Inferring parsimonious preference models in robust ordinal regression

Salvatore Greco

University of Catania, Italy

Vincent Mousseau

Ecole Centrale Paris, France

Roman Słowiński

Poznań University of Technology, Poland

26th European Conference on Operational Research, Rome, 1-4 July, 2013, Preference Learning Stream

Why parsimonious preference models?

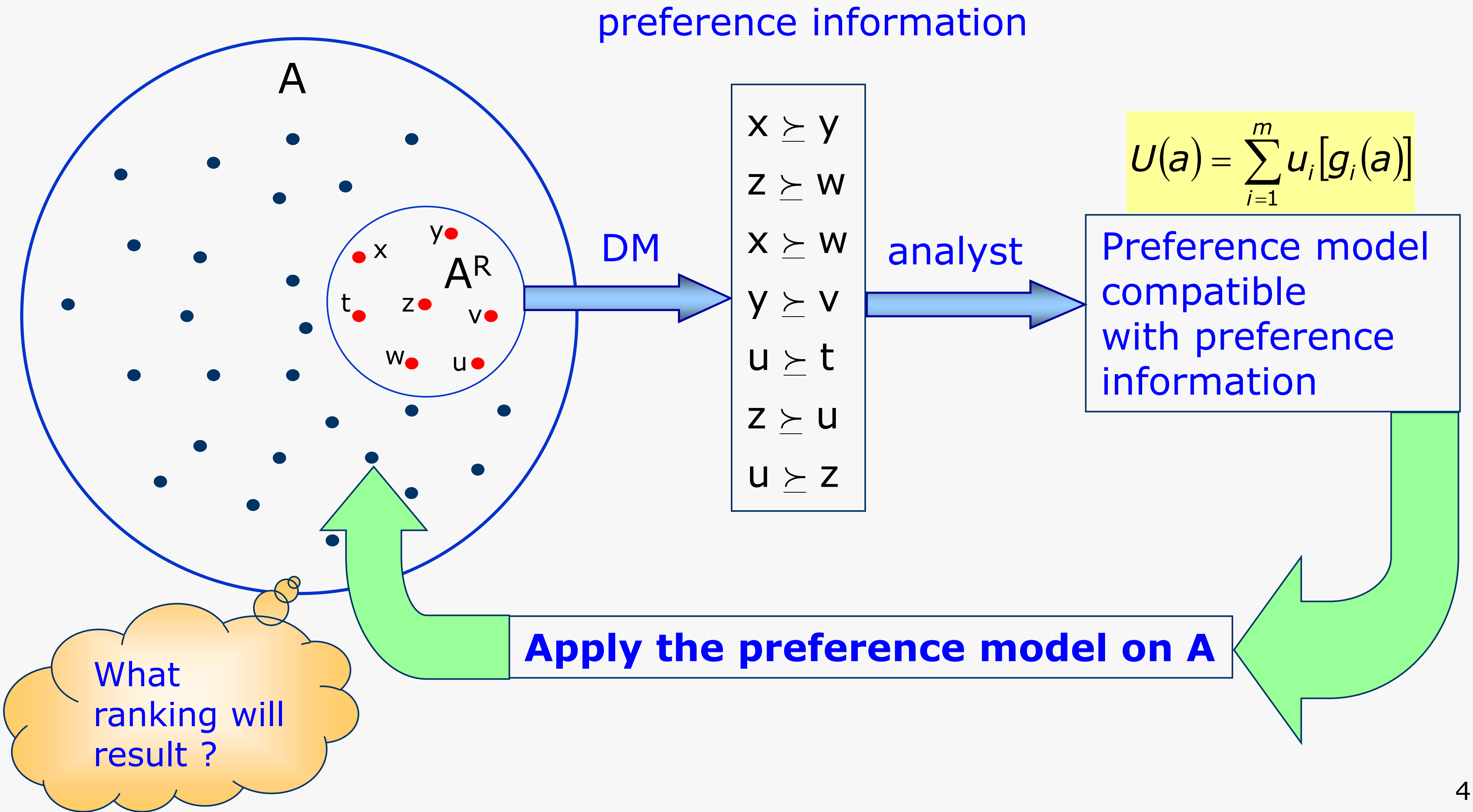
- Ockham’s razor or *lex parsimoniae* – William of Ockham (c. 1287–1347)
- Intuitively: the simplest explanation is most likely the correct one
- Words attributed to Ockham:
  - *entities must not be multiplied beyond necessity*
  - *plurality must never be posited without necessity*
- Various interpretations (e.g. Newton, Kant, Lavoisier, Einstein).
- Another relevant interpretation for our work:
  - *The procedure of induction consists in accepting as true the simplest law that can be reconciled with our experiences* (Ludwig Wittgenstein, Tractatus Logico-Philosophicus, 6.363)

Plan

- Basic idea
- Technical discussion with respect to the additive value function
- Parsimonious decision models and Robust Ordinal Regression
- Didactic example
- Parsimonious non-additive value function models
- Conclusions and further developments

Ordinal regression paradigm (UTA method)

- Ordinal regression paradigm emphasizes the discovery of intentions as an interpretation of actions rather than as *a priori* position



preference information

$x \succsim y$
$z \succsim w$
$x \succsim w$
$y \succsim v$
$u \succsim t$
$z \succsim u$
$u \succsim z$

$$U(a)=\sum_{i=1}^{m}u_i[g_i(a)]$$

Preference model compatible with preference information

Apply the preference model on A

What ranking will result ?

Principle of the ordinal regression – the UTA method

(Jacquet-Lagreze & Siskos 1982)

- Marginal value $u_i(a)$ of action $a \in A$ is approximated by linear interpolation
- The scale of u_i is a conjoint interval scale whatever the scale of g_i

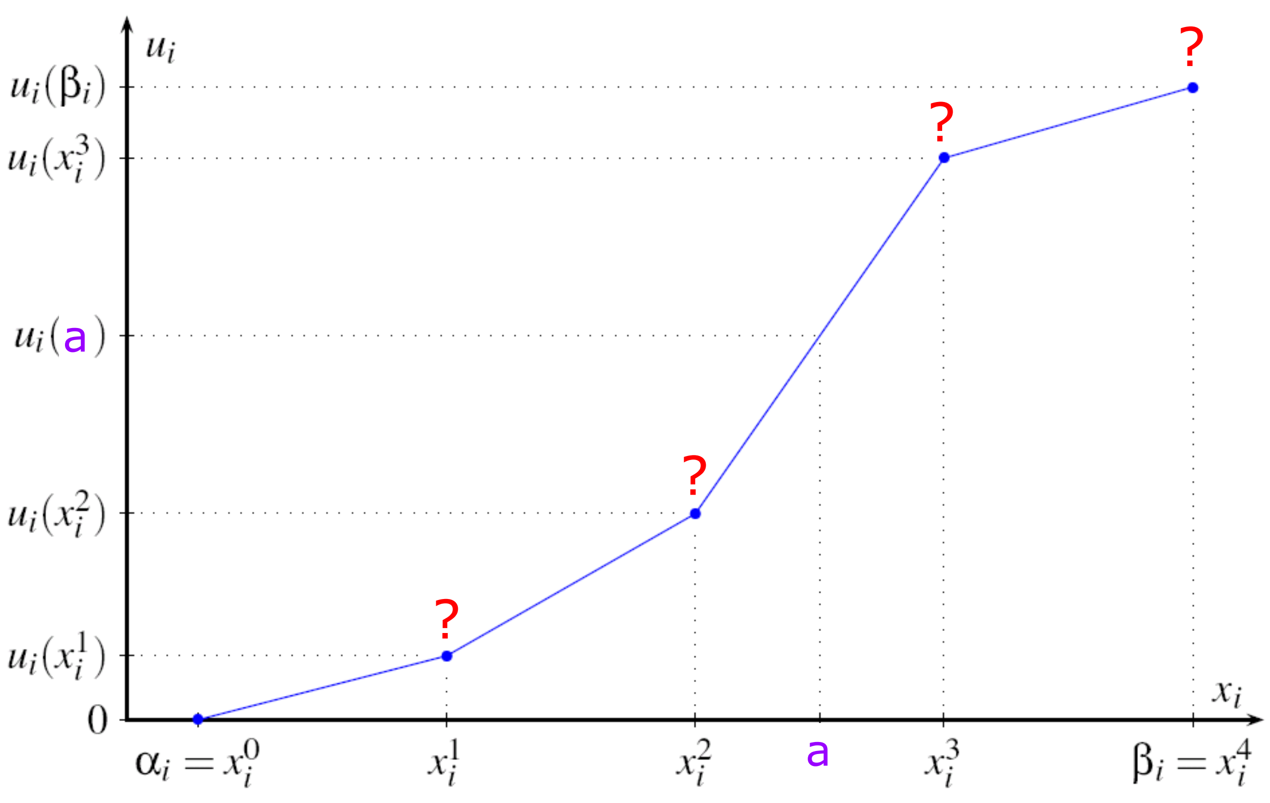
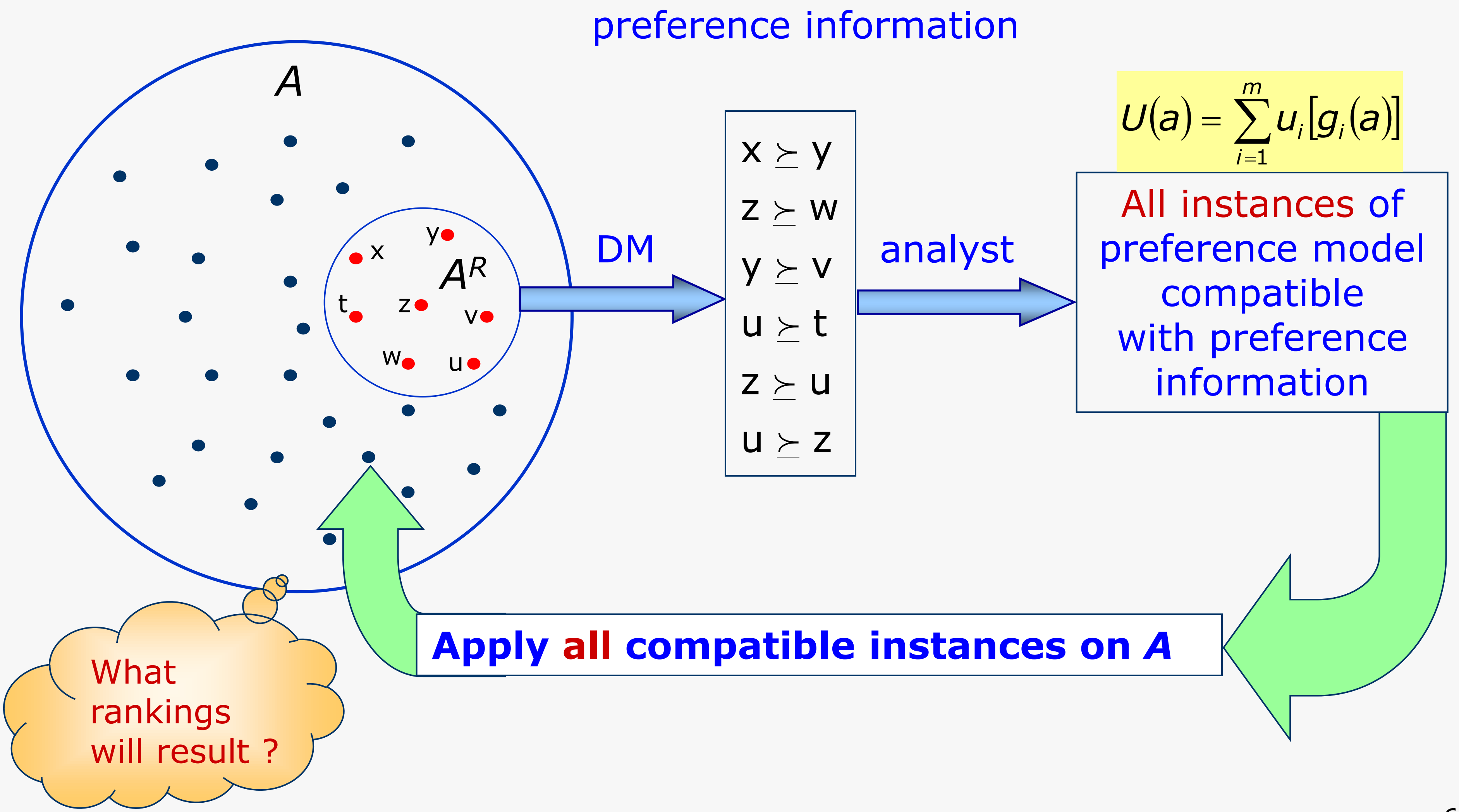


