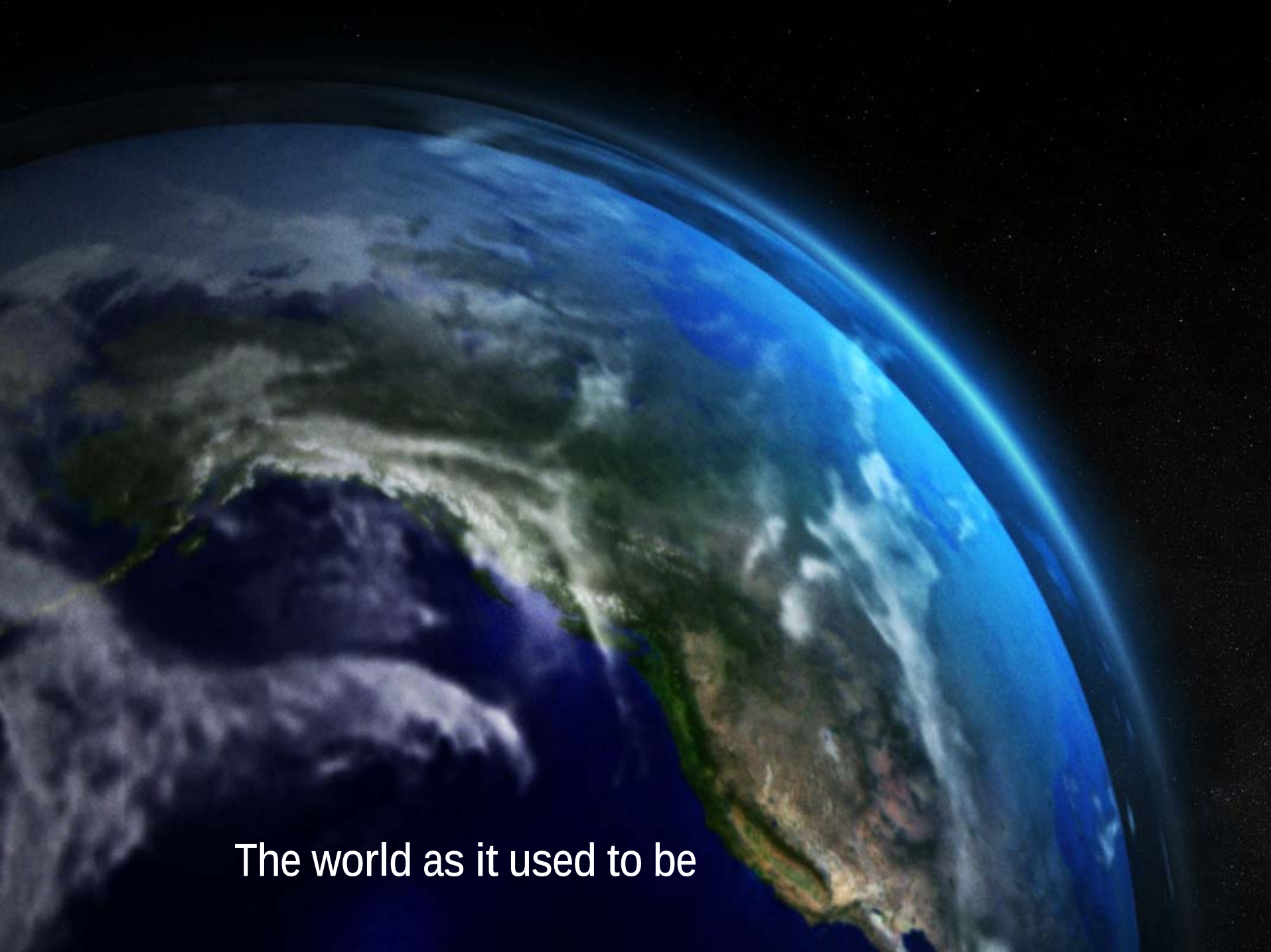
UNIVERSITY OF TWENTE.

