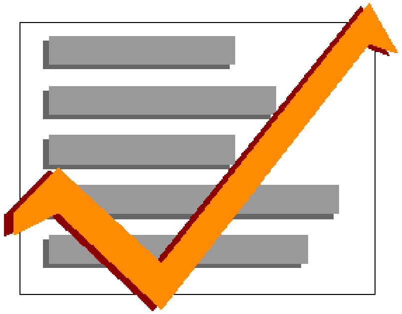
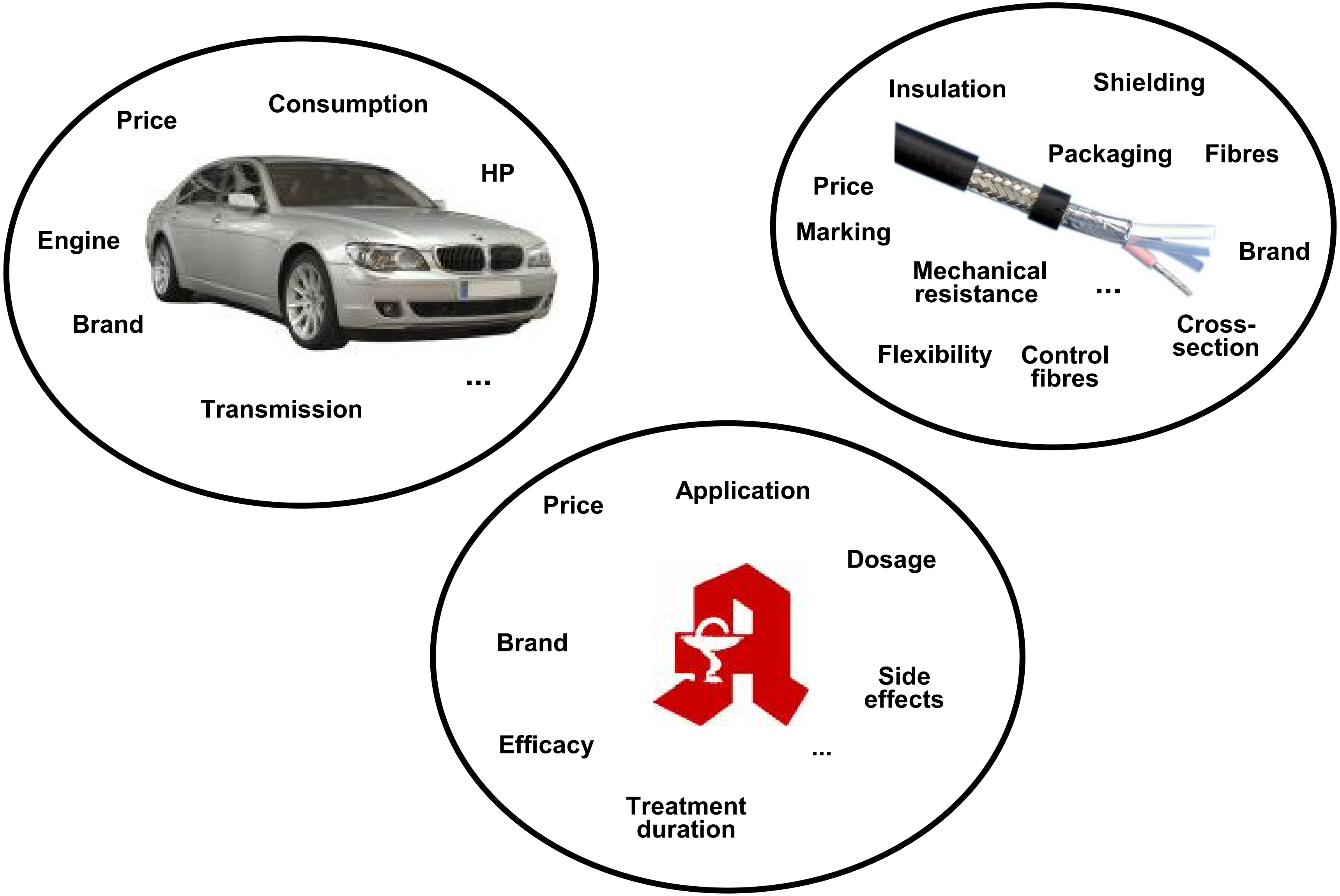
Adaptive discrete choice designs

Presentation held at useR! 2008

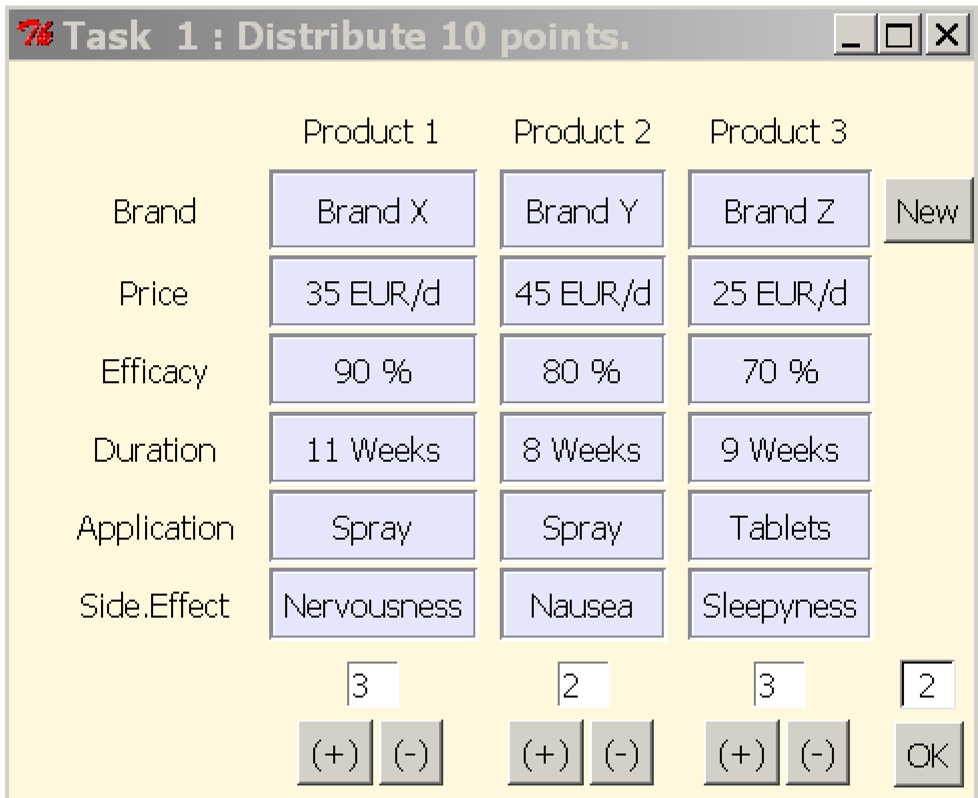
Boris Vaillant
www.quantitative-consulting.eu



The goal of discrete choice analysis in marketing is to assess the influence of product- and service-attributes on customers' choice behaviour



