Identifying *two* piecewise linear additive value functions from anonymous preference information

Vincent Auriau, Khaled Belahcène, Emmanuel Malherbe,
Vincent Mousseau, Marc Pirlot

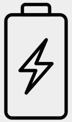
ARTEFACT RESEARCH CENTER
CentraleSupélec
université PARIS-SACLAY
UMONS University of Mons

Outline of the presentation

1. **The Additive Value Function model**
   - Introduction
   - Eliciting an AVF with standard queries
2. **Formalism and motivations**
3. **Elicitation with two AVF from anonymous preference information**
   - Queries geometry: single rectangle & neighbouring rectangle queries
   - Building block of elicitation from several queries
   - General idea of the elicitation algorithm
4. **Experiments**

A motivating example

Comparison of electric cars on different criteria:

 *Autonomy*

 *Sale Price*

> How much does a customer value the increase in autonomy vs increase in price ?

> Is the balance the same for cheap and expensive cars ?

The Additive Value Function (AVF) model

Let x, y be two different alternatives (cars)

Let $\mathcal{I} = \{\text{autonomy, price}\}$ be the criteria ensemble

$$x \succ y \Leftrightarrow \sum_{i \in \mathcal{I}} u_i(x_i) > \sum_i u_i(y_i)$$

$$x \sim y \Leftrightarrow \sum_{i \in \mathcal{I}} u_i(x_i) = \sum_i u_i(y_i)$$

$\succ$ *is asymmetric, transitive*

$\sim$ *is symmetric, transitive*

$\cdot \rightarrow u_i(\cdot)$ *latent marginal utility, defined monotonously increasing*

The ambition is to estimation the u_i in order to compare non-dominated alternatives, e.g. (40,000$; 400km) and (50,000$; 500km)

Estimation of the model

Learning based on examples:

- *Dataset of purchases or expressed preferences*
- *Usual methods formulated as Linear Programming problems*

Elicitation by questioning a "Decision Maker"

- *Which question to ask to efficiently model the preferences ?*
- *Theoretical background to make sure the model is identifiable*

*> A **theoretical** method is to build standard sequences: noise is not considered, not aimed at being used in practice*

Elicitation of the model: building standard sequences

Starting from a base car with *(25k$, 250 km)* we ask **indifference queries**:

(25 k$; 250 km) ~(30 k$; ?)

“ How many more kms should be the autonomy to compensate an increase of 5k$ for a car with 250 km of autonomy that costs 25k$? ”

Elicitation of the model: building standard queries

Starting from a base car with *(25k$, 250 km)* we ask **indifference queries**:

(25 k$; 250 km) ~(30 k$; ?)

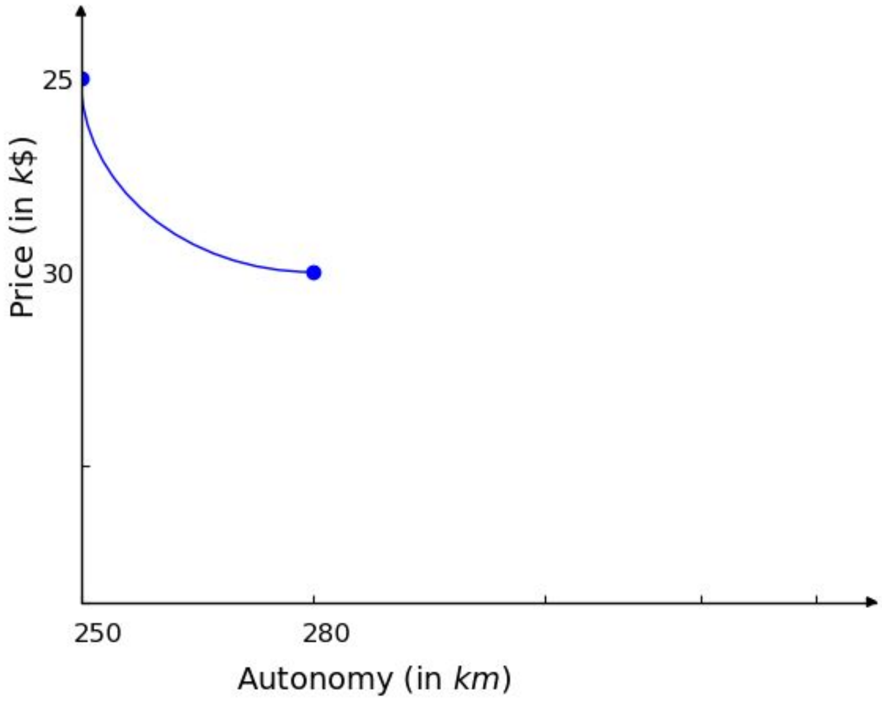
“ How many more kms should be the autonomy to compensate an increase of 5k$ for a car with 250 km of autonomy that costs 25k$? ”

> With an example answer of *280 km* from the DM

Elicitation of the model: building standard queries

We have an **indifference**:

(25 k$; 250 km) ~(30 k$; 280 km)



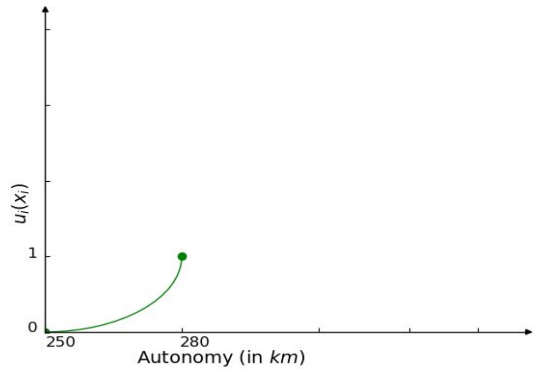
Elicitation of the model: building standard queries

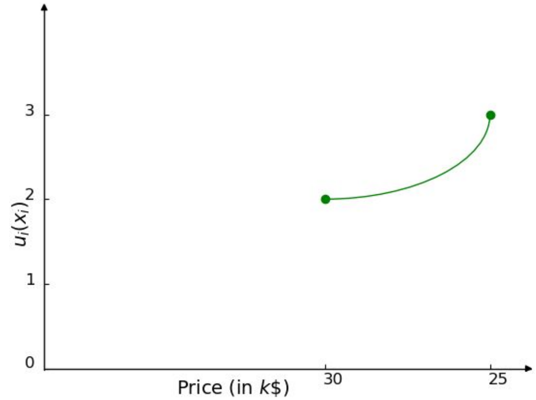
We have an **indifference**:

(25 k$; 250 km) ~(30 k$; 280 km)

or written differently:

$$u_{price}(25) + u_{auto}(250) = u_{price}(30) + u_{auto}(280)$$





ARTEFACT

Elicitation of the model: building standard queries

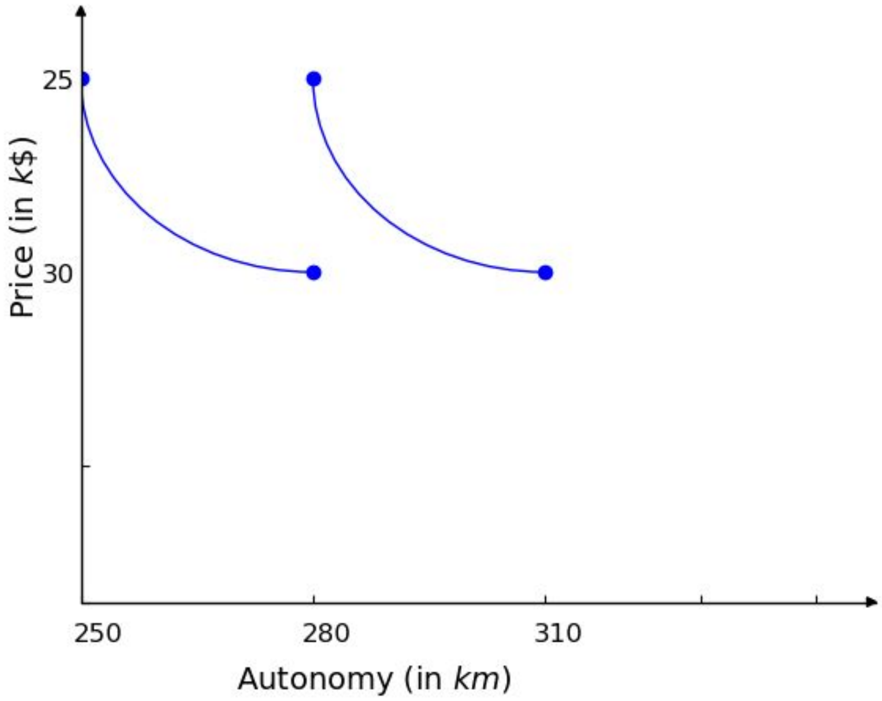
1st query:

(25 k$; 250 km) ~(30 k$; 280 km)

2nd query:

(25 k$; 280 km) ~(30 k$; ?)