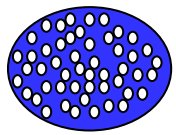
Robust item selection in CAT

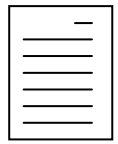
Bernard Veldkamp – RCEC – University of Twente
Mariagiulia Matteucci – University of Bologna


The world as it used to be

Item Writing

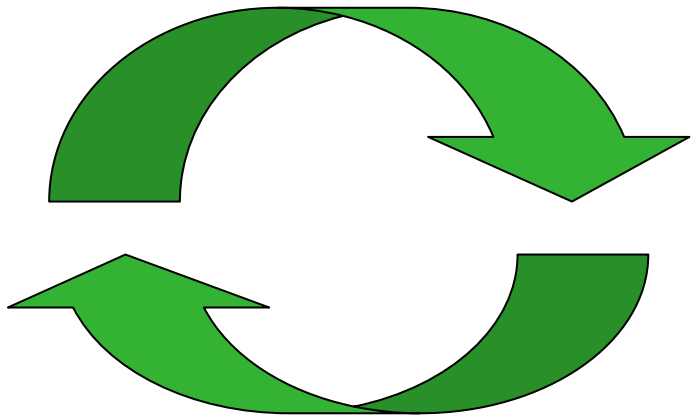
Calibration





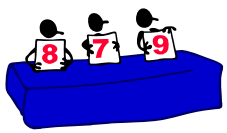
Test assembly

Defining Specifications



Transportation

Commumication of Results



Scoring

Administration

To make educational measurement even more efficient

Optimize item writing and calibration

- Item cloning
- Collateral information
- Informative priors
- Automated item generation

Increased uncertainty in item parameters

Serious consequences for item selection

- Items are stored in an item bank
- Item parameters are assumed to be known.
- Items that provide maximum information are selected for administration

Fundamental critisizm

- Major assumption of known item parameters does not hold (anymore).
- “Please include information about the uncertainty of the item parameters in the test assembly procedure.”

Fundamental critisizm

- Item selection capitalizes on positive estimation errors
- Uncertainty in parameter estimation has to be taken into account

Robust optimization techniques

- Weighted Deviation Model (Stocking & Swanson, 1993) deals with this problem by formulating soft constraints.
- The WDM comes with a cost
- Soyster (1973), Ben-Tal and Nemirovski (1998) and Bertsimas and Sim (2004) offer an alternative that can handle hard constraints.

Robust optimization techniques

- Soyster (1973) developed a method where all parameters in the uncertainty set were fixed at their minimal values.

- Ben-Tal & Nemirovski (2000) proposed a less conservative approach by applying quadratic programming instead of linear programming.

- Bertsimas and Sim (2004) proposed a linear programming technique that is less conservative, but assumes that uncertainty is limited to a number of parameters.

Example 1 Soyster's method

- Connector ability is a test battery that consists of three different subtests: Number series, Figure series, and Raven's matrices.

- The focus is on the Number series (NS) test.

Number series

11 14 20 32 ?

Number series

11 14 20 32 ?

+3 +6 +12 +24

Number series

11 14 20 32 ?

+3 +6 +12 +24

*2 *2 *2

Number series

11 14 20 32 ?

+3 +6 +12 +24

*2 *2 *2

Starting number
Level 1 operation
Level 2 operation

Number series

11 14 20 32 ?