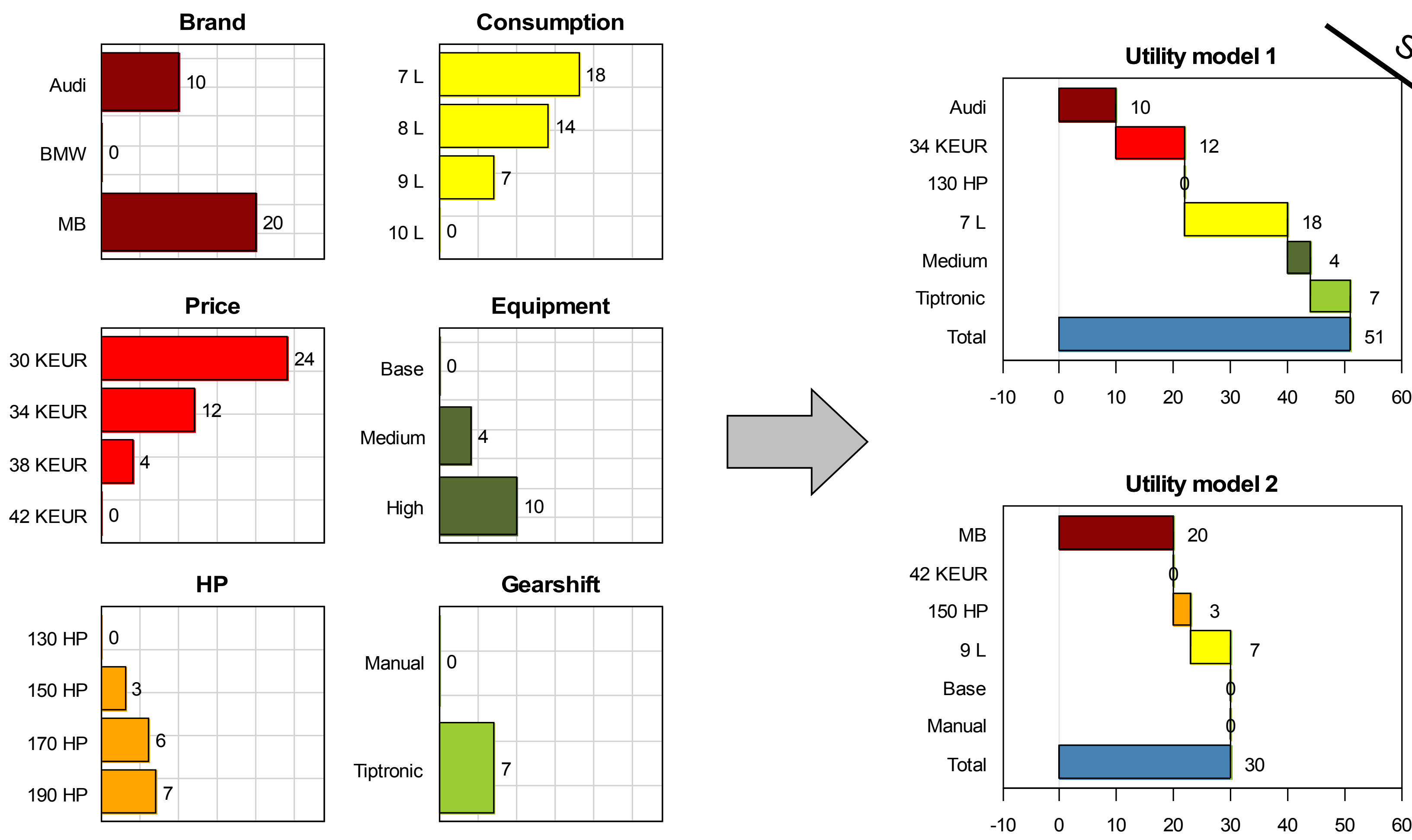
In product comparison questionnaires, respondents indicate their preferred choice or purchase pattern in a series of ‘choice tasks’



Typical example:
Doctors are asked to indicate the prescription share for each of the shown treatments (e.g. for the last 10 patients with the corresponding indication)

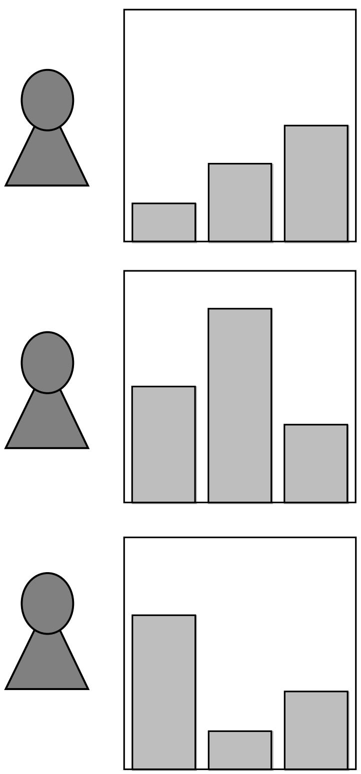
The preference of individual customers or customer groups for the different product elements is parametrised and can be used to calculate the preference for existing or new products

Car example:
Sedan segment



In the marketing practice, discrete choice data often feeds larger market models in which companies can test the outcome of different strategies

Customer preferences



- Customer preferences as measured via the discrete choice questionnaire
- Either individual or group preferences

Product information
- Target segments
- Description of existing and future products

Additional information from the survey
- Price barriers
- Brand value
- Strengths and weaknesses

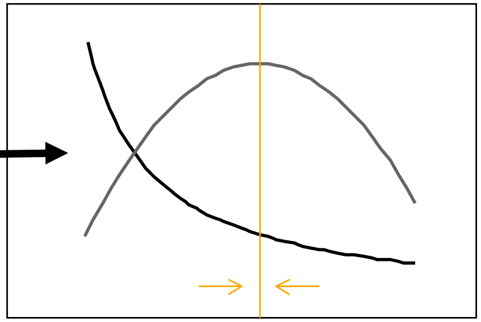
Market simulation

Simulation of customers choice in scenarios
- Price changes
- New products
- Competitive reaction

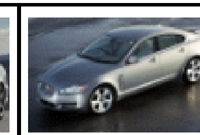
Price information
- Most important products in the market
- Changes in time

Market information
- Market size
- Market shares

Demand curve and profit-optimal price

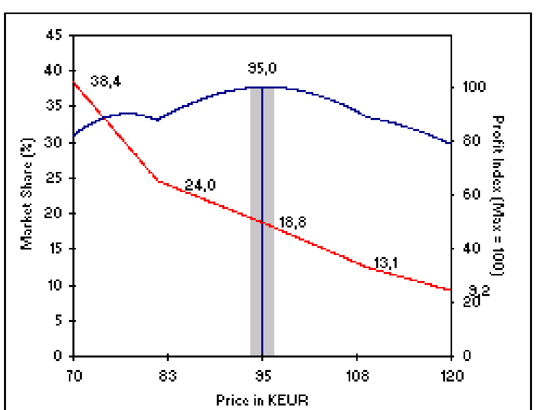
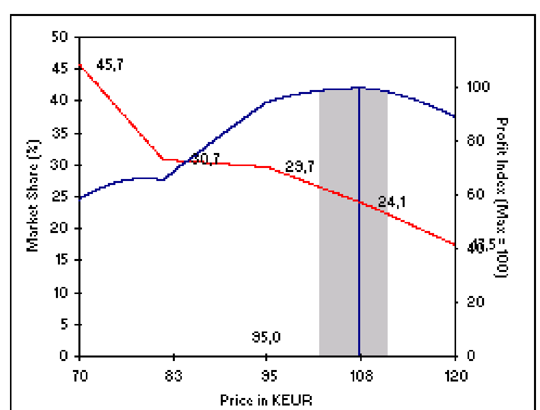
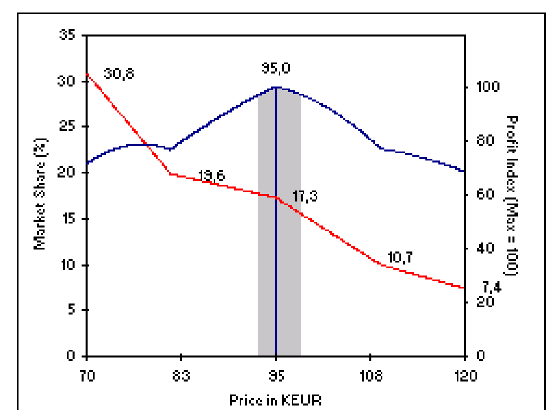


- Group-wise results or aggregation of individual results
- Obtain market effects e.g. for a price change
- Determine optimal new price

