

---

# Do Individual Hit-Rates Matter at All?

Stefan Binner

Design & Innovations Conference  
Munich 27-29 September, 2006



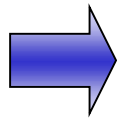
---

# Introduction

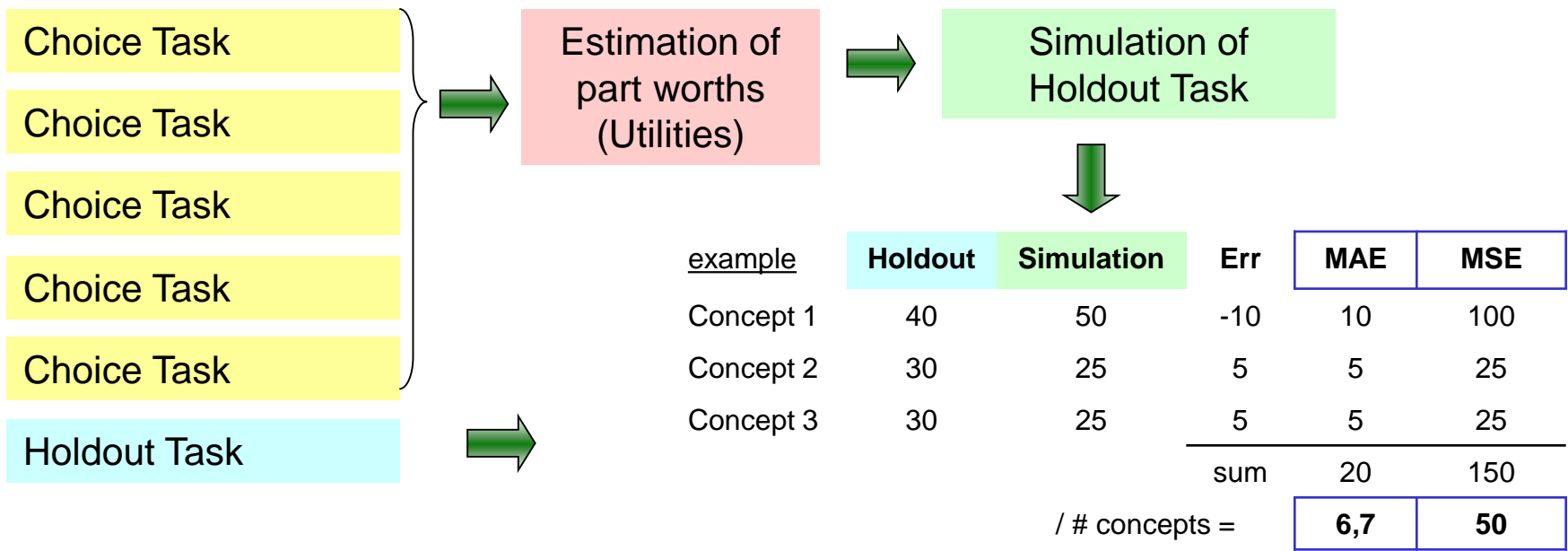
- Conjoint analysis is an excellent tool for modelling and estimating customers choice behaviour in real market situations.
- Especially since the introduction of HB the misfits between model prediction and real market can be really low.
- However, when looking at the data we usually find a phenomenon which made us feel quite uncomfortable in the past.
- Practical paper aimed at researchers and end users.

# How do we measure the validity of CJ studies?

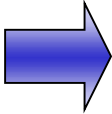
**Standard Solution:**

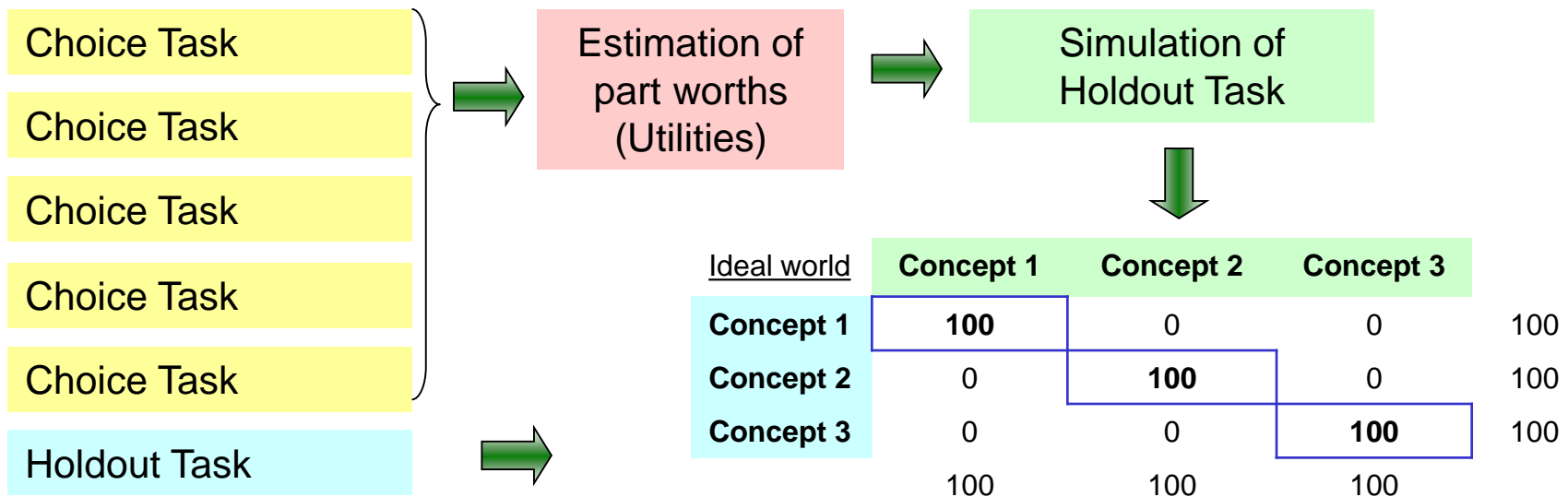


MAE (mean absolute error)  
MSE (mean squared error)