+3 +6 +12 +24

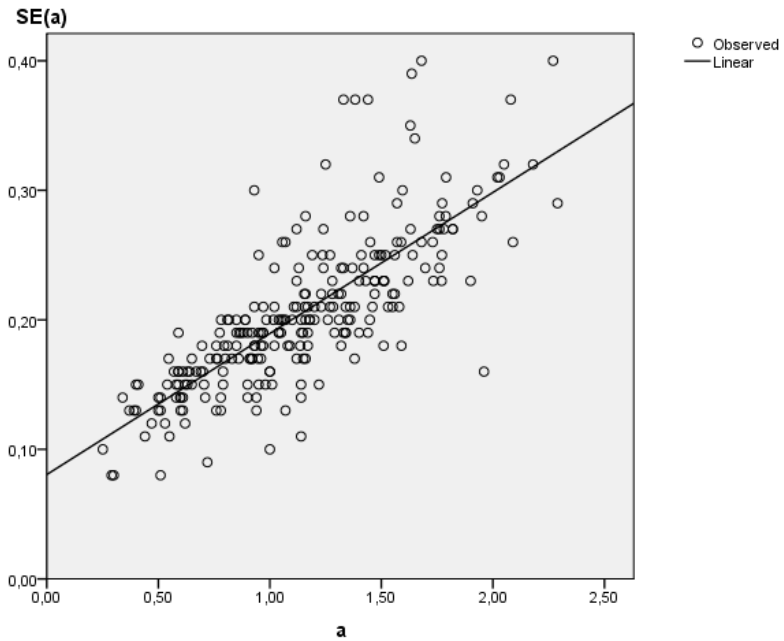
*2 *2 *2

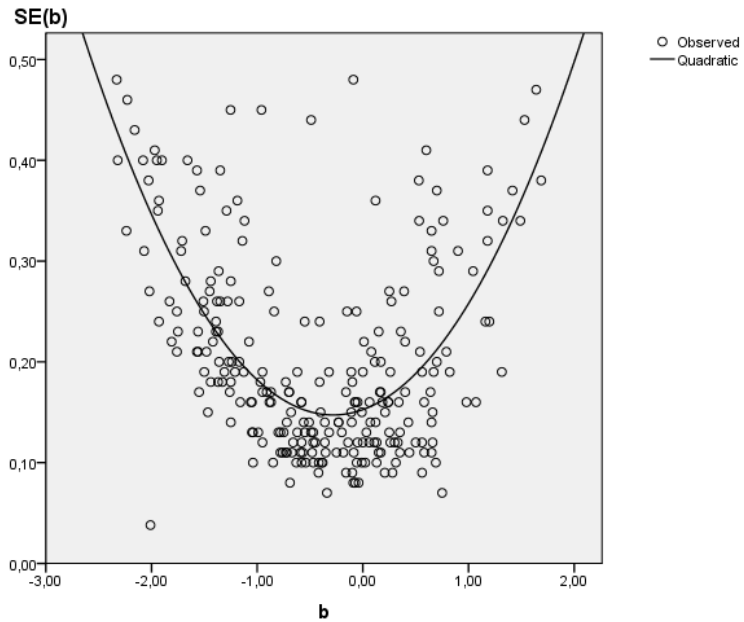
Starting number: 11
Level 1 operation (denoted by i): + 3
Level 2 operation (denoted by j): * 2