Figure 1: Piecewise linear marginal utility function

Robust Ordinal Regression (ROR) - UTA^{GMS} and GRIP methods



preference information

DM

$x \succeq y$
$z \succeq w$
$y \succeq v$
$u \succeq t$
$z \succeq u$
$u \succeq z$

analyst

$$U(a) = \sum_{i=1}^{m} u_i[g_i(a)]$$

All instances of preference model compatible with preference information

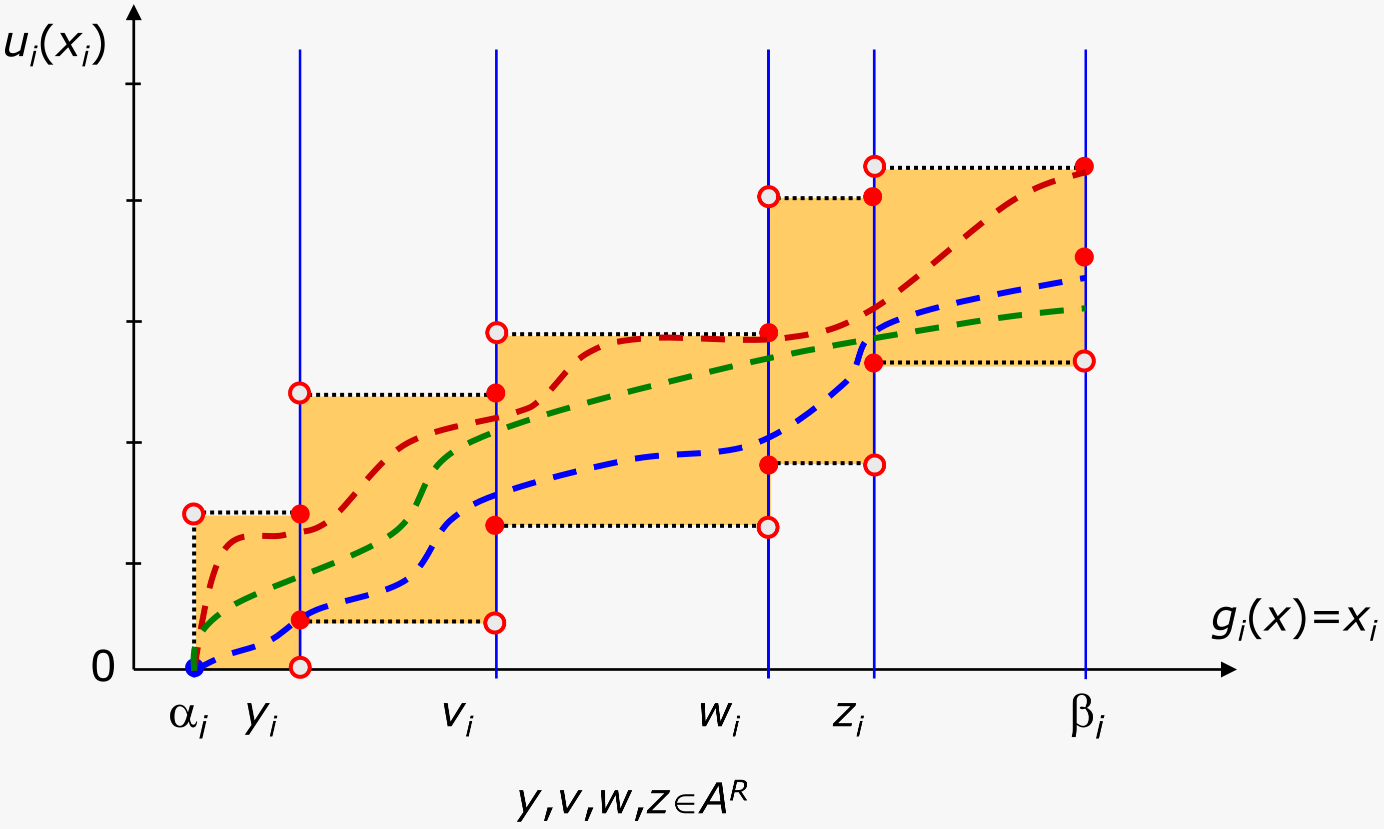
Apply all compatible instances on *A*

What rankings will result ?

Robust ordinal regression - the UTA^{GMS} and GRIP methods

(Greco, Mousseau & Słowiński 2008; Figueira, Greco & Słowiński 2009)

- The marginal value function $u_i(x_i)$

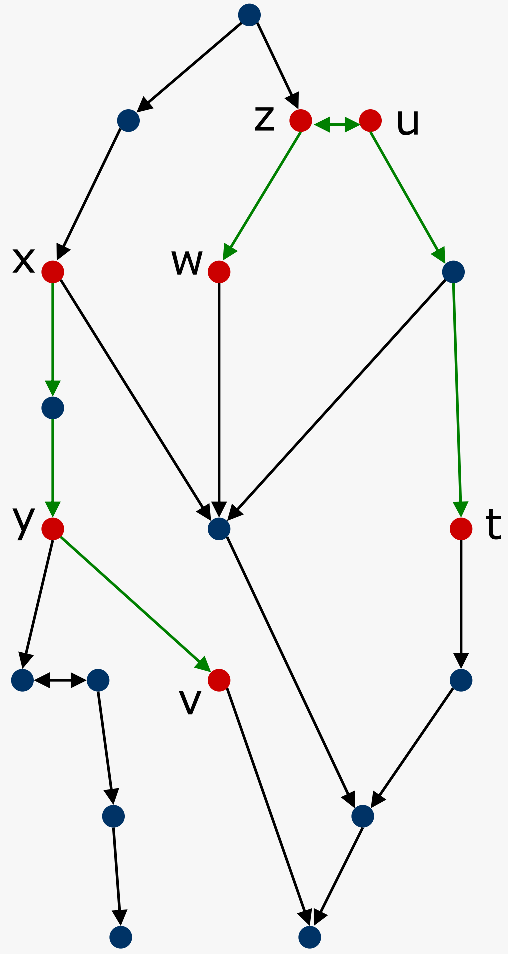


In the 🟠 area, the marginal compatible value functions must be monotone

Two rankings result: necessary and possible

preference information

$x \succsim y$
$z \succsim w$
$y \succsim v$
$u \succsim t$
$z \succsim u$
$u \succsim z$



necessary ranking

Strongly complete,

includes

necessary ranking

and

does not include

the complement of

necessary ranking

possible ranking

Effect of additional preference information

additional preference information

$x \succeq y$

$z \succeq w$

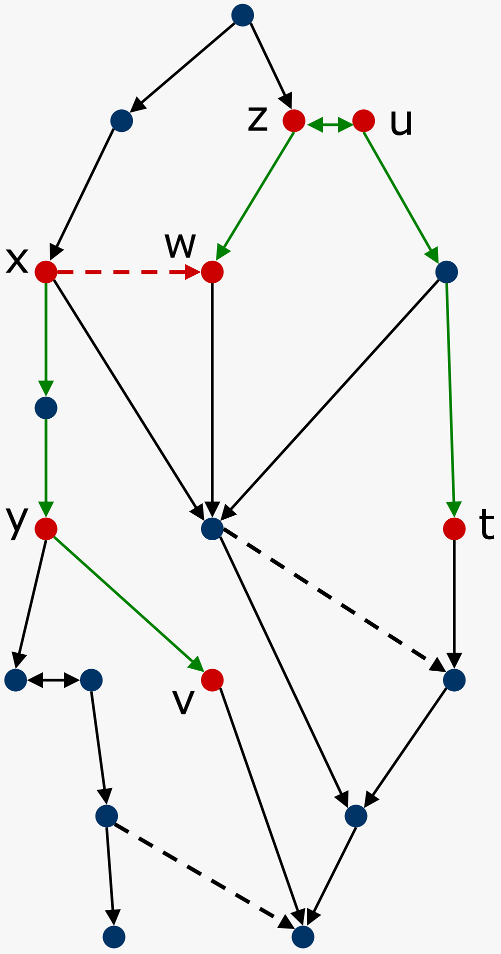
$y \succeq v$

$u \succeq t$

$z \succeq u$

$u \succeq z$

$x \succeq w$



Strongly complete,

includes

necessary ranking

and

does not include

the complement of

necessary ranking

necessary ranking
enriched

possible ranking
impoverished

Properties of necessary $\succeq^{\mathbf{N}}$ & possible $\succeq^{\mathbf{P}}$ preference relations

- $\succeq^{\mathbf{N}}$ is a partial preorder on A (i.e., $\succeq^{\mathbf{N}}$ is reflexive and transitive)
- $x \succeq^{\mathbf{N}} y \;\Rightarrow\; x \succeq^{\mathbf{P}} y$, i.e., $\succeq^{\mathbf{N}} \subseteq \succeq^{\mathbf{P}}$
- $x \succeq^{\mathbf{N}} y$ and $y \succeq^{\mathbf{P}} z \Rightarrow x \succeq^{\mathbf{P}} z, \;\; \forall x,y,z \in A$
- $x \succeq^{\mathbf{P}} y$ and $y \succeq^{\mathbf{N}} z \Rightarrow x \succeq^{\mathbf{P}} z, \;\; \forall x,y,z \in A$
- $x \succeq^{\mathbf{N}} y$ or $y \succeq^{\mathbf{P}} z, \;\; \forall x,y \in A$

Moreover:

- $\succeq^{\mathbf{P}}$ is strongly complete (i.e., for all $x,y \in A$, $x \succeq^{\mathbf{P}} y$ or $y \succeq^{\mathbf{P}} x$) and negatively transitive (i.e., for all $x,y,z \in A$, *not* $x \succeq^{\mathbf{P}} y$ and *not* $y \succeq^{\mathbf{P}} z \Rightarrow$ *not* $x \succeq^{\mathbf{P}} z$), (in general, $\succeq^{\mathbf{P}}$ is not transitive)
- $\succ^{\mathbf{P}}$ is complete, irreflexive and transitive

Parsimonious model in ordinal regression

- What is the simplest additive value function model representing preferences elicited by the DM in UTA, UTA^{GMS} and GRIP ?

$$U(x) = \sum_{i=1}^{m} u_i[g_i(x)]$$

- We believe that the simplest additive value function model is the „most linear”, i.e. the closest to

$$U(x) = \sum_{i=1}^{m} w_i g_i(x)$$

- We envisage two possible interpretations of „distance” from the linear model:
  - deviation from the linearity: the maximal variation in the slope of the value function: δ
  - number of variations of the slope of marginal value functions: η (number of characteristic points)

Basic example

Students	Math	Phys	Lit
s1	4	3	5
s2	5	4	2
s3	4	4	4
s4	4	5	2
s5	3	4	3
s6	3	3	5

Preference information

s3 $\succ$ s2

s3 $\succ$ s4

s5 $\succ$ s6

The principle of the ordinal regression – the UTA method

(Jacquet-Lagreze & Siskos 1982)

- The marginal value functions (values of u_i in characteristic points) are estimated by solving the LP problem

