Analysis of Conjoint Data: Part II: Logit

Aurélie Lemmens

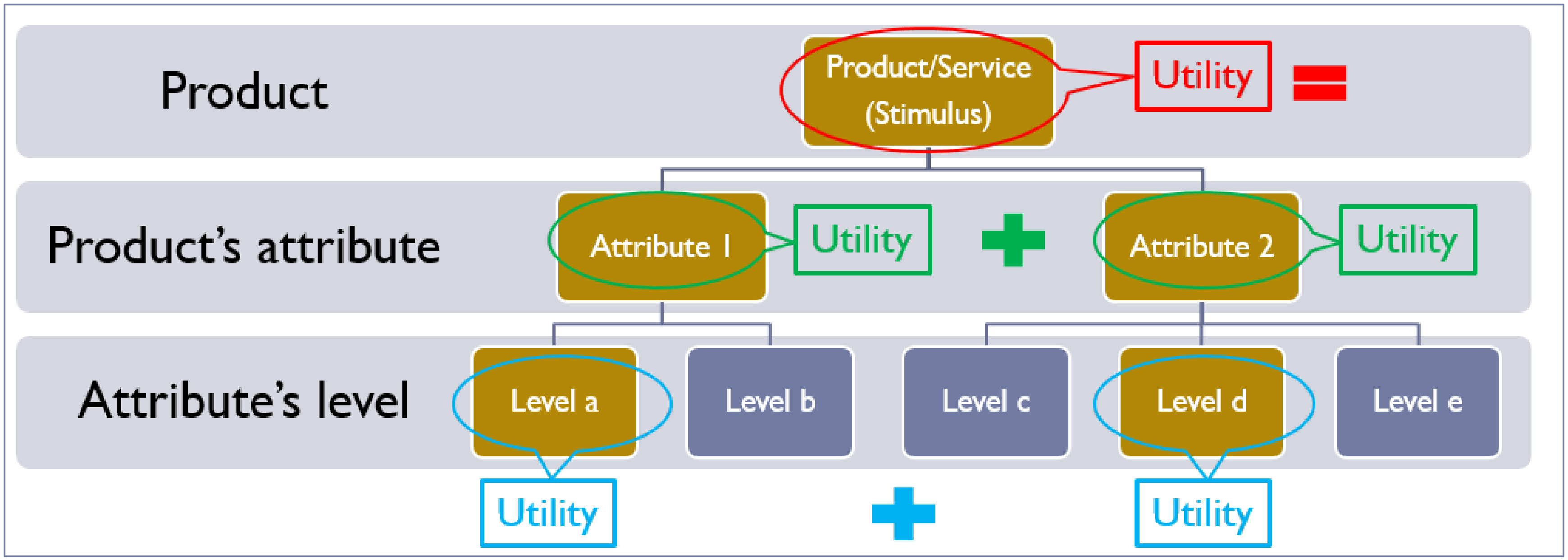
Logistic Regression

Daniel L. McFadden

Dan McFadden developed a method of **logistic regression** to analyze choices people made about such things as transportation.