> *answer: 310*

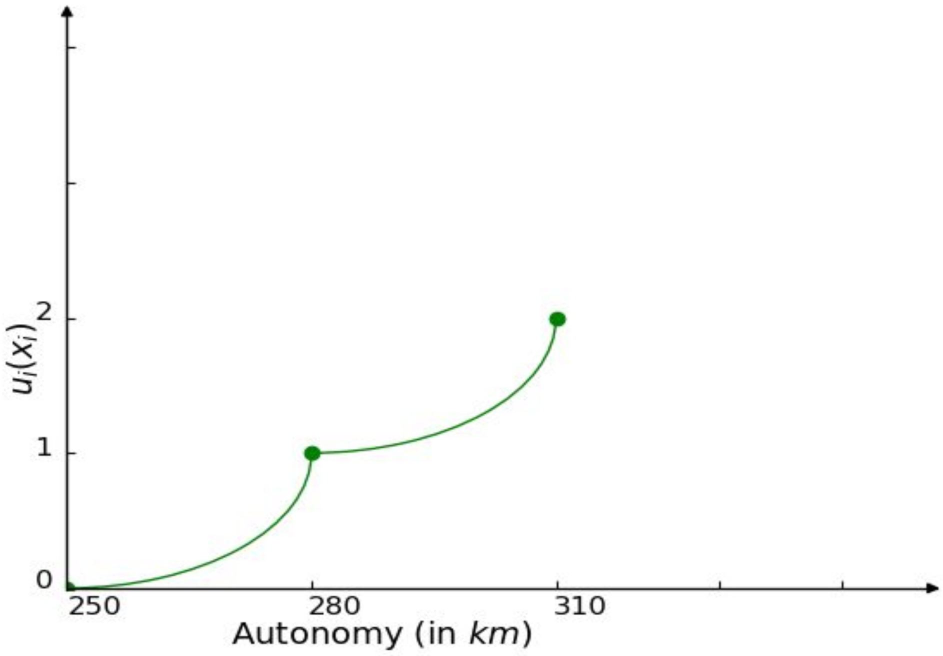


Elicitation of the model: building standard queries

2nd query:

(25 k$; 280 km) ~(30 k$; ?)

> *answer: 310*



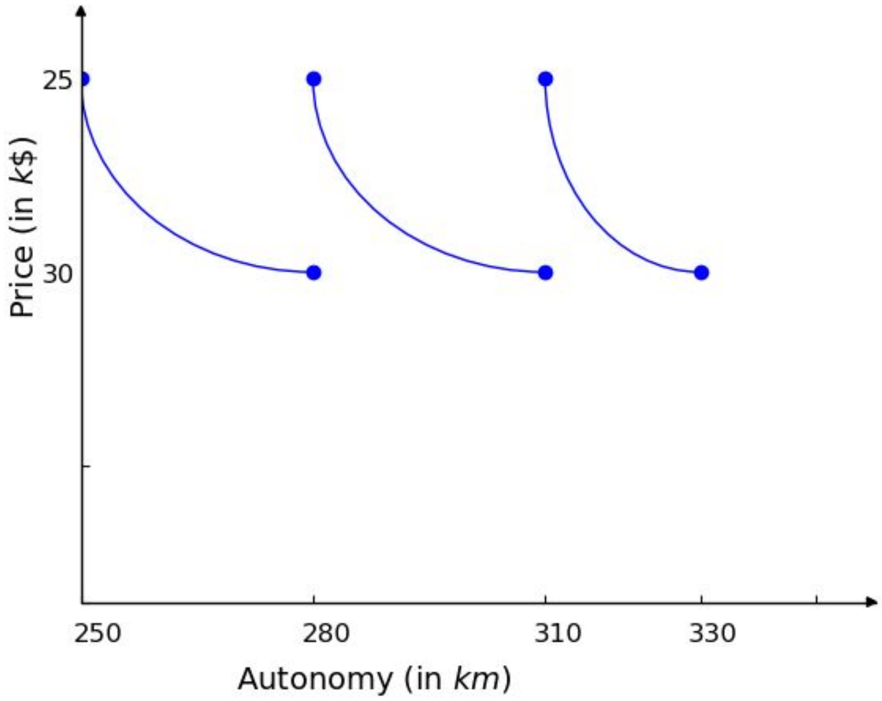
ARTEFACT

Elicitation of the model: building standard queries

3rd query:

(25 k$; 310 km) ~(30 k$; ?)

> answer: 330



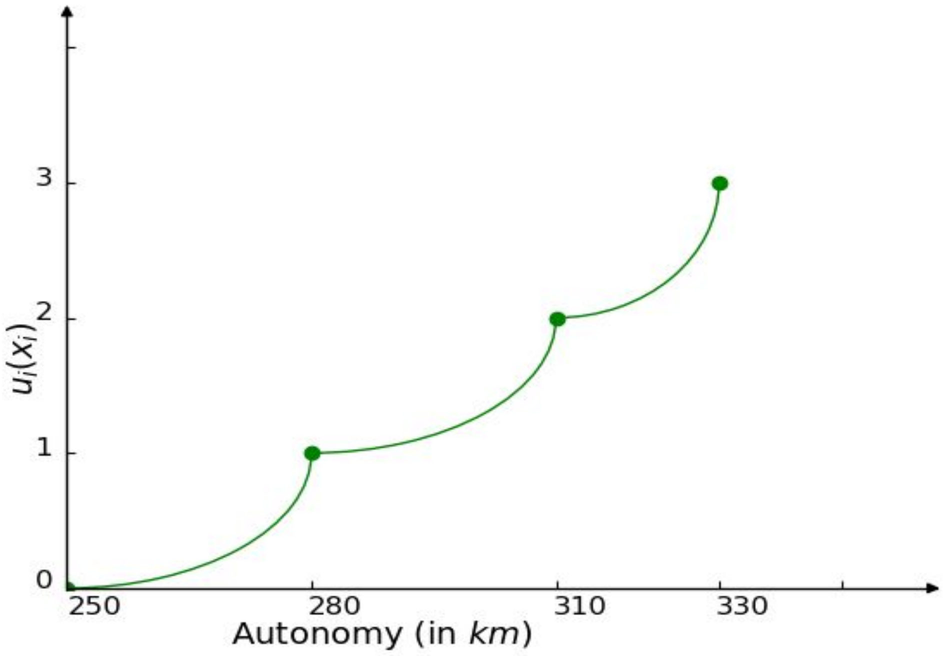
ARTEFACT

Elicitation of the model: building standard queries

3rd query:

(25 k$; 310 km) ~(30 k$; ?)

> *answer: 330*



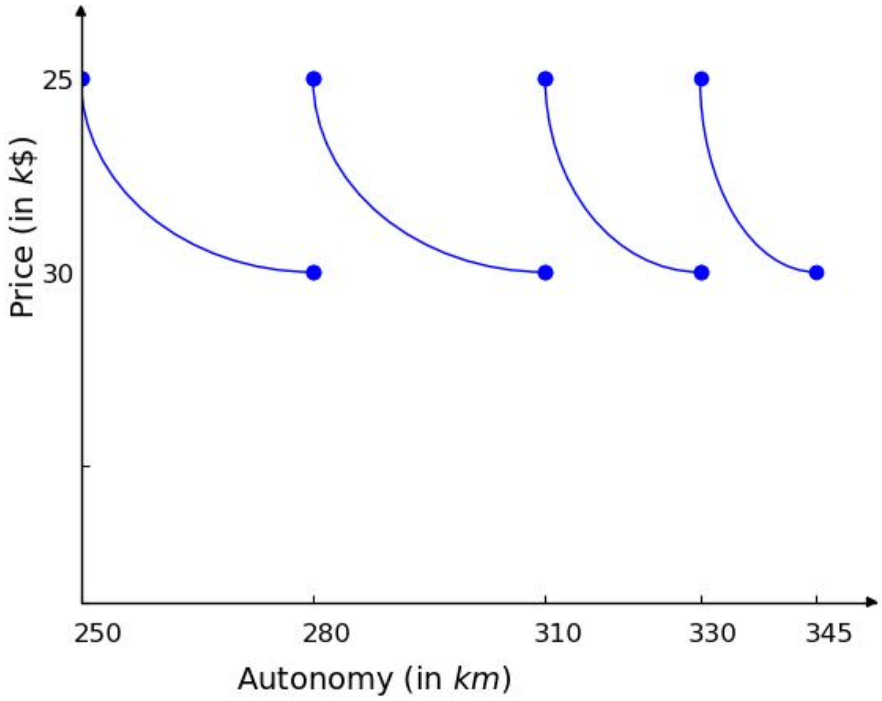
ARTEFACT

Elicitation of the model: building standard queries

*4*th query:

(25 k$; 330 km) ~(30 k$; ?)

