

Kai Kristensen, IMPS '03

# PLS structural Equation Modeling for Customer Satisfaction

## -Methodological and Application Issues-

Kai Kristensen, J. Eskildsen, H.J. Juhl, P. Østergaard  
Centre for Corporate Performance  
The Aarhus School of Business, Denmark

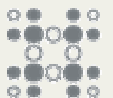


Aarhus School of Business

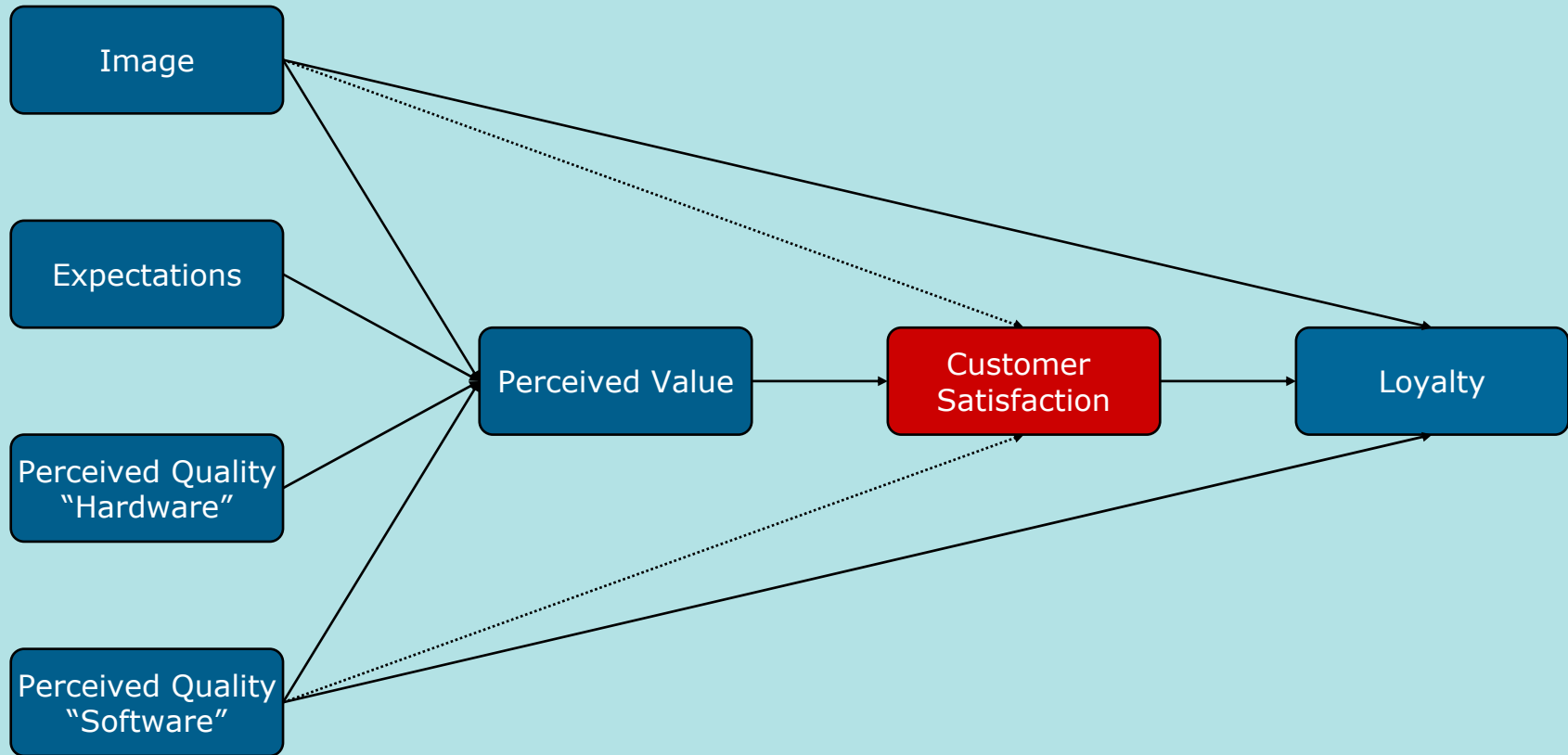
Håndelsakademiet  
i Aarhus

# Agenda

- The EPSI Rating Model
  - Latent Structure
  - Manifests
- A few recent results
  - The Danish car market
  - External validity
- Practical problems and observations
  - The choice of scale
  - Reliability: The choice of manifests
  - Explanatory power
  - Missing values
  - Multicollinearity
- Some results from a simulation study



# The EPSI Rating Model: Latent structure



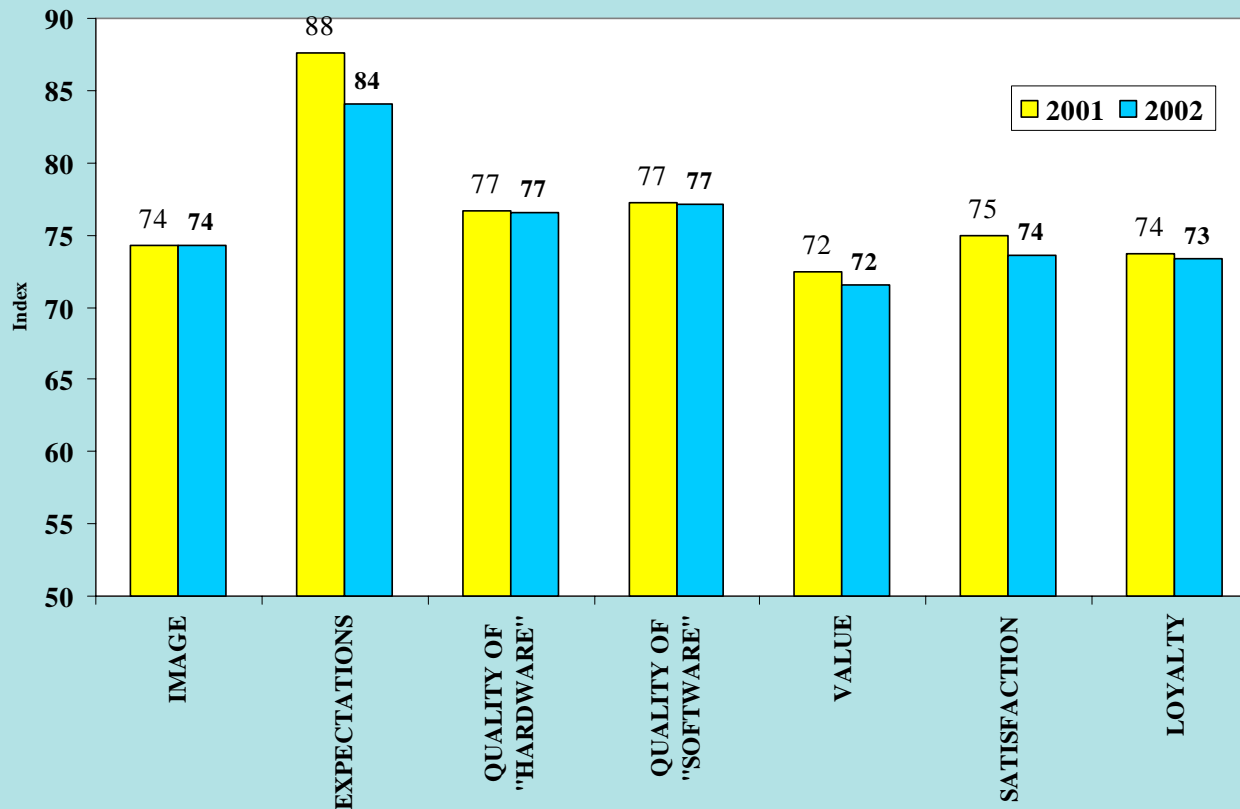
## EPSI Rating model

- Generic model with 7 latent constructs
  - 4 latent exogenous constructs (Image, expectation, quality of "hardware" and "software")
  - 3 endogenous constructs (perception of value, satisfaction and loyalty)
- Each construct is determined by 3-6 manifest measurements.
- The model is estimated by use of PLS (Partial Least Squares estimation techniques).