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2000

Logistic Regression Allows us to Estimate Utilities and Choice Probabilities



Reminder: Random Utility Theory

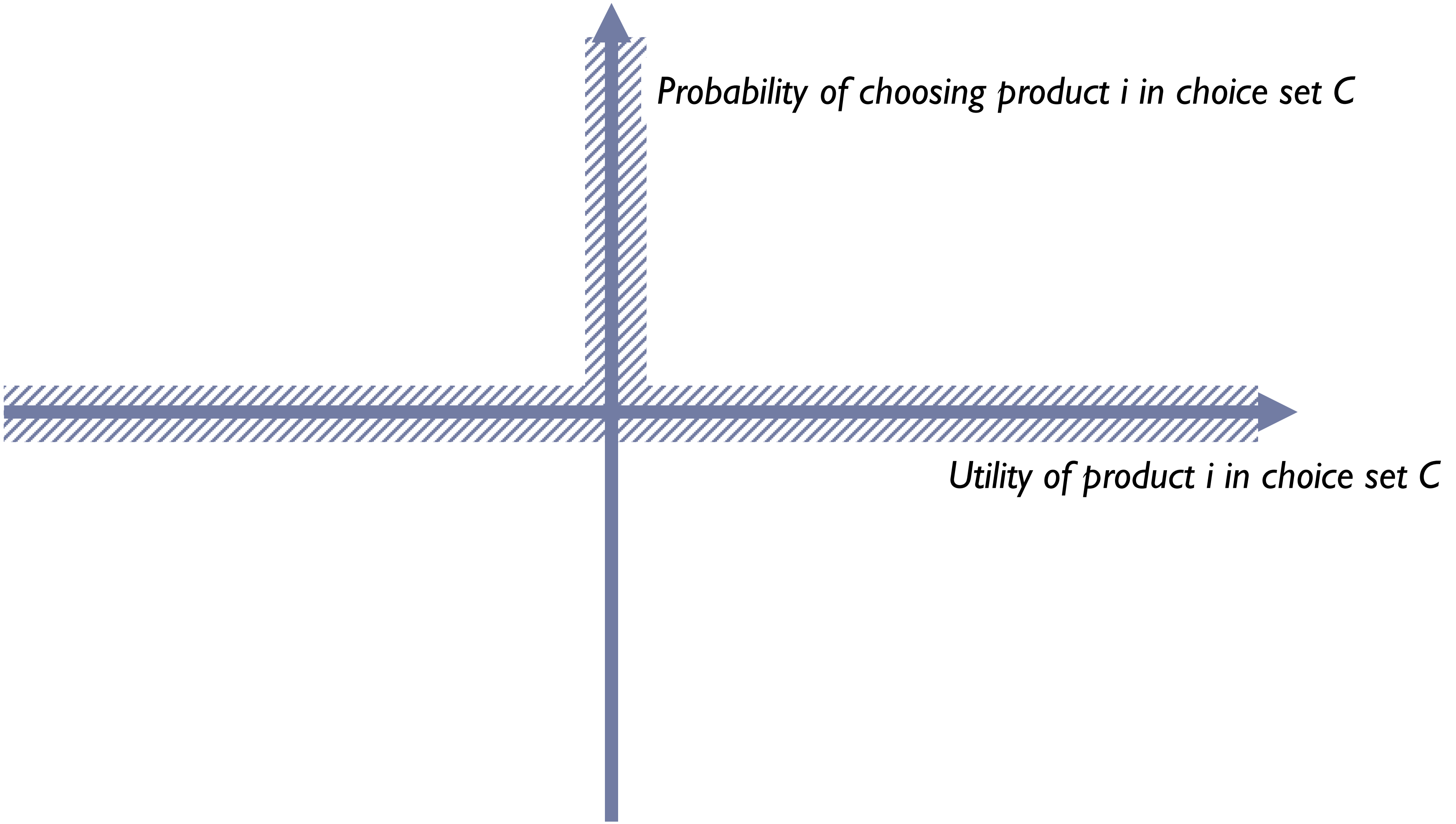
- Respondents are asked to choose a stimulus in a choice set C composed of n stimuli (e.g. products)
- Every stimulus is characterized by a set of k attributes, $x_{i1} \ldots x_{ik}$
- We observe $y_i = 1$ when stimulus i is chosen, $y_i = 0$ otherwise
- Thus, our data are y_i and $x_{i1} \ldots x_{ik}$
- We want to estimate
  - The vector of preferences for each attribute $\beta_1 \ldots \beta_k$ (*part-worths*)