# How do we measure the validity of CJ studies?

**Alternative Solution:**  **Individual Hit Rates**



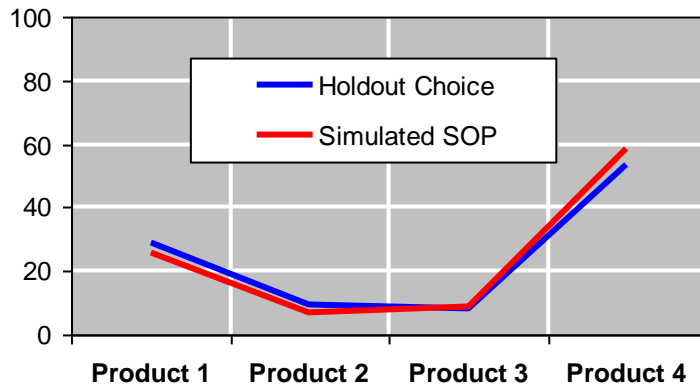
**This is the theory, but how good are these hit rates in reality?**

---

# In practice researchers find very good MAE results, but surprisingly low individual hit rates in CJ studies

Example: (building material study | CBC | N > 200)

Simulation of Holdout 1 (MAE = 3,5)



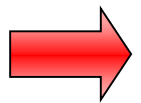
Individual Hit Rate of Holdout 1 (Overall 57,1%)

	Product 1	Product 2	Product 3	Product 4
Product 1	<b>48,02</b>	20,81	13,54	14,06
Product 2	8,19	<b>41,62</b>	5,80	3,62
Product 3	2,37	7,74	<b>59,46</b>	2,95
Product 4	41,42	29,82	21,20	<b>79,36</b>

---

## Questions:

- ▷ Do we really have individual level results?
- ▷ How can it be that the individual hit rate is so low, while simulations seem to be very accurate?
- ▷ Is there anything wrong with the data or with the respondents?
- ▷ Is it a problem of using CBC/HB?



Further investigation on reasons and relevance of this 'phenomenon' to research practice ...

---

## 10 conjoint studies analysed:

- All studies successfully used to improve market share and profit contribution or to optimise product development.
- These projects were conducted in different market segments in Europe, South America and Asia.
- In all studies selected the data was collected with personal interviews (CAPI or CLT) interviews in order to eliminate fraud in the fieldwork and to guarantee that respondents fit to screening criteria.
- All studies selected used Sawtooth Software CBC. Some with alternative specific designs, some with conditional pricing or individual relative pricing.

---

## 10 conducted studies analysed:

#	Market Segment	Target Group	N =	Delivery	Method	Choice Tasks	Attributes	Total Levels
1	Capital Goods	b2b	158	CAPI	CBC STD	16	5	14
2	Building Material	b2b	212	CAPI	CBC RP	16	5	22
3	Chemical	b2c	207	CAPI	CBC RP	16	5	20
4	Office Supply	b2c	399	CAPI / CLT	CBC ASD	16	7	34
5	Capital Goods	b2b	257	CAPI	CBC CP	16	4	14
6	Home Decoration	b2c	465	CAPI	CBC ASD	15	2	15
7	Building Material	b2b	600	CAPI / CLT	CBC STD	16	3	16
8	Consumer Adhesive	b2c	215	CAPI / CLT	CBC ASD	18	2	20
9	DIY - Home Improvement	b2c	155	CAPI	CBC ASD	15	5	23
10	Automotive	b2c	806	CAPI / CLT	CBC STD	10	6	23

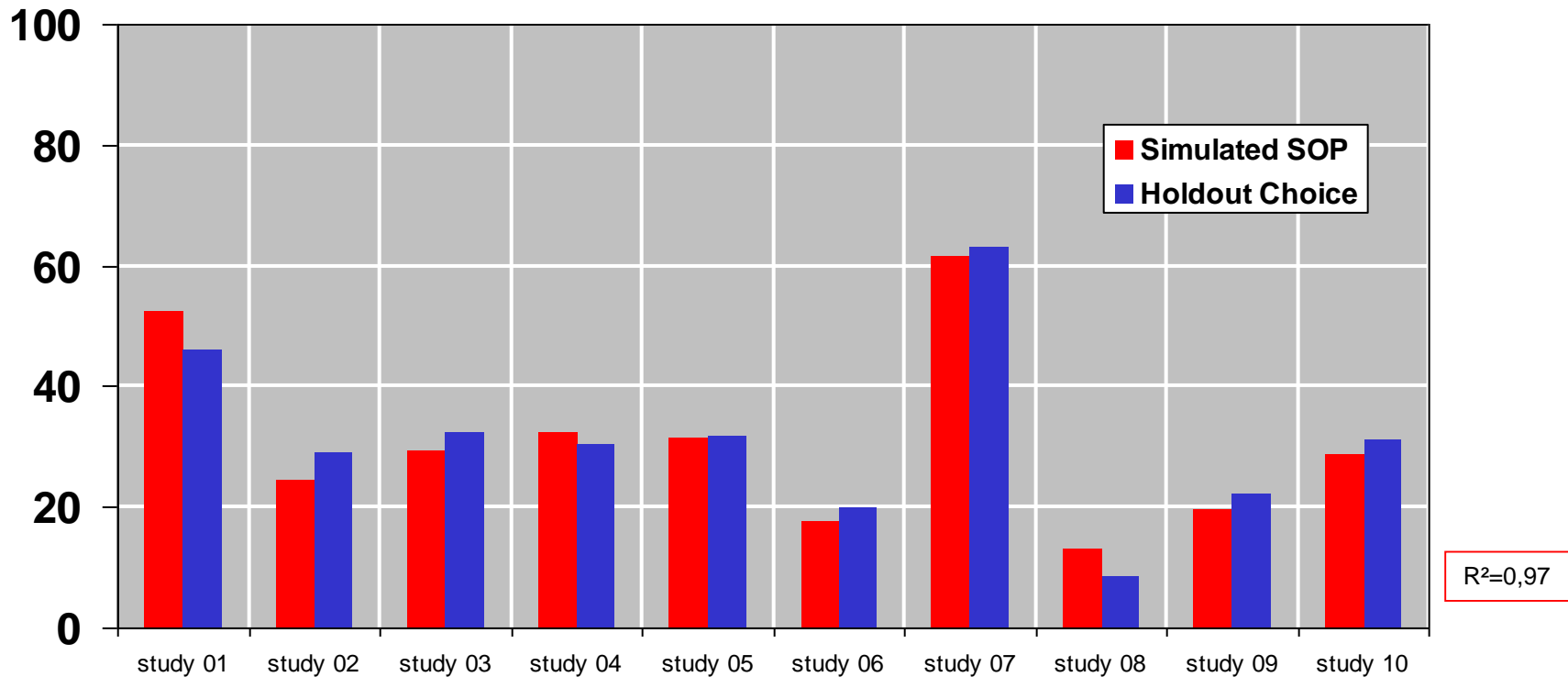
- All studies included between 2 and 5 fixed holdout tasks.
- Holdout tasks were typically very close to current market situation.



