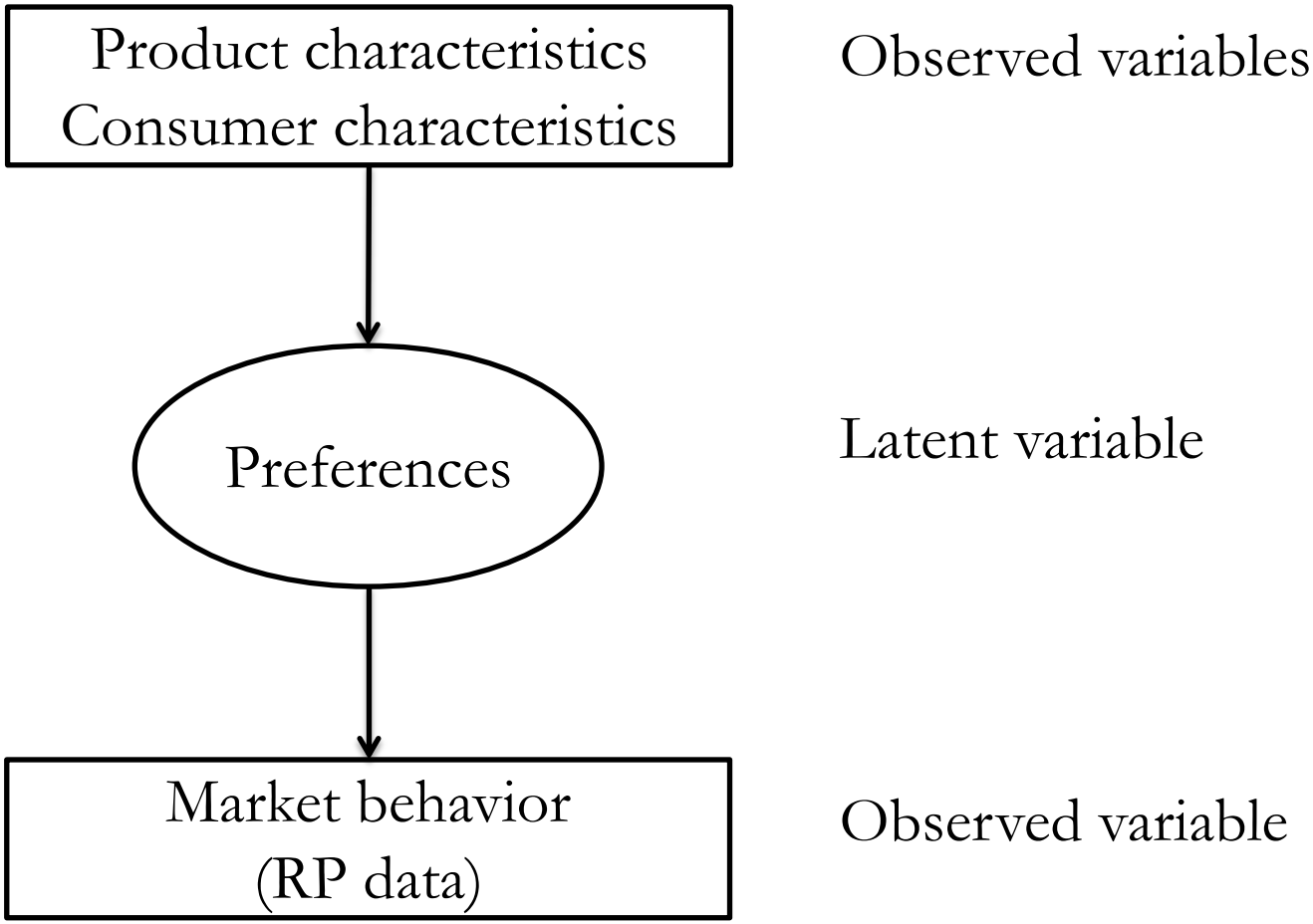
Discrete choice models incorporating revealed preferences and psychometric data

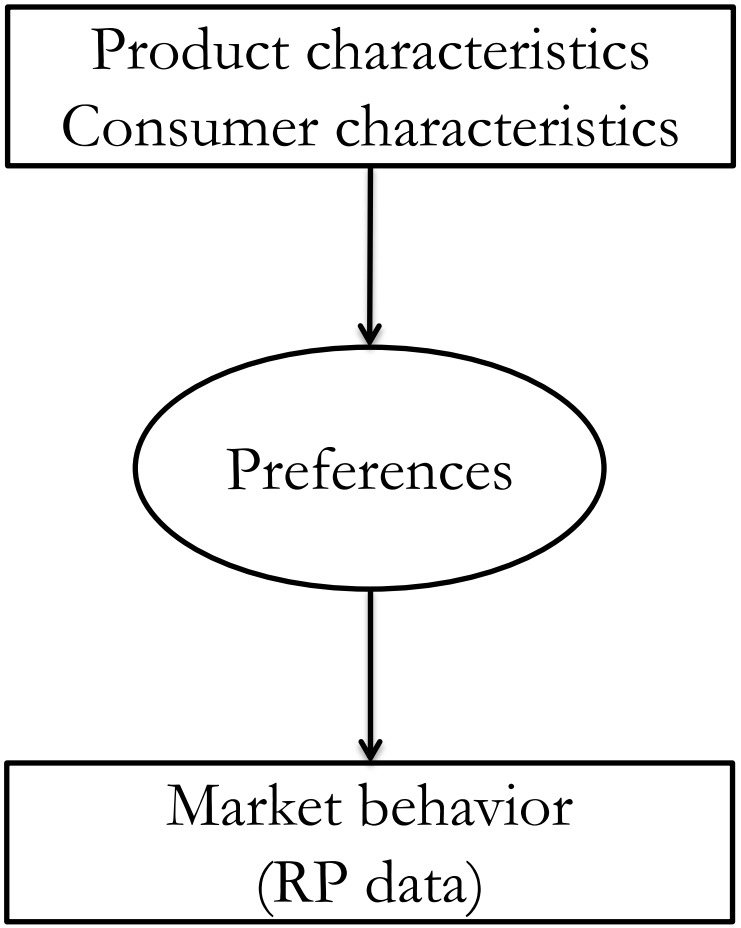
In: Econometric Models in Marketing, 2002