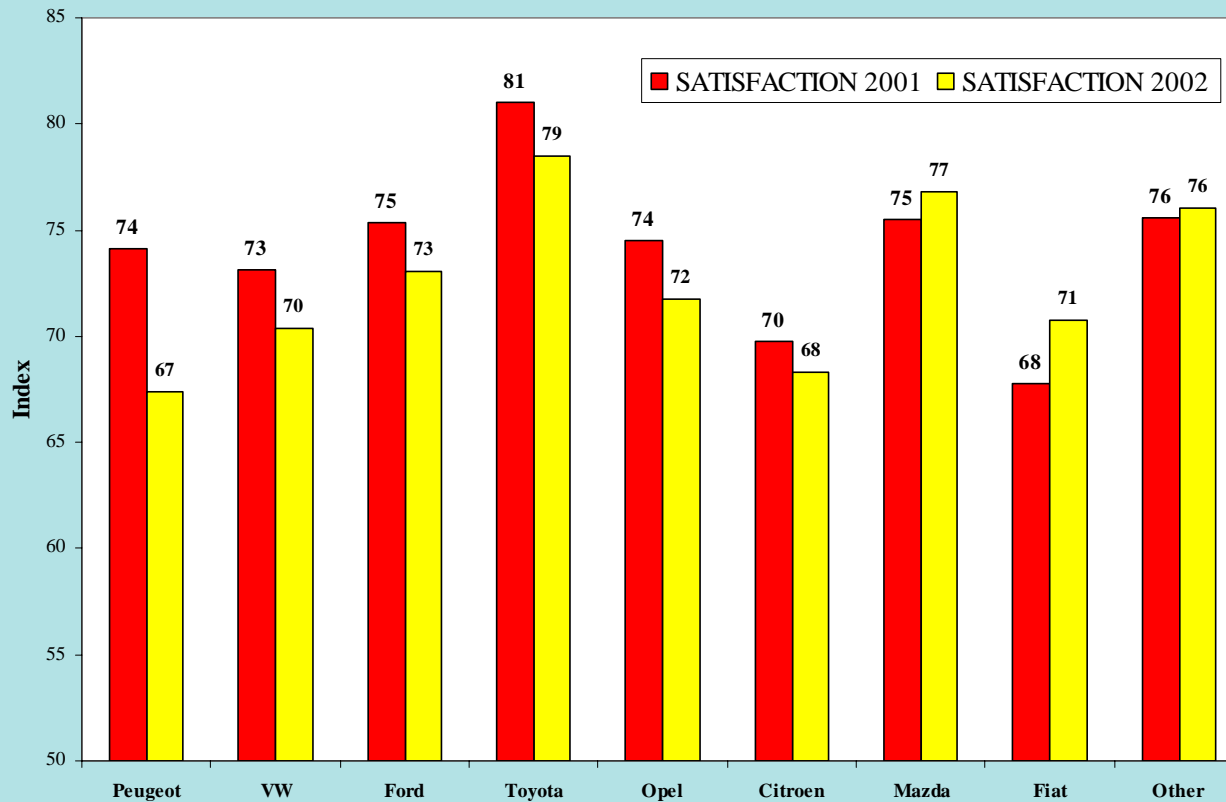
## Examples of manifest measurements

- **Image:** General perception of company image with regard to:
  - Reliability
  - Being customer focussed
  - Giving value for money
  - Innovation in products and services
  - Overall image
- **Satisfaction:**
  - Overall satisfaction
  - Comparison to ideal
  - Disconfirmation
- **Loyalty:**
  - The customer's intention to repurchase,
  - Intention of cross buying (buy another product from the same company),
  - Intention to recommend the brand/company to other consumers.

# Example from the Danish car industry



# Individual brands



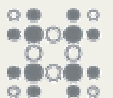
# Inner coefficients for the 2002 model

## UNSTANDARDISED INNER COEFFICIENTS

	IMAGE	EXPECTA- TIONS	QUALITY OF "HARDWARE"	QUALITY OF "HUMAN WARE"	VALUE	SATISFAC- TION	LOYALTY
IMAGE							
EXPECTATIONS							
QUALITY OF "HARDWARE"							
QUALITY OF "HUMAN WARE"							
VALUE	0,36	-0,06	0,35	0,32			
SATISFACTION	0,44	-0,03	0,23				
LOYALTY	0,29	-0,11	0,24	0,27		0,53	

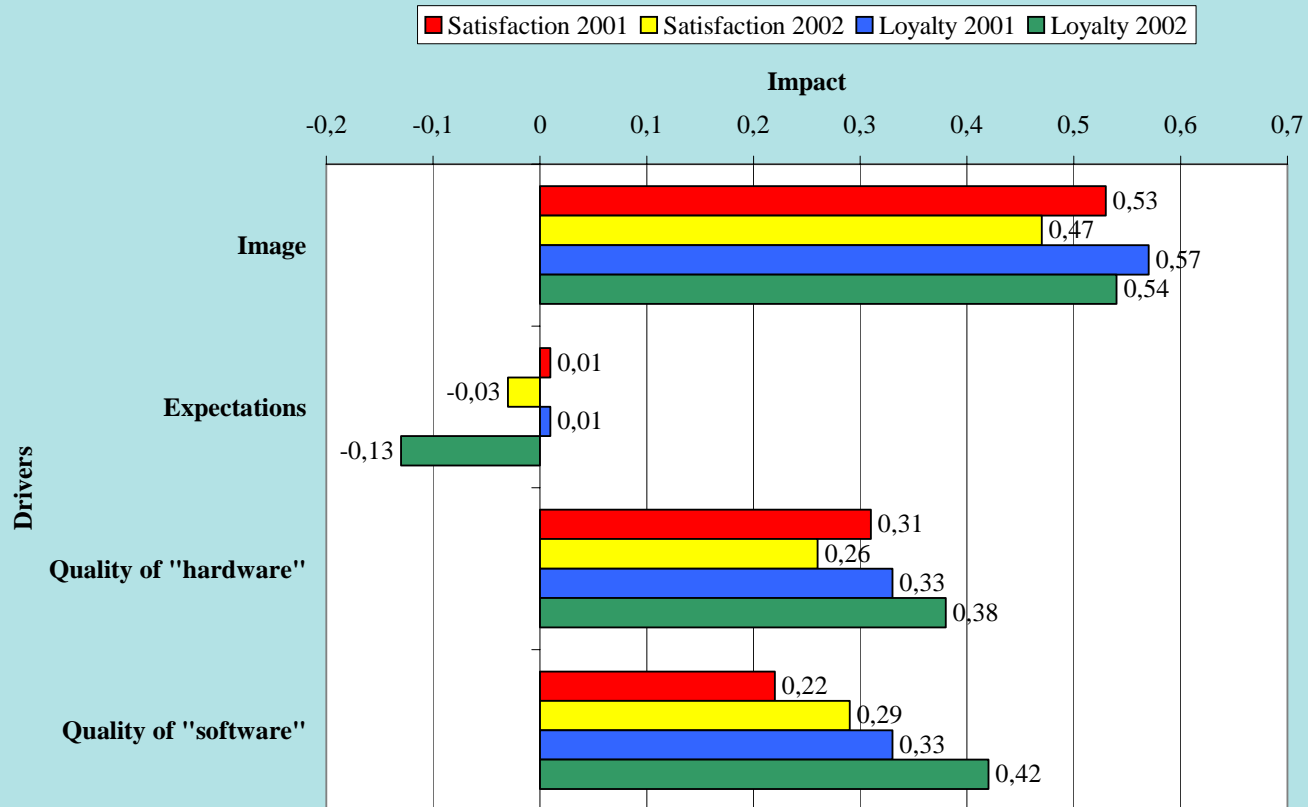
## T-VALUES FOR INNER COEFFICIENTS

	IMAGE	EXPECTA- TIONS	QUALITY OF "HARDWARE"	QUALITY OF "HUMAN WARE"	VALUE	SATISFAC- TION	LOYALTY
IMAGE							
EXPECTATIONS							
QUALITY OF "HARDWARE"							
QUALITY OF "HUMAN WARE"							
VALUE	14,72	-3,38	11,30	11,21			
SATISFACTION	20,93	-1,80	9,06	10,82	5,54		
LOYALTY	7,10	-4,08	5,22	6,10		13,84	