NS items

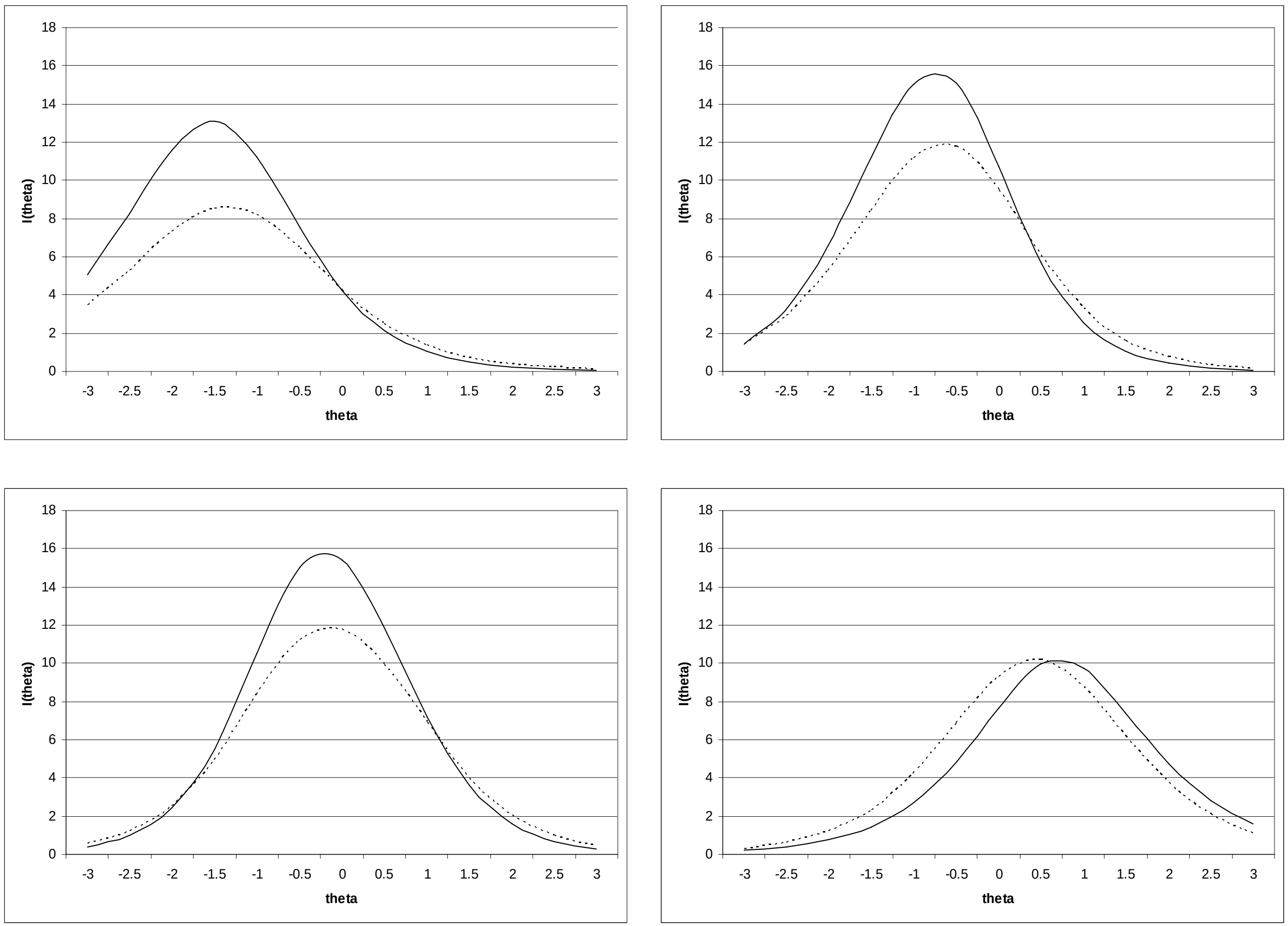
- An item bank of 253 items is available.
- The MIRT software package was applied to calibrated the items with the 2 PL model based on a sample of 3000 respondents, where each of the respondent answered 20 items in a balanced block design.