$$U_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

  - The probability of choosing stimulus i among the choice set C

$$P(i|C) = P[U_i > U_j], \text{for all } j \in C$$

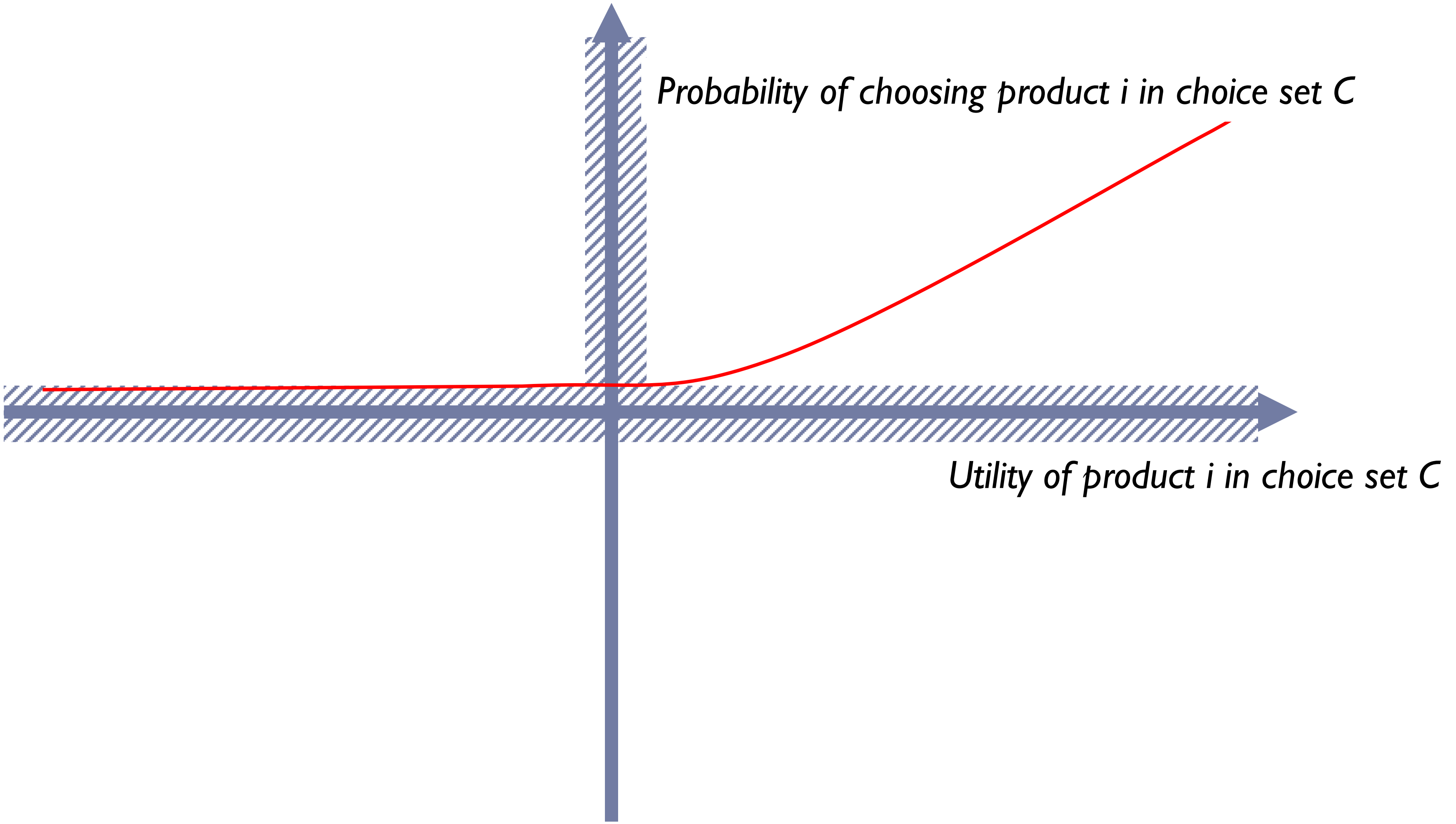
Mapping Utilities into Probabilities



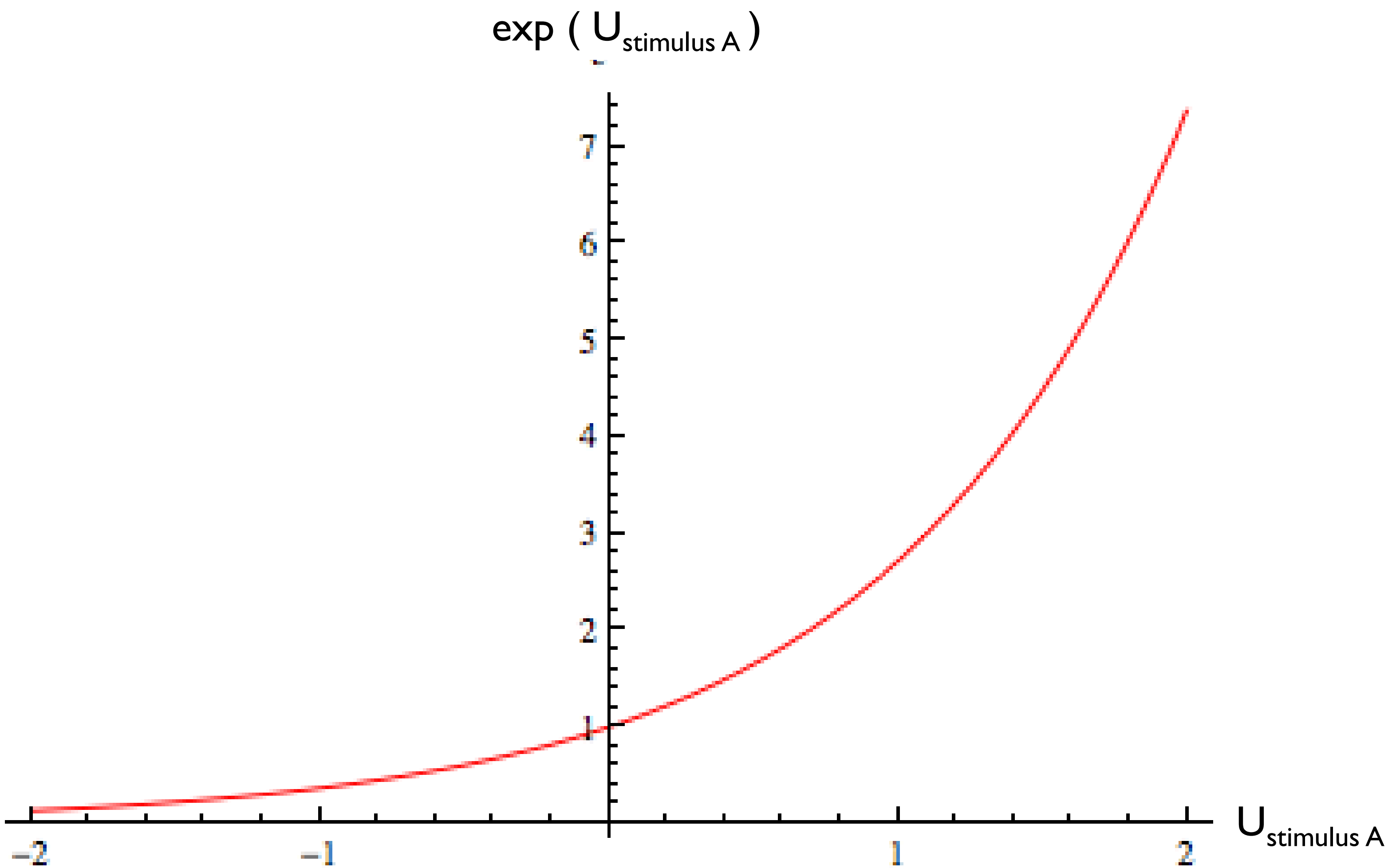
Estimating Utilities in CBC

- *Our data*
  - Response variable: choice
  - Explanatory variables: attributes of the hypothetical products
- *The aim of our analysis*
  - We don't observe utilities, just choices $\Rightarrow$ our model should predict choices (choice probabilities), not utilities
- *How do we transform utilities to choice probabilities?*
  - Taking into account that the choice probabilities should be
    - Positive
    - Between 0 and 1
    - Sum to 1 across all alternatives in a choice set

Mapping Utilities into Probabilities



Exponential function: exp(U)