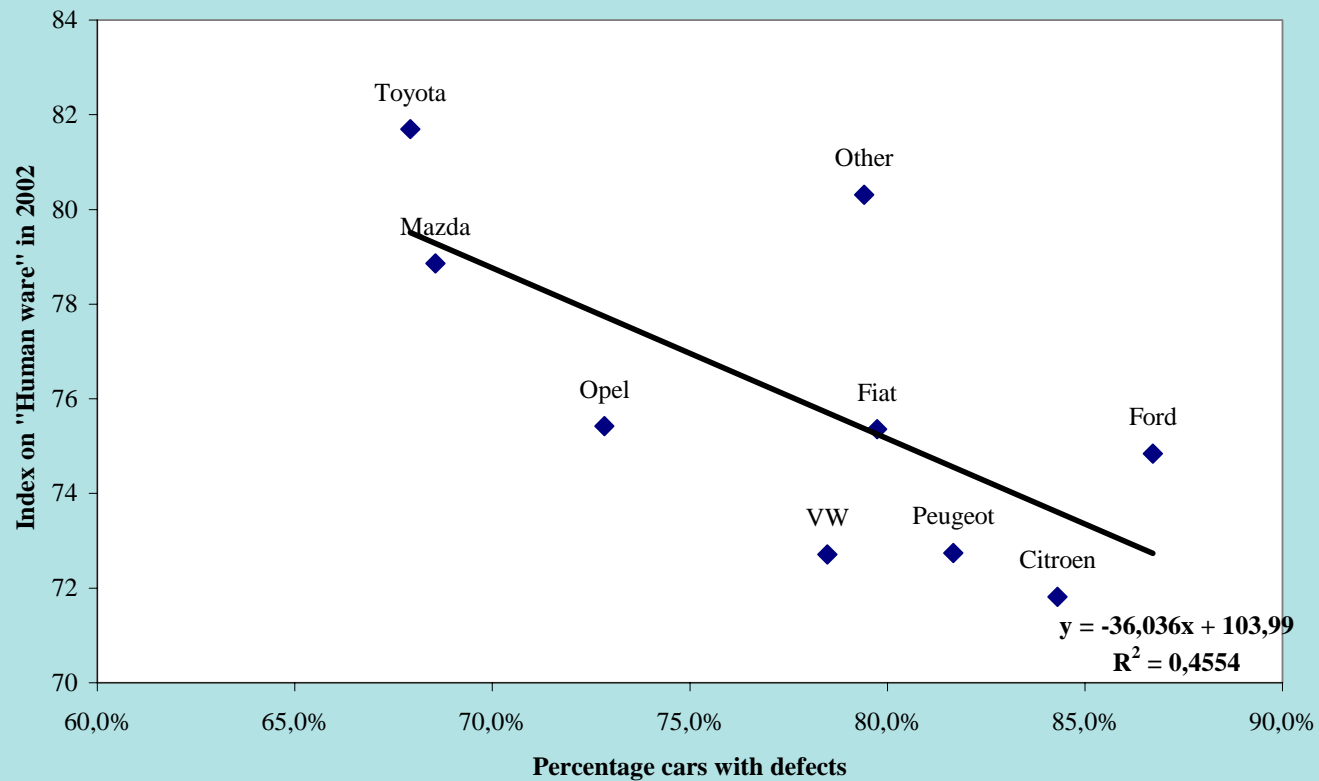




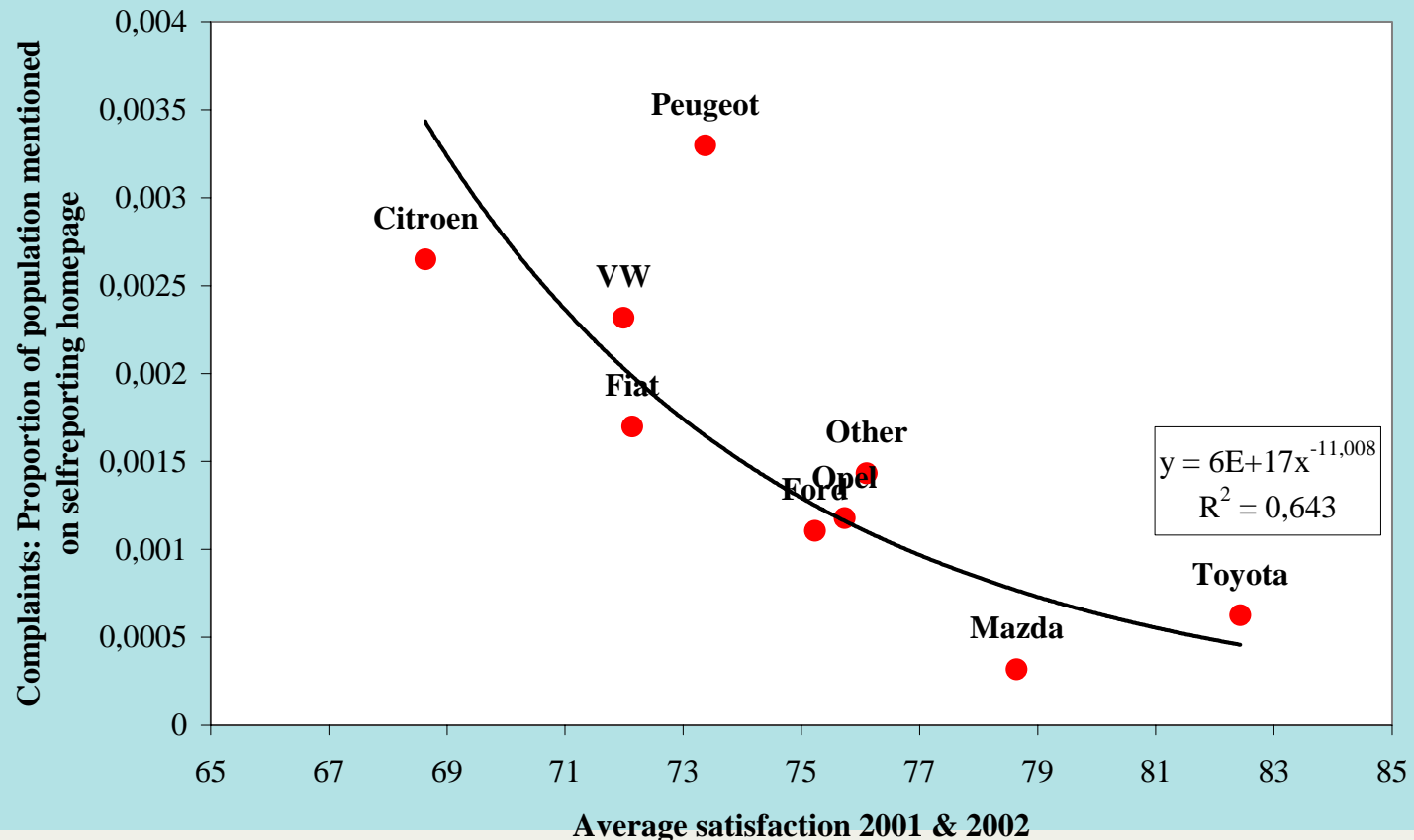
# The impact of drivers on satisfaction and loyalty 2001 & 2002



## External validity: Relation to actual service performance



# External validity: Relationship between satisfaction and complaints



# Practical problems and observations

## The Choice of Scale



## The experiment

- In order to test the effect of scale choice on the results of customer satisfaction studies a controlled experiment was set up.
- Under totally identical conditions two samples were drawn from the population. The only difference between the samples was that in the first sample a 5-point scale was used and in the second a 10-point scale was used.
- The questionnaires were the standard customer satisfaction questionnaires used for a given company.
- The size of the samples was 545 for the 10-point scale and 563 for the 5-point scale.

# Mean value of latent variables

Variable	Data source		Significance, two sided
	Ten points	Five points	
	Mean	Mean	
Expectations	73,3	75,1	0,13
Products	64,2	64,3	0,88
Service	66,9	66,4	0,70
Value	54,4	54,4	0,96
Satisfaction	65,2	65,2	0,97
Loyalty	57,5	58,7	0,36
Image	63,6	64,0	0,74

## Conclusion: Mean values

- There is no significant difference between the mean values of the aggregate variables.
- This means that the choice of scale has no influence on the level of the customer satisfaction index or the loyalty index.

## Standard deviation of aggregate variables

Variable	Data source		Significance
	Ten points	Five points	
	Std Deviation	Std Deviation	
Expectations	19,2	20,1	0,476
Products	19,1	20,5	0,274
Service	21,2	23,4	0,014
Value	19,7	22,4	0,005
Satisfaction	19,3	21,5	0,013
Loyalty	21,7	23,6	0,054
Image	18,1	19,5	0,069

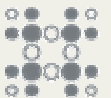


## Conclusion: Latent variable standard deviations

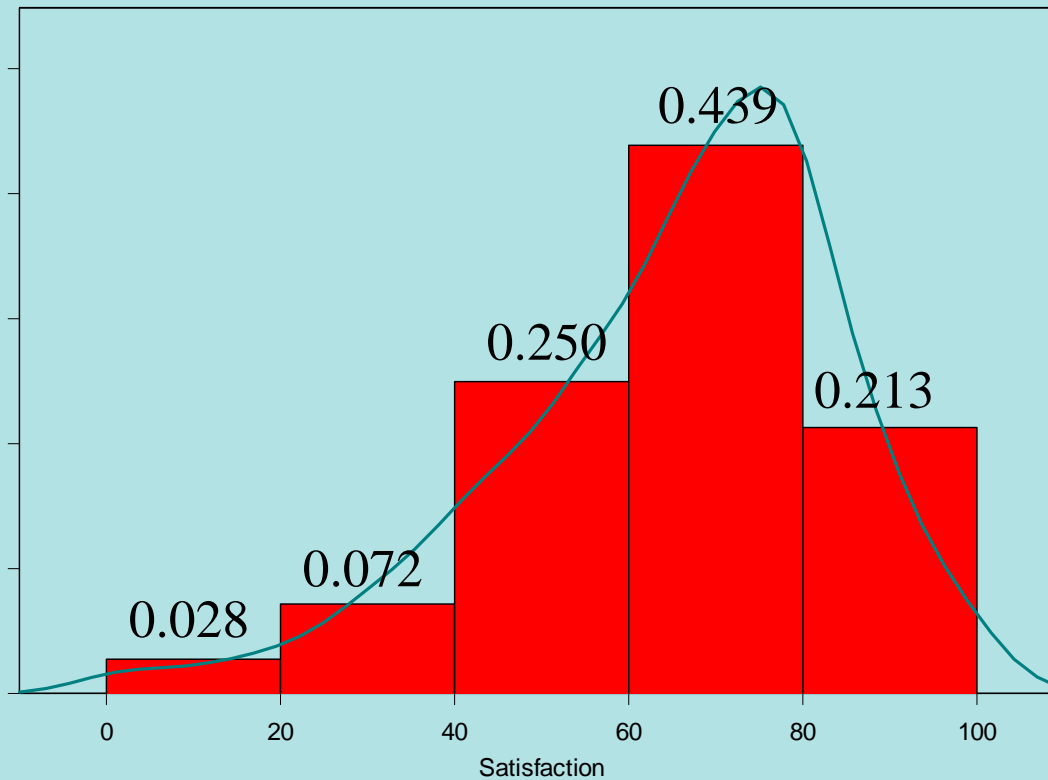
- As expected the standard deviation of the 10-point scale is smaller than the standard deviation of the 5-point scale with Image, Expectations and Products as possible exceptions.
- The difference is on the average app. 10%.
- The reason for this difference is, that the underlying distributions are discrete.