Item parameters

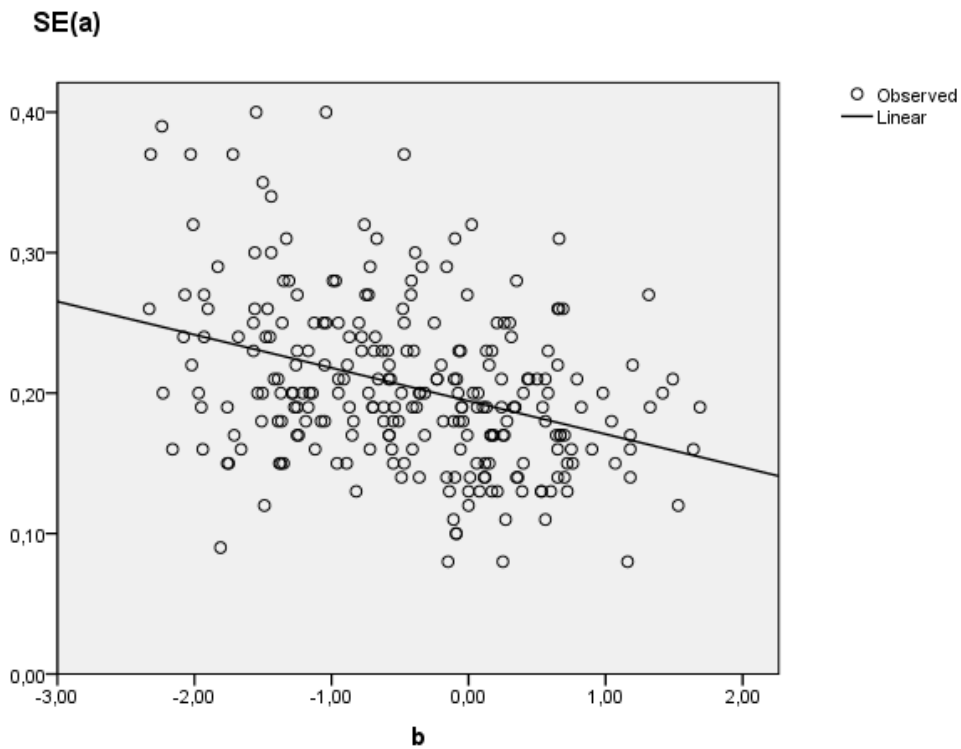




Results


18
16
14
12
10
8
6
4
2
0
I(theta)
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
theta
18
16
14
12
10
8
6
4
2
0
I(theta)
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
theta
18
16
14
12
10
8
6
4
2
0
I(theta)
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
theta
18
16
14
12
10
8
6
4
2
0
I(theta)
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
theta

Results


SE(a)
0,40
0,30
0,20
0,10
0,00
-3,00
-2,00
-1,00
0,00
1,00
2,00
b
Observed
Linear

Example 2: Ben-Tal and Nemirovski

When a ellepsoidal uncertainty set can be defined, the optimization problem results in a quadratic programming problem that could be solved in polynomial time.

Example 2: Ben-Tal and Nemirovski

When a ellepsoidal uncertainty set can be defined, the optimization problem results in a quadratic programming problem that could be solved in polynomial time.

Example 2: Ben-Tal and Nemirovski

- 95% reliability ellipsoid ~ robust optimization can be applied
- Polynomial time implies that the problem is tractable ~ it can be solved
- Unfortunately, quadratic programming is a very complicated optimization method.

Example 2: Ben-Tal and Nemirovski

***theorem*:** For n<15, the 65% reliability cube lies within the 95% reliability ellipsoid

~ linear programming can be applied.

Application

- Marketing
- International survey
- Short form construction
- Consumers' susceptibility to normative influences (SNI; Bearden, Netemeyer, and Teel, 1989)

SNI SCALE

- If I want to be like someone, I often try to buy the same brands that they buy.
- It is important that others like the products and brands I buy.
- I rarely purchase the latest fashion styles until I am sure my friends approve of them.
- I often identify with other people by purchasing the same products and brands they purchase.
- When buying products, I generally purchase those brands that I think others will approve of.
- I like to know what brands and products make good impressions on others.
- If other people can see me using a product, I often purchase the brand they expect me to buy.
- I achieve a sense of belonging by purchasing the same products and brands that others purchase.