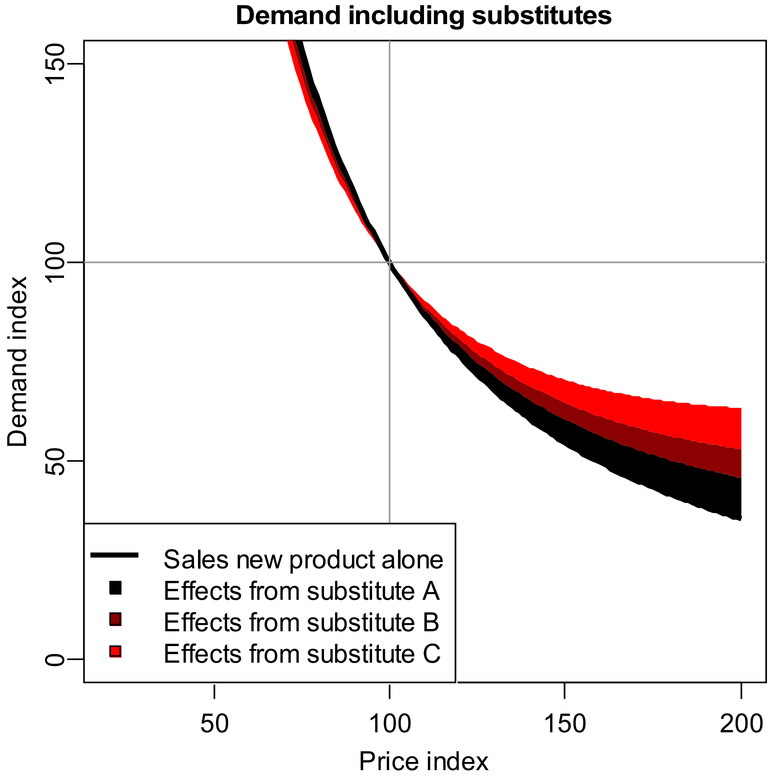
								
	X	**Audi A8 4.2 FSI**	**Audi S6 5.2 quattro**	**BMW M5**	**BMW 750i A**	**MB CLS 500**	**MB E 63 AMG**	**Jaguar XF 4.2 L Super V8**
Preis (EUR)	95000	80200	82250	92500	81500	74137	98651	80820
Leistung (PS)	500	350	435	507	367	388	514	416
Getriebeart	Automatik	Automatik	Automatik	Manuell	Automatik	Automatik	Automatik	Automatik
Ausstattung	Mittel	Mittel	Mittel	Mittel	Mittel	Mittel	Mittel	Mittel

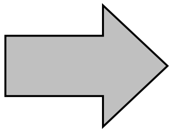
Motor	xxx	4,2 L V8	5,2 L V10	5,0 L V10	4,8 L V8	5,5 L V8	6,2 L V8	4,2 L V8
Hubraum in ccm	xxx	4.163	5.204	4.999	4.799	5.461	6.208	4.196
Gänge	xxx	6	6	7	6	7	7	6
Verbrauch (l/100km)	xxx	10,9	13,4	14,8	11,4	11,6	14,3	12,6

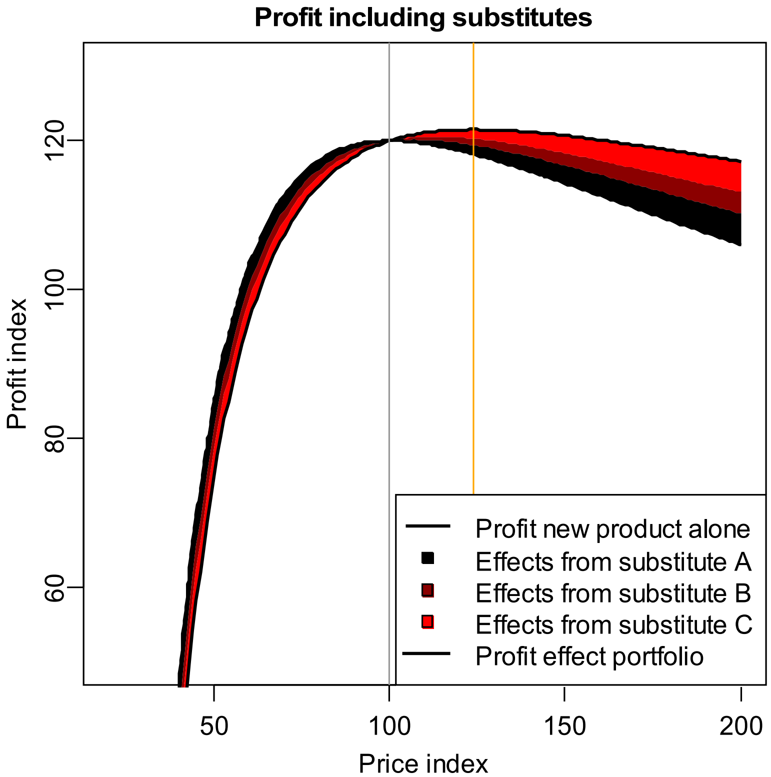
Absatz	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx

Szenario A	Szenario B	Szenario C
		

One practical challenge is the estimation of substitution (or portfolio-) effects. This often requires that the parameters of the model be estimated for each individual customer

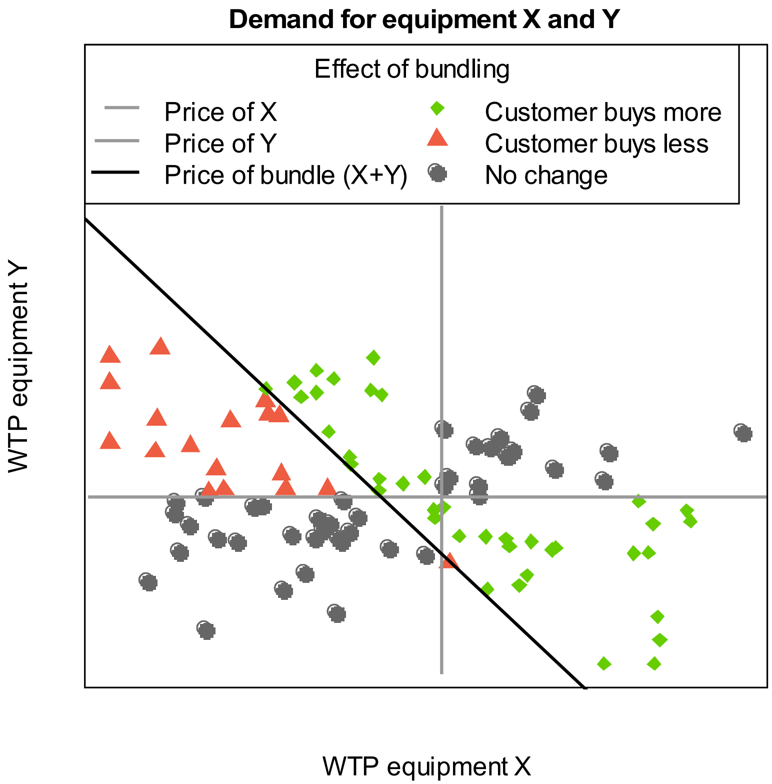


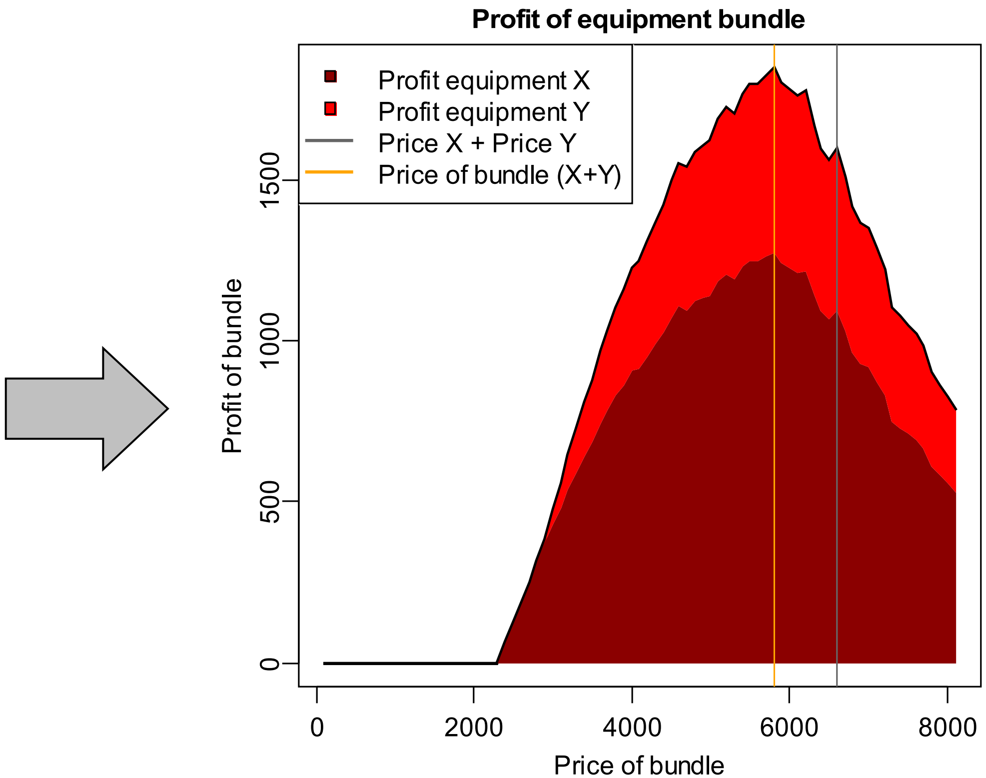




- Substitution effects strongly influence the optimal price decision
- Complexity of substitution decisions is best captured by models based on individual choice data