---

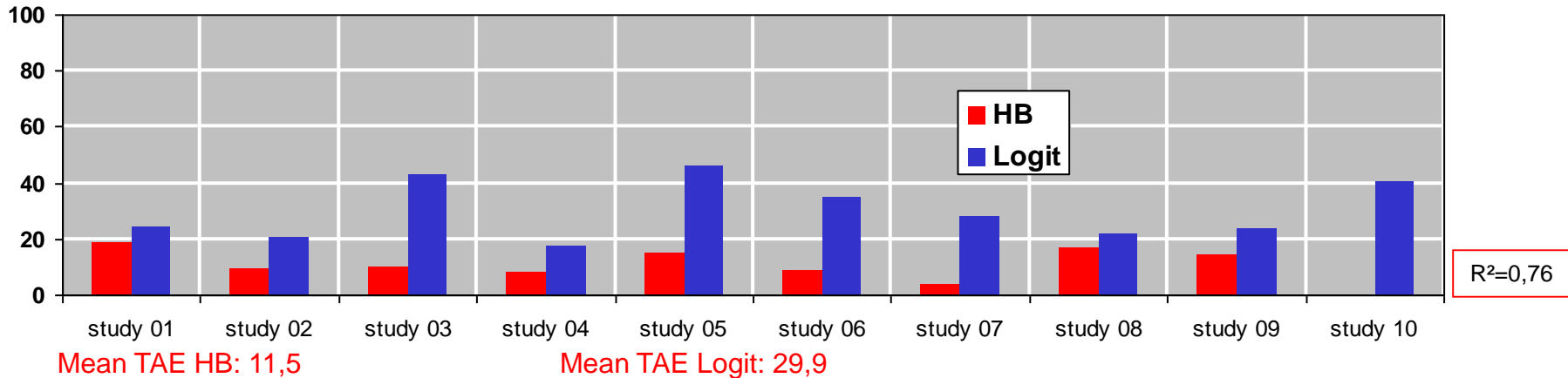
The ability of all of these studies to estimate holdouts was really good:

Simulated SOP (in % ) and real Holdout Choice (in %) of first holdout - concept 1

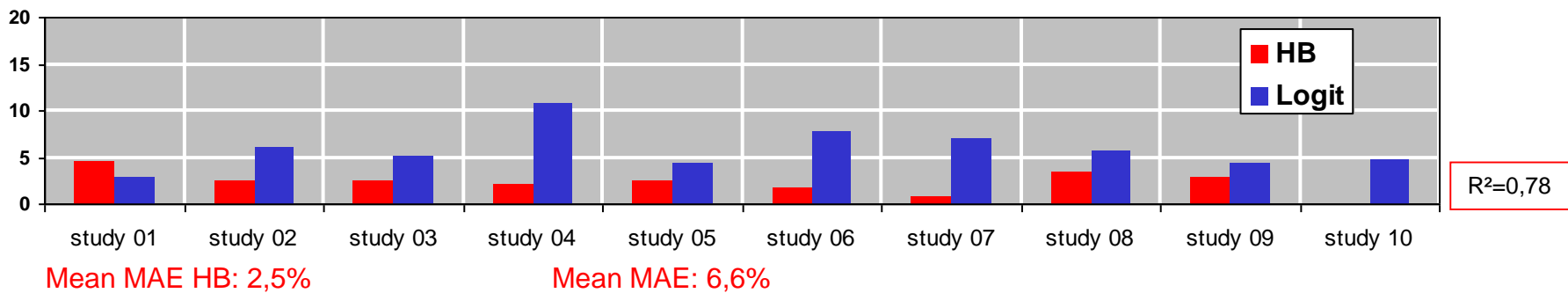


As expected HB is much better in predicting Holdouts Results than Logit:

Average TAE (all holdout questions)



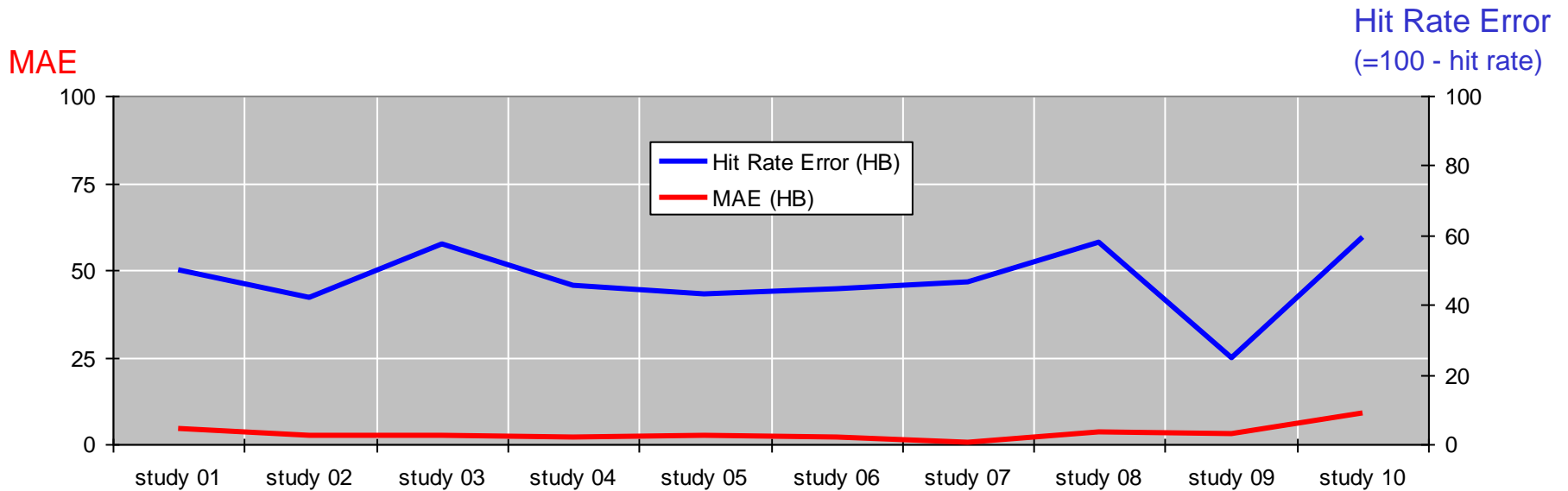
Average MAE (all holdout questions)