## 5- and 10-point scales

### Shape of the distribution

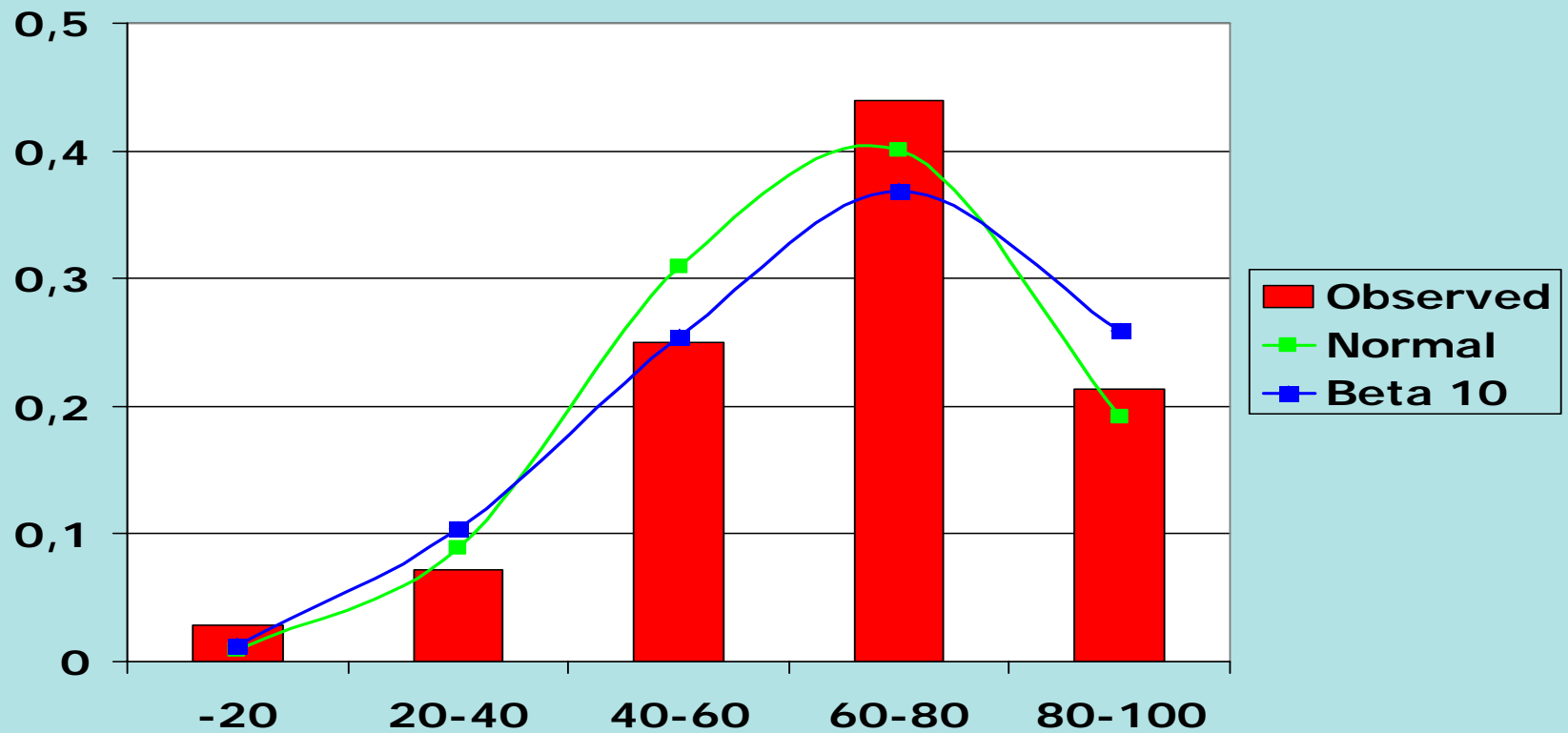


## Satisfaction: Distribution 10 point scale

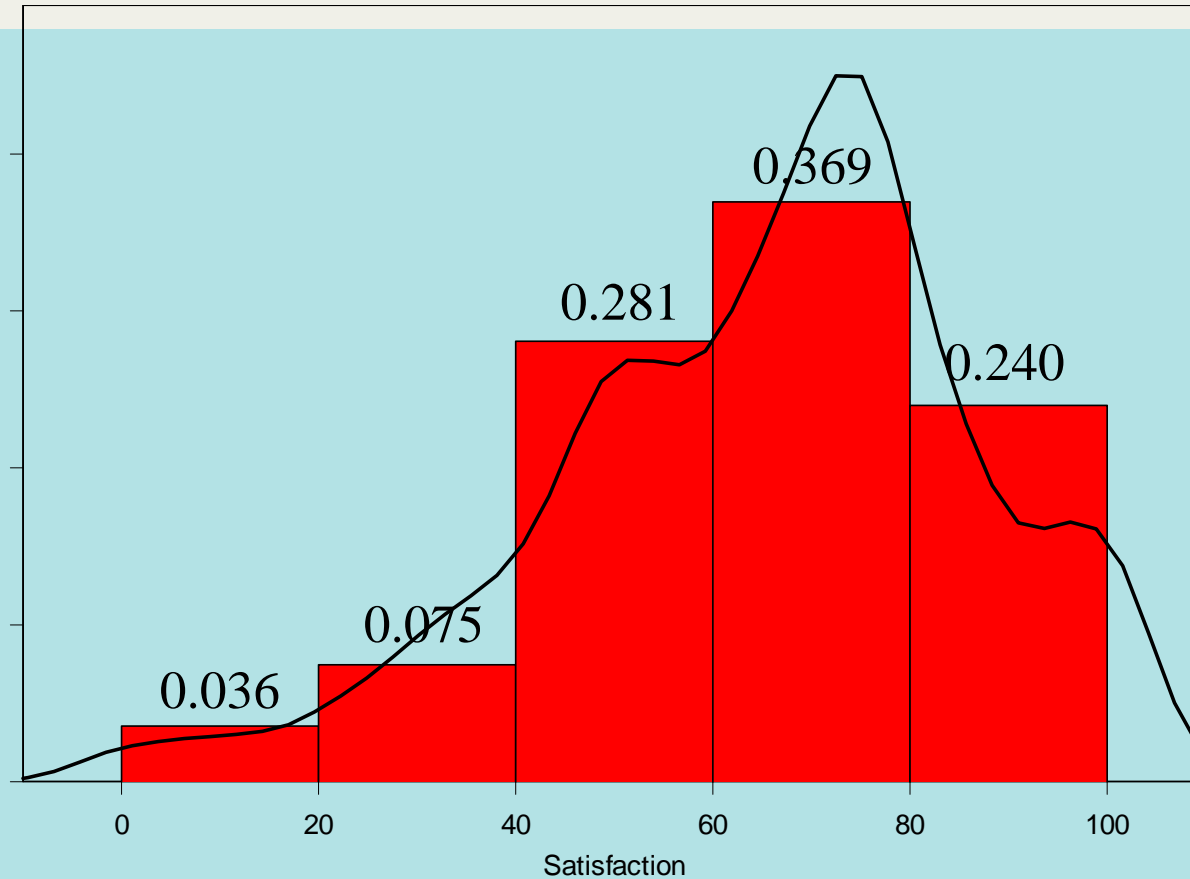


Mean: 65.2  
Std. dev.: 19.2

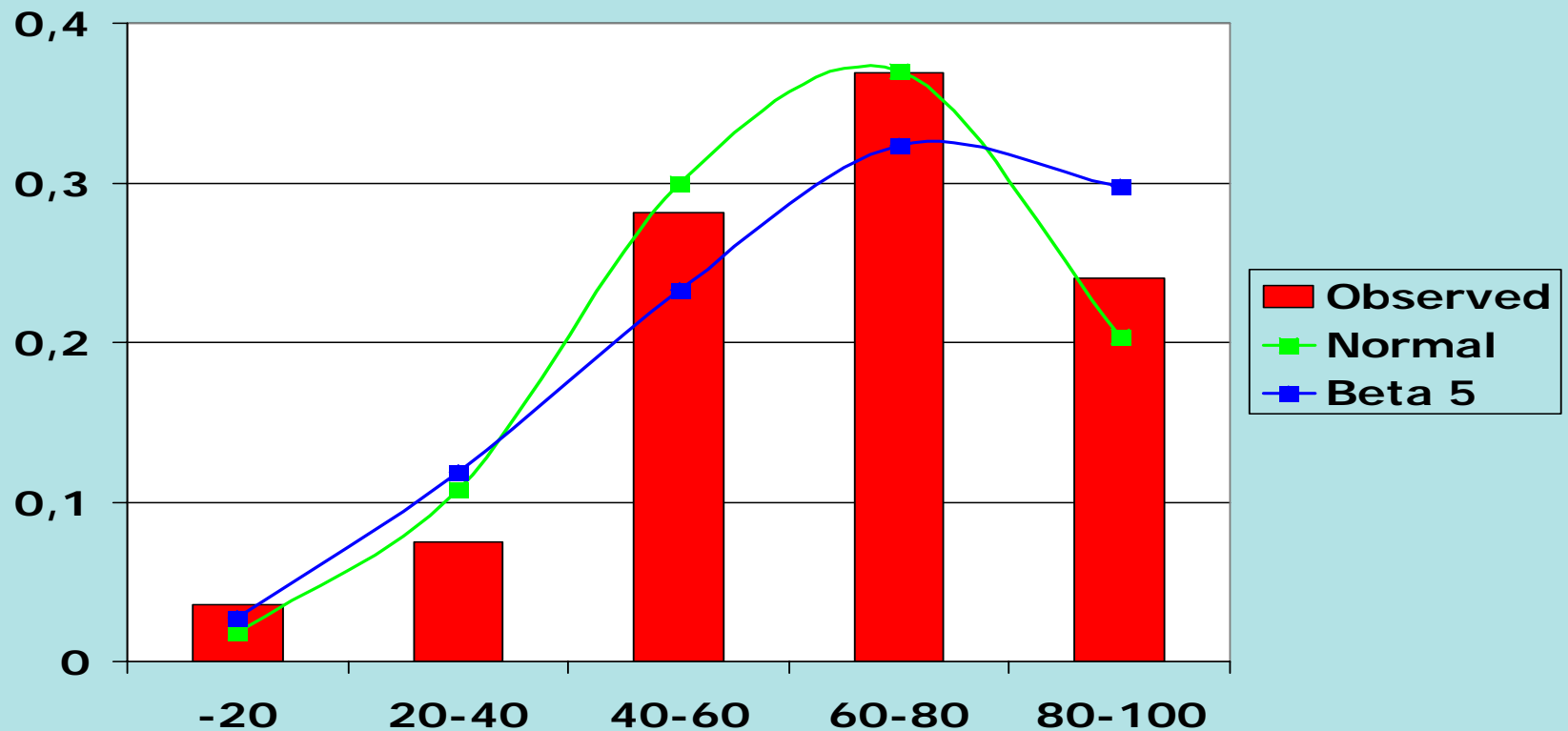
# Comparison of observed and theoretical distributions. (10-point scale)