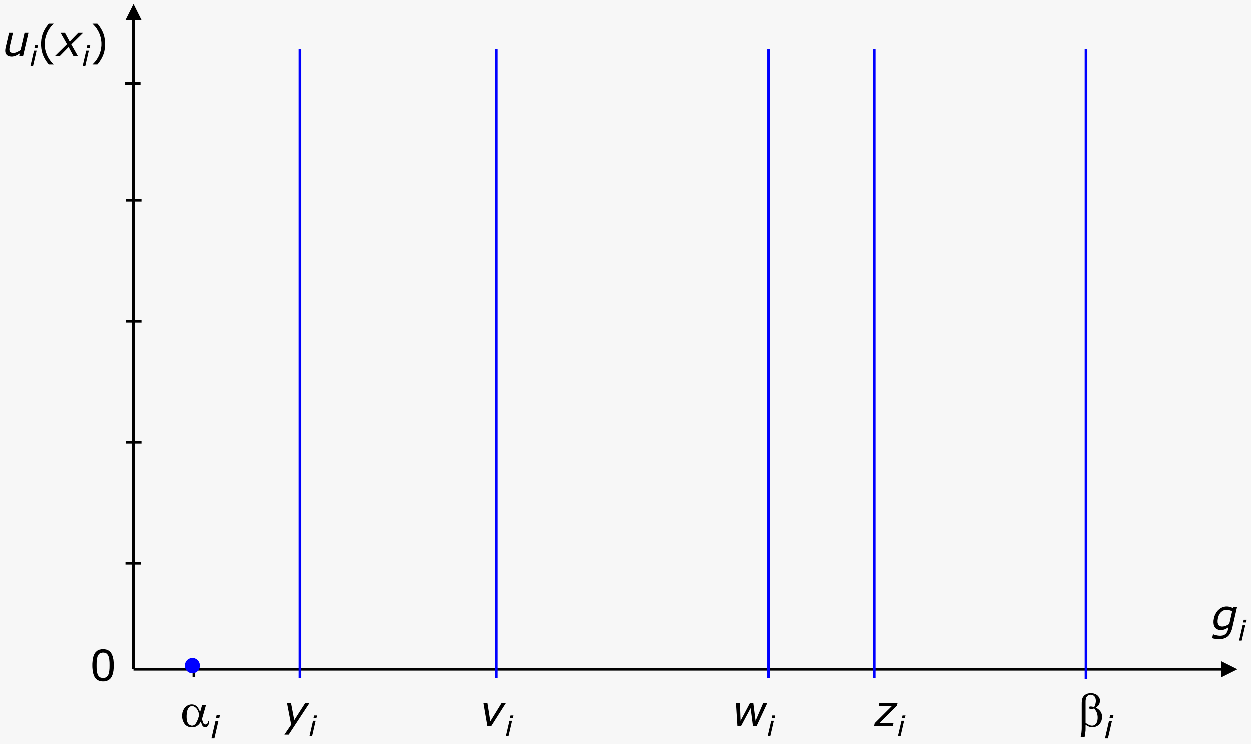
$$\text{Min} \quad \rightarrow \quad E^{UTA} = \sum_{a \in A^R} \left(\sigma^+(a) + \sigma^-(a) \right)$$

$$\left.\begin{array}{l}
\text{subject to} \\
\left.\begin{array}{l}
U(a) + \sigma^+(a) - \sigma^-(a) \geq U(b) + \sigma^+(b) - \sigma^-(b) + \varepsilon \Leftrightarrow \ a \succ b \\
U(a) + \sigma^+(a) - \sigma^-(a) = U(b) + \sigma^+(b) - \sigma^-(b) \Leftrightarrow \ a \sim b
\end{array}\right\} \forall a,b \in A^R \\
u_i(x_i^j) - u_i(x_i^{j-1}) \geq 0 \quad j = 2,\ldots,\gamma_i \ ; \ \forall i \in I \\
\sum_{i=1}^{m} u_i(\beta_i) = 1 \\
u_i(\alpha_i) = 0 \quad \forall i \in I \\
u_i(x_i^j) \geq 0, \ \sigma^+(a) \geq 0, \ \sigma^-(a) \geq 0, \ \forall a \in A^R, \ \forall i \ \text{ and } \ j
\end{array}\right\} (C)$$

where ε is a small positive constant, σ^+ and σ^- are errors of approximation, and γ_i are numbers of characteristic points of u_i

The UTA^{GMS} method (Greco, Mousseau & Słowiński 2004, 2008)

- The marginal value function $u_i(x_i)$



$y,v,w,z \in A^R$

Characteristic points of marginal value functions are fixed on actual evaluations of actions from set A^R

The most discriminant additive value function

- The marginal value functions are estimated by solving the LP problem

$$\text{Max}\left(\underset{a,b\in A^R:\,a\succ b}{\text{Min}}\left(U(a)-U(b)\right)\right)$$

subject to

$$\left.\begin{array}{ll} U(a)>U(b) & \text{if} \quad a\succ b \\ U(a)=U(b) & \text{if} \quad a\sim b \end{array}\right\}\forall a,b\in A^R$$

$$u_i\left(x_i^j\right)-u_i\left(x_i^{j-1}\right)\geq 0 \quad j=2,\ldots,\gamma_i;\ \forall i\in I$$

$$\sum_{i=1}^{m}u_i(\beta_i)=1$$

$$u_i(\alpha_i)=0 \quad \forall i\in I$$

$$u_i\left(x_i^j\right)\geq 0 \quad \forall i \ \text{ and } \ j$$

(C)

The most discriminant additive value function

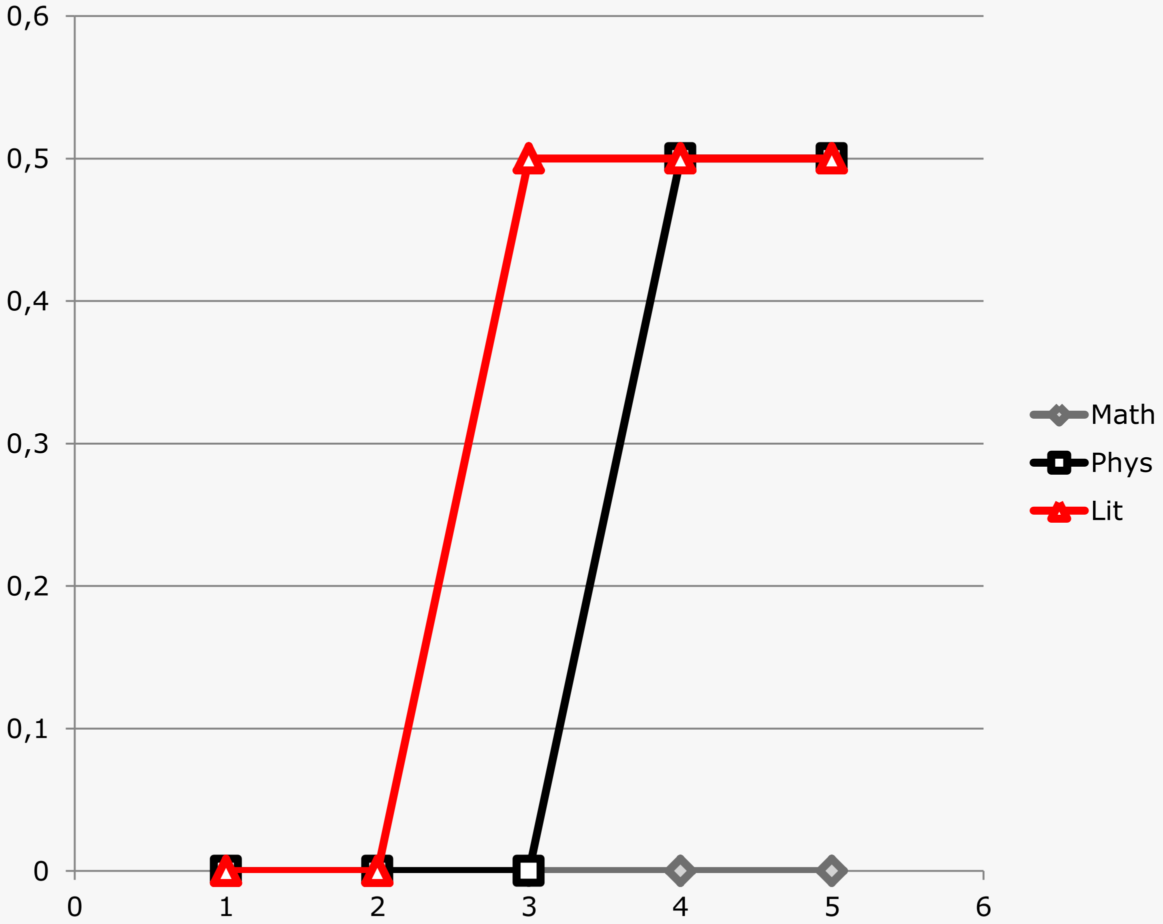
- The marginal value functions are estimated by solving the LP problem

$$\text{Max} \quad \rightarrow \quad \varepsilon$$

subject to

$$\left.\begin{array}{l} \left.\begin{array}{l} U(a) \geq U(b) + \varepsilon \quad \text{if} \quad a \succ b \\ U(a) = U(b) \quad \text{if} \quad a \sim b \end{array}\right\} \forall a,b \in A^R \\ u_i\left(x_i^j\right) - u_i\left(x_i^{j-1}\right) \geq 0 \quad j = 2,\ldots,\gamma_i; \ \forall i \in I \\ \sum_{i=1}^{m} u_i(\beta_i) = 1 \\ u_i(\alpha_i) = 0 \quad \forall i \in I \\ u_i\left(x_i^j\right) \geq 0 \ \forall i \ \text{and} \ j \end{array}\right\} (C)$$

The maximally discriminant value function



The maximally discriminant value function

Students	Math	Phys	Lit	Overall
s1	0 (4)	0 (3)	0.5 (5)	0.5
s2	0 (5)	0.5 (4)	0 (2)	0.5
s3	0 (4)	0.5 (4)	0.5 (4)	1
s4	0 (4)	0.5 (5)	0 (2)	0.5
s5	0 (3)	0.5 (4)	0.5 (3)	1
s6	0 (3)	0 (3)	0.5 (5)	0.5

$\varepsilon=0.5$

Preference information

$s3 \succ s2$

$s3 \succ s4$

$s5 \succ s6$

The additive value function which minimally deviates from linearity

- The marginal value functions result from the LP problem

$$\text{Min}\left(\underset{i\in I,\, j=2,\ldots,\gamma_i}{\text{Max}}\left|\frac{u_i\left(x_i^j\right)-u_i\left(x_i^{j-1}\right)}{x_i^j-x_i^{j-1}}-\frac{u_i\left(x_i^{j-1}\right)-u_i\left(x_i^{j-2}\right)}{x_i^{j-1}-x_i^{j-2}}\right|\right)$$

subject to

$$\left.\begin{aligned}&U(a)\geq U(b)+\varepsilon \quad \text{if} \quad a\succ b\\&U(a)=U(b) \quad \text{if} \quad a\sim b\end{aligned}\right\}\ \forall a,b\in A^R$$

$$u_i\left(x_i^j\right)-u_i\left(x_i^{j-1}\right)\geq 0 \quad j=2,\ldots,\gamma_i\ ;\ \forall i\in I$$

$$\sum_{i=1}^{m}u_i(\beta_i)=1$$

$$u_i(\alpha_i)=0 \quad \forall i\in I$$

$$u_i\left(x_i^j\right)\geq 0 \ \ \forall i \ \ \text{and} \ \ j$$

(C)