> *answer: 345*



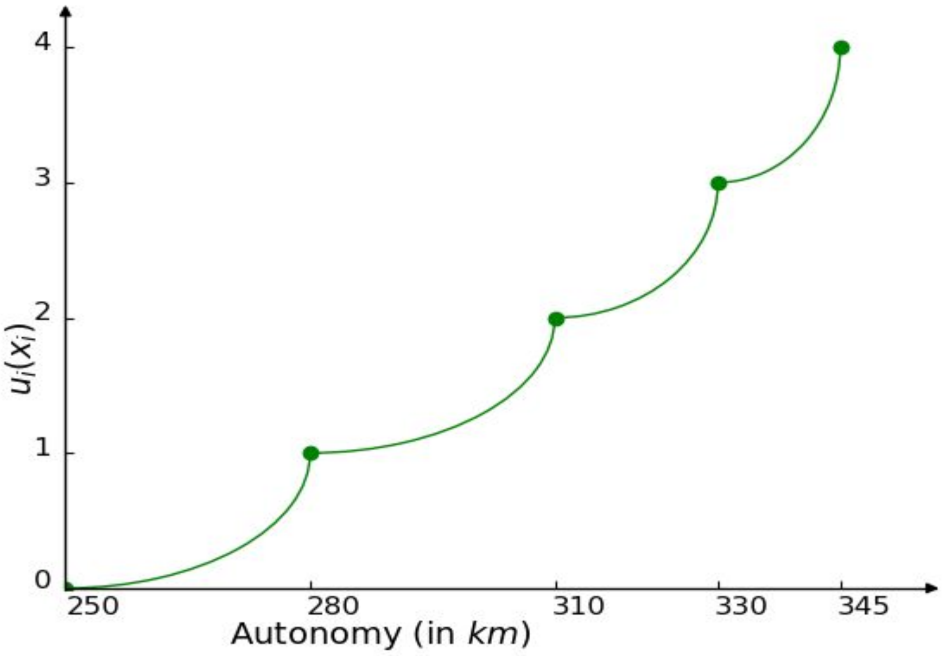
ARTEFACT

Elicitation of the model: building standard queries

*4*th query:

(25 k$; 330 km) ~(30 k$; ?)

> *answer: 345*



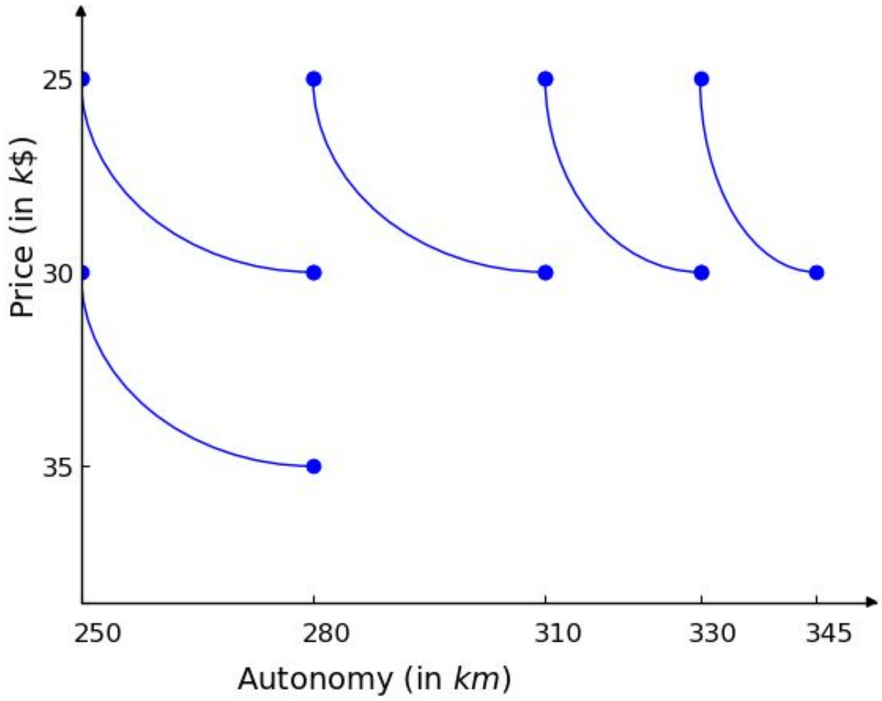
ARTEFACT

Elicitation of the model: building standard queries

5^{th} query:

(30 k$; 250 km) ~(?; 280 km)

>35 k$



ARTEFACT

Elicitation of the model: building standard queries

5th query:

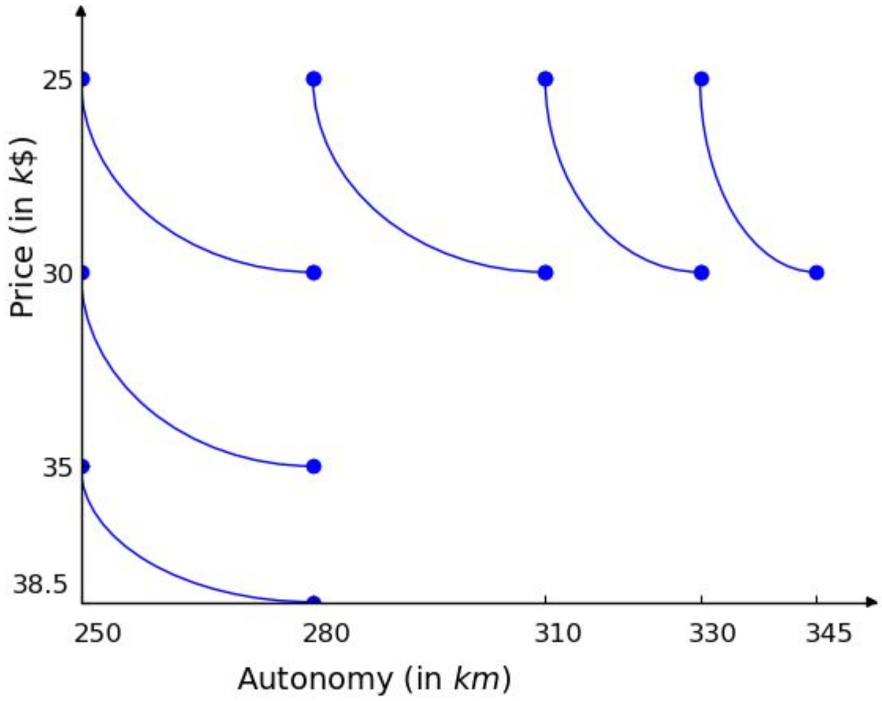
(30 k$; 250 km) ~(?; 280 km)

>35 k$

6th query:

(35 k$; 250 km) ~(?; 280 km)

>38.5 k$



ARTEFACT

Elicitation of the model: Thomsen Condition

(30 k$; 250 km) ~(35 k$; 280 km)

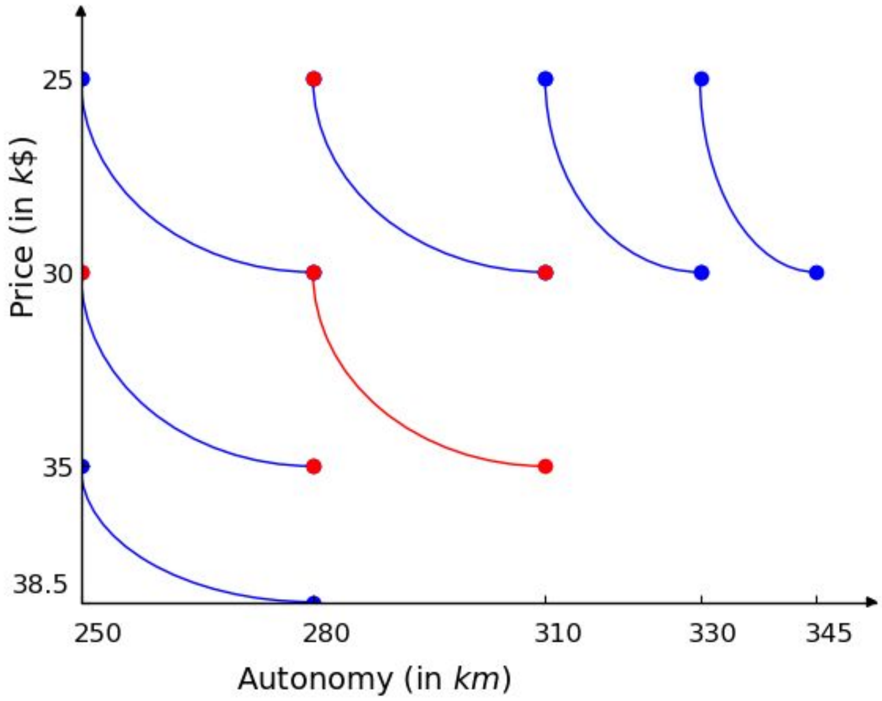
and

(25 k$; 280 km) ~(30 k$; 310 km)

implies that:

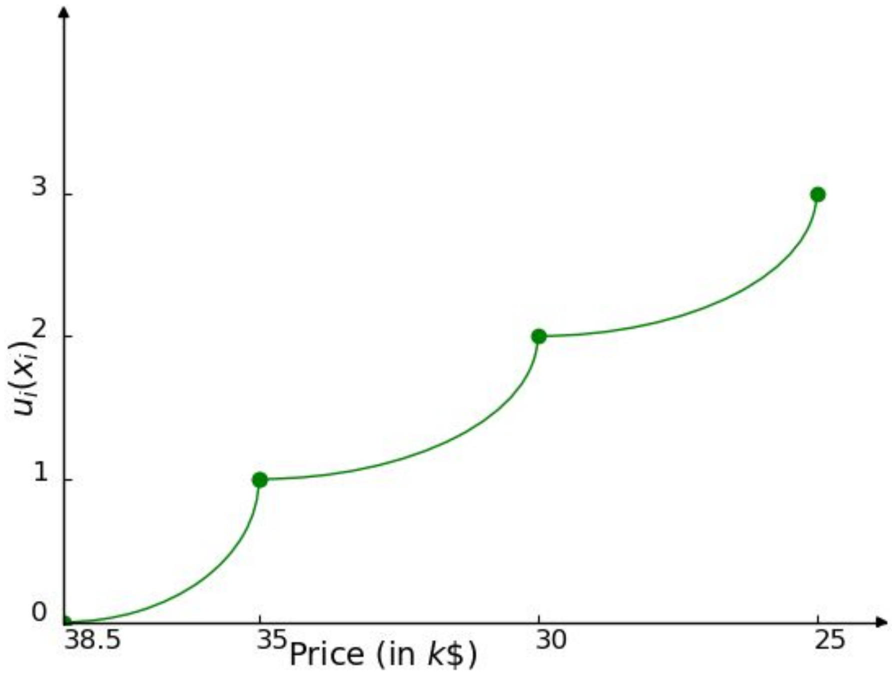
(30 k$; 280 km) ~(35 k$; 310 km)

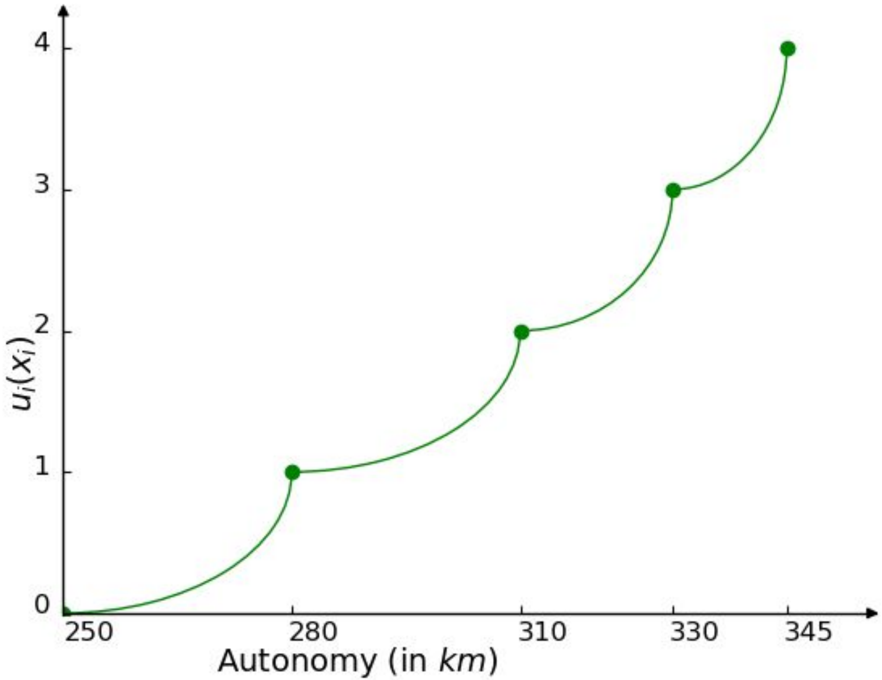
> Completion of the indifferences in the criteria space



ARTEFACT

Elicitation of the model: Thomsen Condition





Interpolation can be used to generalize to any value

ARTEFACT

What happens if we consider several (2) DMs ?

> Consider a **group of two customers** *(Traveler & Commuter)* with different preferences

> The group **can be queried** and both customers answer, **we don’t know which answer comes from which customer**

The indifference query: *(25; 250) ~(30; ?)*

yields two answers: *(280, 310)*

> Inability to directly use the previous method

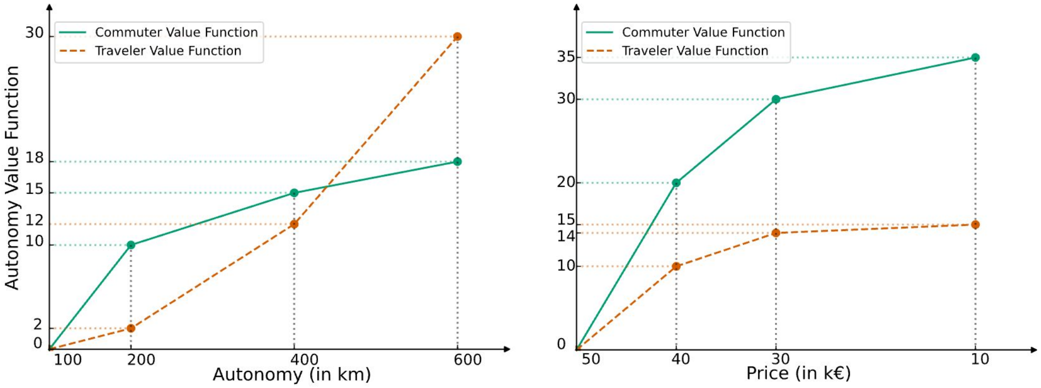
> We make additional assumptions (piecewise linear marginals)

The 2-UTA: Two Piecewise Linear Additive Value Functions

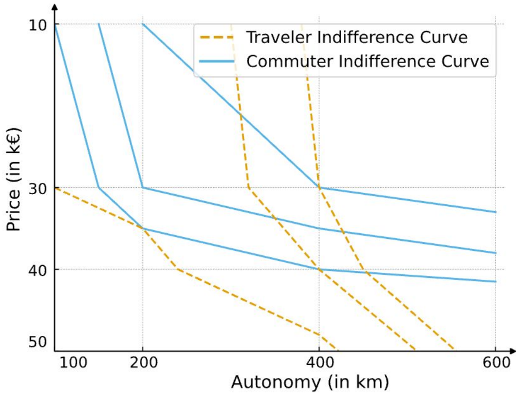
> Piecewise linear value function with a priori-defined intervals (same for both DMs)