Creating optimal bundles is another area which requires knowledge of the individual preferences

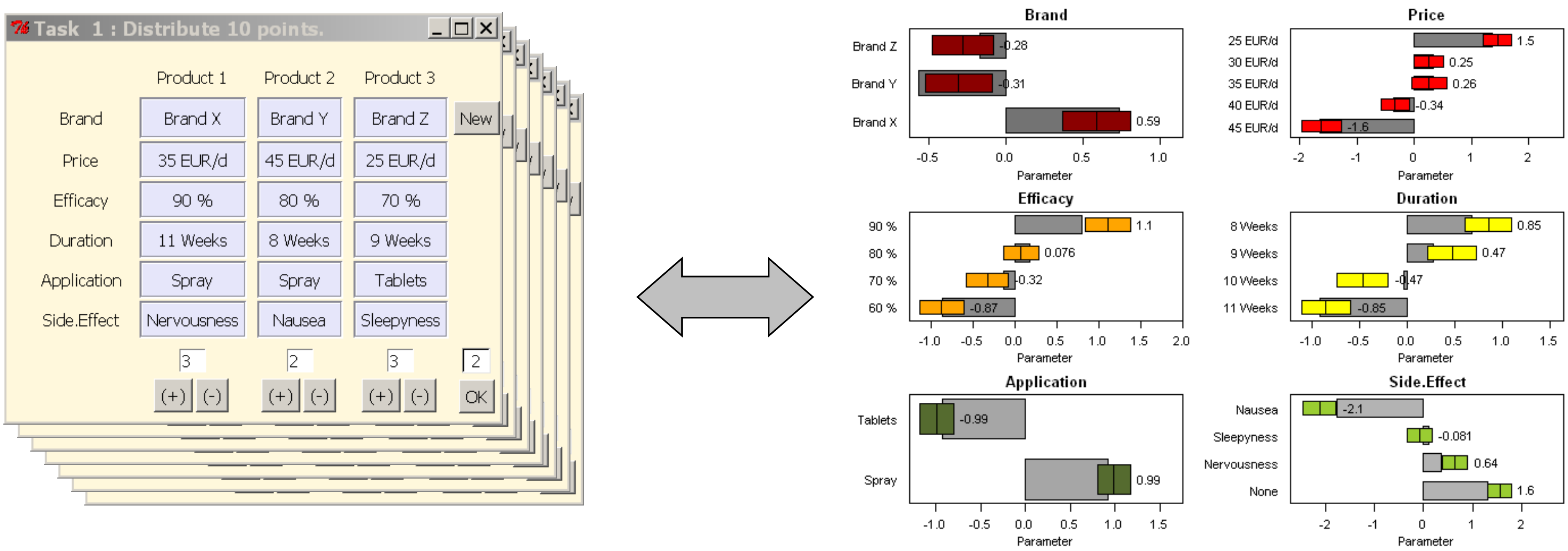


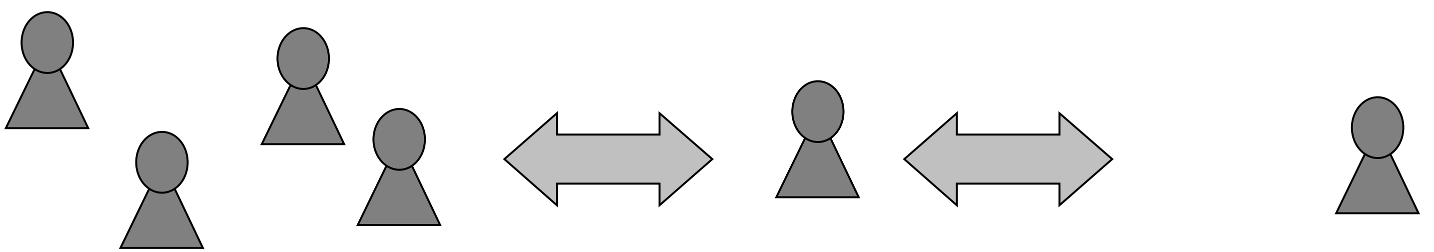


- Individual preferences and WTP for the different products are estimated in a discrete choice study
- From this, one can estimate which products customers will purchase and in which combination

- Testing of bundling scenarios leads to the best bundles to sell and their optimal price

The goal of discrete choice design is to minimise the error of parameter estimates while keeping respondent burden at a minimum





Respondents
- Reduce task complexity
- Reduce number of choice tasks
- Avoid respondent fatigue

Study sponsor
- Reduce cost **and** obtain reliable and valid results

Consultant
- Obtain reliable and valid results
- Have a variety of design options

Probabilistic discrete choice

1. Product attributes: $\mathbf{A} = \{A_1, \ldots, A_{\sharp\mathbf{A}}\}$, e.g. $A_1 = \{$ 'Mercedes', 'Audi', $\ldots\}$
2. Space of products: $\mathbf{M}(\mathbf{A}) = \prod_{A \in \mathbf{A}} A$
3. Probability measure: $\mathbf{q}(\mathbf{n}) = \mathbf{q}(n^{A_1}, \ldots, n^{A_{\sharp\mathbf{A}}}) = \prod_{A \in \mathbf{A}} \mathbf{q}_A(n^A)$
4. Choice probability for choice sets $\mathcal{C} \subset \mathbf{M}(\mathbf{A})$ and $\mathbf{n} \in \mathcal{C}$: $\mathbf{q}(\mathbf{n}|\mathcal{C}) \equiv \mathbf{q}(\mathbf{n})/\mathbf{q}(\mathcal{C})$

Multinomial logit parameters

1. Introduce 'utility'-space $\mathbf{U}(\mathbf{A}) = \prod_{A \in \mathbf{A}} U(A)$ and map $\mathbf{u} \mapsto \mathbf{q}$:

$$U(A) := \left\{\mathbf{u} \in \mathbb{R}^{\sharp A} | \langle \mathbf{1}, \mathbf{u} \rangle = 0 \right\}, \quad \text{and} \quad \mathbf{u}_A \mapsto \mathbf{q}_A = \frac{\exp(\mathbf{u}_A)}{\langle \mathbf{1}, \exp(\mathbf{u}_A) \rangle}$$

2. For $\mathcal{C}$ introduce $\sharp\mathcal{C} \times \sum \sharp A$-'design' matrix $\mathbf{X}_\mathcal{C}$ (row vectors are binary representations of the choice items $\mathbf{n} \in \mathcal{C}$)
3. MNL formula: $\mathbf{q}_\mathcal{C} \equiv \mathbf{q}(\mathbf{u})(\cdot|\mathcal{C}) = \exp(\mathbf{X}_\mathcal{C}\mathbf{u})/\langle \mathbf{1}, \exp(\mathbf{X}_\mathcal{C}\mathbf{u}) \rangle$.