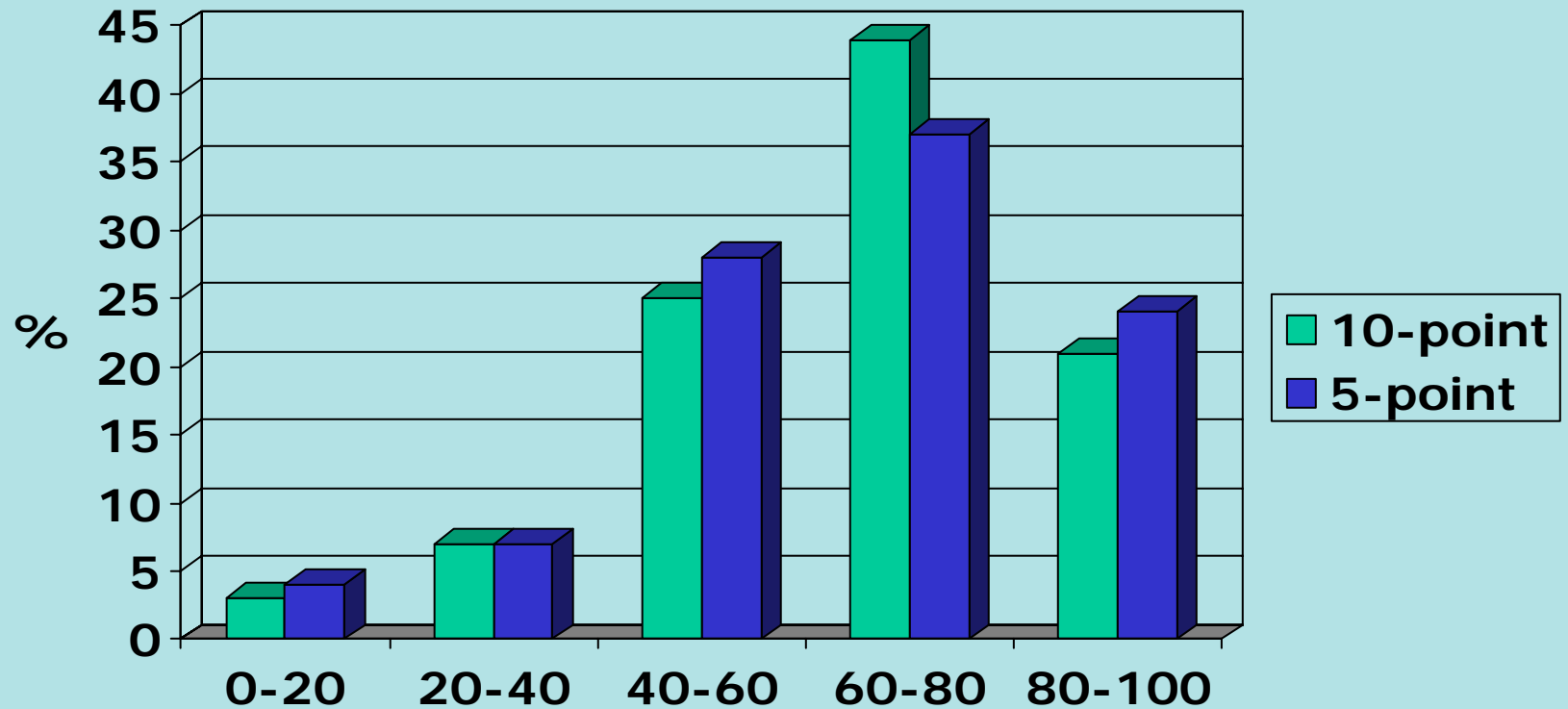
# Satisfaction: Distribution 5 point scale



## Comparison of observed and theoretical distributions. (5-point scale)



# A comparison of satisfaction distributions



## Satisfaction: A general comparison of the distribution of 5- and 10-point scales

	Kolmogorov-Smirnov Z	Asymp. Sig. (2-tailed)
Image	1,08	0,19
Expectations	2,53	0,00
Products	1,41	0,04
Service	1,63	0,01
Value	1,95	0,00
Satisfaction	1,77	0,00
Loyalty	1,53	0,02

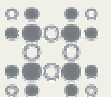


## Conclusion

- In general the standardized distributions are not identical with Image as a possible exception. This is to be expected due to the discrete underlying distributions.
- The beta distribution or the doubly truncated normal distribution seem to give the closest approximation to the distribution but even here we have a significant difference in both cases.

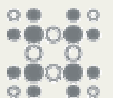
## 5- and 10- point scales

Are demographics and scale interacting?



# Variables and factors for the analysis of variance

- Dependent variables:
  - All aggregate variables
- Explanatory variables
  - Data Source (5-point, 10-point)
  - Age (-25, 26-35, 36-45, 46-55, 56-65, 66-)
  - Education ( 8 groups from high school to university)
  - Gender (Male, Female)
  - Location (Copenhagen, Sealand, Funen, Jutland)



## Analysis of variance (10% significance)

	Image	Expectation	Product	Service	Value	Satisfaction	Loyalty
Main	Age Location Education Gender	Location	Location	Age Location Education Gender	NONE	Age Location Education	Location Education
Two-way interaction	NONE	Age x source	NONE	NONE	Age x source	NONE	NONE

## Conclusion

- In the case of Image, Product, Service, Satisfaction and Loyalty there is no effect from the data source.
- Only in the case of Expectation and Value we can trace an effect. In these cases there is a tendency that the age groups are using the scales differently.
- Based on this our general conclusion is, that the demographic interpretation of customer satisfaction studies will not be seriously affected by the choice of scale.
- When it comes to satisfaction there seems to be a universal main effect of Age, Location and Education.
  - Satisfaction is increasing with age and decreasing with education.
  - Satisfaction is decreasing with the degree of urbanization.

## Conclusions concerning scales

- In general terms a 10-point scale is preferable to a five point scale:
  - Smaller variance.
  - Closer approximation to a continuous variable.
  - 10-point scales are used by all the major national customer satisfaction studies.
- In general it is possible to compare studies using 5 and 10-point scales since the mean values (on a 100-point scale) are not affected.
- Demographics have a small but not very important effect on the results from the scales.

# Practical problems and observations

Reliability and prediction



# Internal reliability and validity of the results: The car example

		IMAGE	EXPECTATIONS	QUALITY OF "HARDWARE"	QUALITY OF "HUMANWARE"	VALUE	SATISFACTION	LOYALTY
2001	R-SQUARE FOR LATENT VARIABLES					0,56	0,76	0,55
	COMPOSITE RELIABILITY	0,92	0,95	0,89	0,86	0,97	0,90	0,91
	AVERAGE VARIANCE EXPLAINED BY LATENT VARIABLES	0,69	0,86	0,73	0,68	0,91	0,75	0,84
2002	R-SQUARE FOR LATENT VARIABLES					0,62	0,72	0,56
	COMPOSITE RELIABILITY	0,92	0,94	0,91	0,90	0,97	0,90	0,91
	AVERAGE VARIANCE EXPLAINED BY LATENT VARIABLES	0,71	0,84	0,77	0,75	0,91	0,74	0,84

$$AVE = \frac{\sum_{i=1}^p \lambda_i^2}{\sum_{i=1}^p (\lambda_i^2 + \Theta_i)}$$

$$\rho_c = \frac{\left( \sum_{i=1}^p \lambda_i \right)^2}{\left( \sum_{i=1}^p \lambda_i \right)^2 + \sum_{i=1}^p \Theta_i}$$



# Reliability and choice of manifests

- Automobiles
  - Reasonable reliability: No reason for changes.
- Petrol stations
  - High reliability: No reason for changes.
- Banks:
  - In the satisfaction construct the “comparison to ideal” may cause a problem. Much lower level than the two other questions.
- Supermarkets
  - In the satisfaction construct the “comparison to ideal” may cause a problem. Much lower level than the two other questions.
  - The value for money indicator and the assortment indicator may cause a problem since they reflect the type of supermarket.
  - The question about opening hours which is classified as belonging to the service block should possibly be re-classified



## Reliability and choice of manifests: Conclusions

- For most of the areas covered by the Danish Customer Satisfaction Index the manifest questions are working well.
- The only area where we have observed a necessity for changes is Supermarkets.
- Other conclusions may apply when we have discussed the problem of missing values.

## Explanatory power

- The general observation is, that the explanatory power of the model is rather good.
- There is no problem in obtaining an  $R^2$  beyond .65 for the satisfaction construct as required by the ECSI Technical Committee. In general  $R^2$  is somewhere between .70 and .80.
- The degree of explanation for value and loyalty is usually a little lower.

# Practical problems and observations

Missing values and  
multicollinearity



# Missing values

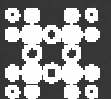
Supermarkets	Banks	Automobiles	Petrol Stations
<ul style="list-style-type: none"><li>• <i>Below 10% for all items</i></li></ul>	<ul style="list-style-type: none"><li>• <i>Relative comparisons are problematic. 40-50% missing values</i></li></ul>	<ul style="list-style-type: none"><li>• <i>19 out of 22 items have missing values below 5%</i></li></ul>	<ul style="list-style-type: none"><li>• <i>13 out of 22 have missing values below 10%.</i></li><li>• <i>8 have missing values between 10% and 20%</i></li></ul>

## Multicollinearity (Latent variables)

- In general the degree of multicollinearity is rather high.
  - Banks: Correlations between .54 (expectation and service) and .82 (*product and service*).
  - Petrol stations: Correlations between .42 (expectations and service) and .69 (*product and service*).
  - Automobiles: Correlations between .48 (expectations and service) and .85 (*product and service*).
  - Mobile telephones: Correlations between .44 (expectations and service) and .76 (*product and service*).
  - Supermarkets: Correlations between .52 (expectations and image) and .71 (image and product).

# Simulation study

A study of some of the implications of the empirical findings



# Background

- To get insight into the consequences of some of the empirical problems based on a true model which is very close to the actual models observed. Our model is reflective for all latent variables.
- To formulate some simple rules of thumb.
- To supplement and verify the simulation study conducted by Cassel, Hackl and Westlund (1999, 2000). These authors investigated the effect of the following factors on the estimation of an EPSI like model with formative exogenous and reflective endogenous latent variables:
  - Skewness of manifest variables
  - Multicollinearity between latent variables
  - Misspecification (omission of relevant regressors or regressands, or manifests within a measurement model)
  - Sample size
  - Size of the path coefficients



## Simulation setup

- STAGE 1(Screening): Orthogonal main effect plan with 7 factors in 27 runs with 25 replications for each run. Each replication has a number of observations varying between 50 and 1000.
  - Exogenous distribution (Beta vs. Normal)
  - **Multicollinearity between latent exogenous variables**
  - Indicator validity (bias)
  - **Indicator reliability (standard deviation within a block)**
  - Structural model specification error
  - **Sample size**
  - **Number of indicators in each block**
- STAGE 2: Full factorial design with 4 factors in 54 runs with 25 replications for each run
  - Multicollinearity, reliability, sample size and number of indicators