Estimating Utilities in CBC

- **Multinomial Logit Model** (MNL):
  Assumes that the probability that an individual will choose one of the *m* alternatives *i* from the choice set *C* is:

$$p(i|C) = \frac{exp(U_i)}{\sum_{j=1}^{m} exp(U_j)} = \frac{exp(x_i\beta)}{\sum_{j=1}^{m} exp(x_j\beta)}$$

  where
  U_i = the utility of alternative *i*,
  x_i = a vector of attribute level dummies for alternative *i*,
  β = a vector with unknown part-worth utilities [to be estimated]

- Golf ball example:

$$U_i = \beta_1 \text{ HIGHFLY}_i + \beta_2 \text{ MAGNUM}_i + \beta_3 \text{ ECLIPSE}_i + \beta_4 \text{ LONGSHOT}_i + \beta_5 \text{ 5YARDS}_i + \beta_6 \text{ 10YARDS}_i + \beta_7 \text{ 15YARDS}_i + \beta_8 \text{ PRICE\_1}_i + \beta_9 \text{ PRICE\_2}_i + \beta_{10} \text{ PRICE\_3}_i + \beta_{11} \text{ PRICE\_4}_i$$

Estimating Utilities in CBC

	HIGHFLY	Beta_1	MAGNUM	Beta_2	ECLIPSE	Beta_3	LONGSHOT	Beta_4	5YARDS	Beta_5	10YARDS	Beta_6	15YARDS	Beta_7
Stimulus 1: High Fly, 10 yards, $10.99	1	0.54	0	0.36	0	-0.37	0	-0.53	0	-0.47	1	0.13	0	0.35
Stimulus 2: Magnum, 5 yards, $8.99	0	0.54	1	0.36	0	-0.37	0	-0.53	1	-0.47	0	0.13	0	0.35
Stimulus 3: Eclipse+, 10 yards, $6.99	0	0.54	0	0.36	1	-0.37	0	-0.53	0	-0.47	1	0.13	0	0.35
Stimulus 4: Long Shot, 15 yards, $4.99	0	0.54	0	0.36	0	-0.37	1	-0.53	0	-0.47	0	0.13	1	0.35
Stimulus 5: High Fly, 15 yards, $10.99	1	0.54	0	0.36	0	-0.37	0	-0.53	0	-0.47	0	0.13	1	0.35

	PRICE_1	Beta_8	PRICE_2	Beta_9	PRICE_3	Beta_10	PRICE_4	Beta_11	UTILITY	EXP(U)	PROBABILITY
Stimulus 1: High Fly, 10 yards, $10.99	0	0.66	0	0.17	0	-0.09	1	-0.74	-0.07	0.93	0.17
Stimulus 2: Magnum, 5 yards, $8.99	0	0.66	0	0.17	1	-0.09	0	-0.74	-0.2	0.82	0.15
Stimulus 3: Eclipse+, 10 yards, $6.99	0	0.66	1	0.17	0	-0.09	0	-0.74	-0.07	0.93	0.17
Stimulus 4: Long Shot, 15 yards, $4.99	1	0.66	0	0.17	0	-0.09	0	-0.74	0.48	1.62	0.30
Stimulus 5: High Fly, 15 yards, $10.99	0	0.66	0	0.17	0	-0.09	1	-0.74	0.15	1.16	0.21
									total:	5.46	1.00

Estimating Utilities in CBC

- **Multinomial Logit Model** (MNL):

  Assumes that the probability that an individual will choose one of the *m* alternatives *i* from the choice set *C* is:

$$p(i|C) = \frac{exp(U_i)}{\sum_{j=1}^{m} exp(U_j)} = \frac{exp(x_i \beta)}{\sum_{j=1}^{m} exp(x_j \beta)}$$

  where

  U_i = the utility of alternative *i*,

  x_i = a vector of attribute level dummies for alternative *i*,

  β = a vector with unknown part-worth utilities [to be estimated]