Maximum likelihood theory

1. For a choice set $\mathcal{C}$ write $Y_\mathcal{C}$ for the multinomial rv of the number of choices:

$$\mathbf{P}(Y_\mathcal{C}|\mathbf{u}) = \binom{N_\mathcal{C}}{Y_\mathcal{C}} \mathbf{q}_\mathcal{C}(\mathbf{u})^{Y_\mathcal{C}} = \binom{N_\mathcal{C}}{Y_\mathcal{C}} \frac{\exp(\langle Y_\mathcal{C}, \mathbf{X}_\mathcal{C}\mathbf{u}\rangle)}{\langle \mathbf{1}, \exp(\mathbf{X}_\mathcal{C}\mathbf{u})\rangle^{N_\mathcal{C}}}, \qquad \binom{N_\mathcal{C}}{Y_\mathcal{C}} = \frac{\langle \mathbf{1}, Y_\mathcal{C}\rangle!}{\prod_{\mathbf{m}\in\mathcal{C}} Y_\mathcal{C}(\mathbf{m})!}$$

2. Likelihood: $\mathcal{L}(\mathbf{u}; \mathbf{Y}) = \mathbf{P}(Y_1, \ldots, Y_K|\mathbf{u}) = \prod_{i=1}^K \binom{N_i}{Y_i} \frac{\exp(\langle Y_i, \mathbf{X}_{\mathcal{C}_i}\mathbf{u}\rangle)}{\langle \mathbf{1}, \exp(\mathbf{X}_{\mathcal{C}_i}\mathbf{u})\rangle^{N_i}}$,

3. Information matrix: $I_\mathbf{U}(\mathbf{u}) = \sum_{i=1}^K N_i \mathbf{X}_{\mathcal{C}_i}^T \left(\mathbf{D}(\mathbf{q}_{\mathcal{C}_i}) - \mathbf{q}_{\mathcal{C}_i}\mathbf{q}_{\mathcal{C}_i}^T\right) \mathbf{X}_{\mathcal{C}_i}$

4. Expected covariance: $\Sigma_\mathbf{U}(\hat{\mathbf{u}}) = \left(I_\mathbf{U}(\hat{\mathbf{u}})|_{\mathbf{U}(\mathbf{A})}\right)^{-1}$.

Optimal discrete choice design: Choose the series $\mathcal{C}_i$ to minimise the variance of the (ML-) estimate, e.g.

(**C1**) '**D-criterion**': Minimise $\det_{\mathbf{U}(\mathbf{A})}(\Sigma)^{\frac{1}{2\mathrm{e}(\mathbf{A})}}$

(**C2**) '**E-criterion**': Minimise $\max_{(}\ \|\Sigma|_{\mathbf{U}(\mathbf{A})}\|)^{1/2}$ $(= \sqrt{\text{largest eigenvalue}})$

Heuristics for optimal discrete choice design:

(**H1**) '**Minimal overlap**': Within each choice set $\mathcal{C}_i$ the levels of each attribute should be repeated as little as possible

(**H2**) '**Level balance**': Across all choice sets $\mathcal{C}_i$, the levels of an attribute $A \in \mathbf{A}$ should appear as equally as possible

(**H3**) '**Utility balance**': Based on a best guess of $\mathbf{u}$, the choice probabilities $\mathbf{q}_{\mathcal{C}}(\mathbf{u})$ for the choice items in $\mathcal{C}$ should be approximately equal.

(**H4**) '**Total level balance**': If partial profiles / choice sets are allowed, all levels across all attributes and all choice tasks $\mathcal{C}_i$ should appear (approximately) equally often.

(**H5**) '**Consistency**': The choice items in $\mathcal{C}$ should have *no* a priori ordering (e.g. cet. par. do not compare a harddrive of '200 GB' for '100 EUR' with a harddrive of '100 GB' for '150 EUR')

Bayesian approach

1. Introduce a prior distribution on $\mathbf{U}(\mathbf{A})$ by $\mathbf{P}(\mathbf{u}) = \mathbf{N}(\mu, \Lambda)|_{\mathbf{U}(\mathbf{A})}$
2. Posterior: $\mathbf{P}(\mathbf{u}|\mathbf{Y}) = \frac{\mathbf{P}(\mathbf{Y}|\mathbf{u})}{\mathbf{P}(\mathbf{Y})}\mathbf{P}(\mathbf{u}) = \frac{1}{\mathbf{P}(\mathbf{Y})}\mathcal{L}(\mathbf{u};\mathbf{Y})\mathbf{P}(\mathbf{u})$
3. Bayesian parameter estimate: $\widehat{\mathbf{u}}_B = \mathbf{E}(\mathbf{u}|\mathbf{Y})$
4. Covariance: $\mathbf{cov}(\mathbf{u}|\mathbf{Y}) \equiv \mathbf{cov}(\mathbf{u},\mathbf{u}|\mathbf{Y}) = \mathbf{E}(\mathbf{u}\mathbf{u}^T|\mathbf{Y}) - \mathbf{E}(\mathbf{u}|\mathbf{Y}) \cdot \mathbf{E}(\mathbf{u}|\mathbf{Y})^T$.
5. Expected entropy: $\mathbf{H}_{\mathbf{Y}}(\mathbf{u}|Z) = -\mathbf{E}(\log_2(\mathbf{P}(\mathbf{u}|Z,\mathbf{Y}))|\mathbf{Y}) = -\mathbf{E}(\mathbf{H}_{Z,\mathbf{Y}}(\mathbf{u})|\mathbf{Y})$.

Bayesian adaptive design: Given the posterior $\mathbf{P}(\mathbf{u}|\mathbf{Y})$, choose the next choice task $Z = Z_{\mathcal{C}}$ according to one of:

(B1) '**Bayesian D-criterion**': Minimise $\det(\mathbf{E}(\mathbf{cov}(\mathbf{u}|Z,\mathbf{Y})|\mathbf{Y}))$,

(B2) '**Bayesian E-criterion**': Minimise $\|\mathbf{E}(\mathbf{cov}(\mathbf{u}|Z,\mathbf{Y})|\mathbf{Y})\|$,

(B3) '**Expected Entropy**': Minimise $\mathbf{H}_{\mathbf{Y}}(\mathbf{u}|Z)$,

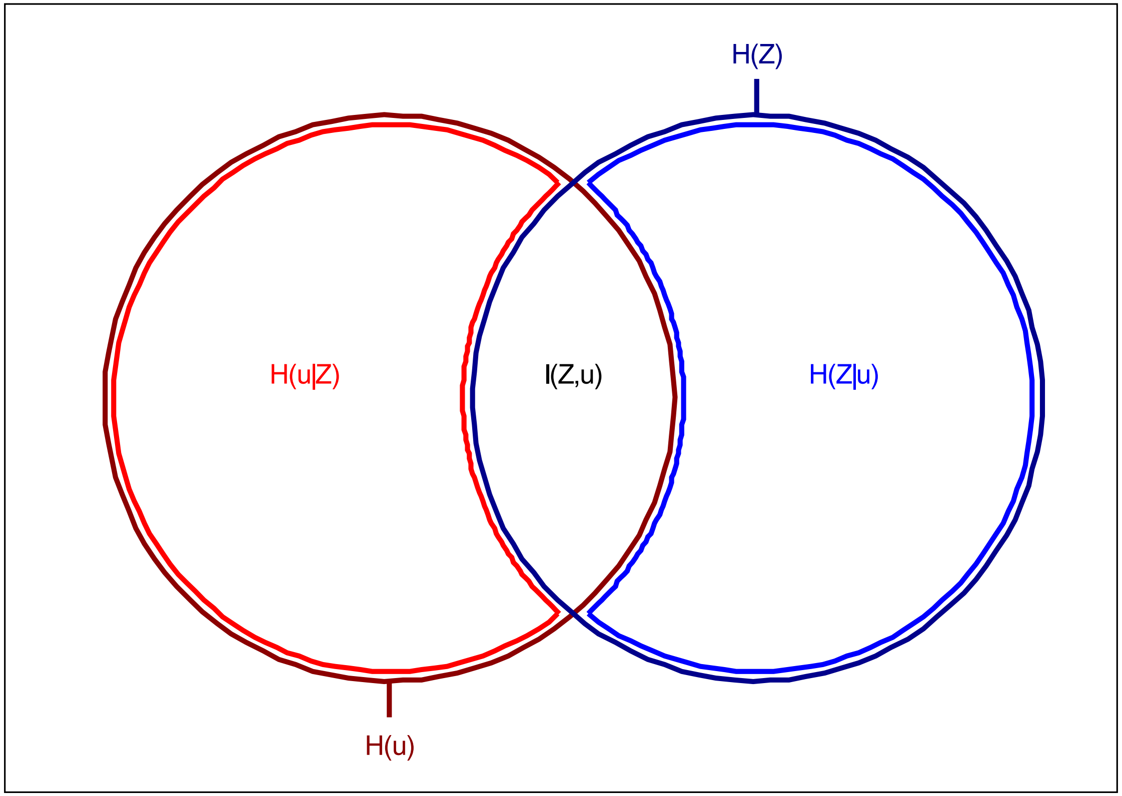
(B4) '**Bayesian utility balance**': Maximise $\mathbf{H}_{\mathbf{Y}}(Z)$,

(B5) '**Simplified utility balance**': Maximise $\mathbf{H}(\mathbf{P}(Z|\widehat{\mathbf{u}}_B)) = \mathbf{H}(\mathbf{P}(Z|\mathbf{E}(\mathbf{u}|\mathbf{Y})))$.