## Stage 2 factor levels and response variables

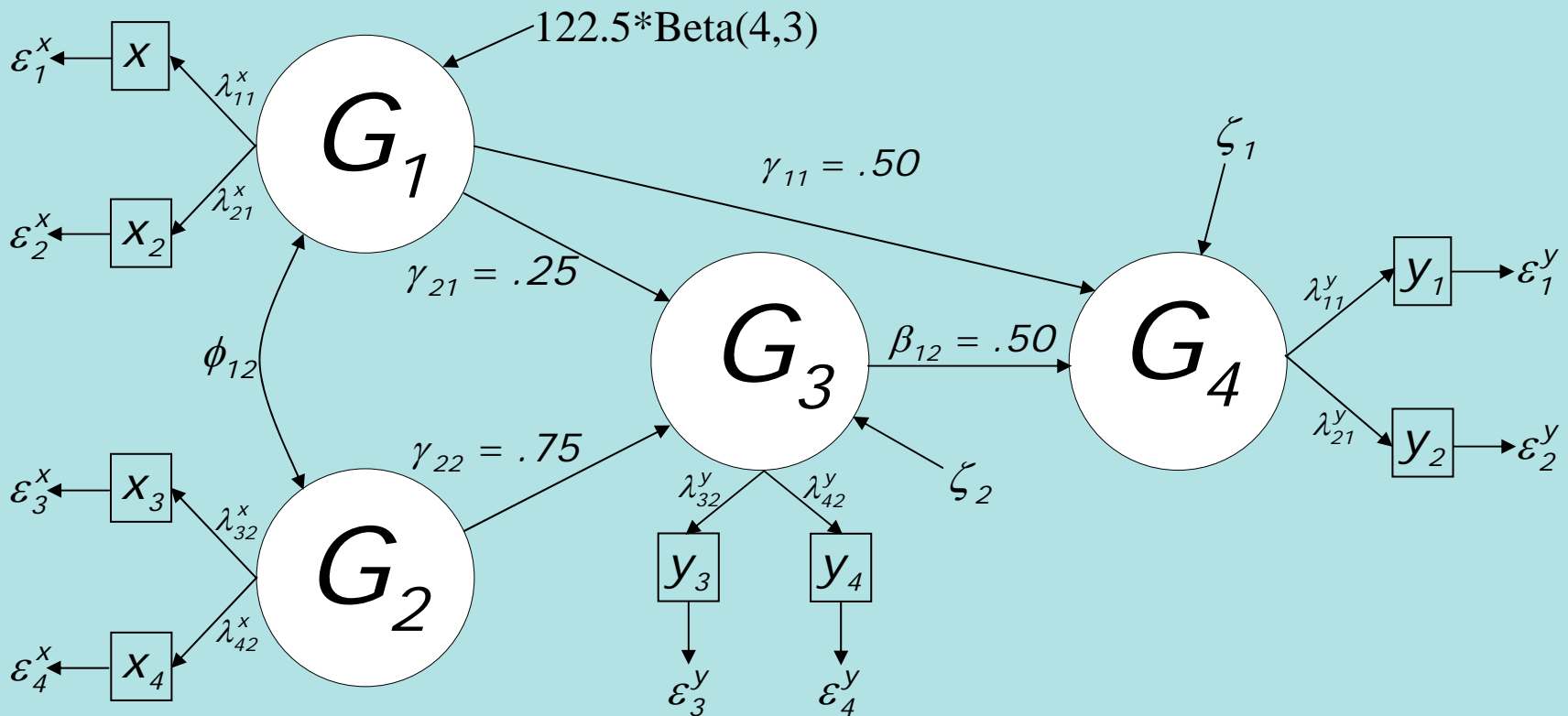
- **FACTOR LEVELS:**

- Multicollinearity:  $\rho = \{0.2; 0.8\}$ .
- Reliability:  $\sigma = \{1; 10; 20\}$ .
- Sample size:  $n = \{50; 250; 1000\}$ .
- Number of indicators:  $p = \{2; 4; 6\}$ .

- **RESPONSE VARIABLES:**

- Absolute bias for indices
- Standard deviation for indices
- Bias for path coefficients
- $R^2$ , AVE and RMSE.

# The simulation model (A simplified customer satisfaction model)

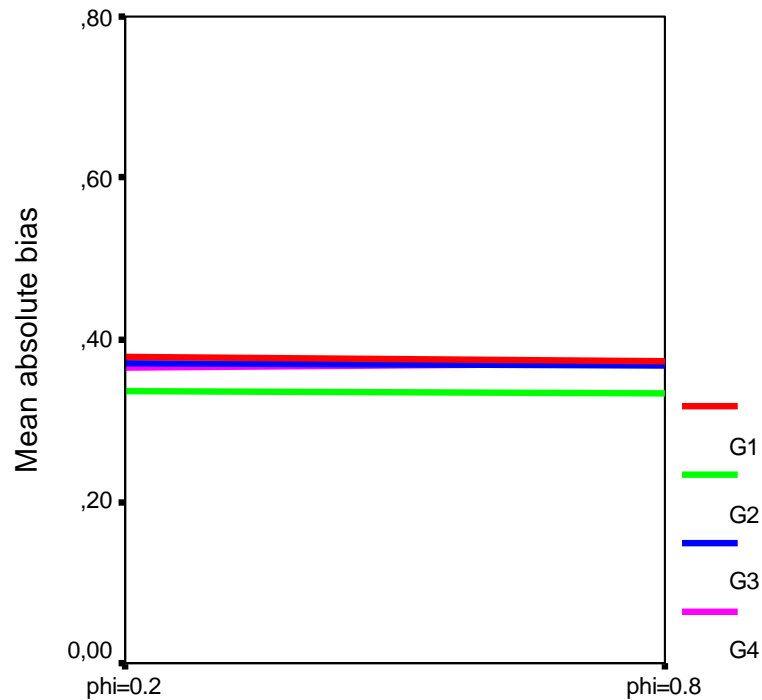


# Simulation results

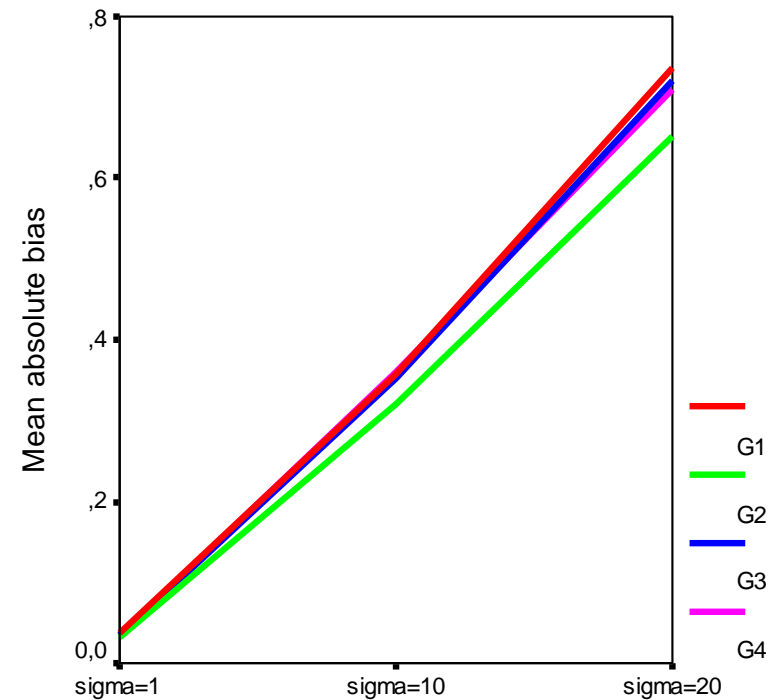
	F1	F2	F3	F4						
	Multicollinearity	Indicator reliability	Sample size	# indicators	F1*F2	F1*F3	F1*F4	F2*F3	F2*F4	F3*F4
g1		**	**	**				**	**	**
g2		**	**	**				**	**	**
g3		**	**	**				**	**	**
g4		**	**	**				**	**	**
stdg1		**	**	**					**	**
stdg2		**		**			**		**	**
stdg3	**	**		**	**				**	**
stdg4	**	**	*	**	**				**	
gamma21	**	**		**	**	*	**		**	
gamma22	**	**		**	**				**	
gamma11	**	**		**	**		**		**	
beta12	**	**		**	**		**		**	
stdgam21	**	**	**	**	**	**		**	**	**
stdgam22	**	**	**	**	**	**		**	**	**
stdgam11	**	**	**	**	**	**		**	**	**
stdbet12	**	**	**	**	**	**		**	**	**
rsq1		**		**	**				**	
rsq2	**	**		**	**				**	
ave1		**		**					**	
ave2		**		**					*	
ave3		**		**	*				*	
ave4		**		**	**				*	
avetot		**		**					*	
rmse	**	**	**	**	**			**	**	**



# Bias of indices: Multicollinearity and indicator reliability

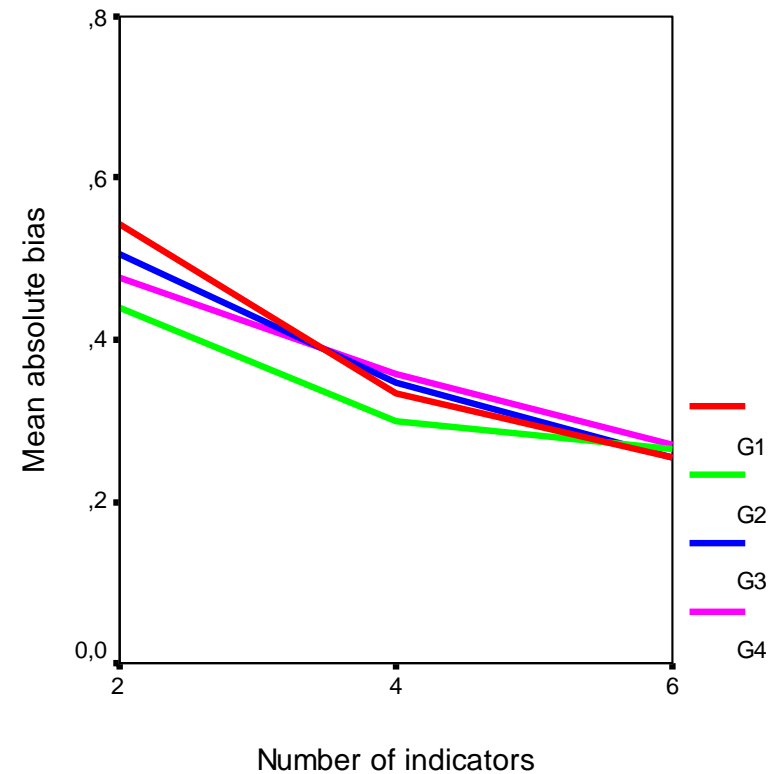
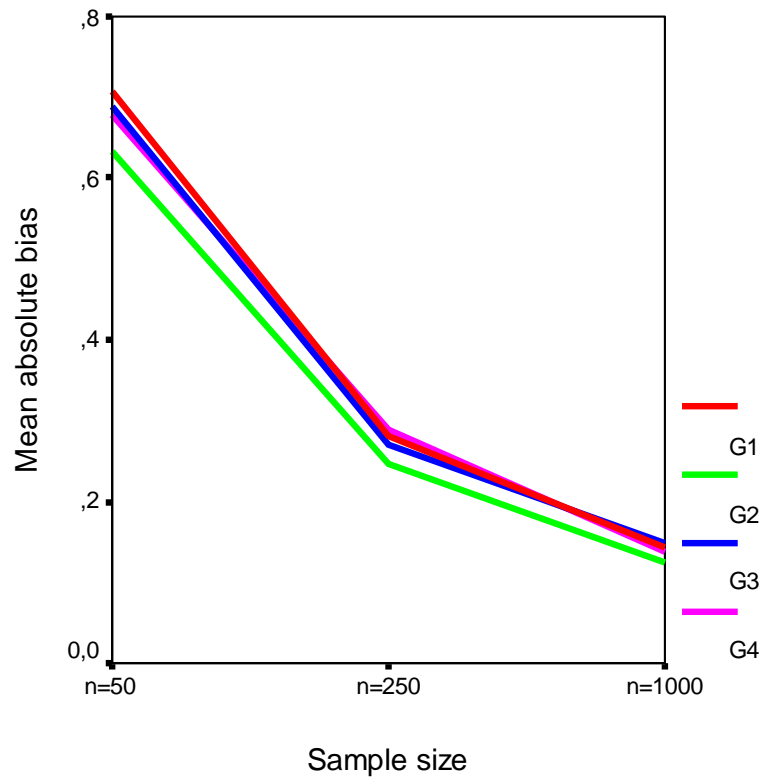