Taka Morikawa, Nagoya Univ.
Moshe Ben-Akiva, MIT
Daniel McFadden, Berkeley

DCM Framework



DCM Framework



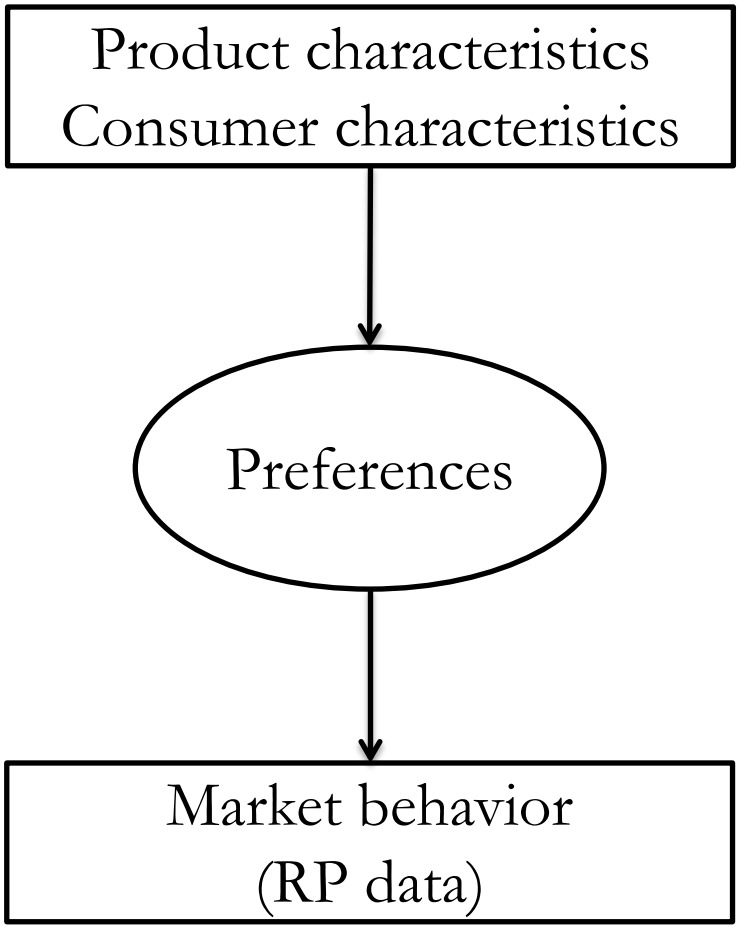
X_j is a set of product *j*'s characteristics

D_i is a set of consumer *i*'s characteristics

$$U_{ij} = V_{ij}(X_j, D_i) + \epsilon_{ij}$$

$$P(y_i = j) = \Gamma(V_i.)$$

DCM Framework



X_j is a set of product *j*'s characteristics

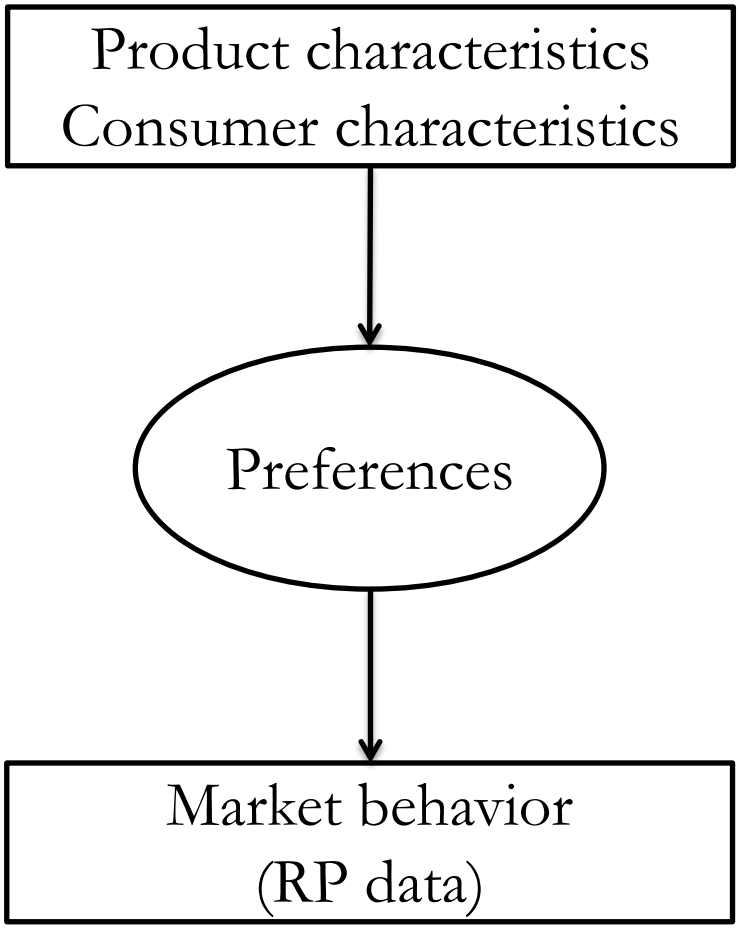
D_i is a set of consumer *i*'s characteristics

$$U_{ij} = X_j\beta + D_i\gamma + \epsilon_{ij}$$

$$\epsilon_{ij} \sim i.i.d.\,\text{EVI}$$

$$P(d_i = j) = \frac{e^{X_j\beta + D_i\gamma}}{\sum_k e^{X_k\beta + D_i\gamma}}$$

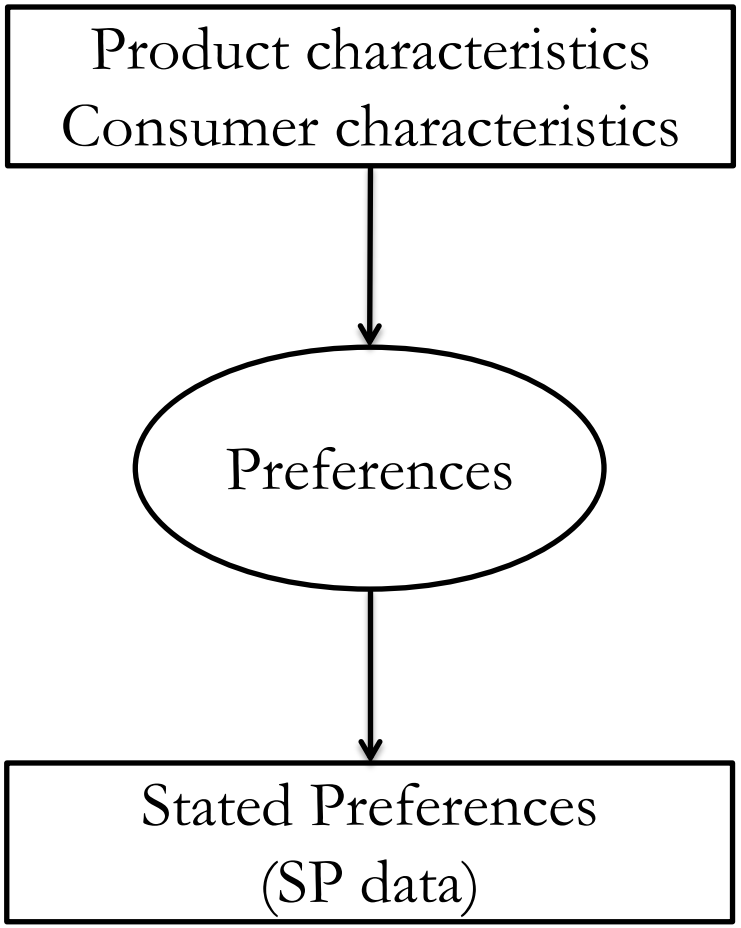
DCM Framework



Challenges with RP data:

- Lack of the data about consumers
- Possible multicollinearity in product characteristics
- No data on the actual choice of only hypothetical alternatives:
  - New characteristics
  - New values of present characteristics
- Unknown actual choice set or consideration set

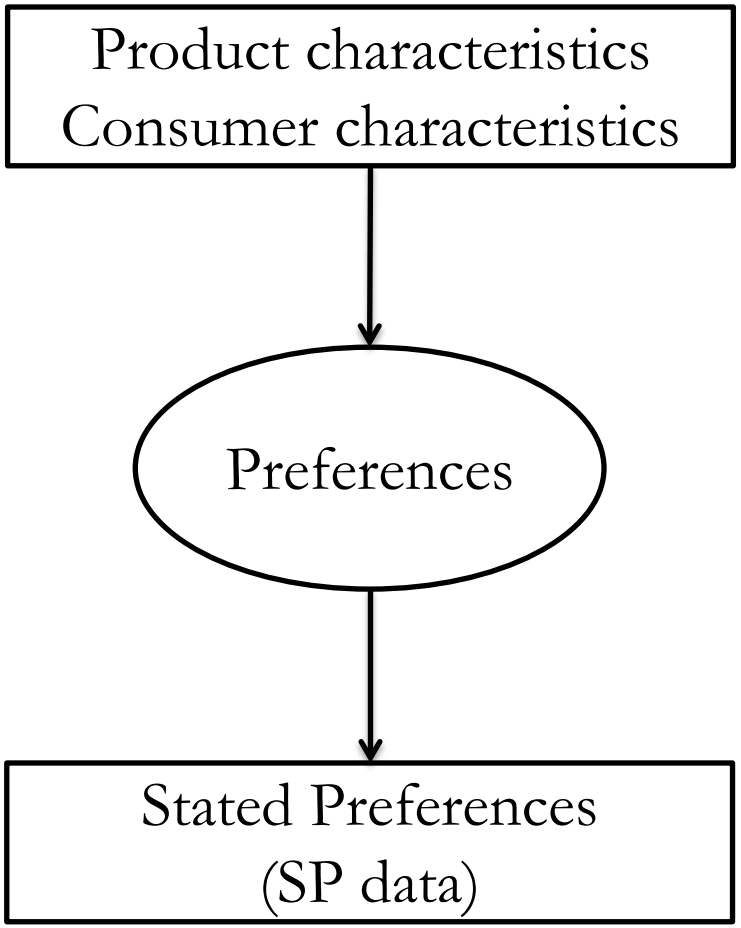
DCM Framework



Opportunities with SP data:

- Preferences for non-existing alternatives or attributes
- The choice set is prespecified
- Multicollinearity is avoided
- Range of attribute values can be extended

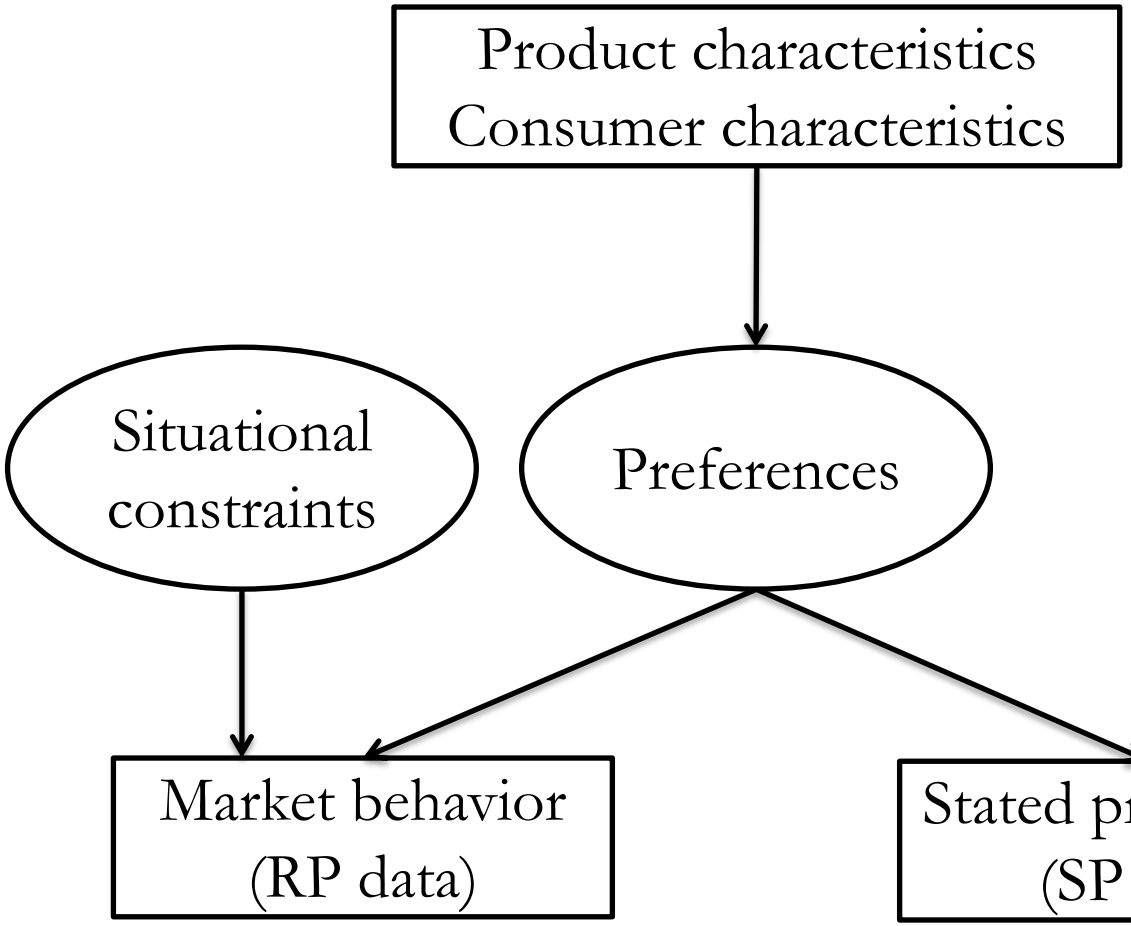
DCM Framework



Challenges with SP data:

- The respondent considers only the most important attribute
- The response is influenced by an 'inertia' of the current actual choice
- Respondent use the survey as an opinion statement for his benefit (overstating)
- Not consider situational constraints
- Ignores or misinterprets an attribute if an attribute value lacks reality