where ε is a small positive constant

The additive value function which minimally deviates from linearity

- The marginal value functions result from the LP problem

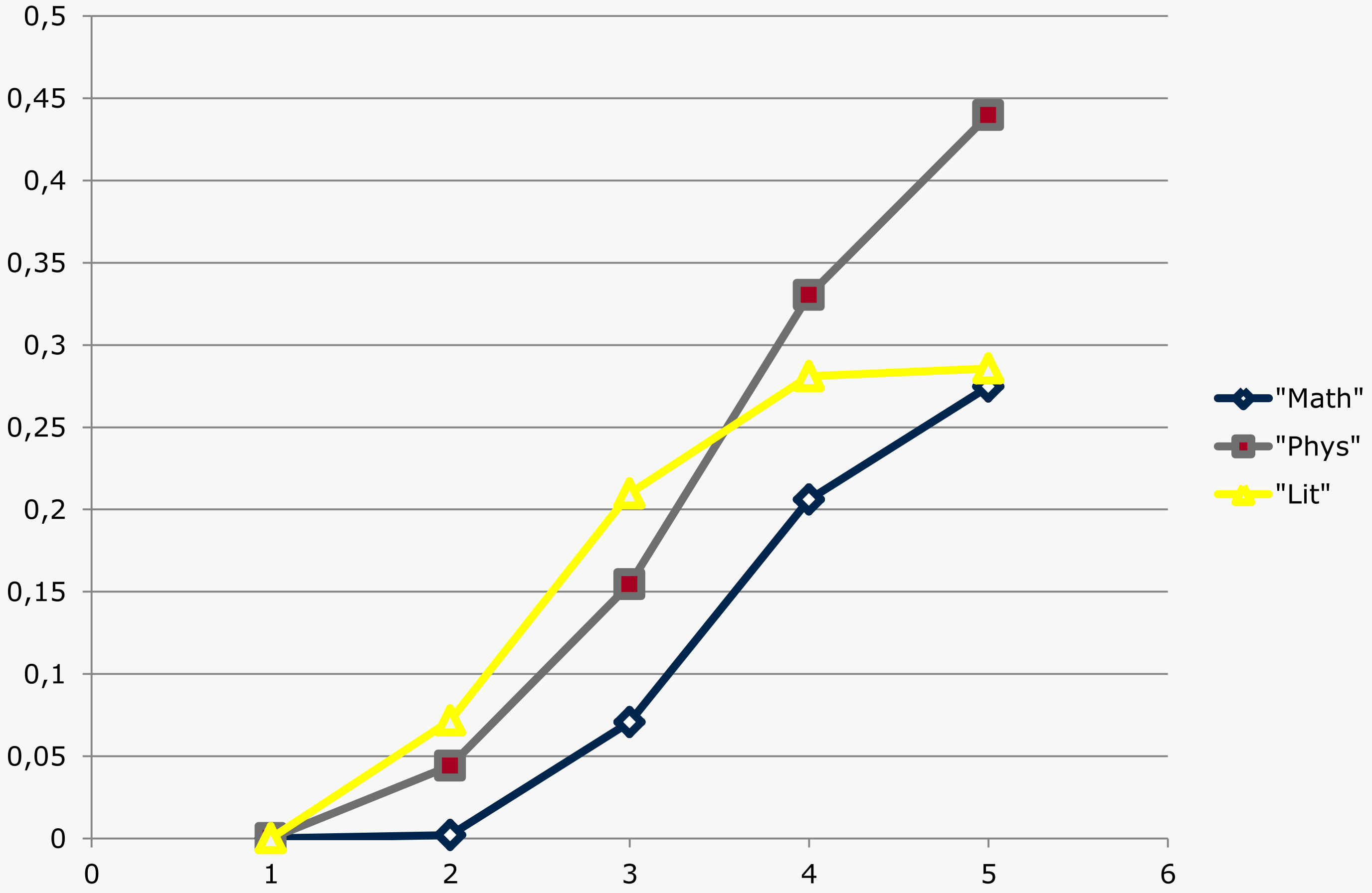
$$\text{Min} \rightarrow \delta$$

subject to

$$\left.\begin{array}{l}
\left.\begin{array}{l} U(a) \geq U(b) + \varepsilon \quad \text{if} \quad a \succ b \\ U(a) = U(b) \quad \text{if} \quad a \sim b \end{array}\right\} \forall a,b \in A^R \\
u_i\left(x_i^j\right) - u_i\left(x_i^{j-1}\right) \geq 0 \quad j = 2,\ldots,\gamma_i \;;\; \forall i \in I \\
\dfrac{u_i\left(x_i^j\right) - u_i\left(x_i^{j-1}\right)}{x_i^j - x_i^{j-1}} - \dfrac{u_i\left(x_i^{j-1}\right) - u_i\left(x_i^{j-2}\right)}{x_i^{j-1} - x_i^{j-2}} \leq \delta \quad j = 3,\ldots,\gamma_i \;;\; \forall i \in I \\
\dfrac{u_i\left(x_i^j\right) - u_i\left(x_i^{j-1}\right)}{x_i^j - x_i^{j-1}} - \dfrac{u_i\left(x_i^{j-1}\right) - u_i\left(x_i^{j-2}\right)}{x_i^{j-1} - x_i^{j-2}} \geq -\delta \quad j = 3,\ldots,\gamma_i \;;\; \forall i \in I \\
\sum_{i=1}^{m} u_i(\beta_i) = 1 \\
u_i(\alpha_i) = 0 \quad \forall i \in I \\
u_i\left(x_i^j\right) \geq 0 \quad \forall i \text{ and } j
\end{array}\right\} \text{(C)}$$

where ε is a small fixed positive constant

The additive value function which minimally deviates from linearity (ε=0.1, δ=0.067)



The additive value function which minimally deviates from linearity

Students	Math	Phys	Lit	Overall
s1	0.206 (4)	0.154 (3)	0.286 (5)	0.645
s2	0.275 (5)	0.330 (4)	0.071 (2)	0.676
s3	0.206 (4)	0.330 (4)	0.281 (4)	0.817
s4	0.206 (4)	0.440 (5)	0.071 (2)	0.717
s5	0.070 (3)	0.330 (4)	0.210 (3)	0.610
s6	0.070 (3)	0.154 (3)	0.286 (5)	0.510

$\varepsilon=0.01, \ \delta=0.067$

Preference information

$s3 \succ s2$

$s3 \succ s4$

$s5 \succ s6$

The additive value function with a minimal number of characteristic points

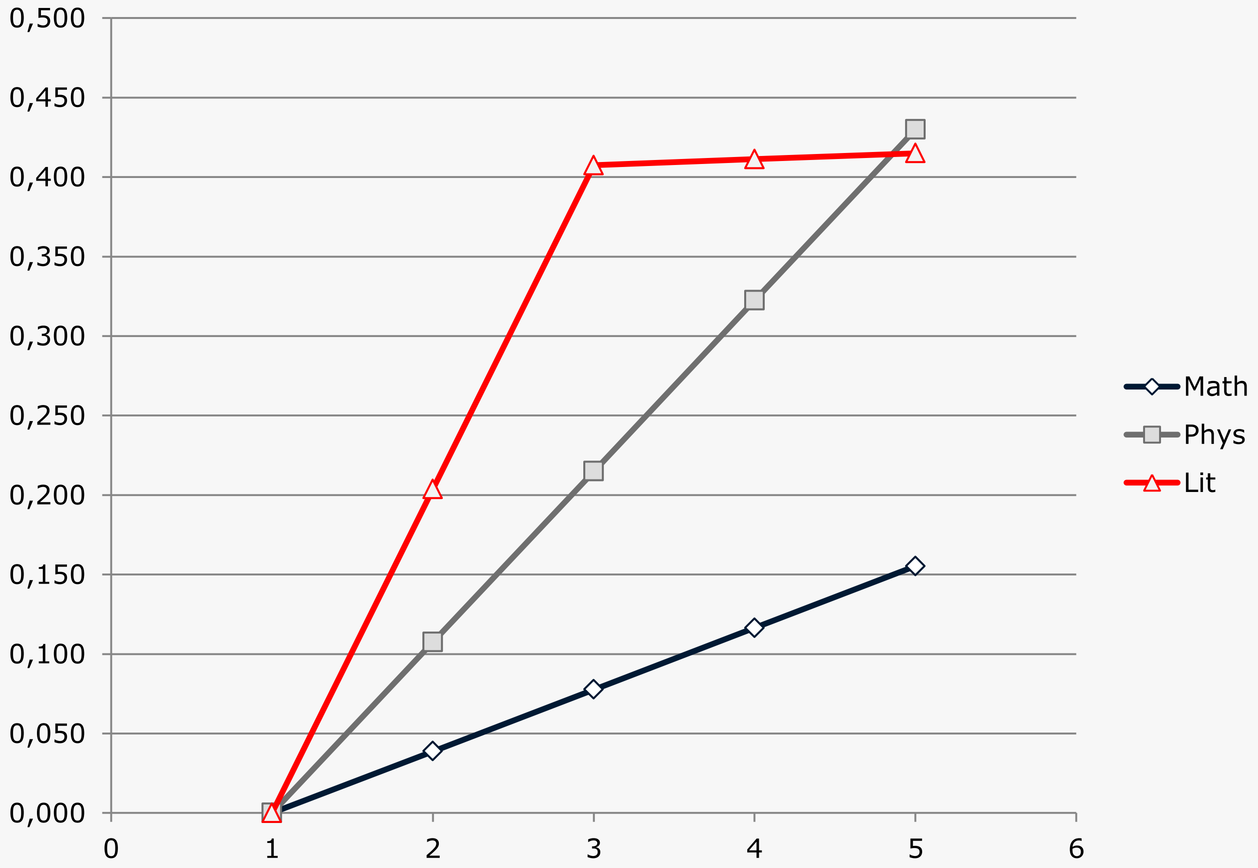
$$\text{Min} \quad \rightarrow \quad \eta = \sum_{i\in I, j=3,\ldots,\gamma_i} \boxed{\delta_{i,j}}$$

subject to

$$\left.\begin{array}{l}
\left.\begin{array}{l}
U(a) \geq U(b) + \varepsilon \quad \text{if} \quad a \succ b \\
U(a) = U(b) \quad \text{if} \quad a \sim b
\end{array}\right\} \quad \forall a,b \in A^R \\
u_i\left(x_i^j\right) - u_i\left(x_i^{j-1}\right) \geq 0 \quad j = 2,\ldots,\gamma_i\ ; \ \forall i \in I \\
\dfrac{u_i\left(x_i^j\right) - u_i\left(x_i^{j-1}\right)}{x_i^j - x_i^{j-1}} - \dfrac{u_i\left(x_i^{j-1}\right) - u_i\left(x_i^{j-2}\right)}{x_i^{j-1} - x_i^{j-2}} \leq M\boxed{\delta_{ij}} \quad j = 3,\ldots,\gamma_i\ ; \ \forall i \in I \\
\dfrac{u_i\left(x_i^j\right) - u_i\left(x_i^{j-1}\right)}{x_i^j - x_i^{j-1}} - \dfrac{u_i\left(x_i^{j-1}\right) - u_i\left(x_i^{j-2}\right)}{x_i^{j-1} - x_i^{j-2}} \geq -M\boxed{\delta_{ij}} \quad j = 3,\ldots,\gamma_i\ ; \ \forall i \in I \\
\sum_{i=1}^{m} u_i(\beta_i) = 1 \\
u_i(\alpha_i) = 0 \quad \forall i \in I \\
u_i\left(x_i^j\right) \geq 0, \quad \delta_{ij} \in \{0,1\} \quad \forall i \in I \ \text{ and } \ j
\end{array}\right\} \text{(C)}$$

where ε is a positive constant and $M > \dfrac{1}{\min\limits_{i\in I, j=2,\ldots,\gamma_i} \left(x_i^j - x_i^{j-1}\right)}$

The model with a minimal number of characteristic points



$\varepsilon=0.01,\ \eta=1$

The additive value function with a minimal number of characteristic points

Students	Math	Phys	Lit	Overall
s1	0.116 (4)	0.215 (3)	0.415 (5)	0.746
s2	0.155 (5)	0.322 (4)	0.204 (2)	0.681
s3	0.116 (4)	0.322 (4)	0.411 (4)	0.850
s4	0.116 (4)	0.430 (5)	0.204 (2)	0.750
s5	0.078 (3)	0.322 (4)	0.407 (3)	0.807
s6	0.078 (3)	0.215 (3)	0.415 (5)	0.707