HB delivers individual level results  
Logit: calculation of subgroup utilities

---

While the accuracy of holdout predictions is quite high.  
How about predicting the individual choices?



On average of all 10 studies only 54% of individual holdout choice predictions were right!

---

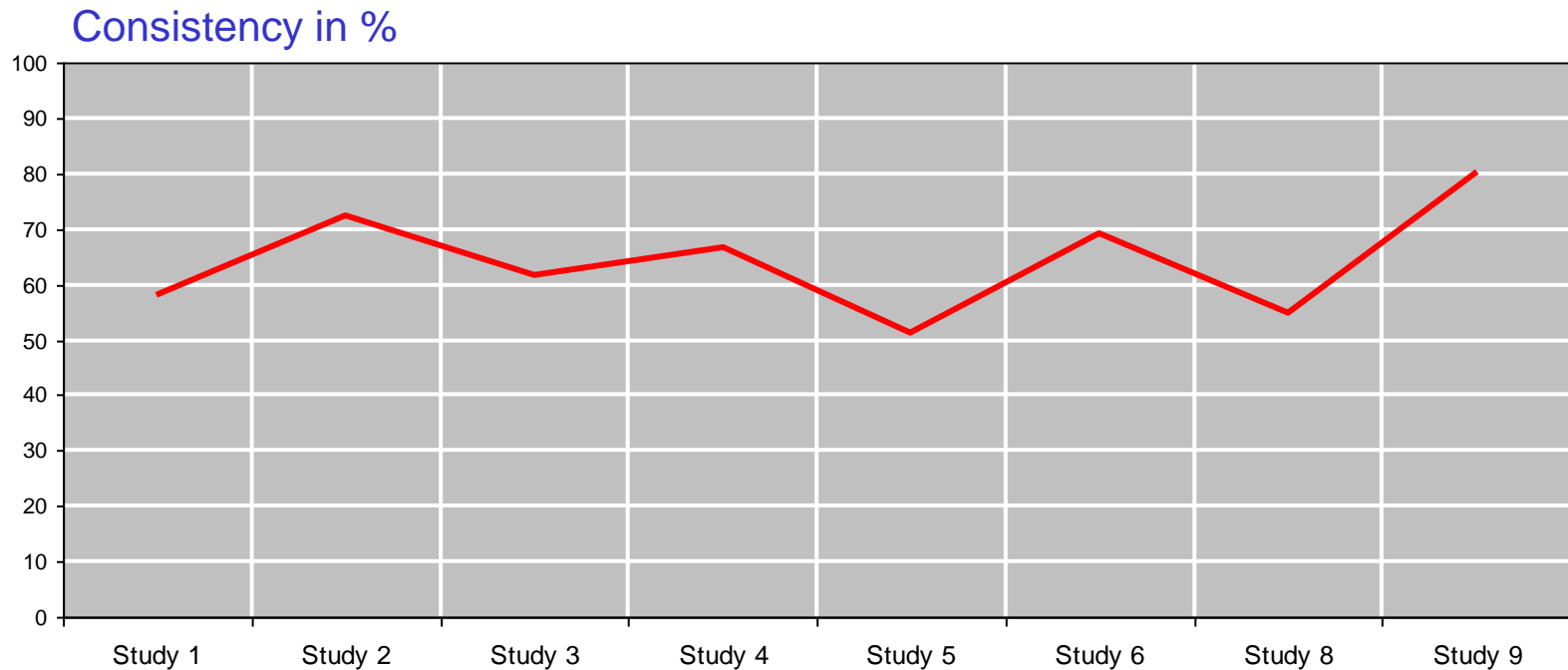
## Hypothesis A:

*There are some respondents who are not consistent in their choices and therefore conjoint models can not provide correct individual hit rates*