> Indifference curves are therefore also piecewise linear

Value Functions



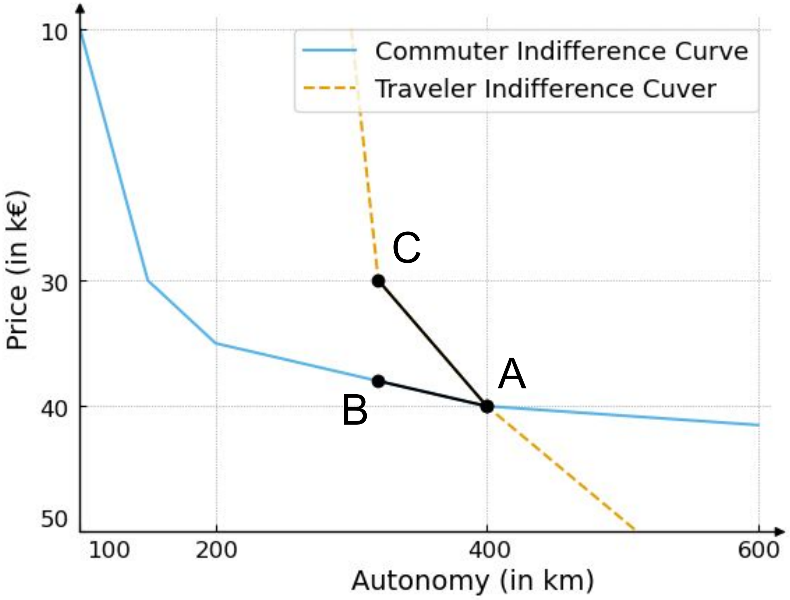
Indifference Curves



We propose an elicitation strategy for such unknown 2-UTA model

ARTEFACT

Defining base indifference queries



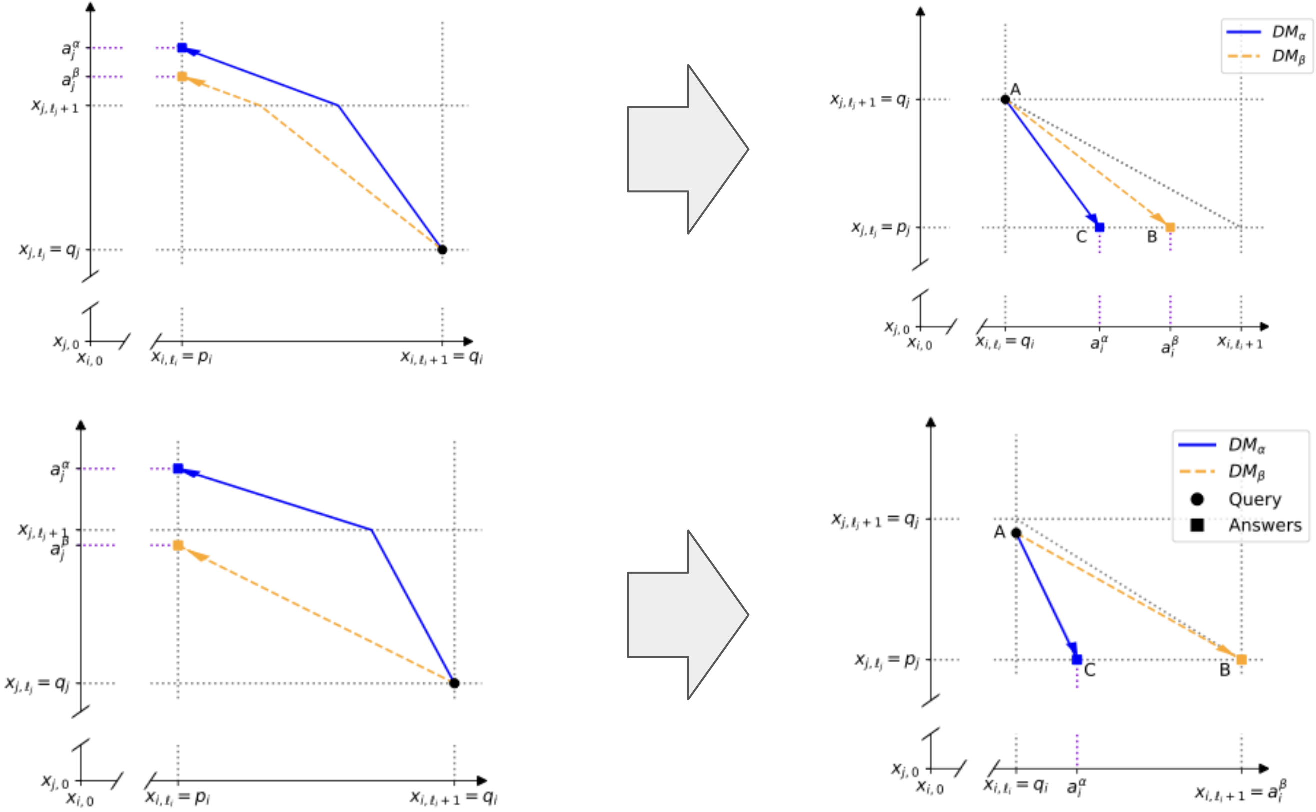
Single Rectangle
Query

Query: $(A_a, A_p) \sim (B_a, ?)$

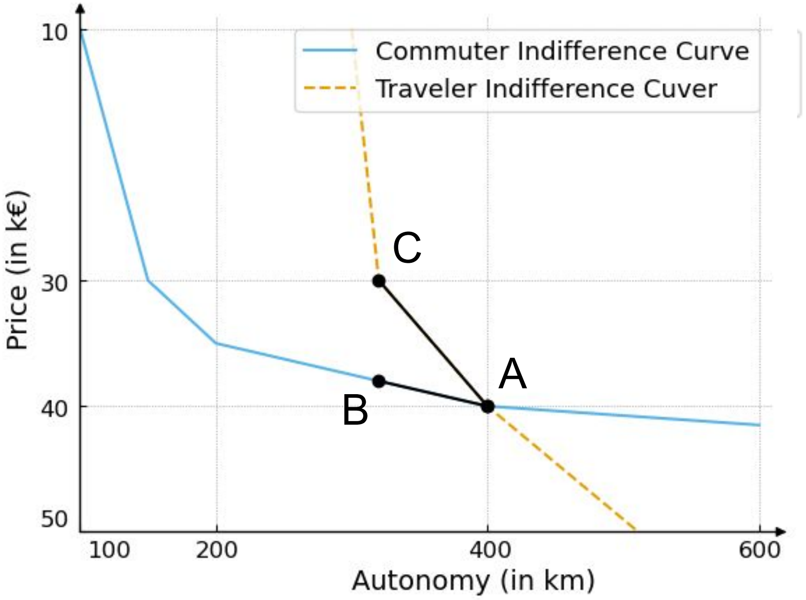
Answers: B_p , C_p

ARTEFACT

Ensuring Single Rectangle Queries



Deducting information from a single rectangle query

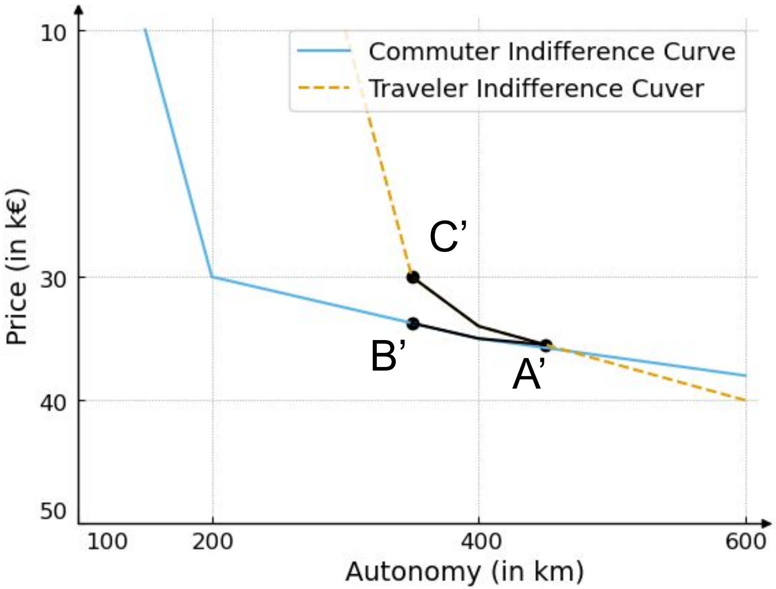


$$\begin{cases}(A_i, A_j) \sim_\kappa (B_i, B_j) \\ (A_i, A_j) \sim_{\kappa'} (C_i, C_j)\end{cases} \Leftrightarrow \begin{cases} u_\kappa((A_i, A_j)) = u_\kappa((B_i, B_j)) \\ u_{\kappa'}((A_i, A_j)) = u_{\kappa'}((C_i, C_j))\end{cases}$$

$$u_{\kappa,j}(A_j) = u_\kappa((B_i, B_j)) - u_{\kappa,i}(A_i)$$
$$u_{\kappa',j}(A_j) = u_{\kappa'}((C_i, C_j)) - u_{\kappa',i}(A_i)$$

> If $u_.((B_i, B_j))$ and $u_{.,i}(A_i)$ are known we can deduce 2 potential couples of values for $u_{.,j}(A_j)$
> Additional information is needed for definite attribution of *(B, C)* to (α, β)