$\varepsilon=0.01$, $\eta=1$

Preference information

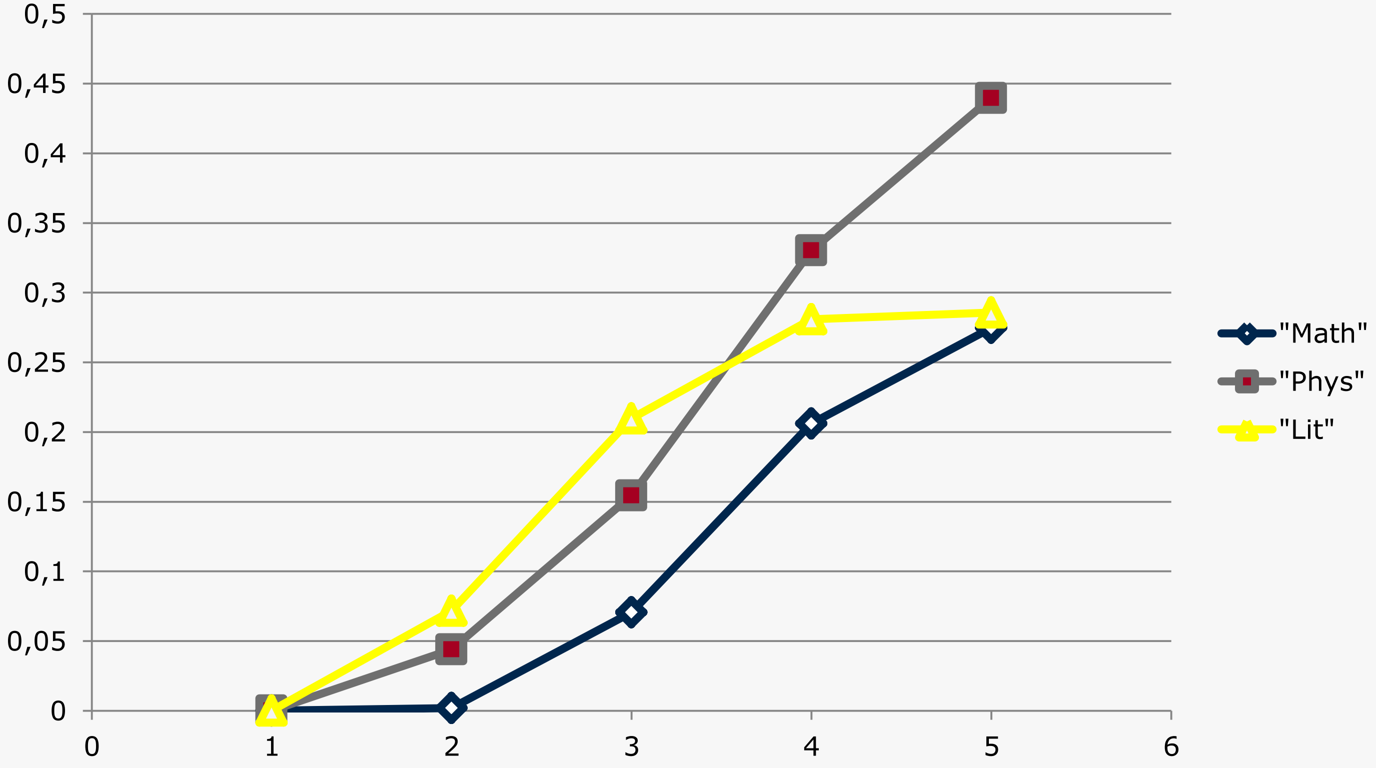
$s3 \succ s2$

$s3 \succ s4$

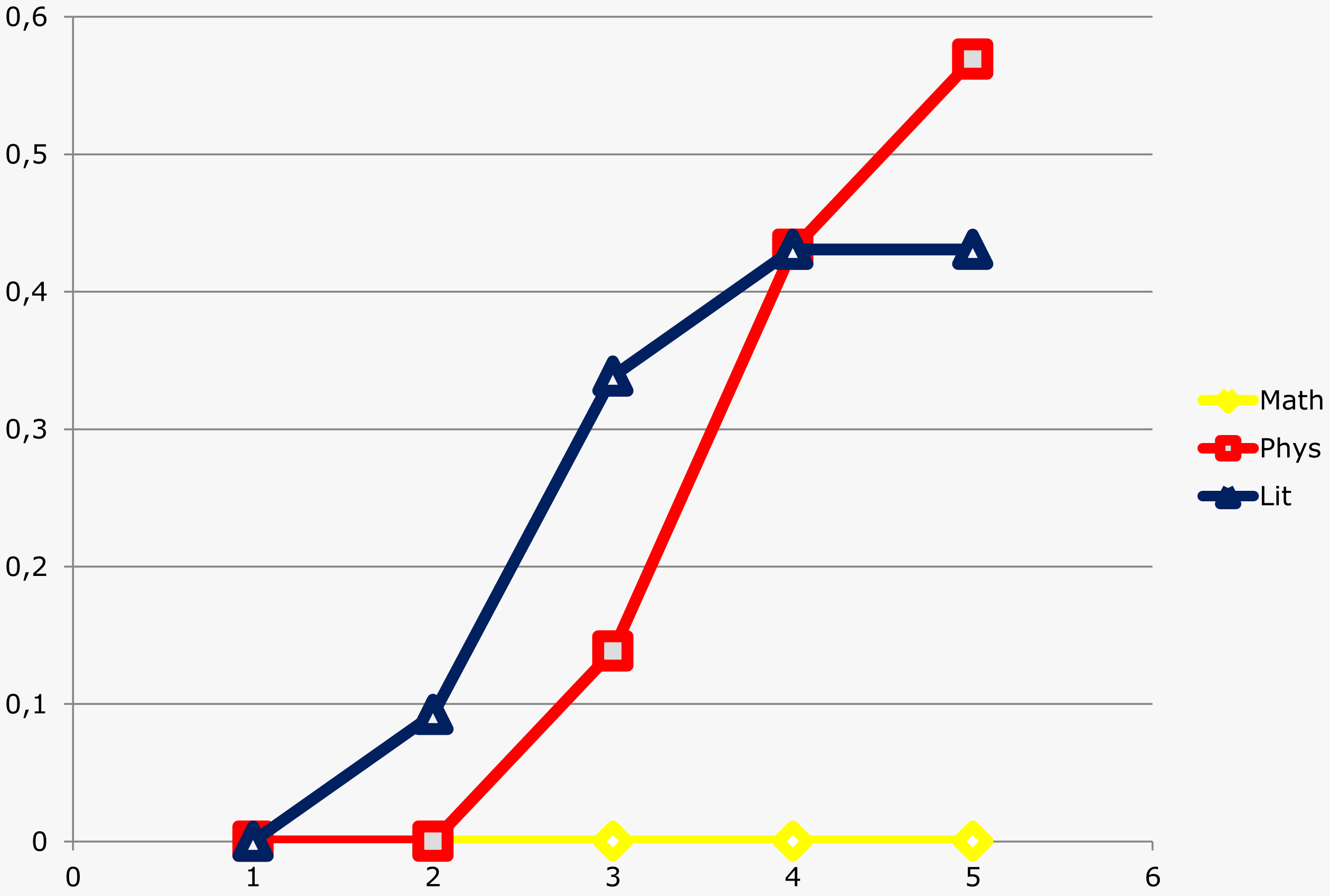
$s5 \succ s6$

What happens when
the discrimination threshold ε changes?

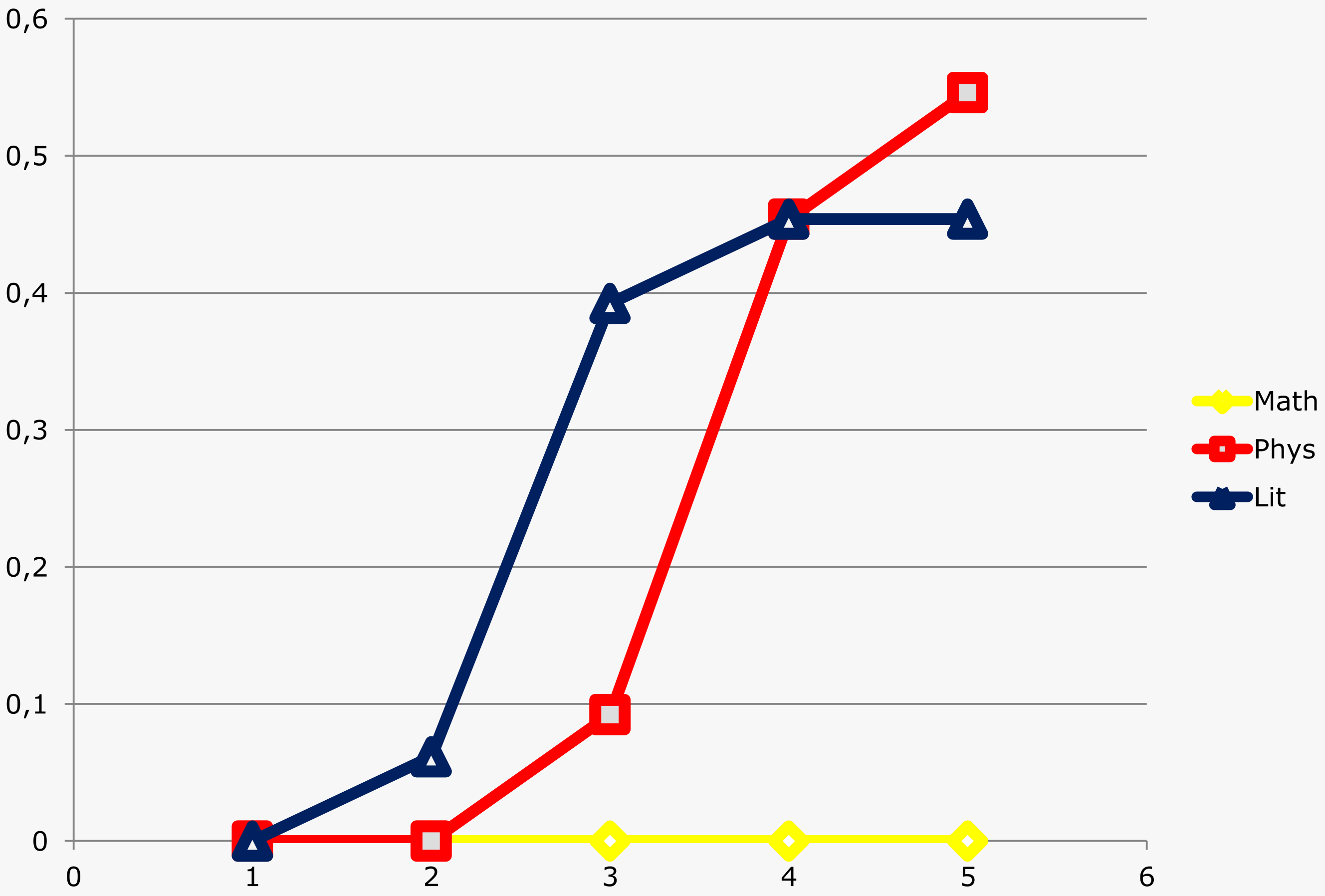
The most linear model (ε=0.1, δ=0.067)