DCM Framework



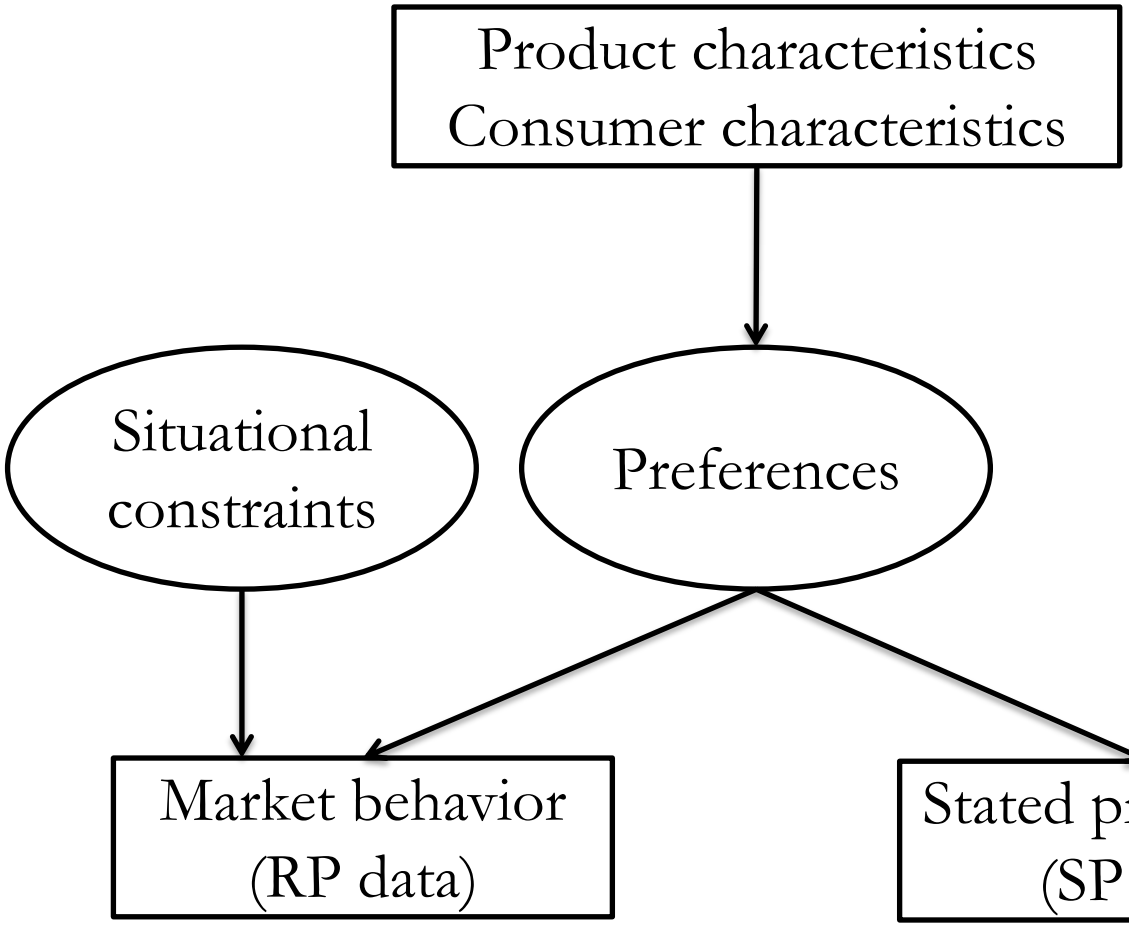
Structural model:

$$U_{ij}^{RP} = V_{ij}(X_j, D_i) + \epsilon_{ij}^{RP}$$
$$U_{ij}^{SP} = \Upsilon_{ij}(X_j, D_i) + + \rho d_{ij}^{RP} + \epsilon_{ij}^{SP}$$

V and Υ may contain the same parameters and different parts

ρd_{ij}^{RP} in SP captures 'inertia'