Entropy, Covariance and Updating

1. Law of total covariance: $\mathbf{E}(\mathbf{cov}(\mathbf{q}|Z,\mathbf{Y})|\mathbf{Y}) = \mathbf{cov}(\mathbf{q}|\mathbf{Y}) - \mathbf{cov}(\mathbf{E}(\mathbf{q}|Z,\mathbf{Y})|\mathbf{Y})$
2. Entropy identity: $\mathbf{H_Y}(\mathbf{u}|Z) = \mathbf{H_Y}(Z|\mathbf{u}) + \mathbf{H_Y}(\mathbf{u}) - \mathbf{H_Y}(Z)$

Th entropy identity provides the intuition that minimising the expected entropy $\mathbf{H_Y}(\mathbf{u}|Z)$ should be better than just providing utility balance (i.e. maximising $\mathbf{H_Y}(Z)$)



Sequential Monte Carlo estimation

To estimate and update the posterior density $\mathbf{P}(\mathbf{u}|Z, \mathbf{Y})$ we use the 'Reweight -Resample -Move' algorithm:

(U1) **Start**: Draw a sequence of unweighted iid-draws $(\mathbf{U}_1, \mathbf{1})$ for the prior $\mathbf{P}(\mathbf{u})$

(U2) **Reweight**: Given a weighted sample $(\mathbf{U}_1^N, \mathbf{W}_1^N)$ for $\mathbf{P}(\mathbf{u}|\mathbf{Y})$, get the updated weighted sample $(\mathbf{U}_2^{N+1}, \mathbf{W}_2^{N+1})$ for $\mathbf{P}(\mathbf{u}|Z, \mathbf{Y})$ by *reweighting*:

$$\mathbf{U}_2^{N+1} = \mathbf{U}_1^N, \quad \mathbf{W}_2^{N+1} = \frac{\mathbf{W}_1^N \cdot \mathbf{P}(Z|\mathbf{U}_1^N)}{\frac{1}{L}\langle \mathbf{W}_1^N, \mathbf{P}(Z|\mathbf{U}_1^N)\rangle} \quad \text{i.e.} \quad W_{2,i}^{N+1} = \frac{W_{1,i}^N \mathbf{P}(Z|\mathbf{u}_{1,i}^N)}{\frac{1}{L}\sum_{i=1}^L W_{1,i}^N \mathbf{P}(Z|\mathbf{u}_{1,i}^N)}$$

(U3) **Resample**:

(a) Sample $\mathbf{U}_3^{N+1}$ from $\mathbf{U}_2^{N+1}$ according to the weights $\mathbf{W}_2^{N+1}$,

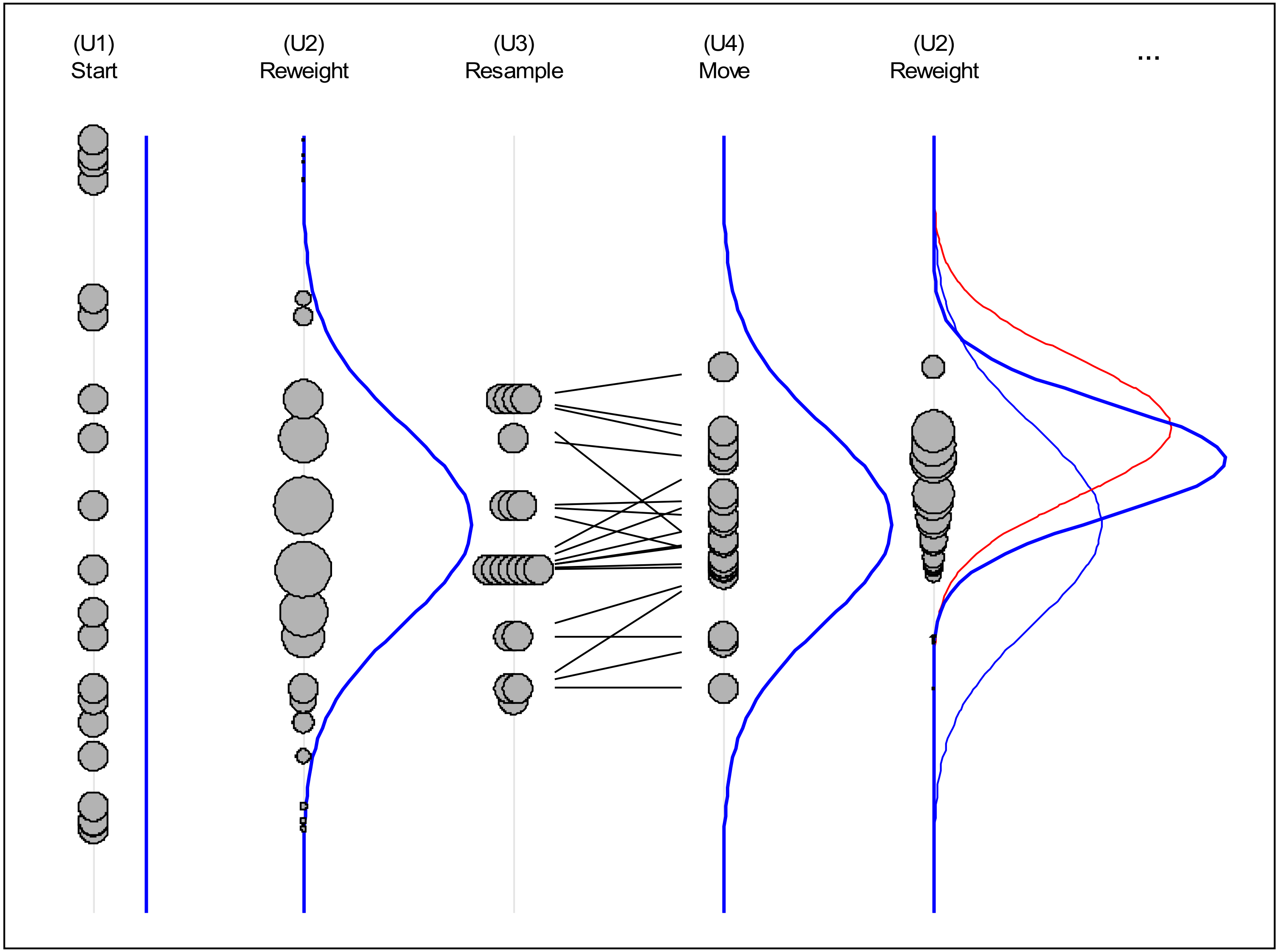
(b) Set $\mathbf{W}_3^{N+1} \equiv 1$

This is still a sample of $\mathbf{P}(\mathbf{u}|Z, \mathbf{Y})$, but with lower variance of the weights.

(U4) **Move**: Diversify the sample using a Markov kernel $\mathbf{K}$ as follows

(a) Sample $\mathbf{U}_1^{N+1}$ via $\mathbf{u}_{1,i}^{N+1} \sim \mathbf{K}(\mathbf{u}_{3,i}^{N+1})(\cdot)$,

(b) $\mathbf{W}_1^{N+1} \equiv \mathbf{W}_3^{N+1}$ (i.e. no change)


(U1)
Start
(U2)
Reweight
(U3)
Resample
(U4)
Move
(U2)
Reweight
...