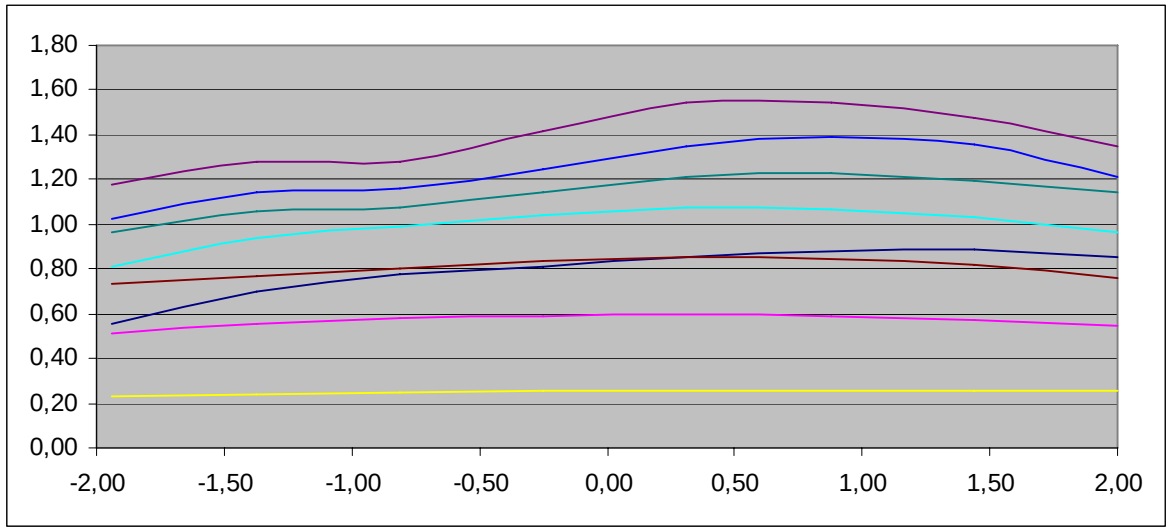
Design

- 28 countries
- 400-600 respondents per country

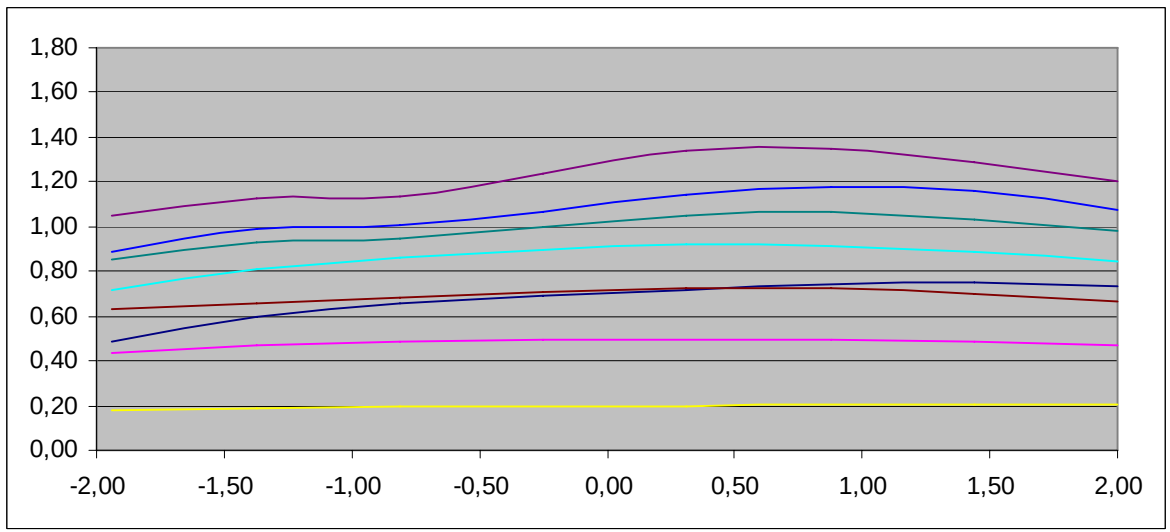
Test assembly problem:

- Target information function
- Minimize the number of items per country

Information functions
8 SNI items ~ Netherlands

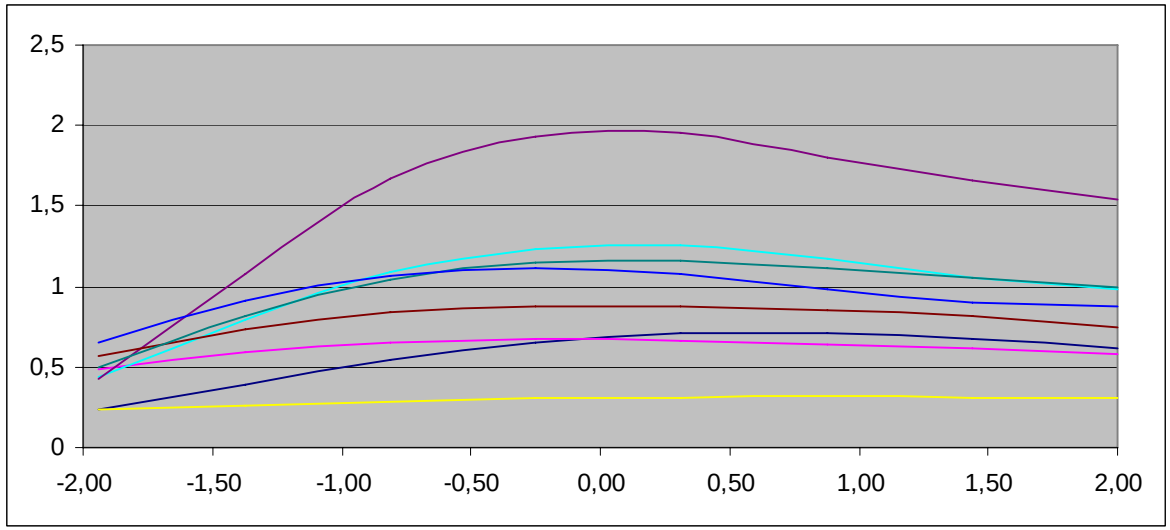


Fisher

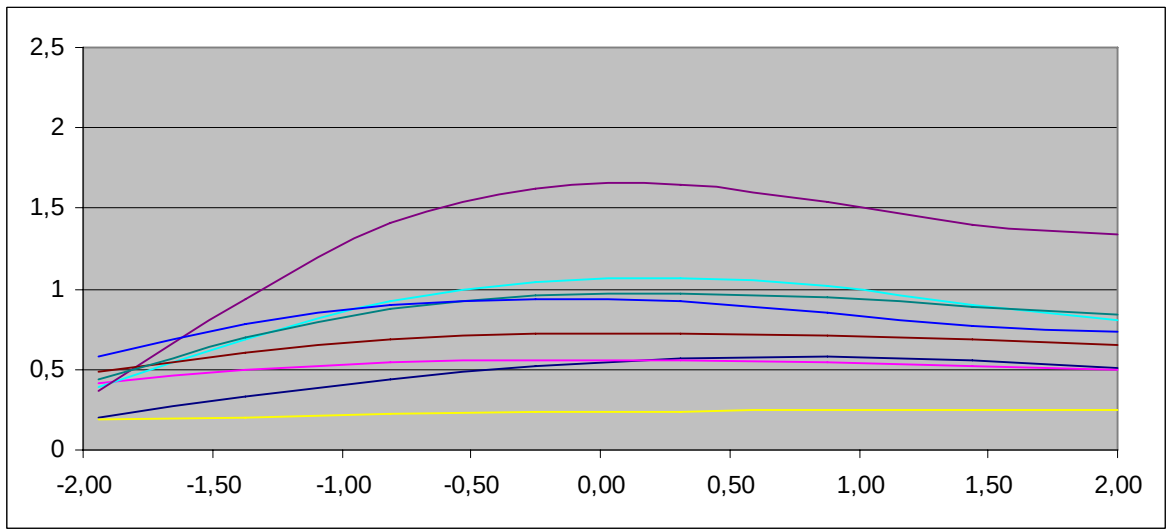


Robust

Information functions
8 SNI items ~ Hungary



Fisher



Robust

Selected items for short forms

	i1	i2	i3	i4	i5	i6	i7	i8
U.K.				X	X		X	X
Germany					X	X	X	X
Ireland				X		X	X	X
France		X		X	X	X	X	X
Austria					X		X	X
Hungary					X	X	X	X
Romania	X	X		X	X	X	X	X
U.S.					X		X	X

On average 0.62 (15%) more items

Conclusion

- Robust optimization is able to handle estimation errors
- The differences are substantial
- Be very careful with item banking
- New techniques for increasing efficiency can be applied.