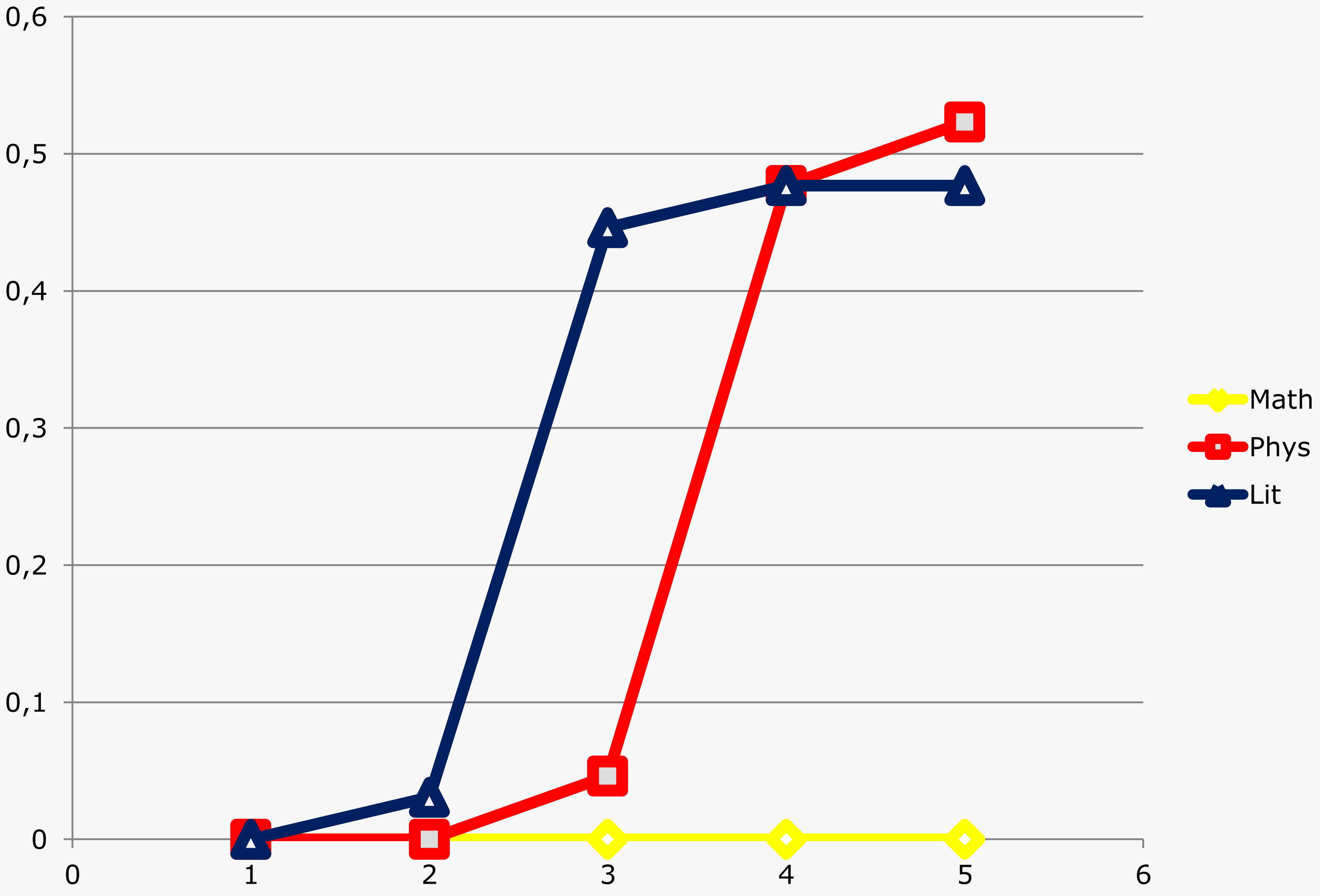
The most linear model (ε=0.2, δ=0.154)



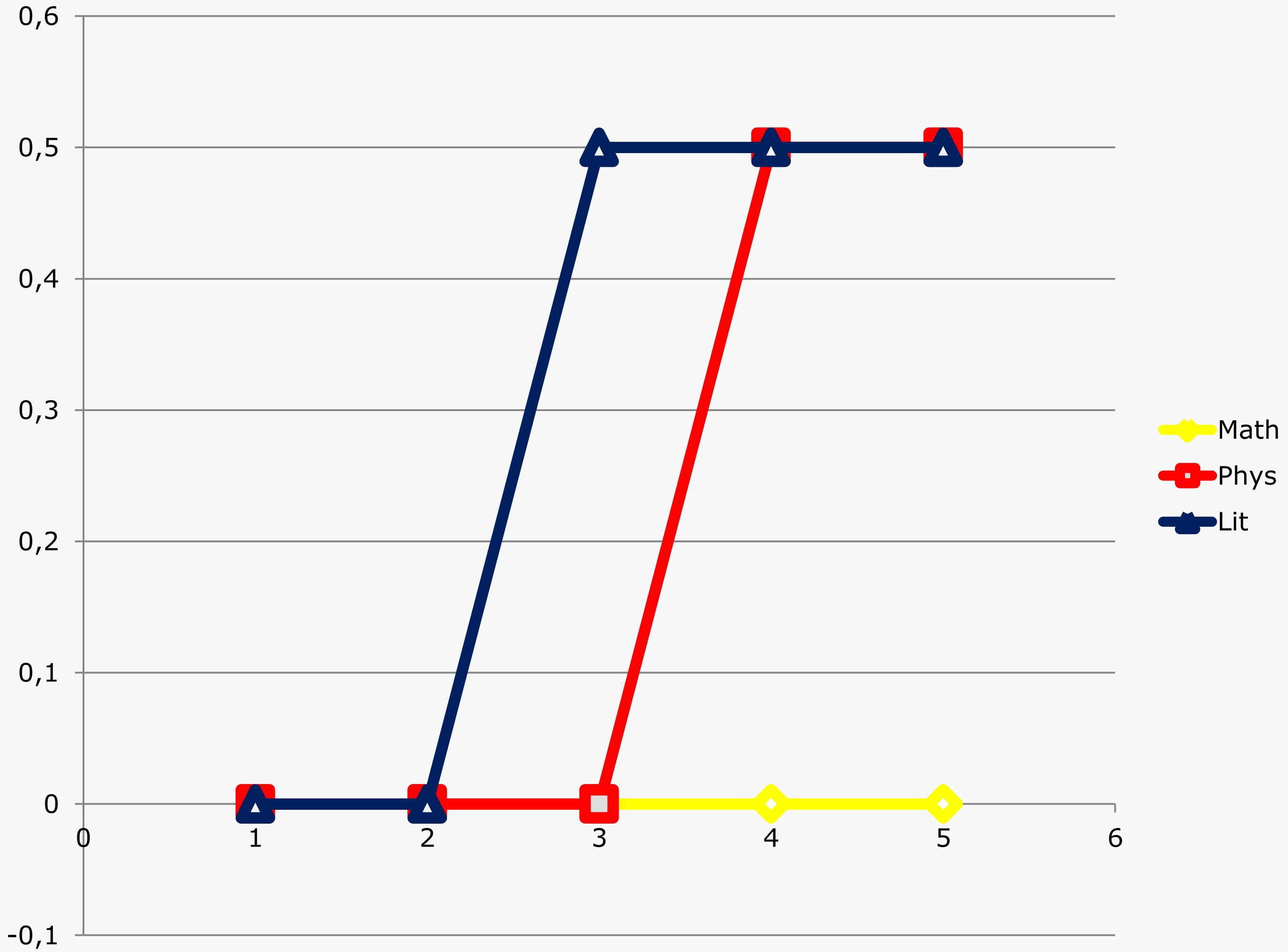
The most linear model (ε=0.3, δ=0.269)



The most linear model (ε=0.4, δ=0.385)

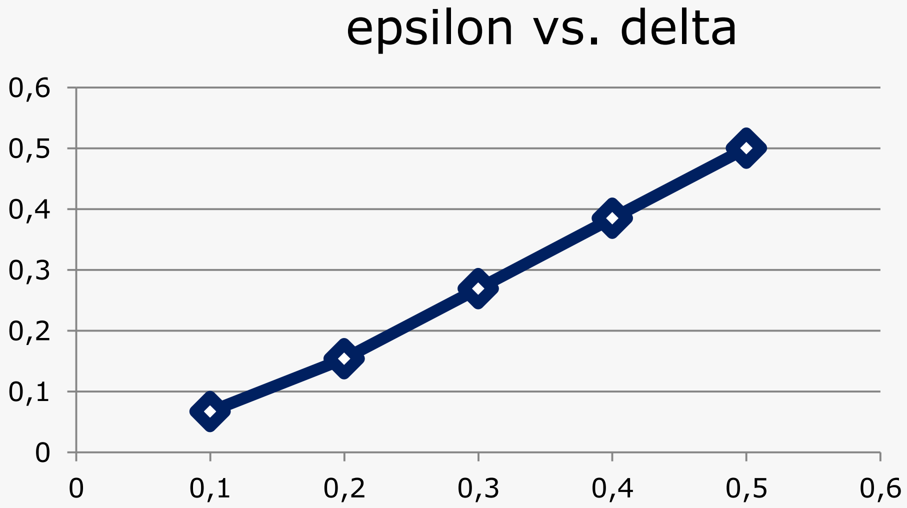


The most linear model (ε=0.5, δ=0.5)