We created a test-suite in R to compare the different design strategies in simulations. Our tests are based on a typical scenario from pharmaceutical market research

PRODUCT DESCRIPTION

Attribute	Levels	Ordered
Brand	3	No
Price	5	Yes
Efficacy	4	Yes
Duration	4	Yes
Application	2	No
Side-effects	4	No

DESIGN PARAMETERS

Screens with full profiles
3 products per screen
No fixed levels or attributes

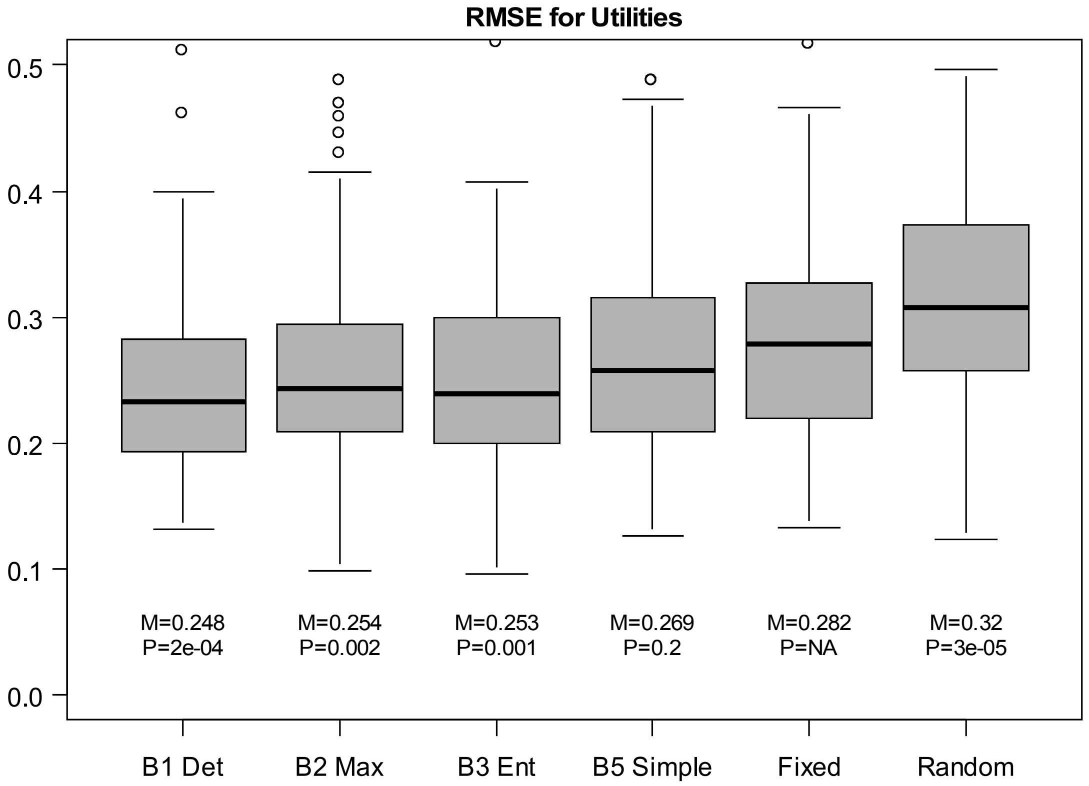
STUDY PARAMETERS

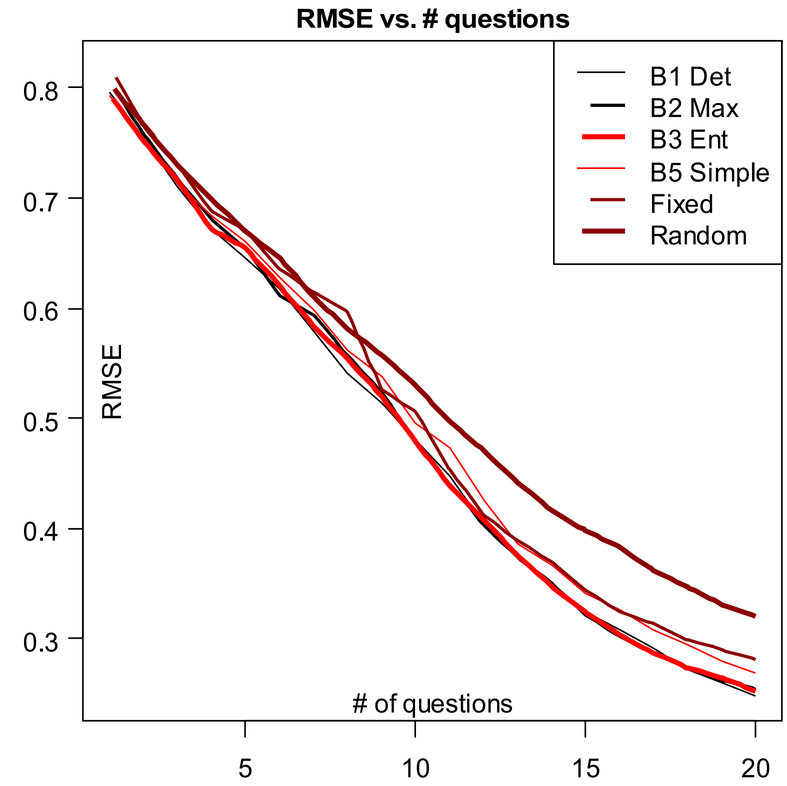
150 virtual respondents
20 questions
Distribution of 10 points

SCENARIOS

Algorithm	(**B1**), (**B2**), (**B3**), (**B5**), Fixed, Random
Prior parameters	$\mu = 0, \quad \Lambda = \mathbf{D}(\mathbf{5})$
Virtual respondents	(ordered) draws with $\mu = 0, \quad \Lambda = \mathbf{D}(\mathbf{2})$
Move-step	MH normal random walk
Size of SMC sample	5.000, 10.000, 20.000, 40.000

This is a typical scenario from pharmaceutical market research. Here, doctors are asked to estimate the prescription share of each of the presented treatments. Compared to simple choice questions, this provides sufficient information for individual-level estimation