Multicollinearity

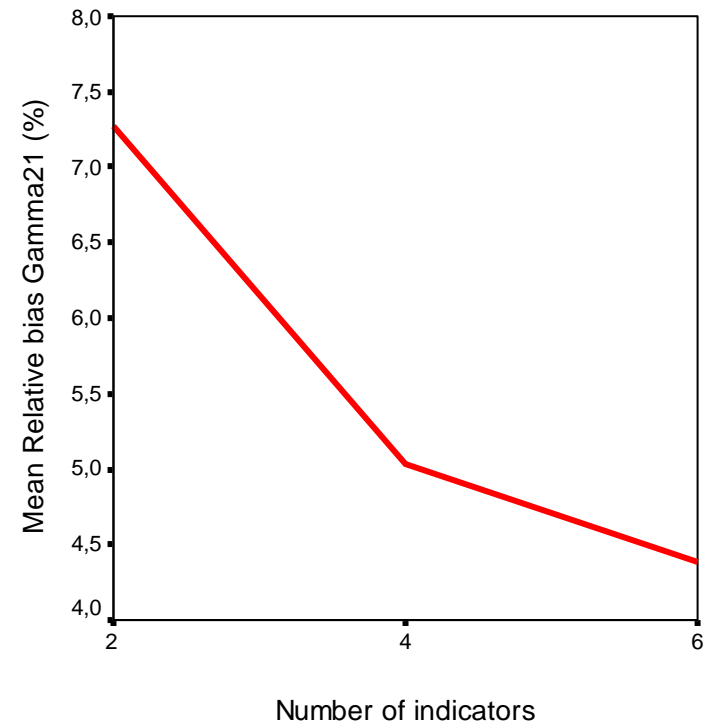
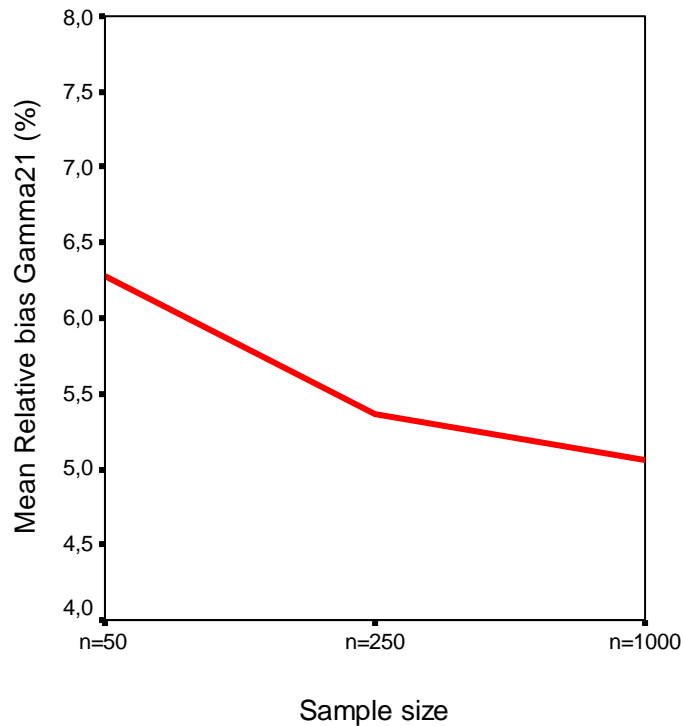


Indicator reliability

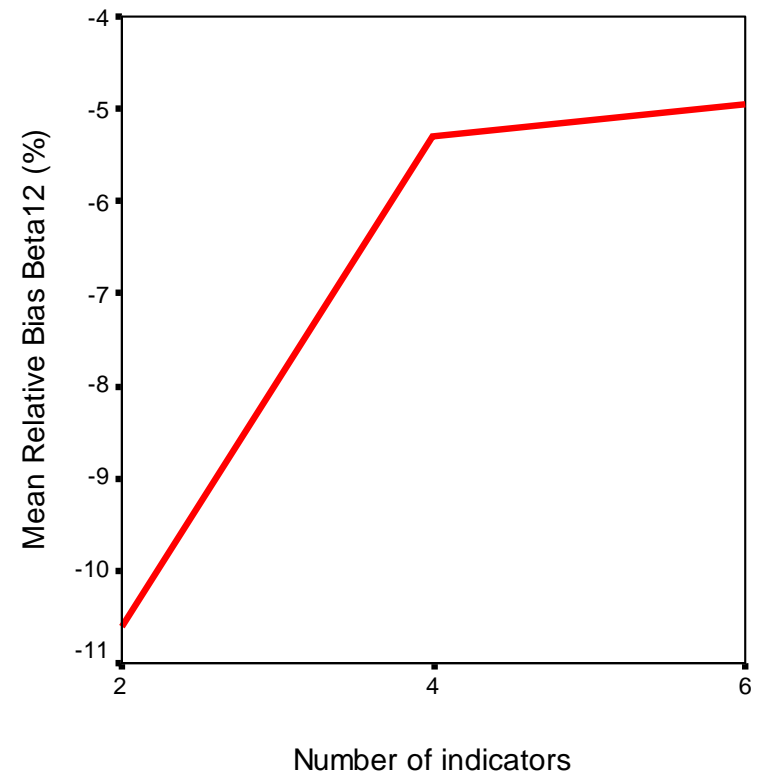
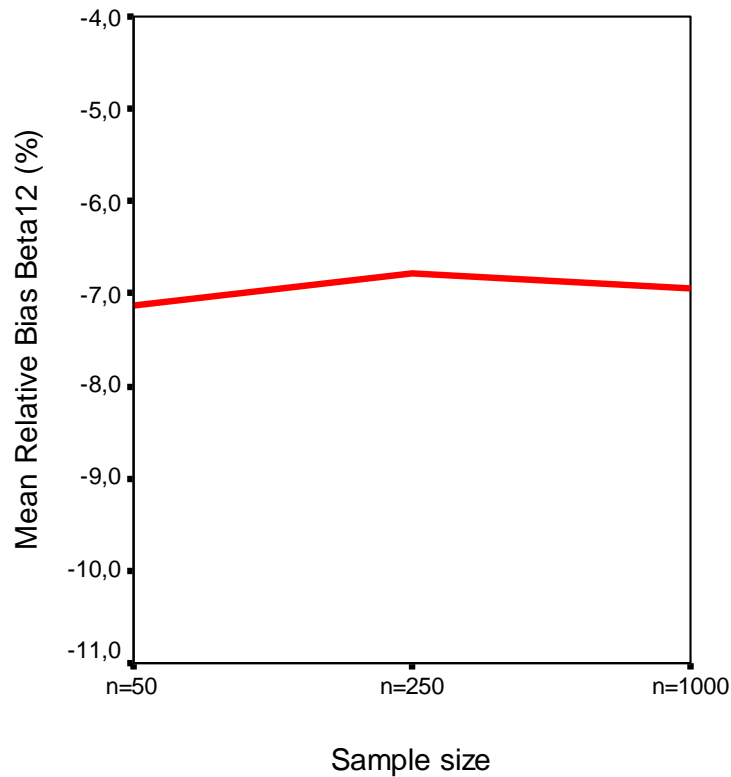
# Bias of indices: Sample size and number of indicators