Defining base indifference queries



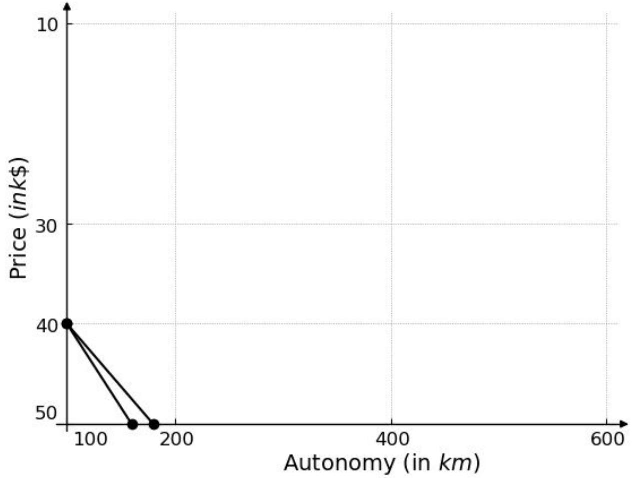
Single Rectangle
Query

Query: $(A'_a, A'_p) \sim (B'_a, ?)$

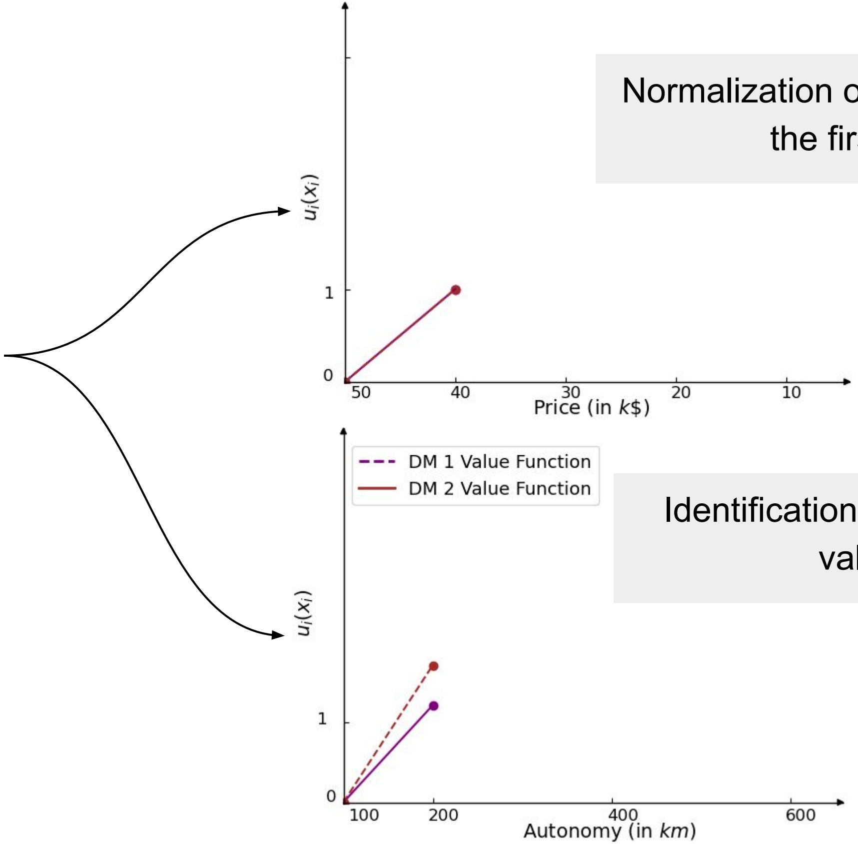
Answers: B'_p , C'_p

ARTEFACT

Elicitation algorithm: Query #1



Query in the criteria space

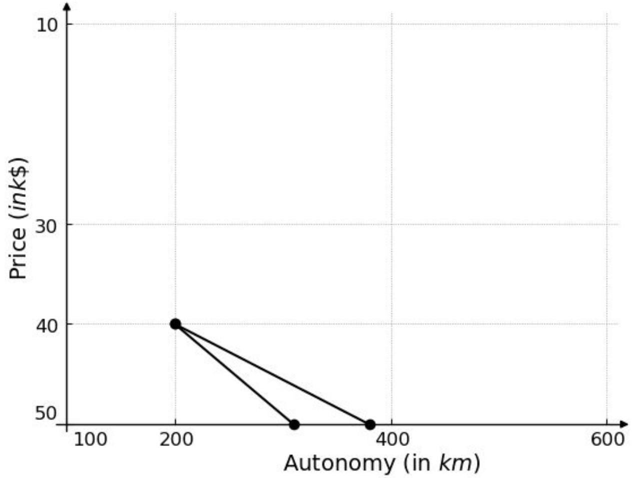


Normalization on the first interval of the first criterion

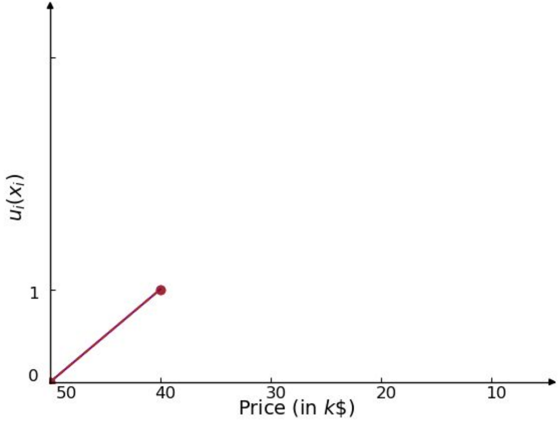
Identification of the slopes in the value space

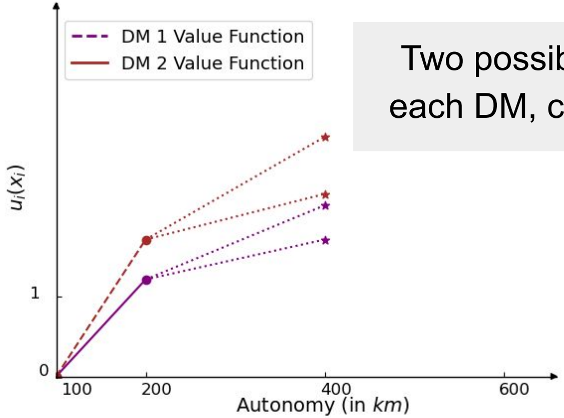
ARTEFACT

Elicitation algorithm: Query #2



Query in the criteria space

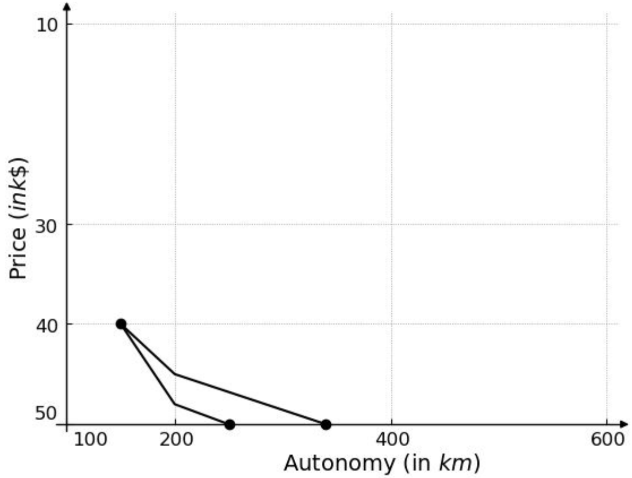




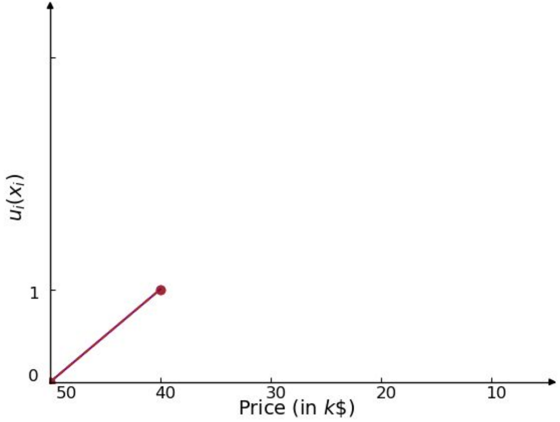
Two possible value functions for each DM, cannot be identified yet

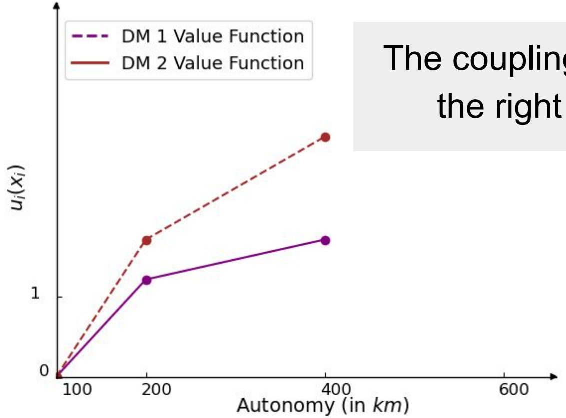
ARTEFACT

Elicitation algorithm: Query #2.5



Coupling Query in the criteria space





The coupling query lets us identify the right slope for each DM

Elicitation: General algorithm

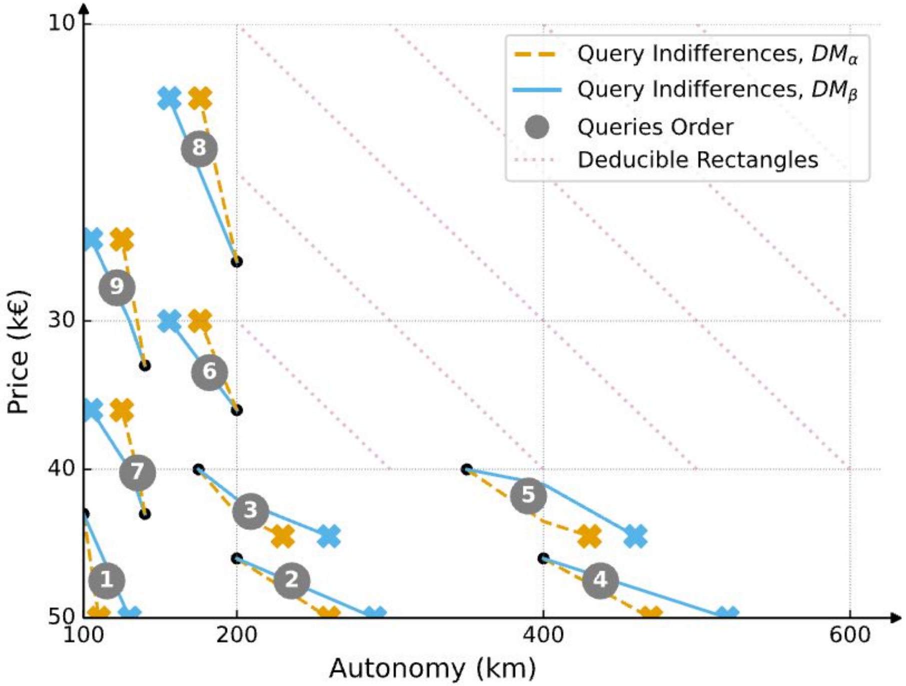
Algorithm 5: General Strategy for the Elicitation