---

---

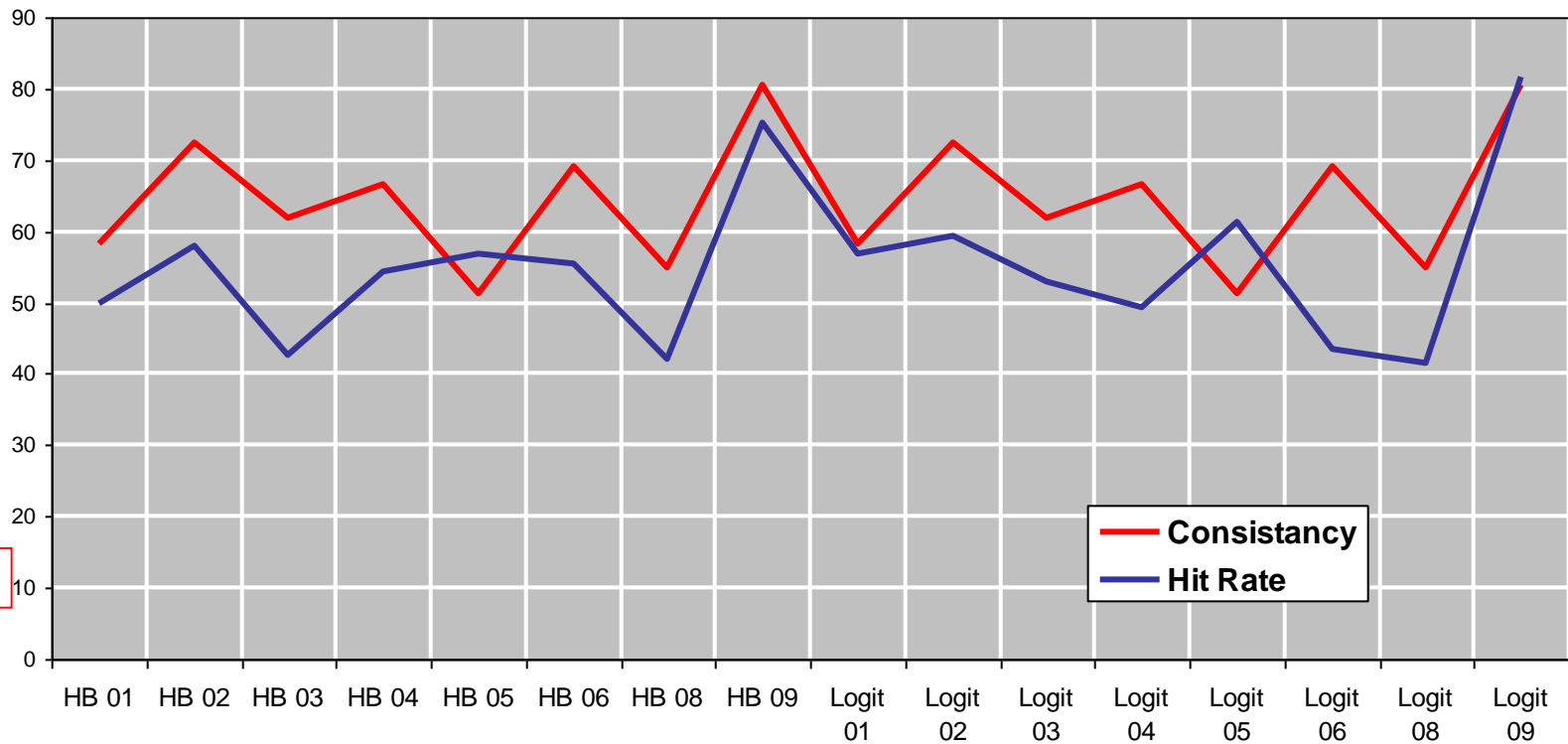
The consistency\* of respondents answering the same choice tasks twice is also lower than many clients expected.



\*) Consistency was defined as percentage of respondents who answered two identical holdout questions in the same way.

# Furthermore there seems to be a relationship between the individual hit rate and consistency\*

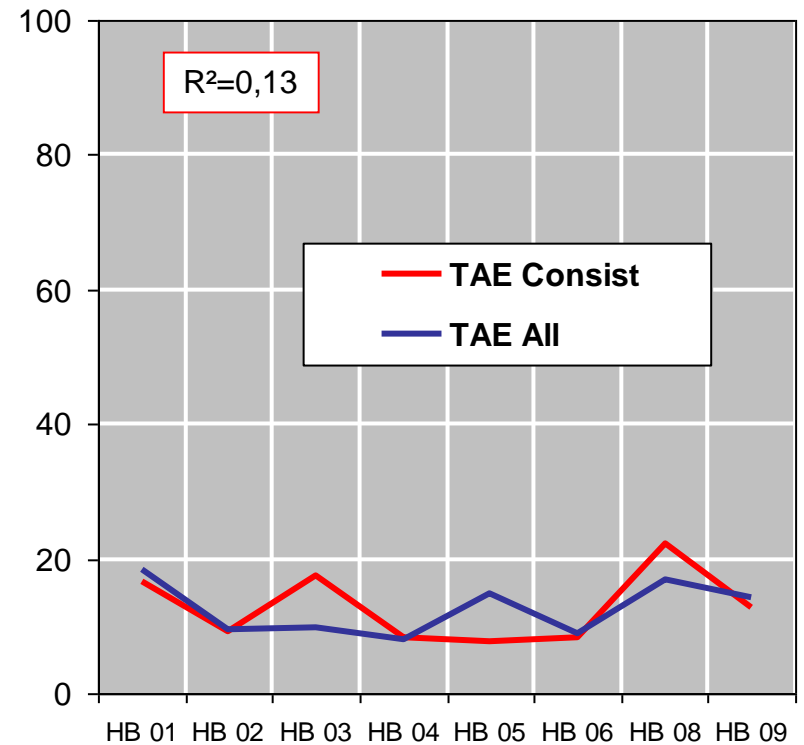
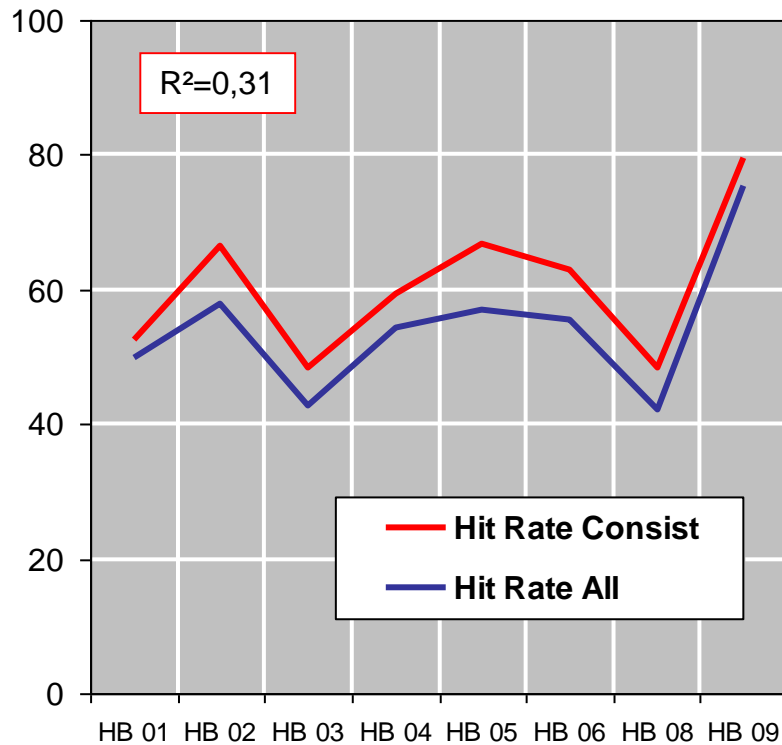
Consistency and Individual Hit Rate in %



$R^2=0,60$

- 
- What happens if we get rid of those “inconsistent” respondents?
  - When we exclude all respondents which had not identical holdout results. Will the individual hit rates improve?