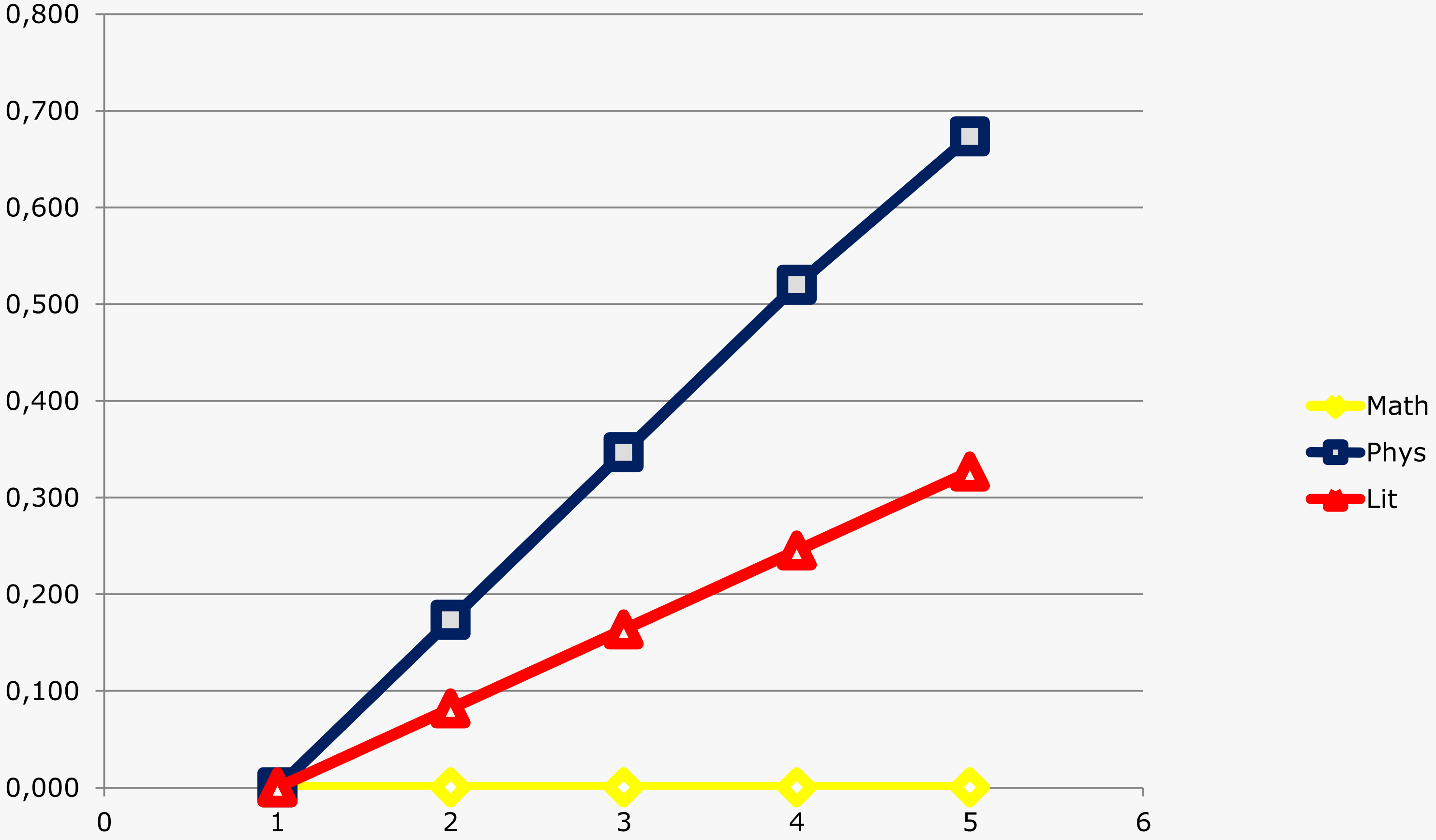


Relationship between discrimination and deviation from linearity

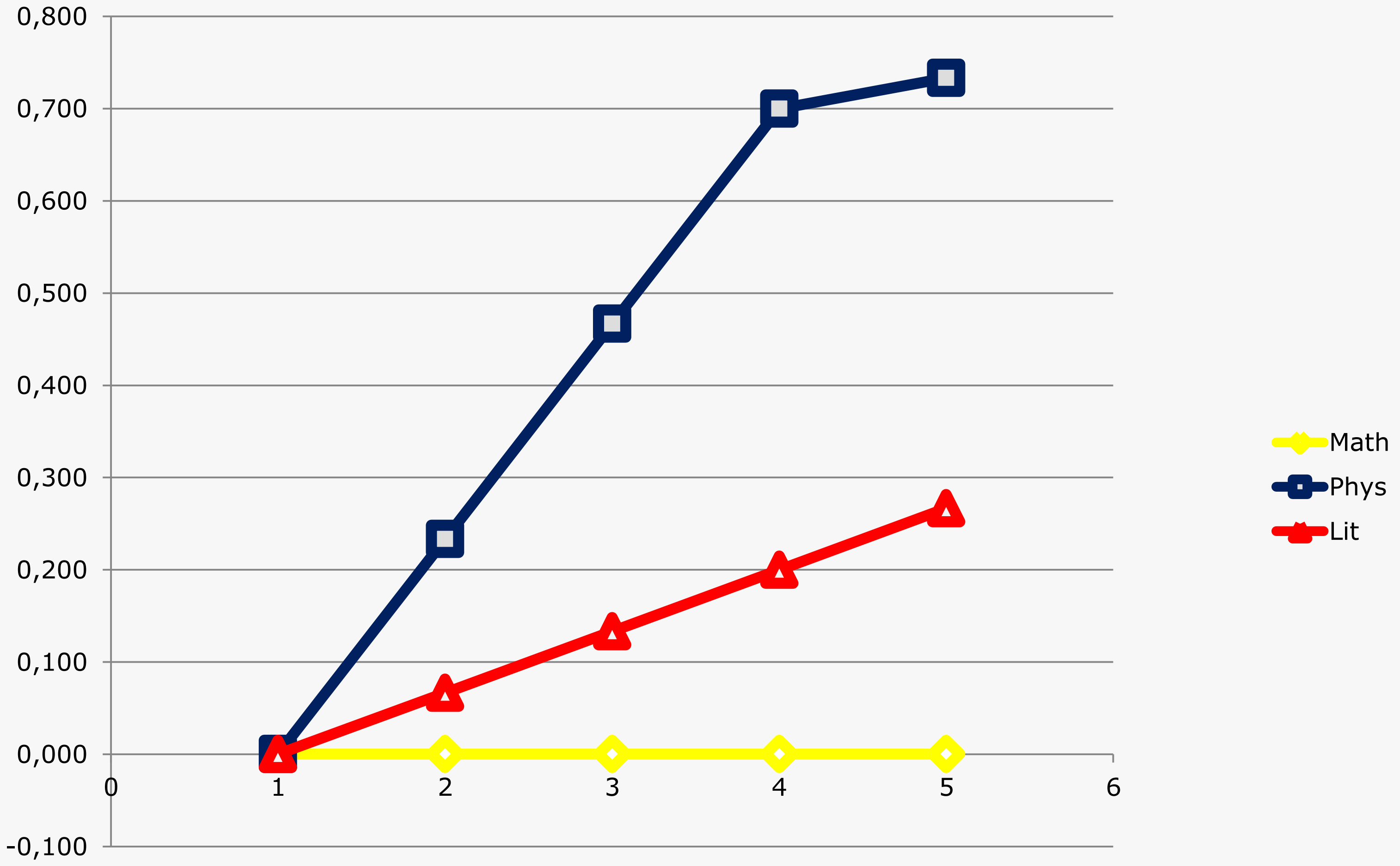
ε	**δ**
0.1	**0.066667**
0.2	**0.153846**
0.3	**0.269231**
0.4	**0.384615**
0.5	**0.5**



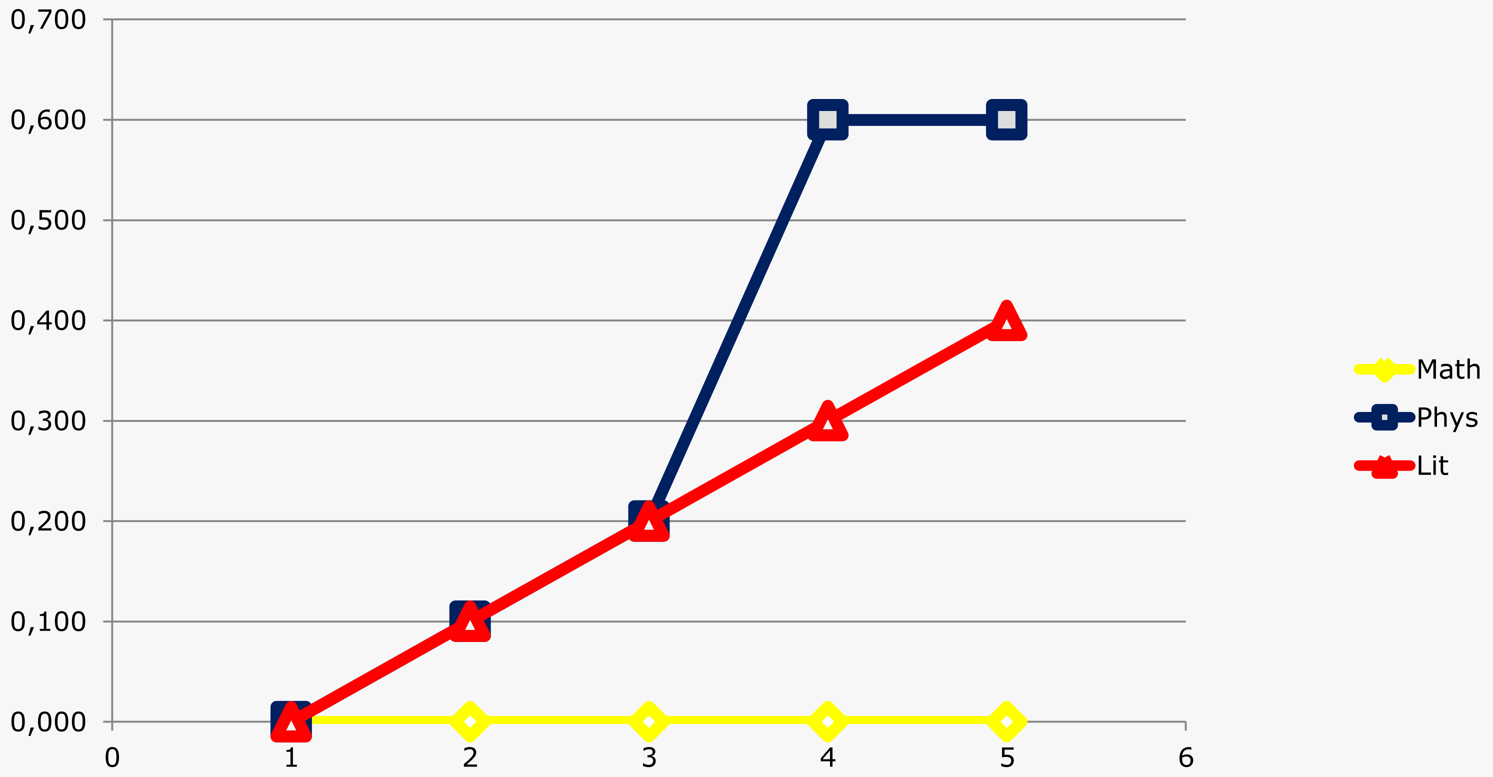
Model with a minimal number of characteristic points (ε=0.01, δ=0.02, η=1)



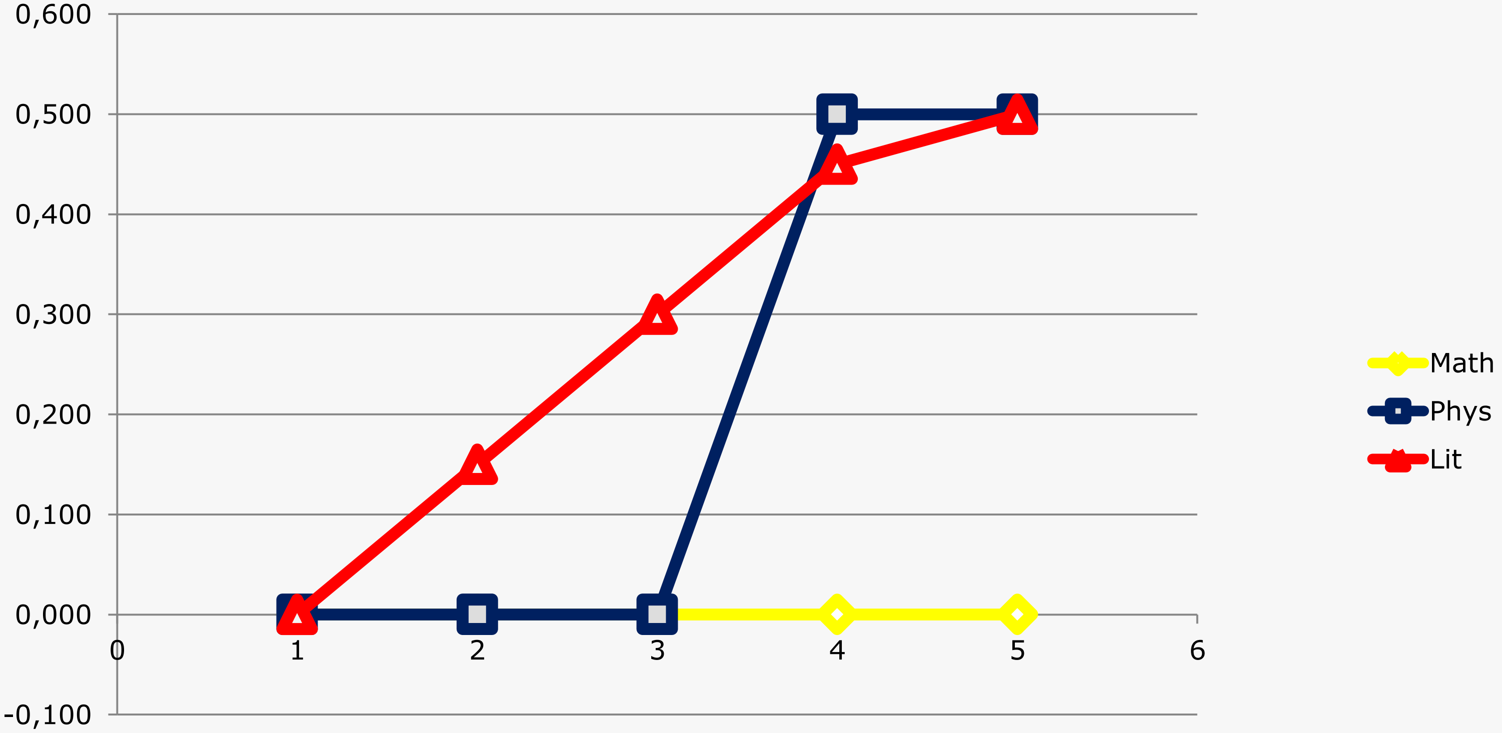
Model with a minimal number of characteristic points ($\varepsilon=0.1$, $\delta=0.2$, $\eta=1$)