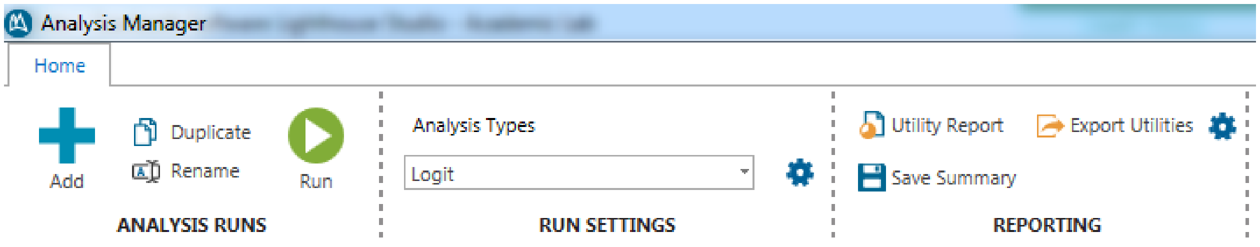
Estimation

Seek partworths (beta's) such that the predicted probabilities of chosen alternatives are maximized

Golf Ball Data - Estimation

- Logit model in Sawtooth
- Estimation output
  I. Summary of model fit
  II. Part-worth estimates and t-statistics
  III. Attribute importances

Summary of Model Fit



Analysis run 1

	A	B	C	D	E	F	G
3	Iteration	Chi-Square	RLH				
4	1	1166.44198	0.29207				
5	2	1196.33425	0.29323				
6	3	1196.41919	0.29324				
7	4	1196.41919	0.29324				
8	*Converged after 0.15 seconds.						
9							
10	Log-likelihood for this model	-4600.39426					
11	Log-likelihood for null model	-5198.60385					
12	Difference	598.20959					
13							
14	Percent Certainty	11.50712					
15	Akaike Info Criterion	9218.78852					
16	Consistent Akaike Info Criterion	9283.85412					
17	Bayesian Information Criterion	9274.85412					
18	Adjusted Bayesian Info Criterion	9246.25643					
19	Chi-Square	1196.41919					
20	Relative Chi-Square	132.93547					

Measure of model fit, higher values (less negative) better

Measure of model fit, higher values better

Chi-square relative to that of null (fully random) model

Part-worths: Estimates & T-stats

Variable	Effect	Std Error	t Ratio
High-Flyer Pro, by Smith and Forester	0.54407	0.03526	15.42997
Magnum Force, by Durango	0.36260	0.03523	10.29098
Eclipse+, by Golfers, Inc.	-0.37368	0.04088	-9.14155
Long Shot, by Performance Plus	-0.53299	0.04274	-12.46984
Drives 5 yards farther than the average ball	-0.47256	0.03278	-14.41561
Drives 10 yards farther than the average ball	0.12703	0.02914	4.35978
Drives 15 yards farther than the average ball	0.34553	0.02845	12.14726
$4.99 for package of 3 balls	0.65849	0.03540	18.59904
$6.99 for package of 3 balls	0.17237	0.03668	4.69901
$8.99 for package of 3 balls	-0.09275	0.03867	-2.39866
$10.99 for package of 3 balls	-0.73811	0.04475	-16.49244
NONE	0.00751	0.04141	0.18131

- 95% significance when |t|>1.96
- All but "NONE" significant

Part-worths: Interpretation

Variable	Effect	Std Error	t Ratio
High-Flyer Pro, by Smith and Forester	0.54407	0.03526	15.42997
Magnum Force, by Durango	0.36260	0.03523	10.29098
Eclipse+, by Golfers, Inc.	-0.37368	0.04088	-9.14155
Long Shot, by Performance Plus	-0.53299	0.04274	-12.46984
Drives 5 yards farther than the average ball	-0.47256	0.03278	-14.41561
Drives 10 yards farther than the average ball	0.12703	0.02914	4.35978
Drives 15 yards farther than the average ball	0.34553	0.02845	12.14726
$4.99 for package of 3 balls	0.65849	0.03540	18.59904
$6.99 for package of 3 balls	0.17237	0.03668	4.69901
$8.99 for package of 3 balls	-0.09275	0.03867	-2.39866
$10.99 for package of 3 balls	-0.73811	0.04475	-16.49244
NONE	0.00751	0.04141	0.18131

Effects coding: the last level is dropped and is estimated as minus the sum of all other levels of that attribute

(*Dummy coding*: the last level is dropped and constrained to zero.)