Initialize criterion i:
- Set $\gamma^{\alpha}_{i,\ell_i} = \gamma^{\beta}_{i,\ell_i} = 1$ for all DMs on $[x_{i,\ell_i-1}, x_{i,\ell_i}]$

Initialize criterion j:
- Play Pattern 1 on $[x_{i,\ell_i-1}, x_{i,\ell_i}] \times [x_{j,\ell_j}, x_{j,\ell_j+1}]$;
- Attribute slopes $\gamma^{\alpha}_{j,\ell_j}, \gamma^{\beta}_{j,\ell_j}$ on $[x_{j,\ell_j-1}, x_{j,\ell_j}]$;

Fully Elicit Criterion i, j:
- **for** $\ell'_j \in [\![\ell_j, L_j]\!]$
  - Play Pattern 4 on $[x_{j,\ell'_j}, x_{j,\ell'_j+1}] \times [x_{i,\ell_i-1}, x_{i,\ell_i}]$;
  - Solve Eq. 5 and deduct slopes $\gamma^{\alpha}_{j,\ell'_j}, \gamma^{\beta}_{j,\ell'_j}$;
- **for** $\ell'_i \in [\![\ell_i, L_i]\!]$
  - Play Pattern 4 on $[x_{i,\ell'_i}, x_{i,\ell'_i+1}] \times [x_{j,\ell_j-1}, x_{j,\ell_j}]$;
  - Solve Eq. 5 and deduct slopes $\gamma^{\alpha}_{i,\ell'_i}, \gamma^{\beta}_{i,\ell'_i}$;

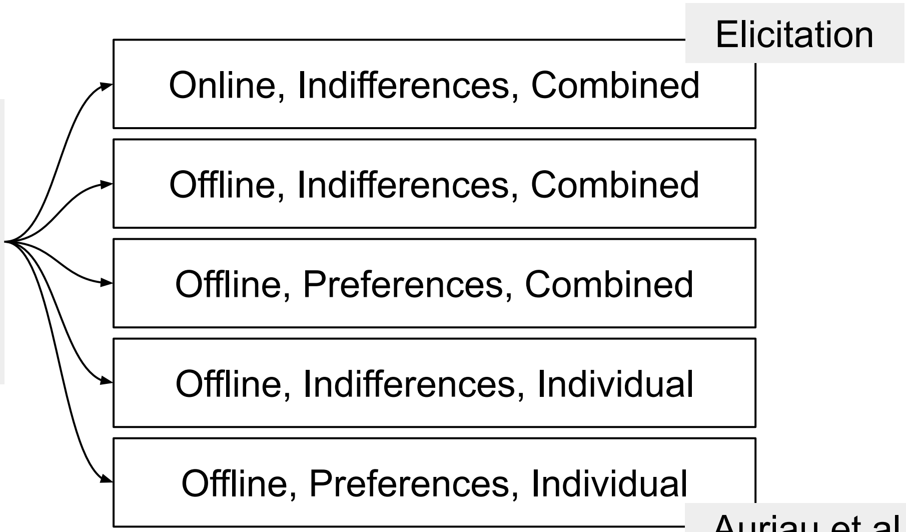
for *criterion* $j' \in \mathcal{N}, j \neq i, j$
- **Initialize criterion j':**
  - Play Pattern 3 on $[x_{j',0}, x_{j',1}] \times [x_{i,\ell_i}, x_{i,\ell_i+1}]$;
  - Solve Eq. 4 and deduct slopes $\gamma^{\alpha}_{j',1}, \gamma^{\beta}_{j',1}$;
- **Fully Elicit Criterion j':**
  - **for** $\ell_{j'} \in [\![1, L_{j'}]\!]$
    - Play Pattern 4 on $[x_{j',\ell_{j'}}, x_{j',\ell'_j+1}] \times [x_{i,\ell_i}, x_{i,\ell_i+1}]$;
    - Solve Eq. 5 and deduct slopes $\gamma^{\alpha}_{j',\ell'_j}, \gamma^{\beta}_{j',\ell'_j}$;



ARTEFACT

From Conjoint Measurement to Preference Learning: Measuring the gap with real-world use cases

> **Offline** vs online learning
>
> **Preference** vs indifference statements
>
> **Individual** vs combined answers



Auriau et al.
ADT 2024

Synthetic Experiments to measure how such different scenarios lead to different performances of a learning procedure

ARTEFACT

Compatible Space given a set of observations

Let Ω, be the space of UTA models, and $\mathcal{I} = \{(x, y) \in \mathcal{X}, x \sim y\}$ a set of observed indifferences

We define the **compatible UTA space** $\Omega_{\mathcal{I}} \subset \Omega$ as the ensemble of UTA models that respect the indifferences, ie

$$u \in \Omega_{\mathcal{I}} \iff \begin{cases} u \in \Omega \\ \forall (x, y) \in \mathcal{I}, u(x) = u(y) \end{cases}$$

Similarly we define the **2-UTA compatible space**:

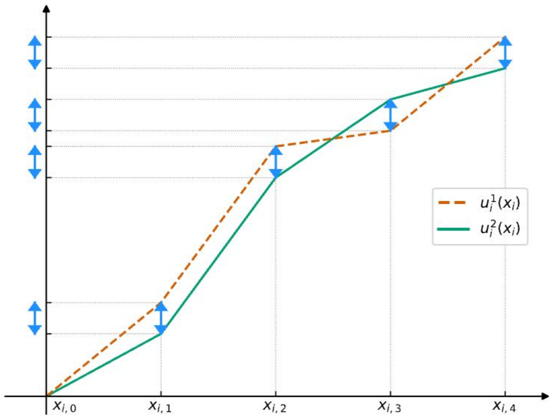
$$(u, v) \in \Omega^2_{\mathcal{P}} \iff \begin{cases} u \in \Omega, v \in \Omega \\ \forall (x, y) \in \mathcal{P}, u(x) > u(y) \text{ or } v(x) > v(y) \end{cases}$$

> **Objective:** Measuring the diamater of such compatible space for our different scenarios

Measuring the diameter of a compatible space

Distance between two UTA models, u^1 and u^2:

$$\delta(u^1, u^2) = \frac{1}{2L} \sum_{i=1}^{n} \sum_{\ell=1}^{L} |u_i^1(x_i^\ell) - u_i^2(x_i^\ell)|$$



Diameter of a compatible space:

For a subset **ω** of the space of UTA models, we define its diameter as:

$$\forall \omega \subset \Omega, D(\omega) = \max_{(u,v)\in\Omega} \delta(u, v)$$

ARTEFACT

Case with 2-UTA

Diameter of a 2-UTA space:

$$D(\Omega^2) = \frac{1}{2} \cdot \max_{(u^1,u^2),(v^1,v^2)} \min \left[\delta(u^1,v^1) + \delta(u^2,v^2), \delta(u^1,v^2) + \delta(u^2,v^1)\right]$$