**Not really!** Without inconsistent respondents the individual hit rates seem to be higher, but total absolute errors are sometimes even worse





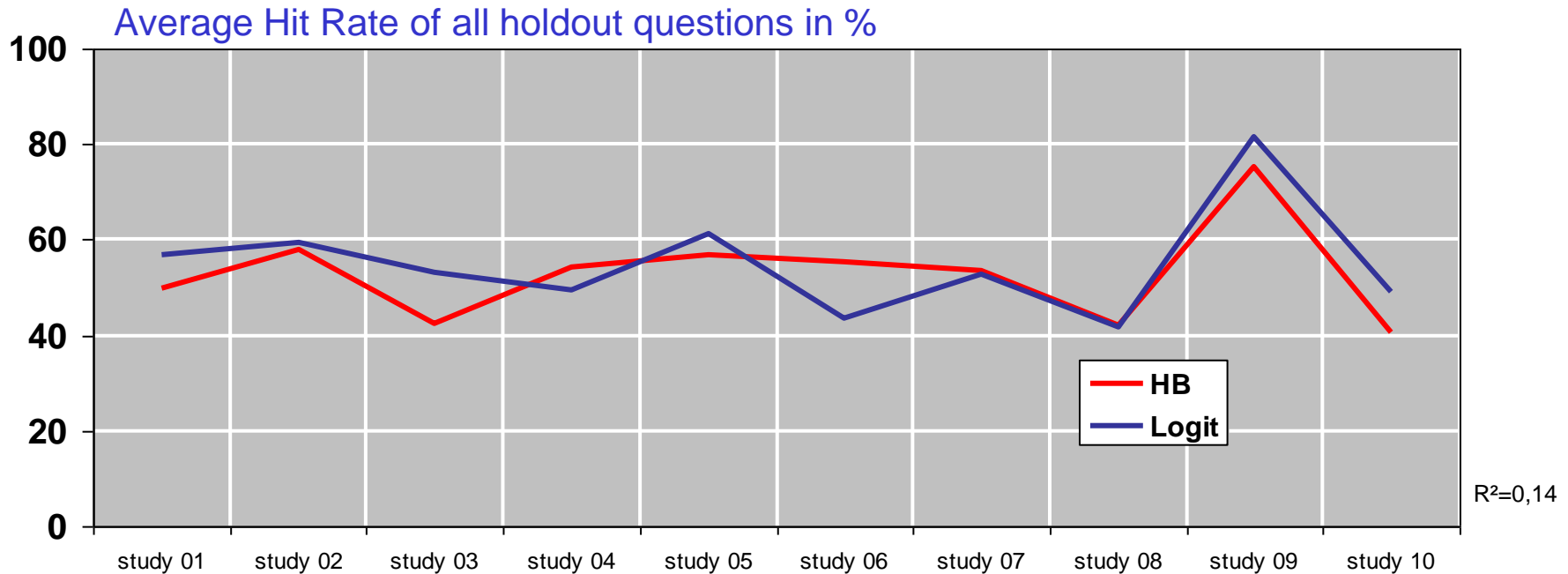
---

## Hypothesis B:

*The individual hit rate is a result of using the hierarchical HB model for individual choice estimation which “lends data” for individuals from the total population.*

---

# No! The hit rates of HB and Logit are quite the same

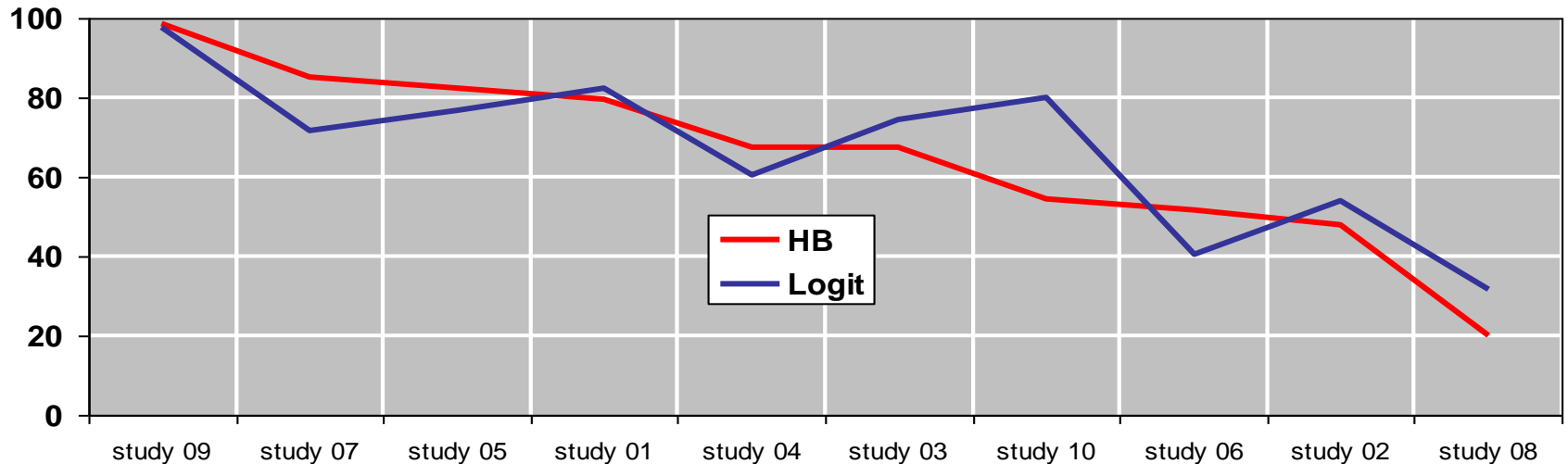


Note: Logit hit rates were calculated based on aggregated utilities of those respondents who selected the respective alternative in the holdout tasks (i.e. one utility set for each answer alternative)

---

# But why are some studies doing better?

Hit rate of first concept in the first holdout task in %



- In some studies the individual hit rate seems to be better than others.
- Are there any methodical factors to drive this?