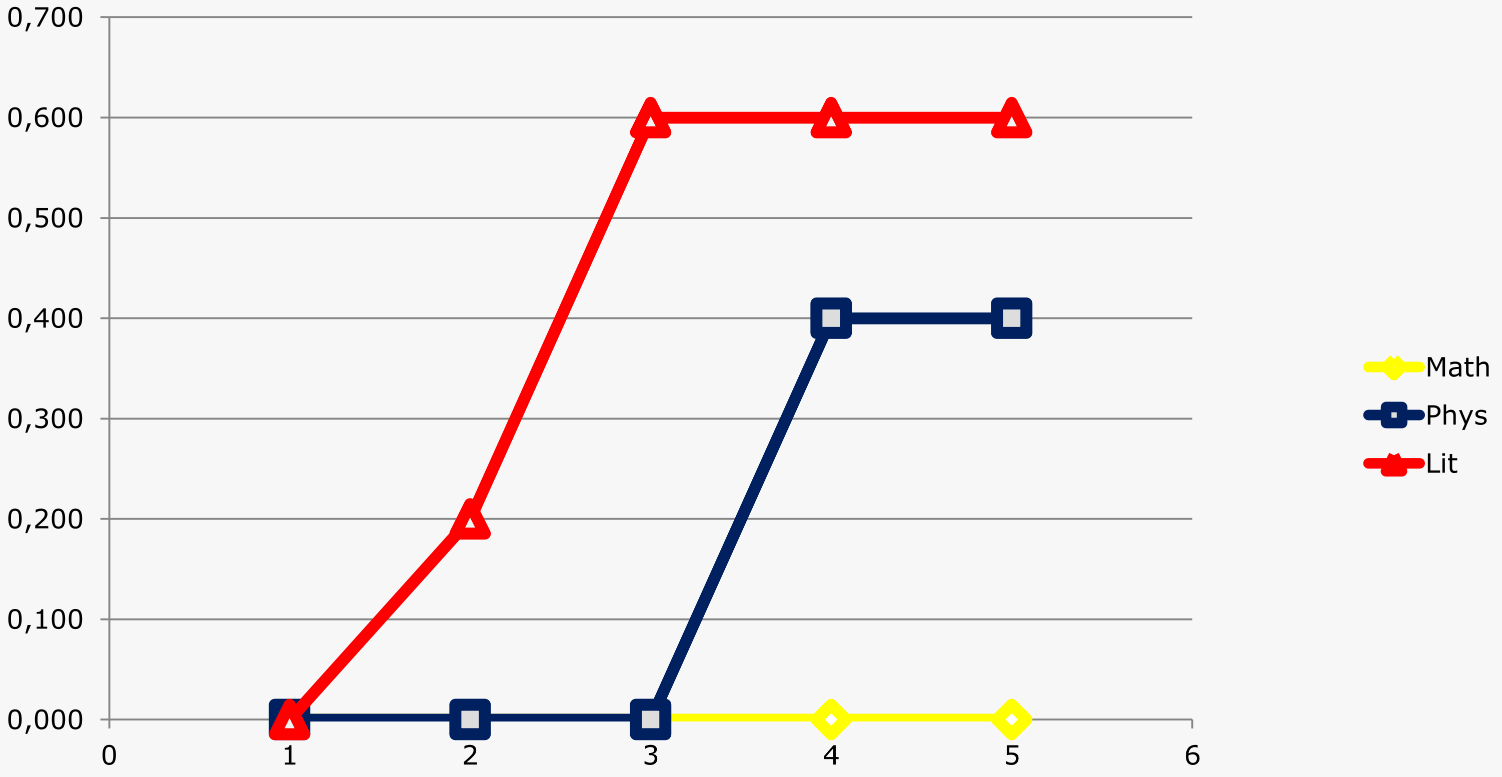
Model with a minimal number of characteristic points (ε=0.2, δ=0.4, η=2)



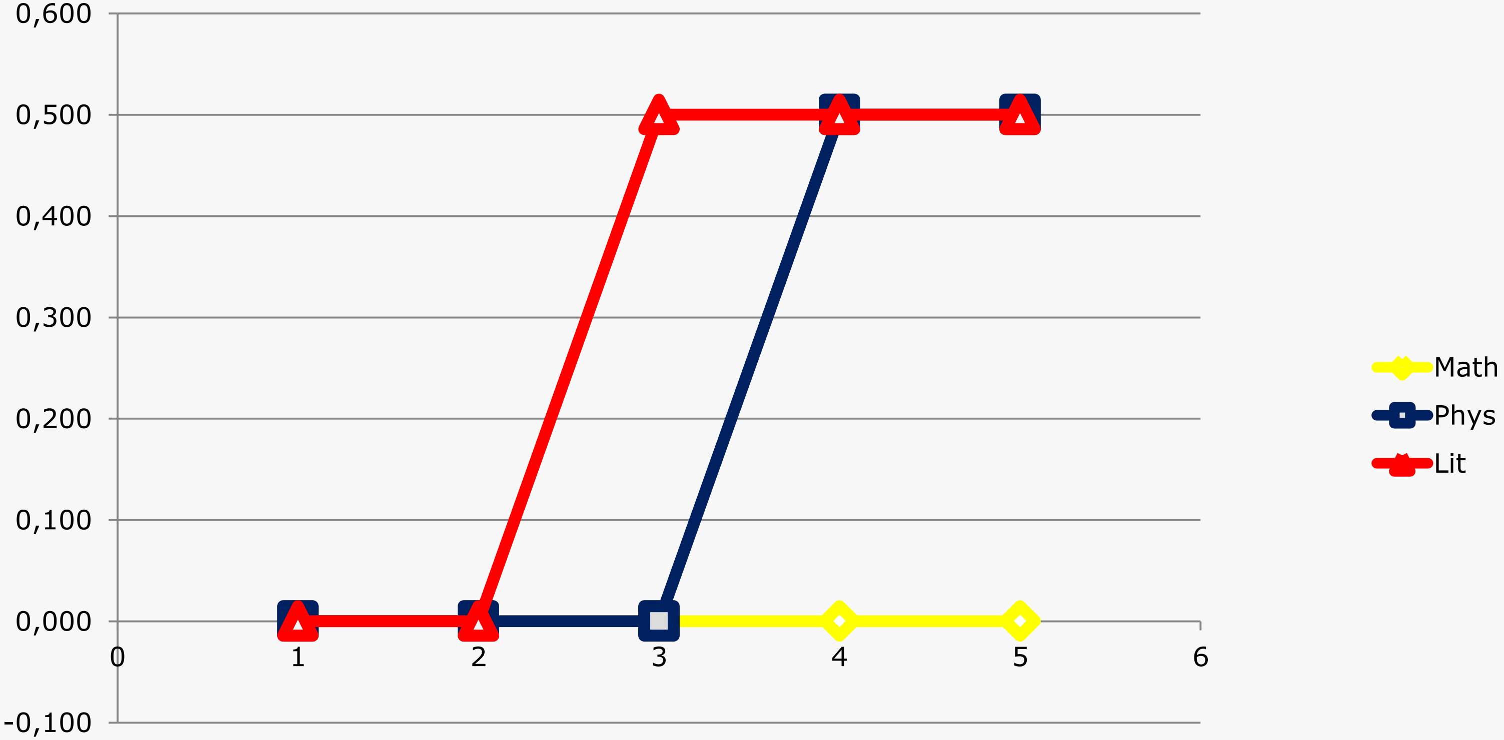
Model with a minimal number of characteristic points (ε=0.3, δ=0.4, η=3)



Model with a minimal number of characteristic points (ε=0.4, δ=0.4, η=4)



The model with the minimal number of characteristic points (ε=0.5, δ=0.5, η=4)



Relationship between discrimination, deviation from linearity and number of characteristic points

Discrimin (ε)	Dev-lin (δ)	Num ch-p (η)
0.01	0.02	1
0.1	0.2	1
0.2	0.4	2
0.3	0.4	3
0.4	0.4	4
0.5	0.5	4

Parsimony in Robust Ordinal Regression

- Remark 1. If there exists one value function representing DM’s preferences, in general, there exist infinitely many other compatible value functions
- Remark 2. In general, each one of these infinitely many compatible value functions gives a different ranking of actions from *A*.
- **It is fair to consider all compatible value functions**

Necessary preference relations under constraints of discrimination, deviation from linearity and number of characteristic points

New necessary preference relations

	s1	s2	s3	s4	s5	s6
s1	N	*	*	*	*	N
s2	*	N	*	*	*	*
s3	N	N	N	N	N	N
s4	N	N	*	N	*	N
s5	*	*	*	*	N	N
s6	*	*	*	*	*	N

$\varepsilon \geq 0.04, \ \delta \leq 0.1, \ \eta = 1$

The idea of the „representative" value function
(Figueira, Greco, Slowinski 2008; Kadzinski, Greco, Slowinski 2010, 2011)

- The idea is to select among compatible value functions that value function which better highlights the necessary ranking, i.e., maximizes the difference of values for pairs of actions *a* and *b*, such that *a* is necessarily preferred to *b* while *b* is not necessarily preferred to *a*
- As secondary objective, we minimize the difference of values for pairs of actions for which the possible preference relation is symmetric (no necessary relation holds)

Procedure to determine the „representative” value function

1) Determine the necessary and the possible preference relations in the considered set of actions
2) Add constraints on discrimination, deviation from linearity and minimal number of characteristic points
3) For all pairs of actions (a,b), such that a is necessarily preferred to b ($a \succeq^N b$ but *not* $b \succeq^N a$), add the following constraints to the linear programming constraints of the ordinal regression: $U(a) \geq U(b) + \rho$.
4) Maximize ρ
5) Add the constraint $\rho = \rho^*$, where $\rho^* = \text{Max } \rho$ from point 3), to the linear programming constraints of point 2)
6) For all pairs of actions (a,b), such that neither a is necessarily preferred to b nor b is necessarily preferred to a (*not* $a \succeq^N b$ and *not* $b \succeq^N a$), add the following constraints to the linear programming constraints of the ordinal regression : $U(a) - U(b) \leq \phi$ and $U(b) - U(a) \leq \phi$
7) Minimize ϕ

Parsimonious non-additive value function models

- The value functions considered by UTA, UTA^{GMS} and GRIP are additive:

$$U(x) = \sum_{i=1}^{m} u_i[g_i(x)]$$

- What happens if it is not possible to represent preferences with additive value functions ?

Illustrative example: violation of preference independence

Students	Mathematics	Physics	Literature
s1	~~High~~	Medium	Low
s2	~~High~~	Low	Medium
s3	~~Medium~~	Medium	Low
s4	~~Medium~~	Low	Medium

Preferences: $s2 \succ s1$ and $s3 \succ s4$

Choquet integral is not able to represent these preferences

We propose to enrich the additive value function...

- We consider a value function of the type

$$U^{int}(a) = \sum_{i=1}^{m} u_i\left(g_i(a)\right) + $$

$$\underbrace{+ \sum_{(i_1,i_2)\in Syn^+} syn^+_{i_1,i_2}\left(g_{i_1}(a), g_{i_2}(a)\right)}_{\text{„bonus”}} \; \underbrace{- \sum_{(i_1,i_2)\in Syn^-} syn^-_{i_1,i_2}\left(g_{i_1}(a), g_{i_2}(a)\right)}_{\text{„penalty”}}$$

Syn^+ is the set of pairs of criteria in a positive interaction

Syn^- is the set of pairs of criteria in a negative interaction

$syn^+_{i_1,i_2}(\cdot,\cdot), syn^-_{i_1,i_2}(\cdot,\cdot)$ are non-decreasing functions in the two arguments

- We allow any criterion to interact with one another only

UTA^{GMS}-INT (Greco, Mousseau, Slowinski 2013),

MUSA-INT (Angilella, Corrente, Greco, Slowinski 2013)

Illustrative example

Students	*math*	*phys*	*lit*	$syn^{+}_{math,lit}(\cdot,\cdot)$	Total score
s1	*Good* 0.03	*Medium* 0.28	*Bad* 0	*Good,Bad* 0	0.31
s2	*Good* 0.03	*Bad* 0	*Medium* 0.26	*Good,Medium* **0.03**	0.32
s3	*Medium* 0.02	*Medium* 0.28	*Bad* 0	*Medium,Bad* 0	0.30
s4	*Medium* 0.02	*Bad* 0	*Medium* 0.26	*Medium,Medium* 0	0.28

$s2 \succ s1$ and $s3 \succ s4$

(0.32 > 0.31 and 0.30 > 0.28)

Conclusions

- We introduced the idea of parsimonious preference model in ordinal regression for additive value functions that minimally deviates from the linearity and has minimal number of characteristic points
- The idea of parsimonious preference model works well with Robust Ordinal Regression
- The same idea has been applied to Non-Additive Robust Ordinal Regression, minimizing the number of interacting pairs of criteria
- Further developments of the idea of parsimonious preference models are related to:
    - UTADIS^{GMS} and UTADIS^{GSS}
    - ELECTRE^{GKMS} and PROMETHEE^{GMS}
    - UTA^{GMS}-GROUP, UTADIS^{GMS}-GROUP, ELECTRE^{GKMS}-GROUP and PROMETHEE^{GMS}-GROUP
    - UTADIS^{GMS}-INT

Thank you