Effect Coding vs Dummy Coding

- Suppose the attribute brand with 4 levels: High-Flyer, Magnum, Eclipse, Long Shot
- One level is always considered as reference level. The parameter for this level is held constant.
  - Example: Long Shot is our reference level.
- We estimate the parameters for the other levels.
  - Example: we create 3 dummy variables for the attribute brand and estimate one parameter for each dummy.

- *Effect coding*: the dummy variables take value -1 for the reference level (see next slide)
- *Dummy coding*: the dummy variables take value 0 for the reference level

Effect Coding

If stimulus is :

	Dummy 1	Dummy 2	Dummy 3
Highfly	1	0	0
Magnum	0	1	0
Eclipse	0	0	1
Longshot	-1	-1	-1

$$U_i = \beta_1 \text{ HIGHFLY}_i + \beta_2 \text{ MAGNUM}_i + \beta_3 \text{ ECLIPSE}_i + \beta_4 \text{ LONG}_i$$

➔ we estimate β_1 , β_2 and β_3

Variable	Effect
High-Flyer Pro, by Smith and Forester	0.54407
Magnum Force, by Durango	0.36260
Eclipse+, by Golfers, Inc.	-0.37368
Long Shot, by Performance Plus	-0.53299

$$\beta_4 = - .54407 - .36260 - (-.37368) = -.53299$$

Part-worths: Interpretation

Variable	Effect	Std Error	t Ratio
High-Flyer Pro, by Smith and Forester	0.54407	0.03526	15.42997
Magnum Force, by Durango	0.36260	0.03523	10.29098
Eclipse+, by Golfers, Inc.	-0.37368	0.04088	-9.14155
Long Shot, by Performance Plus	-0.53299	0.04274	-12.46984
Drives 5 yards farther than the average ball	-0.47256	0.03278	-14.41561
Drives 10 yards farther than the average ball	0.12703	0.02914	4.35978
Drives 15 yards farther than the average ball	0.34553	0.02845	12.14726
$4.99 for package of 3 balls	0.65849	0.03540	18.59904
$6.99 for package of 3 balls	0.17237	0.03668	4.69901
$8.99 for package of 3 balls	-0.09275	0.03867	-2.39866
$10.99 for package of 3 balls	-0.73811	0.04475	-16.49244
NONE	0.00751	0.04141	0.18131

Most preferred brand

Least preferred brand

Interpreting Trade-Offs between Attributes

Variable	Effect
High-Flyer Pro, by Smith and Forester	0.54407
Magnum Force, by Durango	0.36260
Eclipse+, by Golfers, Inc.	-0.37368
Long Shot, by Performance Plus	-0.53299
Drives 5 yards farther than the average ball	-0.47256
Drives 10 yards farther than the average ball	0.12703
Drives 15 yards farther than the average ball	0.34553
$4.99 for package of 3 balls	0.65849
$6.99 for package of 3 balls	0.17237
$8.99 for package of 3 balls	-0.09275
$10.99 for package of 3 balls	-0.73811
NONE	0.00751

Trade-off performance vs price?

Utility gain from driving 10 instead of 5 yards
= 0.12697 - - 0.47255 = 0.59952

Utility loss from paying $6.99 instead of $4.99
= 0.17234 - 0.65858 = -0.48624

Consumers would be willing to pay $6.99 instead of $4.99 when they can drive 10 yrds farther instead of 5

Interpreting Trade-Offs between Attributes

Variable	Effect
High-Flyer Pro, by Smith and Forester	0.54407
Magnum Force, by Durango	0.36260
Eclipse+, by Golfers, Inc.	-0.37368
Long Shot, by Performance Plus	-0.53299
Drives 5 yards farther than the average ball	-0.47256
Drives 10 yards farther than the average ball	0.12703
Drives 15 yards farther than the average ball	0.34553
$4.99 for package of 3 balls	0.65849
$6.99 for package of 3 balls	0.17237
$8.99 for package of 3 balls	-0.09275
$10.99 for package of 3 balls	-0.73811
NONE	0.00751

Trade-off brand vs price?