---

## Hypothesis C:

*The accuracy of individual hit rates is depending on study design and model methodology*

---

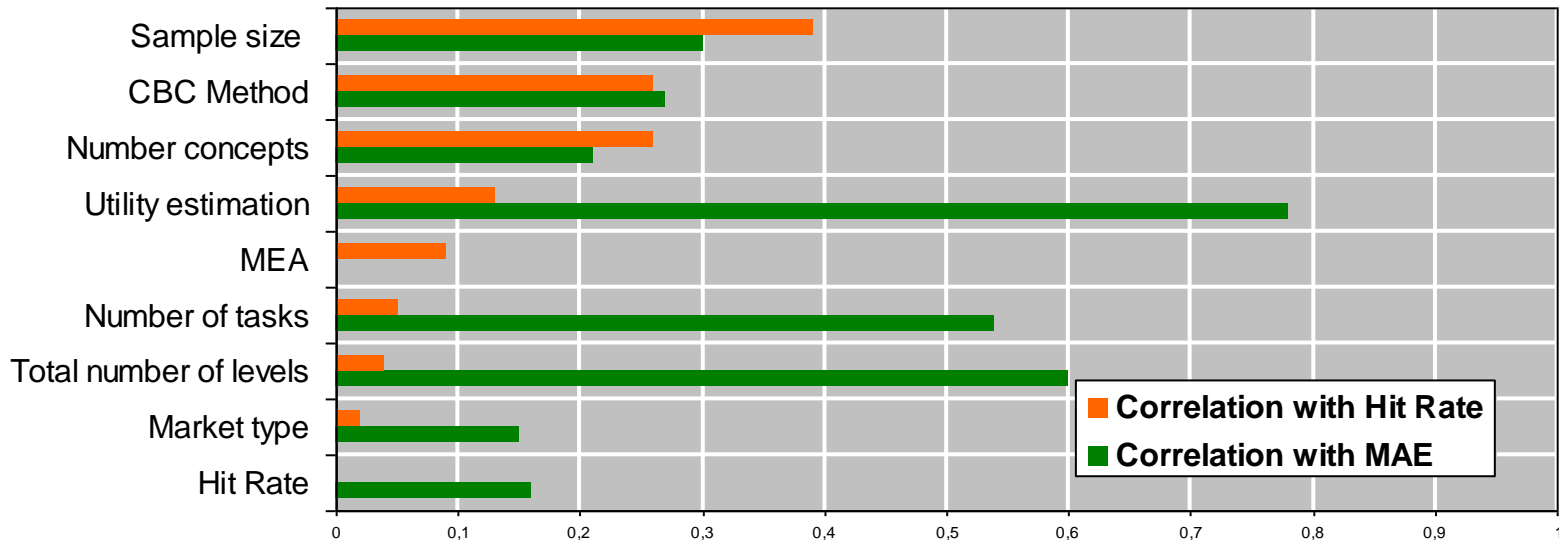
---

Following design and method factors were used for the analysis:

<b>CBC Method</b>	Standard CBC, Alternative Specific Designs, Condition or Individual Relative Pricing
<b>Hit Rate</b>	Individual Hit Rate of Holdout Tasks
<b>Market type</b>	B2B or B2C?
<b>MEA</b>	Mean Absolute Error of Holdout Simulation
<b>Number concepts</b>	Number of concepts shown in Choice Tasks
<b>Number of tasks</b>	Number of Choice Tasks
<b>Sample size</b>	Number of interviews
<b>Total number of levels</b>	Sum of all attribute levels
<b>Utility estimation</b>	HB-Reg or Logit

---

## Rather Not! Here are the correlations ( $R^2$ ) between Hit Rate and selected methodical factors



There is no significant correlation between methodical drivers and hit rate. However, the influence on MAE is different and driven by utility estimation method (HB vs. Logit) and number of tasks.

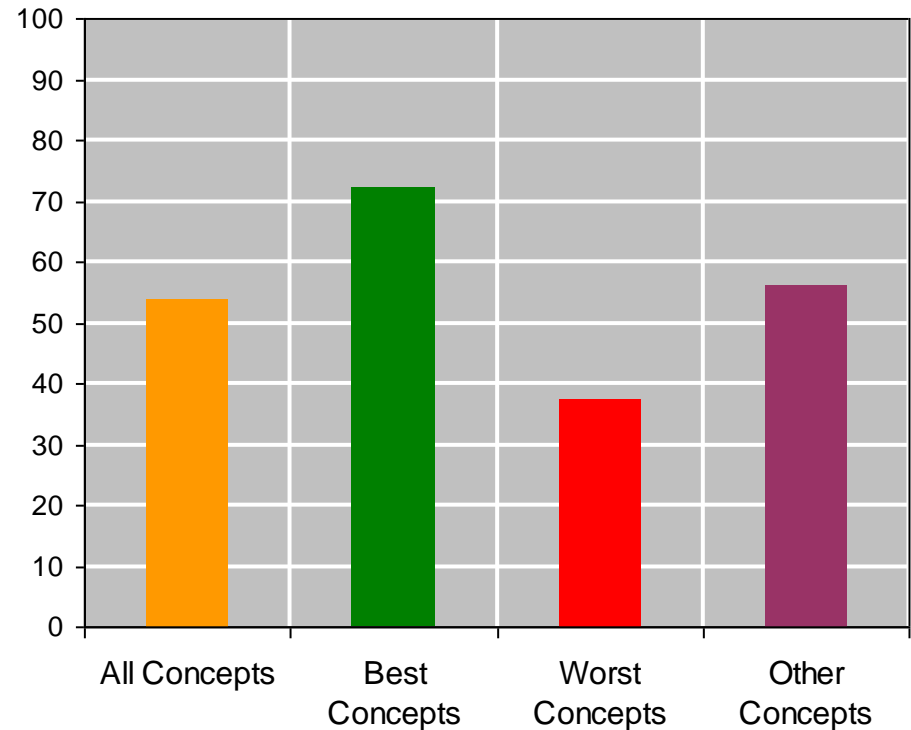
(The presence of a “None” should have an influence. However, there was not enough data to analyse this)