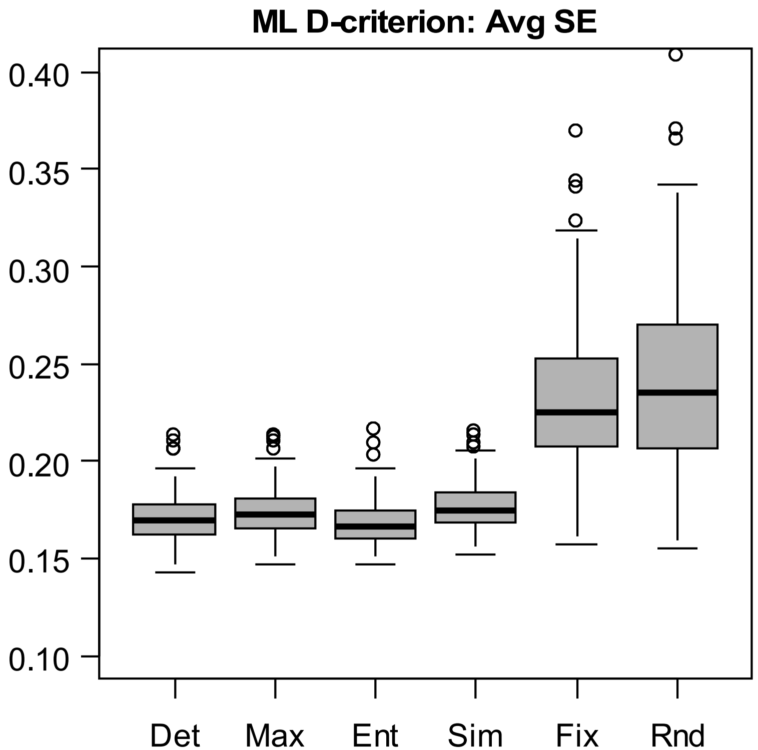


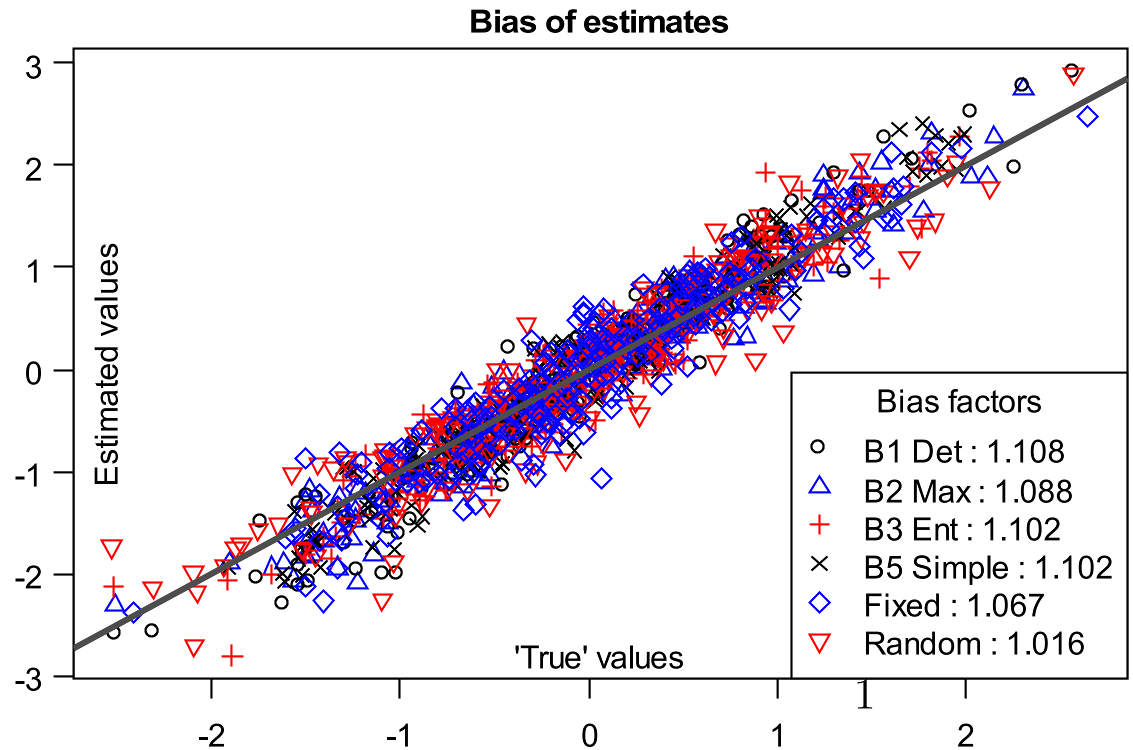
Results:

(a) Methods (**B1**), (**B2**), (**B3**) peform significantly and consistently better than the best fixed design in terms of RMSE. On average, the RMSE of these methods was 10%-15% better than the RMSE from the fixed design.

(b) The 'simplified utility balance' criterion (**B5**) performed as well as the fixed design in terms of RMSE, and thus significantly worse than the 'fully Bayesian' methods (**B1**), (**B2**), (**B3**).

Results (2)

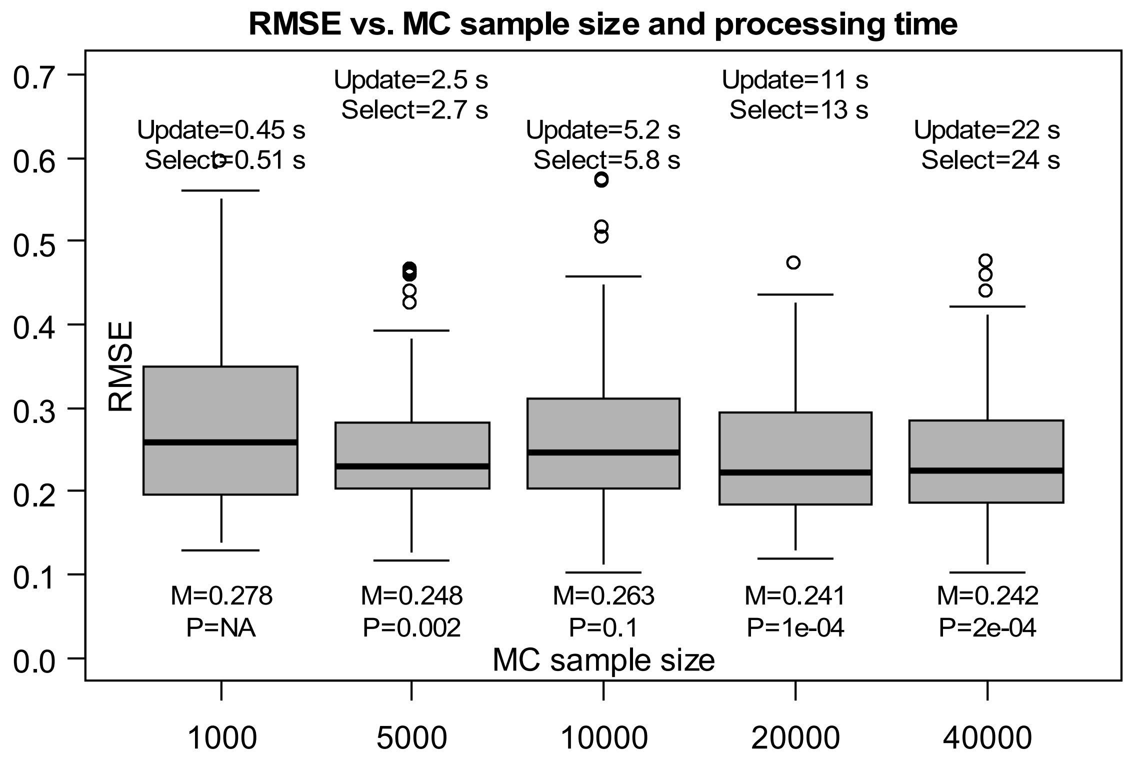




(a) The adaptive methods also perform well compared to the classical criteria.

(b) As is to be expected, there is considerable small-sample bias, which is strongly influenced by the choice of prior. We hope to alleviate this by using a hierarchical Bayes approach in the future.

(c) The results also indicate that there is an upwards oriented endogeneity bias.



1

(a) We tested the time-performance of the adaptive algorithms. The indicated times are for 1 core of an Intel Core2Duo at 1.83 GHz.

(b) The current R-implementation is not yet fast enough for online implementation, but tweaking of parameters, like reduction of the SMC sample size, and parallelisation will bring it there.

[CV] Chaloner, K., I. Verdinelli (1995): "Bayesian Experimental Design: A Review", Statistical Science **10 (3)**, 273-304

[Ch] Chopin, N. (2004): "Central limit theorem for sequential monte carlo methods and its application to bayesian inference", Annals of Statistics **32, 6** 2385-2411

[DDJ1] Del Moral, P., A. Doucet, A. Jasra (2006): "Sequential monte carlo samplers", Journal of the Royal Statistical Society B **68, 3**, 411–436

[DDJ2] Del Moral, P., A. Doucet, A. Jasra (2006): "Sequential Monte Carlo for Bayesian computation", in S. Bayarri, J.O. Berger, J.M. Bernardo, A.P. Dawid, D. Heckerman,A.F.M. Smith, M. West (eds.): Bayesian Statistics **8**

[GB] Gilks, W. R., C. Berzuini (2001): "Following a moving targetMonte Carlo inference for dynamic Bayesian models", Journal of the Royal Statistical Society B **63, 1**, pp. 127-146

[HZ] Huber, J., K. Zwerina (1996): "The Importance of Utility Balance in Efficient Choice Designs", Journal of Marketing Research **33**, 307–317

[Ke] Kessels, R. (2006): "Optimal Designs for the Measurement of Consumer Preferences", PhD-thesis, KU Leuven

[KTG] Kuhfeld, W.F. , R.D. Tobias, M. Garratt (1994): "Efficient Experimental Design with Marketing Research Applications", Journal of Marketing Research **31, 4**, 545–557.

[NMil] Nguyen, N-K. , A.J. Miller (1992): "A review of some exchange algorithms for constructing discrete D-optimal designs", Computational Statistics & Data Analysis **14**, 489–498

[R] R Development Core Team (2008) : "R: A language and environment for statistical computing", R Foundation for Statistical Computing", Vienna/Austria

[RAC] Rossi, P.E., G. Allenby, R. McCulloch (2005): "Bayesian Statistics and Marketing", John Wiley and Sons

[S] Sandor, Z. (2001): "Computation, Efficiency and Endogeneity in Discrete Choice Models", PhD-thesis, Labyrint Publication

[Sa] Sawtooth Software (1991-2007): "The ACA/Web v6.0", Sawtooth Software Technical paper series

[TSHD] Toubia, O., D. Simester, J.R. Hauser, E. Dahan (2003): "Fast Polyhedral Adaptive Conjoint Estimation", Marketing Science **22,3**, 273–303

[THS] Toubia, O., J.R. Hauser, D.I. Simester (2003): "Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis", Journal of Marketing Research, **41**, 116–131

[T] Train, K.E. (2003): "Discrete Choice Methods with Simulation", Cambridge University Press