DCM Framework

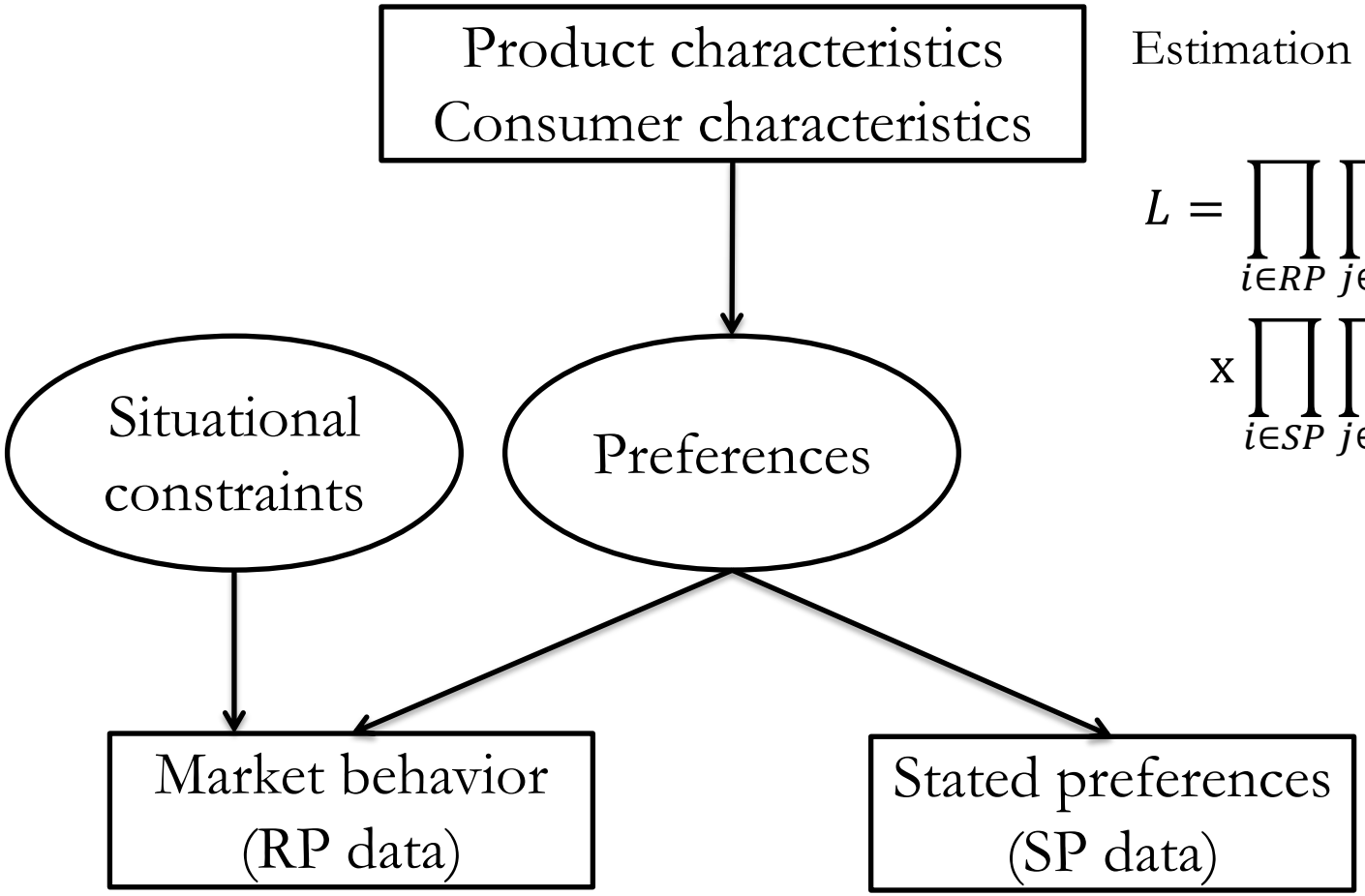


Measurement (binary) model:

$$d_{ij}{}^{RP} = \begin{cases} 1, U_{ij}{}^{RP} \geq 0 \\ 0, U_{ij}{}^{RP} < 0 \end{cases}$$

$$d_{ij}{}^{SP} = \begin{cases} 1, U_{ij}{}^{SP} \geq 0 \\ 0, U_{ij}{}^{SP} < 0 \end{cases}$$

DCM Framework

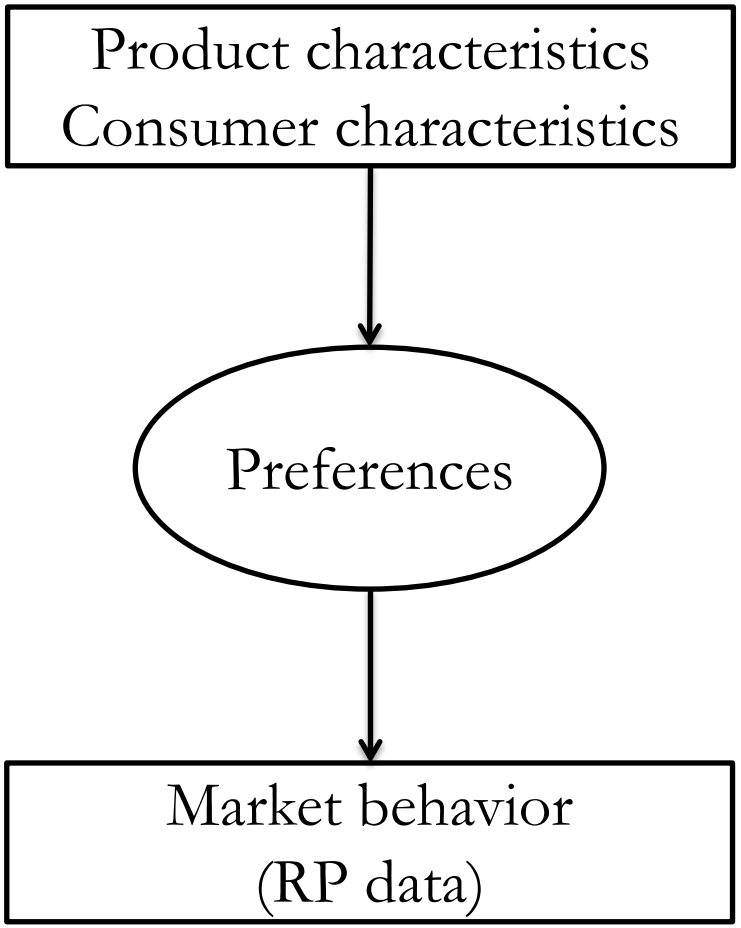


Estimation technique:

$$L = \prod_{i \in RP} \prod_{j \in C_i} [\Pr(d_{ij}{}^{RP})]^{d_{ij}{}^{RP}} \text{ x}$$

$$\text{x} \prod_{i \in SP} \prod_{j \in C_i} [\Pr(d_{ij}{}^{SP})]^{d_{ij}{}^{SP}}$$

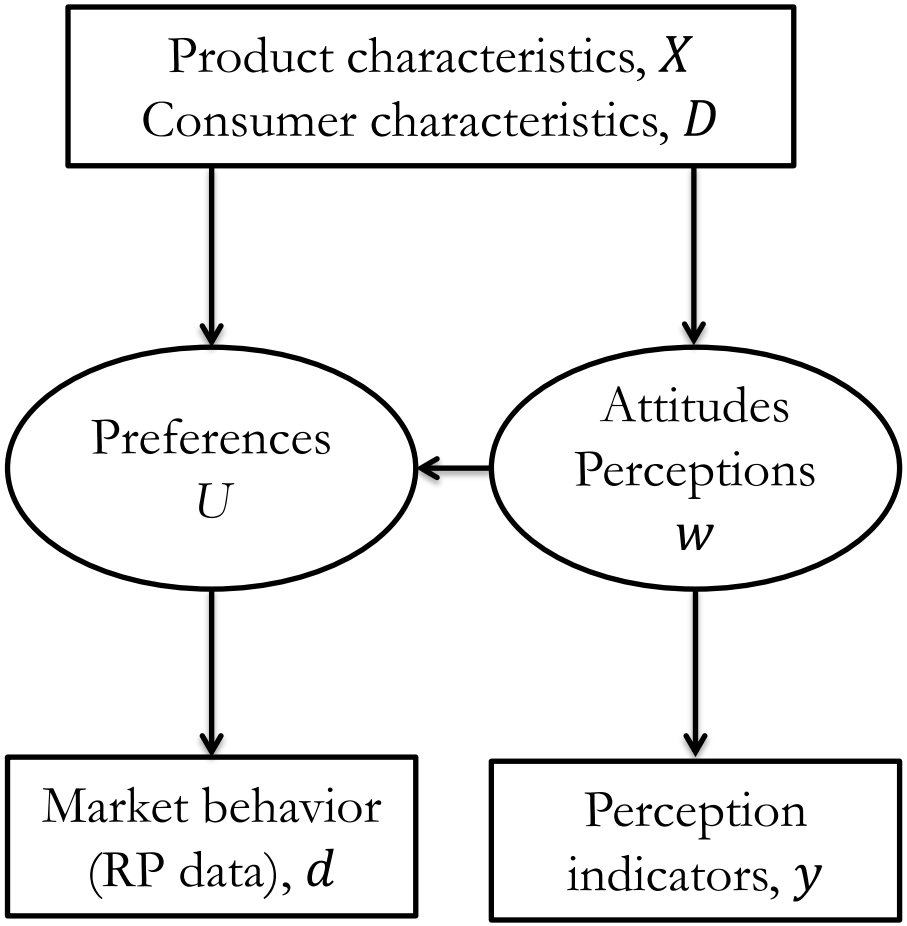
DCM Framework



Challenges with RP data:

- Heterogeneity with respect to latent consumer attributes
  - Perceptions
  - Attitudes
- Ex. in transport choice:
  - Convenience
  - Comfort
- Ex. in culture:
  - Beauty
  - Point of interest
  - Breathtaking

DCM Framework

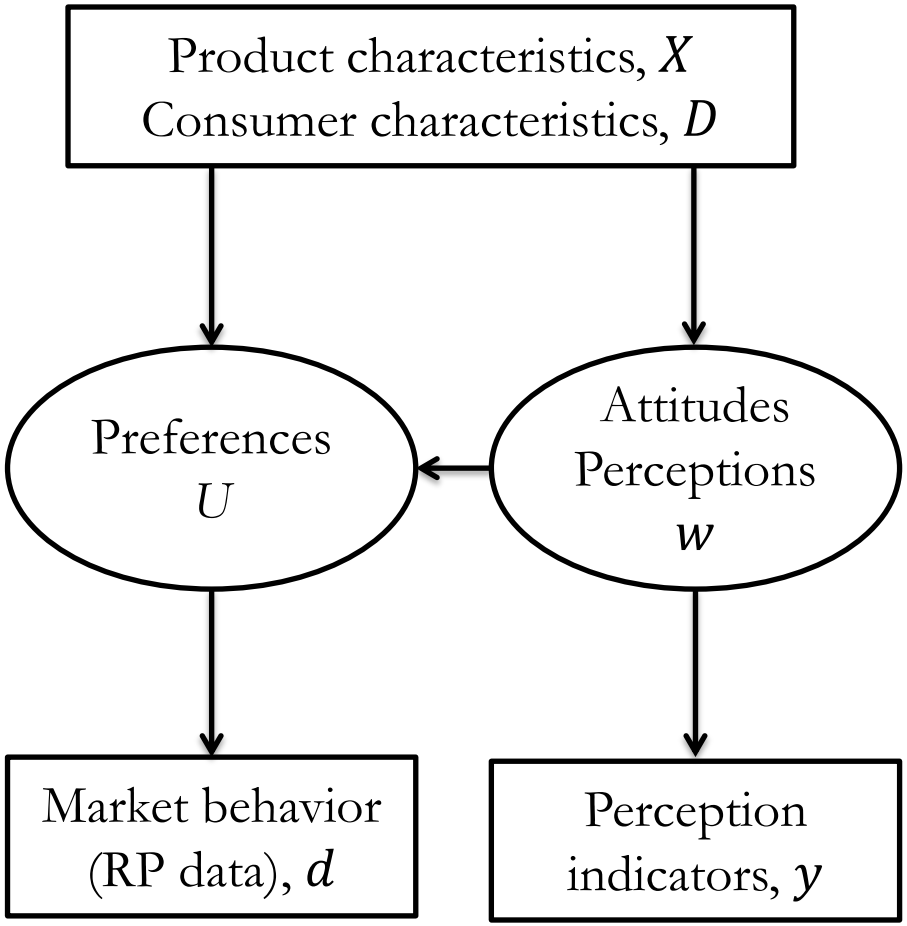


Structural model:

$$U_{ij}{}^{RP} = V_{ij}(X_j, D_i, w_i) + \epsilon_{ij}{}^{RP}$$
$$w_i = BD_i + \varepsilon_i$$

w_i are latent perceptions for alternative or its characteristics

DCM Framework



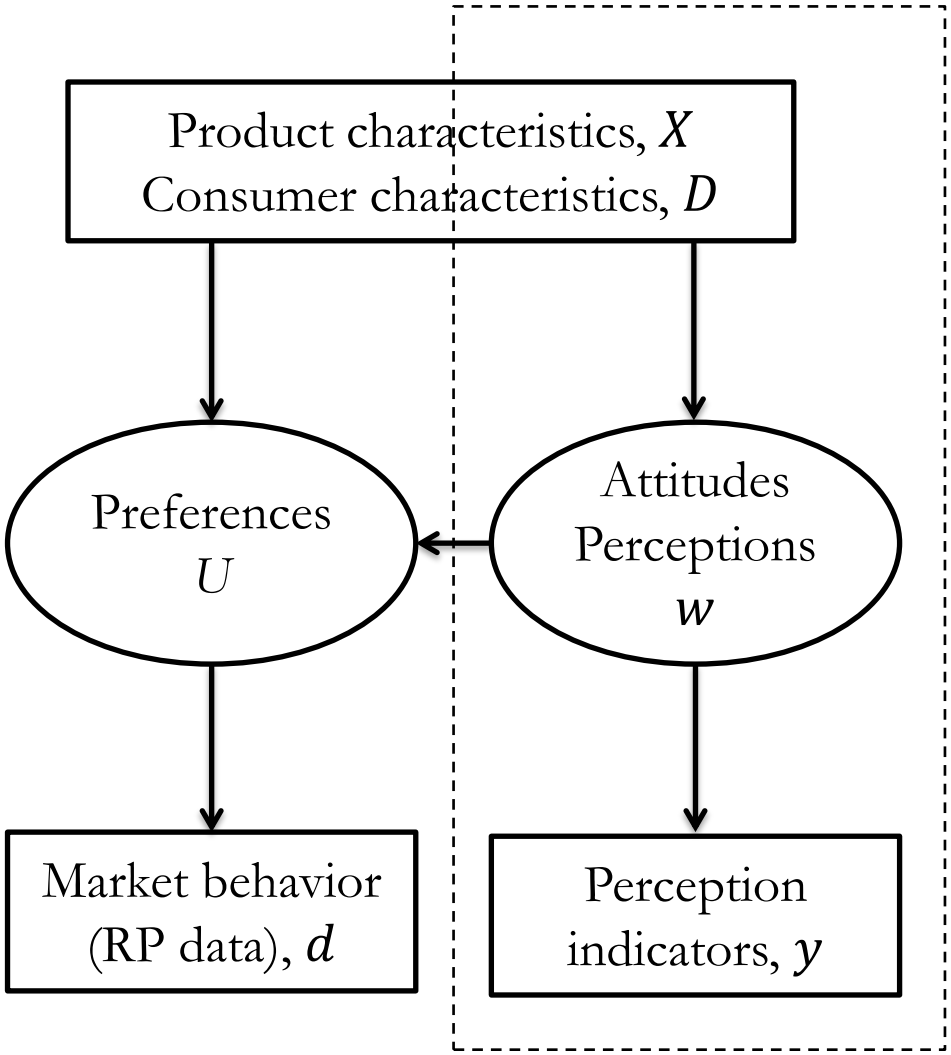
Measurement (binary) model:

$$d_{ij}{}^{RP} = \begin{cases} 1, U_{ij}{}^{RP} \geq 0 \\ 0, U_{ij}{}^{RP} < 0 \end{cases}$$

$$y_i = \Lambda w_i + \nu_i$$

w_i are latent perceptions
y_i are perception indicators

DCM Framework



Estimation technique:

First stage (LISRES):

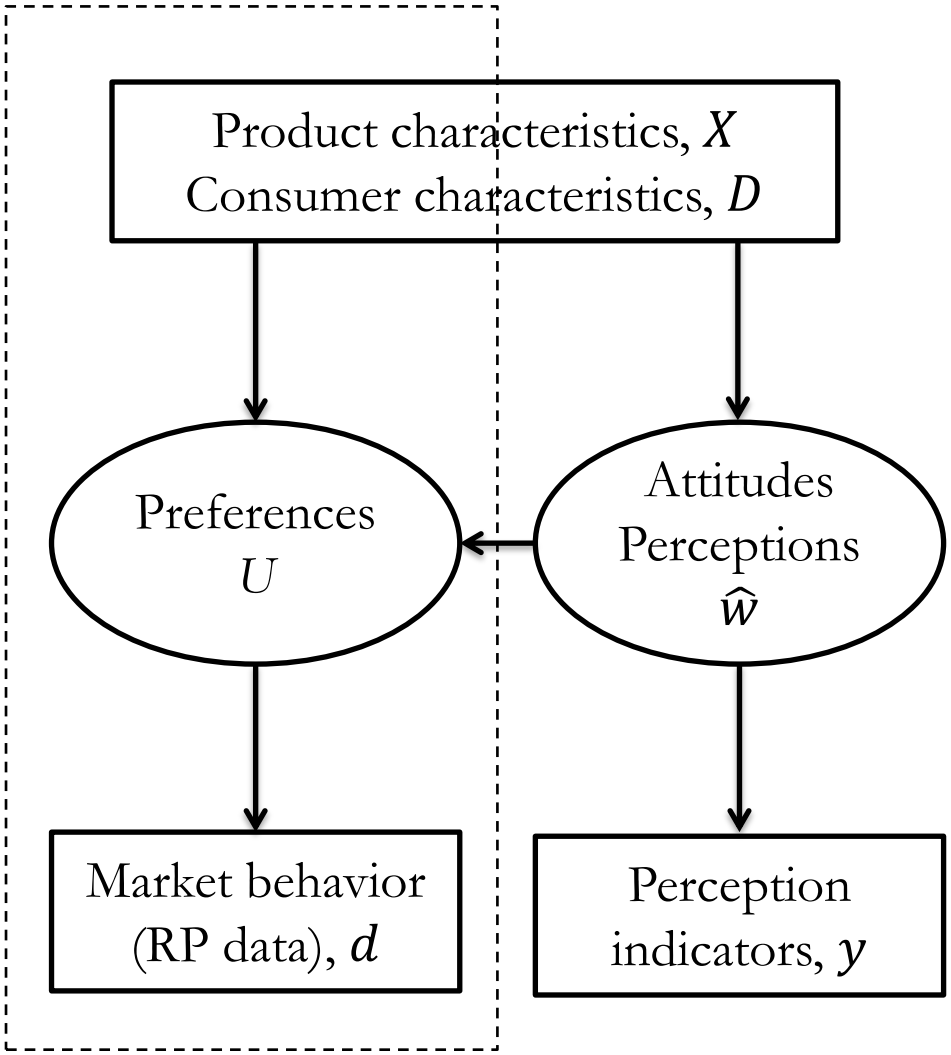
$$w_i = BD_i + \varepsilon_i$$
$$y_i = \Lambda w_i + \nu_i$$

$$y_i = \Lambda(BD_i + \varepsilon_i) + \nu_i$$

Obtain

$$\widehat{w}_i = \widehat{\Lambda}^{-1} y_i$$

DCM Framework

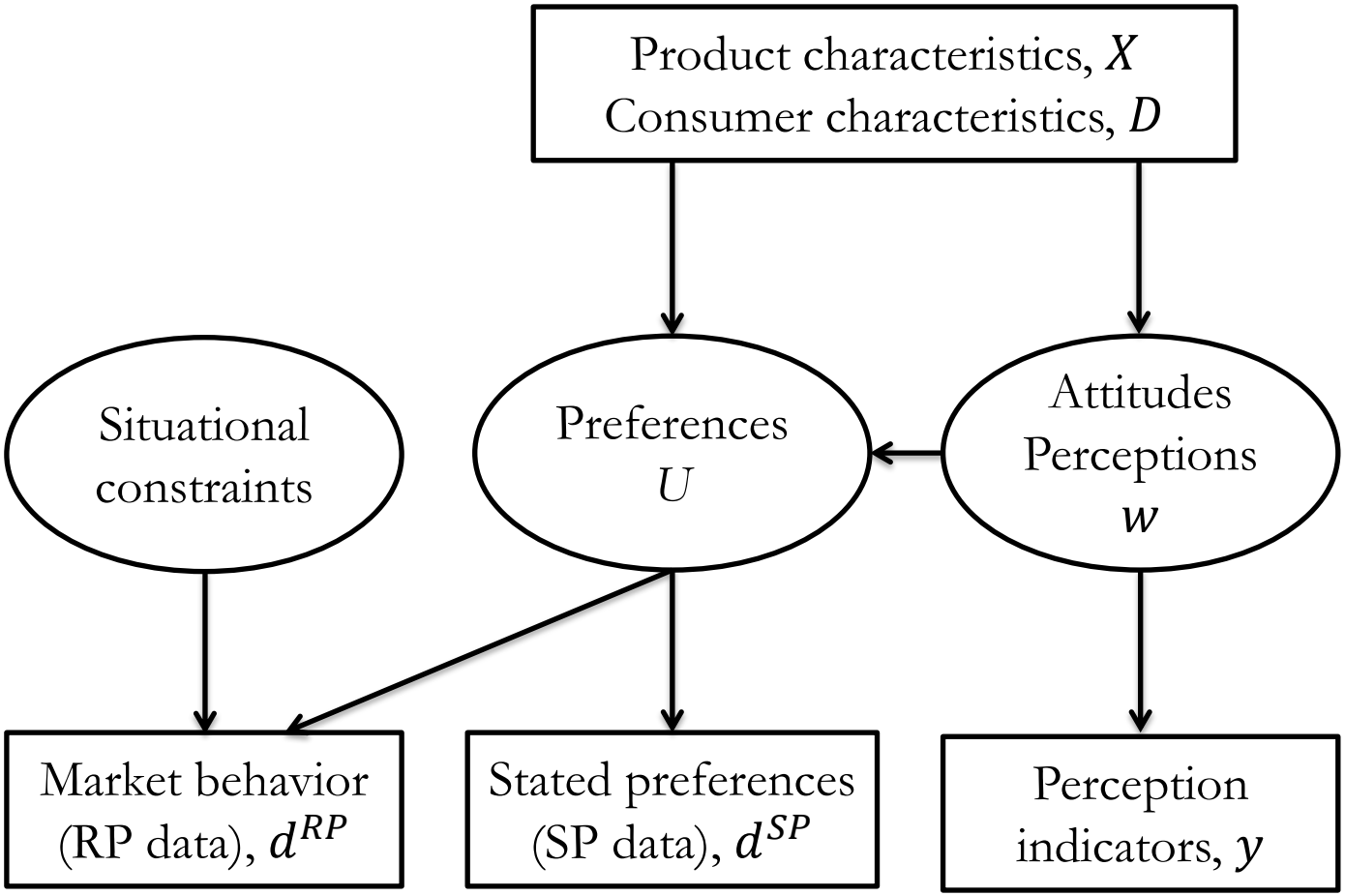


Estimation technique:

Second stage:

$$L = \prod_{i \in RP} \prod_{j \in C_i} [\Pr(d_{ij}{}^{RP} | X_j, D_i, \widehat{w}_i)]^{d_{ij}{}^{RP}}$$

DCM Framework



Application: train *vs* car choice

- Nijmegen – the city of interview
- Travel to Randstad (Amsterdam, Hague, Rotterdam)
  - By rail or by car, both approximately 2 hours
- Home conducted interview (228 respondents)
- Actual choice of intercity trip to Randstad during previous 3 months (RP data)
  - Level of service attributes (travel time, cost etc.)
  - Socio-economic characteristics (age, sex) and trip goal
  - Subjective rating of latent travel characteristics
- SP experiment of a choice between two different rail services (SP1 data, 2875 comparisons, ordered choice data)
- SP experiment of a choice between rail and car (SP2 data, 1577 comparisons, ordered choice data)

SP data: choice between two options

- Pairwise comparison:
  - SP1: two rail services
  - SP2: rail *vs* car
- Attributes:
  - Travel cost
  - Travel time
  - Number of transfers (for trains)
  - Luxury level of train (for trains)
- Answers:
  - Definitely choose the alternative 1
  - Probably choose the alternative 1
  - Not sure
  - Probably choose the alternative 2
  - Definitely choose the alternative 1

SP data: choice between two options

	RP	SP1	SP2	RP + SP1	RP + SP2	RP + SP1 + SP2
Rail constant (RP)	0.501			0.455	0.702	0.718
	(1.8)			(1.8)	(3.0)	(3.4)
Rail constant (SP)			–0.970		–3.82	–3.82
			(–9.8)		(–4.0)	(–4.0)
Cost per person	–0.0270	–0.0828	–0.0111	–0.0279	–0.0338	–0.0337
	(–4.4)	(–25.4)	(–5.6)	(–5.2)	(–6.5)	(–6.8)
Line-haul time	–0.342	–0.967	–0.156	–0.327	–0.401	–0.394
	(–1.4)	(–11.6)	(–1.9)	(–4.9)	(–2.1)	(–6.1)
Terminal time	–1.61		–0.272	–1.60	–1.46	–1.47
	(–4.83)		(–1.9)	(–4.9)	(–4.63)	(–4.77)
Number of transfers	–0.139	–0.140	0.0433	–0.0478	–0.0348	–0.0569
	(–1.0)	(–4.3)	(0.8)	(–3.4)	(–0.3)	(–3.8)
Comfort		0.493		0.166		0.201
		(14.4)		(4.9)		(6.24)
Business trip dummy	0.902		–0.115	0.887	0.358	0.363
	(3.2)		(–1.2)	(3.2)	(1.74)	(1.78)
Female dummy	0.488		–0.102	0.488	0.230	0.232
	(2.4)		(–1.5)	(2.4)	(1.4)	(1.5)
Inertia dummy			1.60		5.68	5.70
			(18.7)		(4.7)	(4.8)

Predicting the latent attributes

- For both chosen and unchosen modes
- Perceptional indicators
  - Relaxation during the trip (relax)
  - Reliability of arrival time (relia)
  - Flexibility of choosing departure time (flex)
  - Ease of travelling with children or heavy luggage (ease)
  - Safety during the trip (safety)
  - Overall rating of the mode
- Each indicator is valued by 5-point scale
- Overall rating is values by 10-point scale
- Two latent attributes:
  - Ride comfort (w_1)
  - Convenience (w_2)

Predicting the latent attributes

- Two latent attributes:
  - Ride comfort (w_1)
  - Convenience (w_2)
- w affected by consumer attributes D through B
- w affect perceptional indicators y through Λ

$$\hat{\mathbf{B}}' = \begin{bmatrix} (w_1^*) & (w_2^*) & \\ -0.427(-2.4) & 0.378(2.4) & (aged) \\ -0.323(-1.7) & 0 & (lhtime) \\ 0 & -1.98(-9.0) & (trmtime) \\ 0.281(0.9) & 0 & (first) \\ 0 & -0.396(-3.7) & (xfern) \\ 0 & 0.482(3.5) & (freepark) \\ -0.339(-1.3) & 0 & (aged \times lhtime) \end{bmatrix}$$

$$\hat{\Lambda} = \begin{bmatrix} (w_1^*) & (w_2^*) & \\ 0.433(7.6) & 0.280(3.2) & (relax) \\ 0.527(12.5) & 0.661(10.2) & (relia) \\ 0 & 0.815(14.7) & (flex) \\ 0 & 0.794(14.2) & (ease) \\ 0.462(11.6) & 0.311(5.2) & (safe) \\ 0.784(8.5) & 1.76(14.1) & (overall) \end{bmatrix}$$

RP model with latent attributes

	Model w/o Latent Attributes	Sequential Estimation Model	Simultaneous Estimation Model
Rail constant	0.583	0.322	–1.81
	(2.0)	(1.0)	(–0.9)
Cost per person	–0.0268	–0.0338	–0.0379
	(–4.2)	(–4.1)	(–4.3)
Line-haul time	–0.405	0.0751	0.379
	(–1.6)	(0.2)	(0.9)
Terminal time	–1.57	–1.18	–0.818
	(–4.2)	(–2.6)	(–2.3)
Number of transfers	–0.195	–0.316	–0.230
	(–1.3)	(–1.7)	(–1.2)
Business trip dummy	0.942	1.33	1.28
	(3.6)	(3.6)	(3.3)
Female dummy	0.466	0.652	0.700
	(2.3)	(2.6)	(2.9)
w_1^* *(comfort)*		0.882	1.29
		(2.7)	(1.8)
w_2^* *(convenience)*		1.39	1.10
		(4.1)	(4.7)

Conclusion

- RP+SP+Latent variables give:
  - Identification of preferences for new alternatives/attributes (SP *vs* RP)
  - Bias correction for SP (SP+RP)
  - Efficiency (SP+RP+Latent variables)