---

## However, the analysis showed another relationship: Between Individual hit rate and “success” of concepts

- While **overall almost half** of individual estimates **were wrong...**
- The prediction of **Winning Concepts worked best**
- And it was most **difficult to predict Least Liked Concepts**

Average Hit Rate of all holdout questions in %



$R^2=0,9$

---

## Hypothesis D:

*The accuracy of individual hit rates is depending other (soft) factors*

---



---

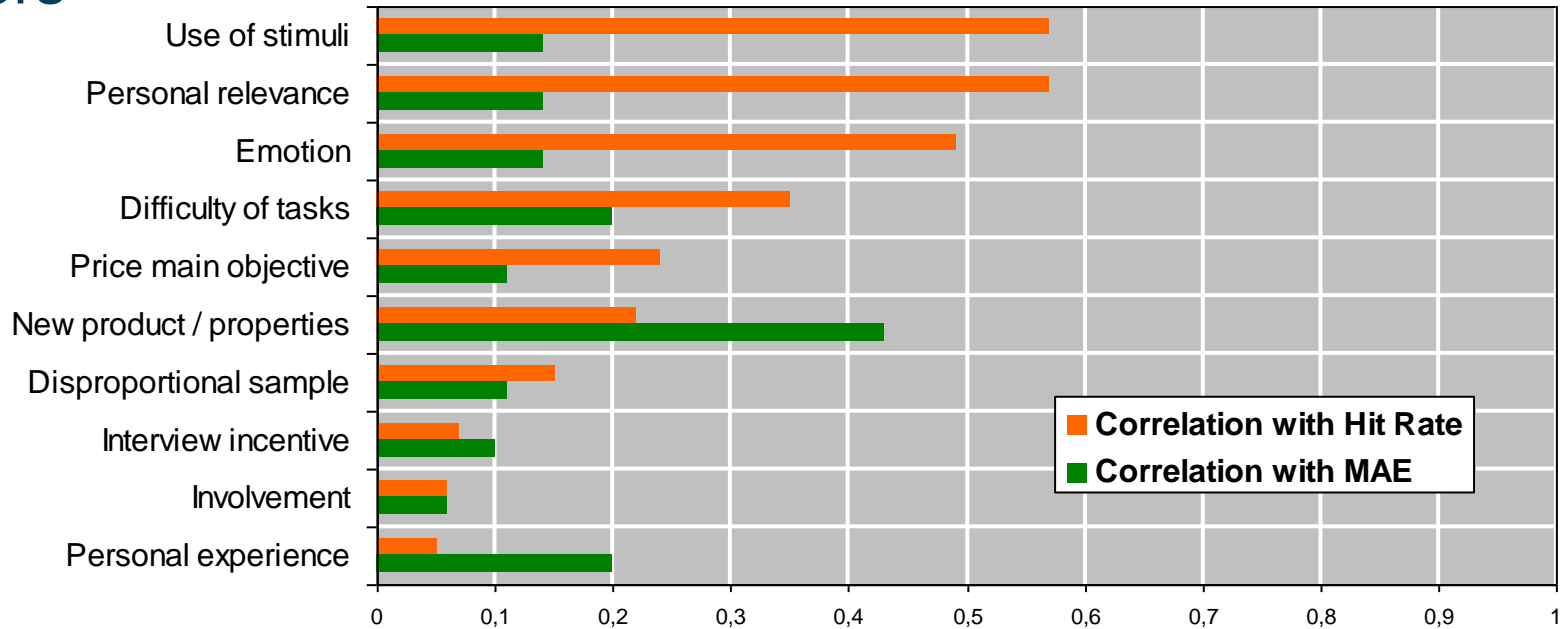
## Following (soft) factors were defined and each study categorized\*:

<b>Difficulty of tasks</b>	How difficult was it to answer the choice tasks?
<b>Disproportional sample</b>	Did we force a disproportionate sample
<b>Emotion</b>	Were there emotional elements?
<b>Interview incentive</b>	Did the interviewee receive an incentive?
<b>Involvement</b>	Were interviewees personally involved?
<b>New product / properties</b>	Were there completely new products or properties
<b>Personal experience</b>	Did the respondent have personal experience?
<b>Personal relevance</b>	Was the product of personal relevance to the resp.?
<b>Price main objective</b>	Was price the main research objective?
<b>Use of stimuli</b>	Did we use stimuli (pictures, mock ups, etc) for the interview?

\*) classification made by the author

---

## Yes! Here is the Correlation ( $R^2$ ) between Hit Rate these soft factors



The use of stimuli and personal relevance with the products in question seems to have the highest influence. Furthermore emotional connection with the researched product category.

# Conclusions

---

- Individual Hit Rates are not the best indicator of study success! (i.e. validity and accuracy of market simulations)
- Individual Hit Rates are neither related to the MAE of simulations nor to relevant factors such as:
  - Sample size
  - Utility estimation method
  - Number of tasks
  - Number of levels

which all have otherwise significant influence on ability of accurate market simulation.

# Conclusions

---

- The consistency of respondents is naturally not perfect. Individual hit rates are even lower. But they are not caused by inconsistent respondents only.
- Drivers of Individual Hit Rates include furthermore
  - the success of individual concepts
  - the use of stimuli
  - personal relevance of the choices for the respondent
  - emotional connection with product category

# Conclusions

---

- If we can trust that we excluded fake interviews or dubious respondents, then the misfit of Individual Hit Rates is just the typical share of customers who select products randomly or even wrong.
- This could be because of
  - perfect level balance
  - lack of personal relevance
  - ignorance of product

# Conclusions

---

- The author therefore concludes that individual hit rates do not matter so much.
- They should be used as indicator for subject relevance and personal involvement of respondents.
- But Low Individual Hit Rates are not the ultimate indicator for measuring validity of conjoint simulations.