How much extra $ would people give for Eclipse + compared to Long Shot?

The difference in utility is about .16.

Such a difference would be compensated by a price increase of ???

Depending on the price range, a $2 increase costs at least .25 in utility, meaning that a change in brand name should not be associated with more than $1 price increase.

Adding the No-Choice Option

- We add a new dummy variable (called "NONE" in Sawtooth)

If stimulus is :

	Dummy 1	Dummy 2	Dummy 3	Dummy 4
Highfly	1	0	0	**0**
Magnum	0	1	0	**0**
Eclipse	0	0	1	**0**
Longshot	-1	-1	-1	**0**
None	**0**	**0**	**0**	**1**

$$U_i = \quad \ldots \quad + \beta_5 \ \text{NONE}_i$$

No-Choice: Interpretation

Variable	Effect
High-Flyer Pro, by Smith and Forester	0.54407
Magnum Force, by Durango	0.36260
Eclipse+, by Golfers, Inc.	-0.37368
Long Shot, by Performance Plus	-0.53299
Drives 5 yards farther than the average ball	-0.47256
Drives 10 yards farther than the average ball	0.12703
Drives 15 yards farther than the average ball	0.34553
$4.99 for package of 3 balls	0.65849
$6.99 for package of 3 balls	0.17237
$8.99 for package of 3 balls	-0.09275
$10.99 for package of 3 balls	-0.73811
NONE	0.00751

How do we interpret the no-choice part-worth?

.00751 is the threshold utility for buying. That is below this utility, customers would prefer not to buy.

Let's take an example: what should be the price and performance level of a Long Shot ball to convince customers to buy (rather than not buying at all)?

Utility Long Shot = -.53299
Utility None = .00751

Difference = -.52548

The difference can be overcome with e.g. a price of $4.99 and a 10 yards performance (.65849+.12703)

Attribute Importances

- Suppose $Range_m$ indicates the *range in absolute value* of partworths for attribute m (=1,…,K)
- Then

$$\text{Importance of attribute } m = |Range_m| / (|Range_1| + |Range_2| + \ldots + |Range_K|)$$

Variable	Effect
High-Flyer Pro, by Smith and Forester	0.54407
Magnum Force, by Durango	0.36260
Eclipse+, by Golfers, Inc.	-0.37368
Long Shot, by Performance Plus	-0.53299
Drives 5 yards farther than the average ball	-0.47256
Drives 10 yards farther than the average ball	0.12703
Drives 15 yards farther than the average ball	0.34553
$4.99 for package of 3 balls	0.65849
$6.99 for package of 3 balls	0.17237
$8.99 for package of 3 balls	-0.09275
$10.99 for package of 3 balls	-0.73811
NONE	0.00751

<u>Importance</u>

$Range_1$ = 0.54 - - 0.53 = 1.08 = 1.08/(1.08+0.82+1.40) = **32.72%**

$Range_2$ = 0.34 - - 0.47 = 0.82 = 0.82/(1.08+0.82+1.40) = **24.85%**

$Range_3$ = 0.66 - - 0.74 = 1.40 = 1.40/(1.08+0.82+1.40) = **42.43%**

Attribute Importances

Brand:	32.72267
Performance:	24.85227
Price:	42.42505

Transforming Utilities in Probabilities

- Suppose two alternative stimuli offered to consumers

Stimuli 1		Stimuli 2	
High-Flyer Pro	.54	Eclipse +	-.37
Drives 10 yards	.13	Drives 15 yards	.35
$6.99 for 3 balls	.17	$6.99 for 3 balls	.17
Total Utility_1	.84	Total Utility_2	.15
$\text{Exp}(\text{Utility}_1)$	2.32	$\text{Exp}(\text{Utility}_2)$	1.16

$$\text{Exp}(\text{Utility}_1) + \text{Exp}(\text{Utility}_2) = 3.48$$

Probability_1	2.32/3.47 = 66.8%	Probability_2	1.16/3.48 = 33.2%