Prob[\boldsymbol{\alpha}_k^{\prime}\boldsymbol{w}_{kin} - \vartheta_{kin} \geq \mu_k] \\ &= Prob[\vartheta_{kin} \leq \boldsymbol{\alpha}_k^{\prime}\boldsymbol{w}_{kin} - \mu_k] \end{aligned}$$

Choice Set Formation Model

Assuming ϑ_{kin} to be logistically distributed,

$$q_{kn}(i) = \frac{1}{1+e^{-(\alpha'_k w_{kin} - \mu_k)}}$$

$$q_n(i) = \prod_{k=1}^{K} \frac{1}{1+e^{-(\alpha'_k w_{kin} - \mu_k)}}$$

the probability of n's choice set being C is given by,

$$Q_n(C|G) = \frac{1}{1-Q_n(\emptyset)} \prod_{i \in M} \left\{ q_n(i)^{d_{i_C}} \cdot \left(1 - q_n(i)\right)^{1-d_{i_C}} \right\}$$

where

$Q_n(\emptyset)$: probability of the random constrain model yielding the empty choice set

$$d_{i_C} = \begin{cases} 1\text{: if alternative i is an element of choice set C} \\ 0\text{: otherwise} \end{cases}$$

Choice Probability of PCS Model

If we assuming a logit model with a compensatory utility function in the second stage, the choice probability of the PCA model is given by,

$$P_n(i) = \sum_{C \in G} P_n(i|C) \cdot Q_n(C|G)$$

$$= \frac{1}{1 - Q_n(\emptyset)} \sum_{C \in G} \left[\frac{e^{V_{in}}}{\sum_{j \in C} e^{V_{jn}}} \prod_{j \in M} \left\{ q_n(j)^{d_{jC}} \cdot \left(1 - q_n(j)\right)^{1-d_{jC}} \right\} \right]$$

Alternative Derivation of PCS Model

Alternative i being preferred to alternative j:

① i and j are included in the choice set and i has a greater utility value than j

or

② i is included in the choice set but j is cut off at the first stage (choice set formation).

(Morikawa, 1995)

Alternative Derivation of PCS Model

$$P_n(i) = \frac{1}{1-Q_n(\varnothing)} \times \text{Prob}\left(i \in C_n\right) \times \text{Prob}\begin{bmatrix} \left\{\left(1 \in C_n\right) \cap \left(U_{in} \geq U_{1n}\right)\right\} \cup \left\{1 \notin C_n\right\} \\ \text{and} \\ \left\{\left(2 \in C_n\right) \cap \left(U_{in} \geq U_{2n}\right)\right\} \cup \left\{2 \notin C_n\right\} \\ \text{and} \\ \vdots \\ \text{and} \\ \left\{\left(J \in C_n\right) \cap \left(U_{in} \geq U_{Jn}\right)\right\} \cup \left\{J \notin C_n\right\} \end{bmatrix}$$

$$= \frac{1}{1-Q_n(\varnothing)} \times q_n(i) \times \text{Prob}\left[\bigcap_{j \in M, j \neq i} \left\{\left(j \in C_n\right) \cap \left(\varepsilon_{jn} \leq U_{in} - U_{jn} + \varepsilon_{in}\right)\right\} \cup \left\{j \notin C_n\right\}\right]$$

Alternative Derivation of PCS Model

Taking the conditional probability on ε_{in},

$$P_n(i) = \frac{1}{1-Q_n(\emptyset)} \times q_n(i) \times \int_{-\infty}^{+\infty} f(\varepsilon_{in}) \prod_{j\in M, j\neq i} \left\{ q_n(j)\text{Prob}\left(V_{in}-V_{jn}+\varepsilon_{in} \geq \varepsilon_{jn}\right) + \left(1-q_n(j)\right) \right\} d\varepsilon_{in}$$

For the logit type second stage choice model,

$$P_n(i) = \frac{q_n(i)}{1-Q_n(\emptyset)} \int_{-\infty}^{\infty} \mathrm{e}^{-\varepsilon_{in}} \cdot \mathrm{e}^{-\mathrm{e}^{-\varepsilon_{in}}} \times \prod_{j\in M, j\neq i} \left[q_n(j) \cdot \mathrm{e}^{-\mathrm{e}^{-V_{in}+V_{jn}-\varepsilon_{in}}} + \left\{ 1-q_n(j) \right\} \right] d\varepsilon_{in}$$

$$= \frac{q_n(i)}{1-\prod_{j\in M}\left(1-q_n(j)\right)} \int_{-\infty}^{\infty} \mathrm{e}^{-\varepsilon_{in}} \cdot \mathrm{e}^{-\mathrm{e}^{-\varepsilon_{in}}} \times \prod_{j\in M, j\neq i} \left[q_n(j) \cdot \mathrm{e}^{-\mathrm{e}^{-V_{in}+V_{jn}-\varepsilon_{in}}} + \left\{ 1-q_n(j) \right\} \right] d\varepsilon_{in}$$

- Requires a single integral with respect to ε_{in}.
- Does not need to evaluate 2^J-1 possible choice sets.

Estimation Methods

- Simultaneous Estimation

$$L = \prod_{n=1}^{N} P_n(i_n)$$ Eq. (1)

i_n: individual n's chosen alternative

- Sequential Estimation
(In case that information on individual choice set is available)

STEP 1: Estimate parameters of the choice set formation model.

$$L = \prod_{n=1}^{N} Q_n(C_n|G)$$

STEP 2: Substitute parameter estimates of STEP 1 into Eq. (1), then estimate parameters of the discrete choice model.

Concluding Remarks

- Discrete choice modeling has been developed under the traditional economic assumption that assumes rational individuals.
- Stated preference data are powerful in eliciting the preference for non-existing alternatives but we need to pay attention to various biases.
- Other psychometric data are also useful to incorporate latent factors such as latent perception and attitude.
- Those models are estimable even with weak computing power of PCs in 1980’s.

Concluding Remarks

- With the advanced computing power and development of estimation methods, any types of discrete choice models can be estimated these days.
- Machine learning is even more powerful to forecast human behavior by using big data.
- But we cannot “understand” or “explain” the behavior only by the machine learning methods.