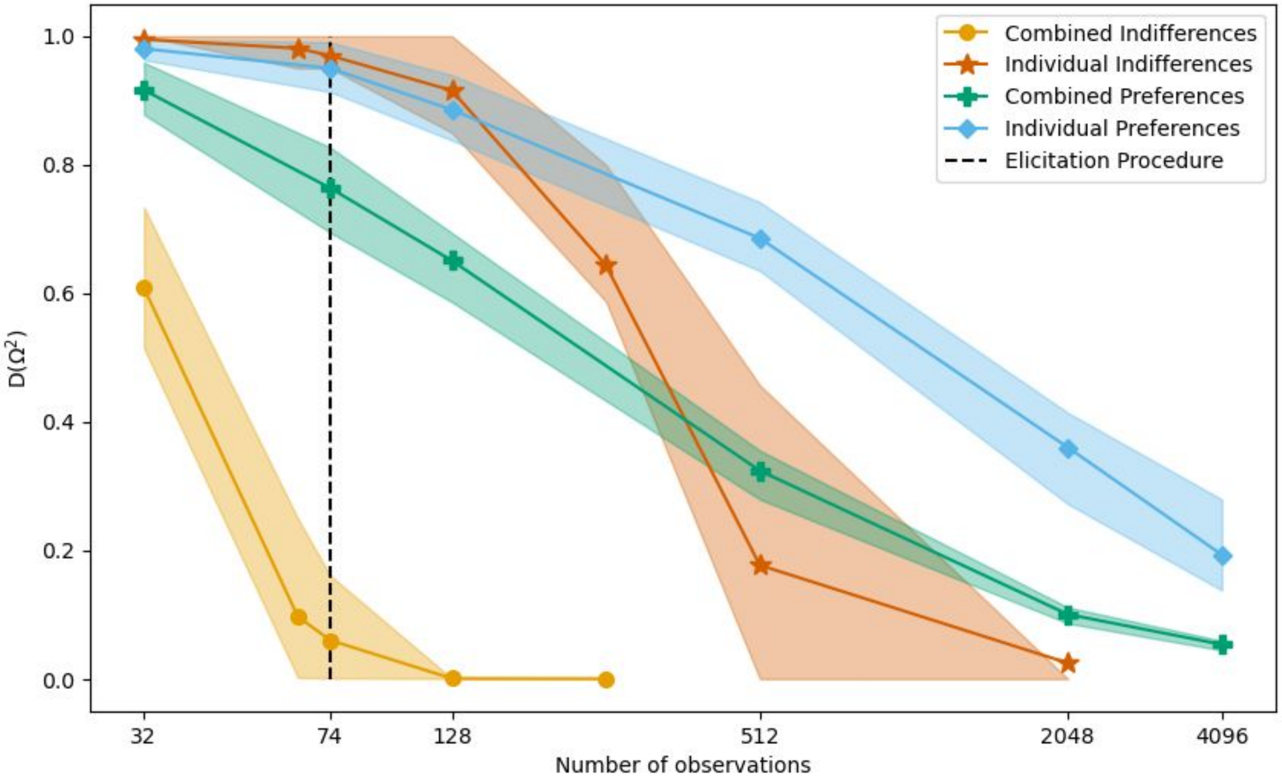
Experiment

- Draw a 2-UTA ground truth
- Draw N observations from the 2-UTA according to a scenario
- Compute the diameter of the 2-UTA space compatible with the observations

- 4 criteria & 5 linear pieces
- 20 repetitions
- Absence of noise in the data

> Computation of the Diameter formulated as MILO
Implemented in Python with Gurobi

The results !



Combined Indifferences
Individual Indifferences
Combined Preferences
Individual Preferences
Elicitation Procedure
1.0
0.8
0.6
0.4
0.2
0.0
D(Ω²)
32
74
128
512
2048
4096
Number of observations

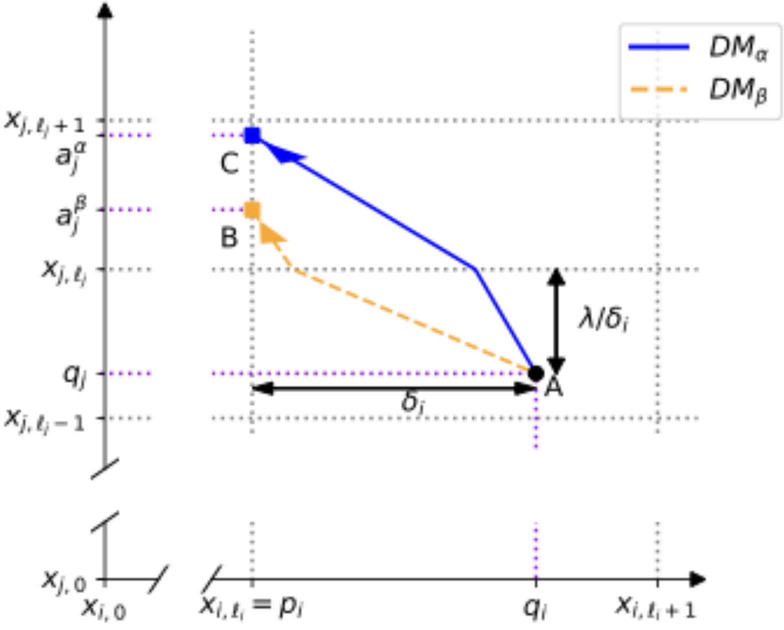
Next Steps & Future Work

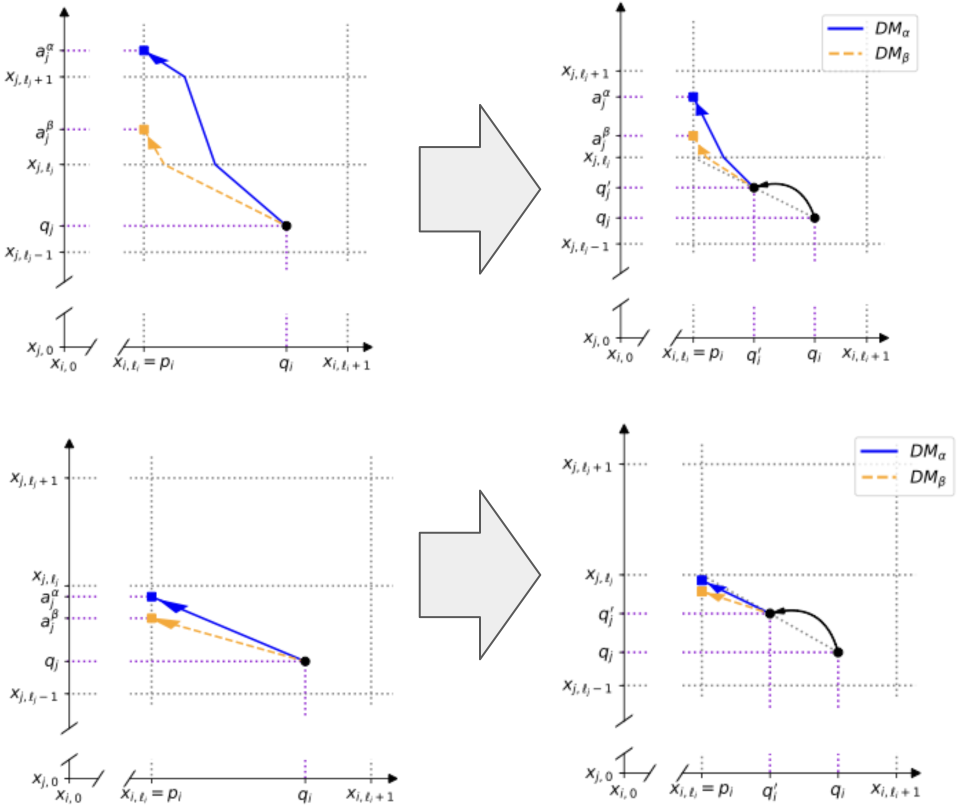
> Handle the case of generic additive value functions

> Handle the case with N>2 DMs

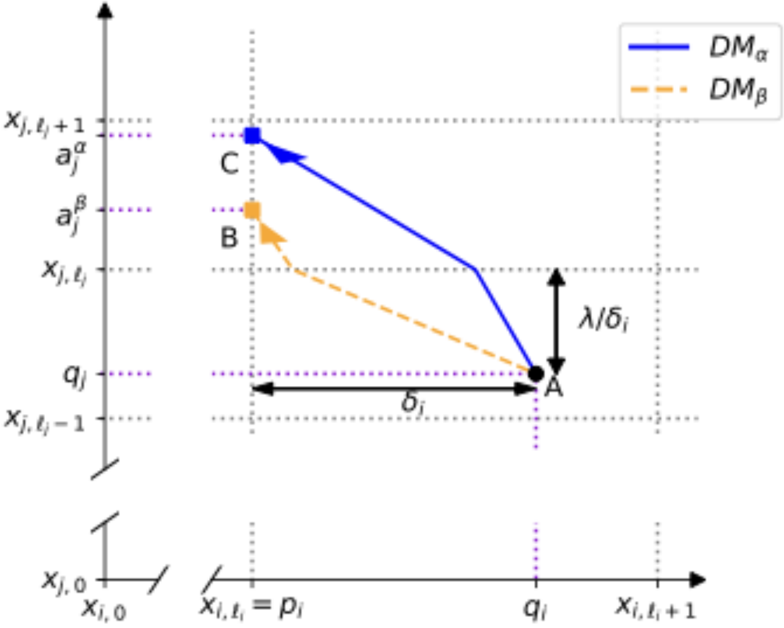
Thank you for your listening !

Neighboring Rectangles Preference Information





Exploiting Neighboring Rectangles Preference Information



$$\begin{cases} \gamma^{\kappa}_{j,\ell_j+1} = \gamma^{\kappa}_{i,\ell_i} \cdot \Theta_\kappa + \gamma^{\kappa}_{k,\ell_j} \cdot \Phi_\kappa \\ \gamma^{\kappa'}_{j,\ell_j+1} = \gamma^{\kappa'}_{i,\ell_i} \cdot \Theta'_\kappa + \gamma^{\kappa'}_{j,\ell_j} \cdot \Phi_{\kappa'} \end{cases}$$

$$\text{with: } \begin{cases} \Theta_\kappa = \dfrac{A_i - B_i}{B_j - x_{j,\ell_j}};\ \Theta_{\kappa'} = \dfrac{A_i - C_i}{C_j - x_{j,\ell_j}} \\ \Phi_\kappa = \dfrac{A_j - x_{j,\ell_j}}{B_j - x_{j,\ell_j}};\ \Phi_{\kappa'} = \dfrac{A_j - x_{j,\ell_j}}{C_j - x_{j,\ell_j}} \end{cases}$$