# Example of mean relative bias for Gamma 21

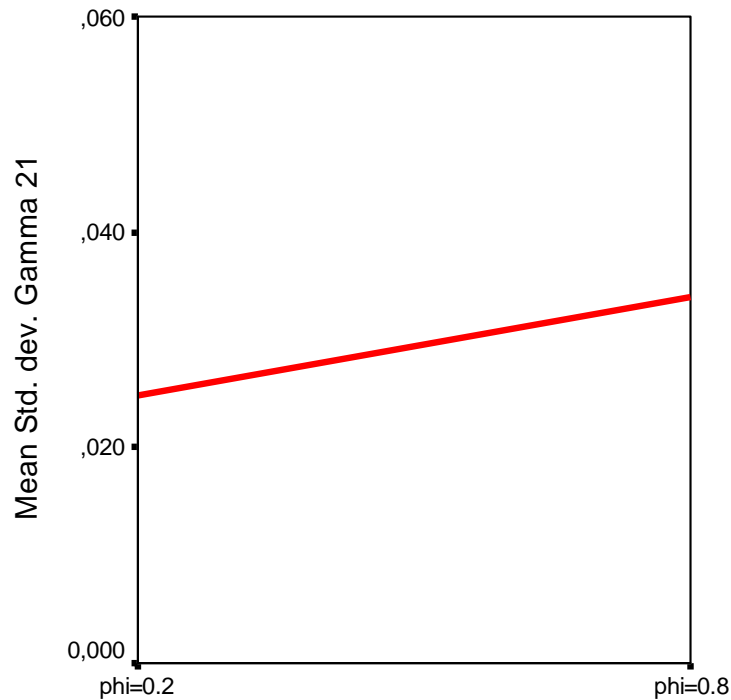


## Example of mean relative bias for Beta 12

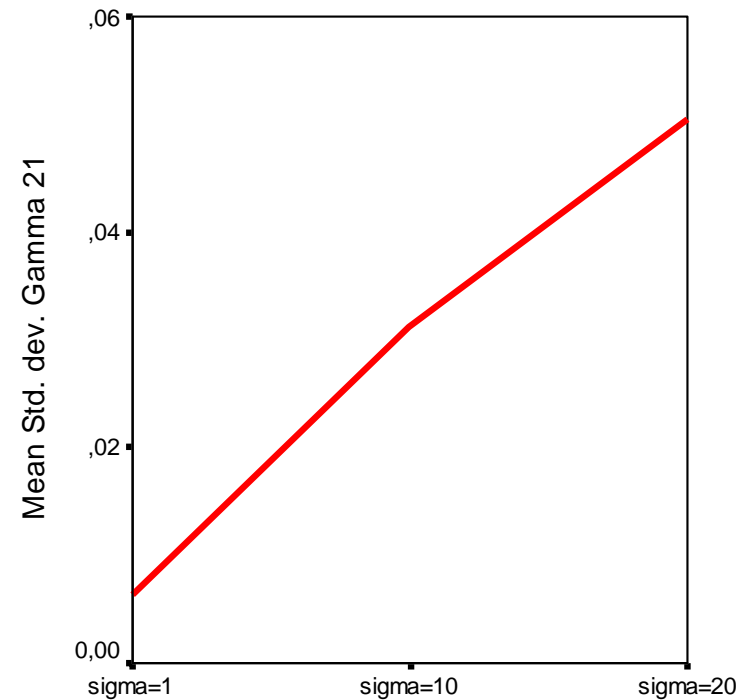




# Standard deviation of Gamma 21 as a function of multicollinearity and indicator reliability

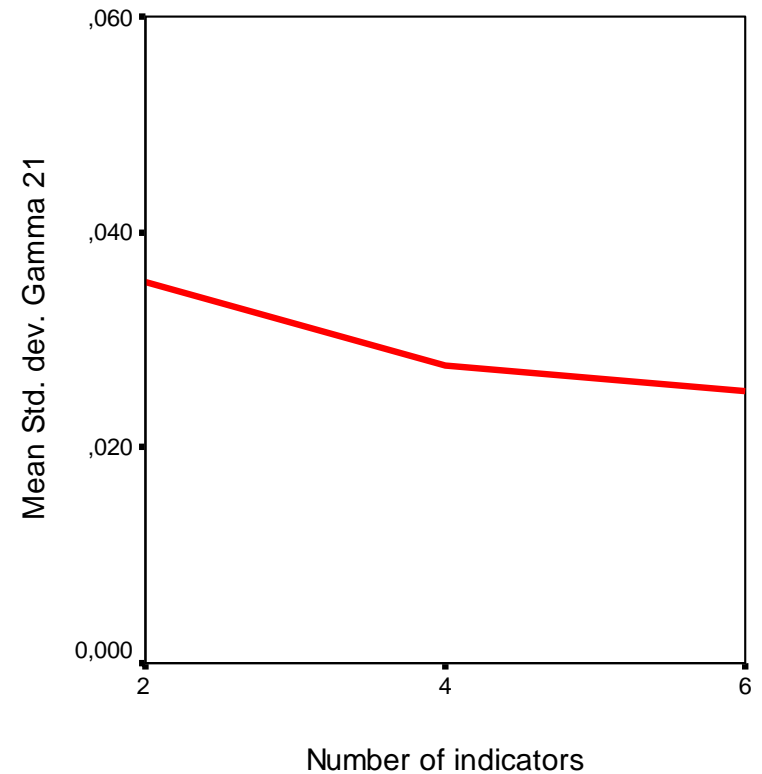
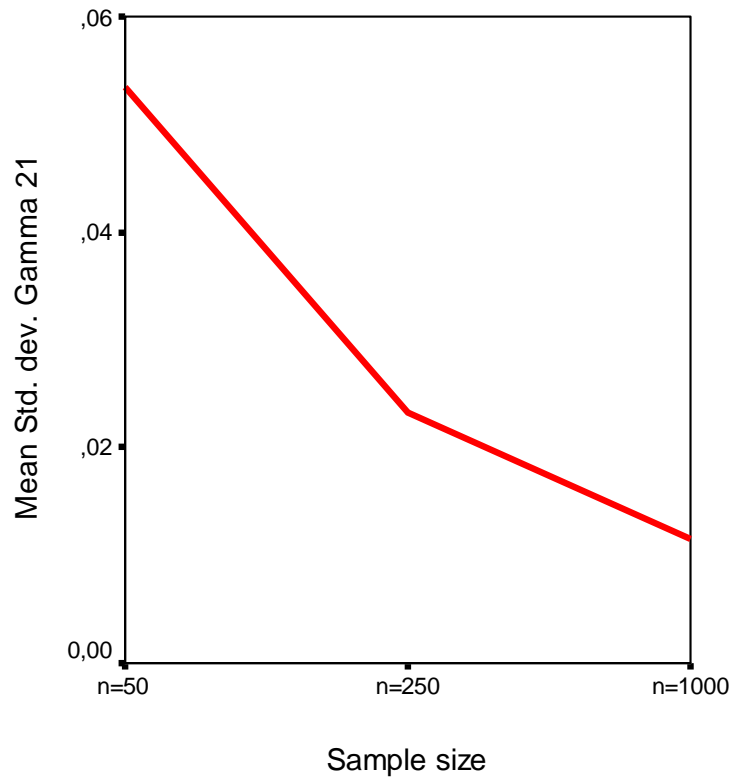


Multicollinearity

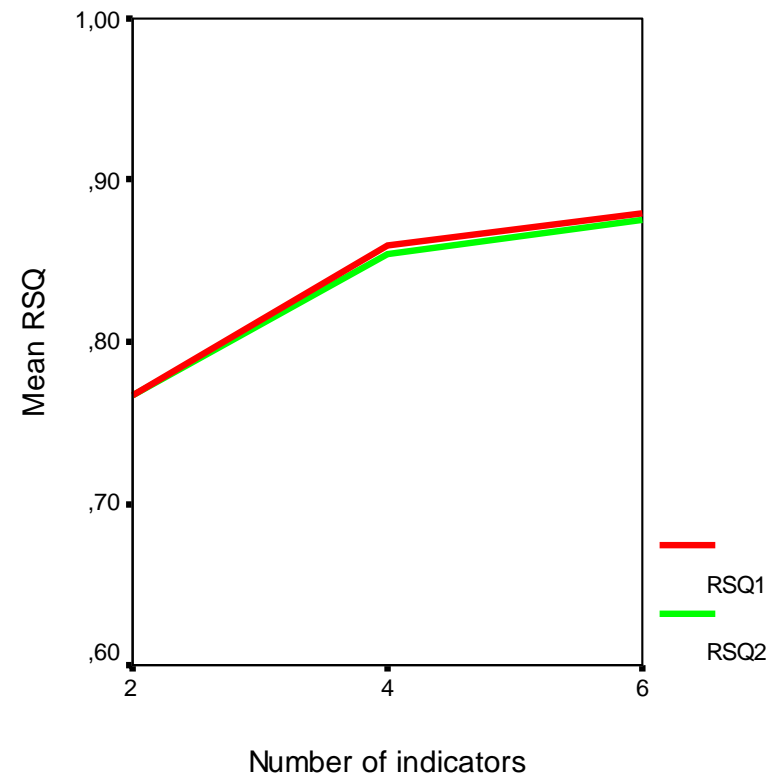
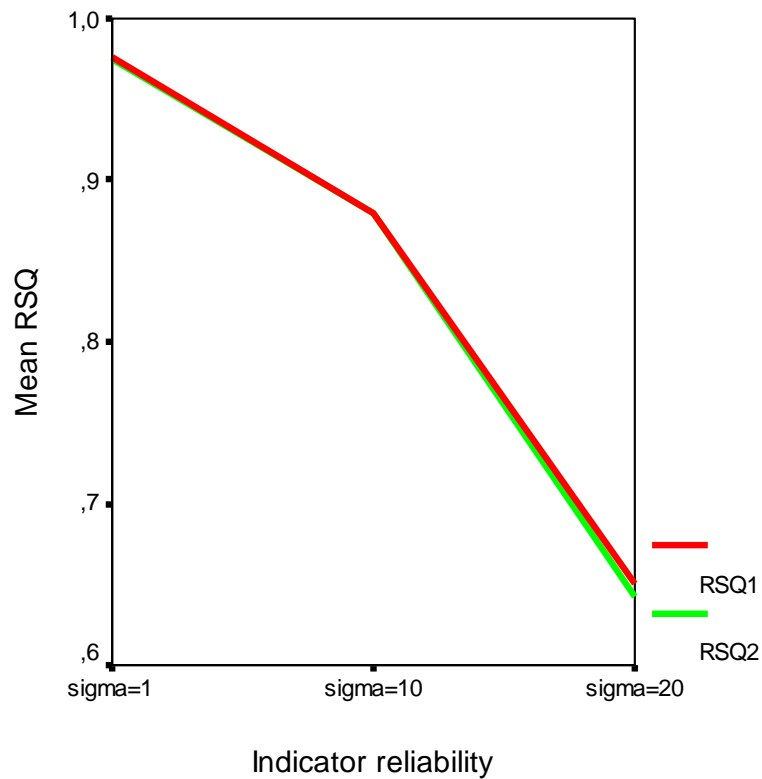


Indicator reliability

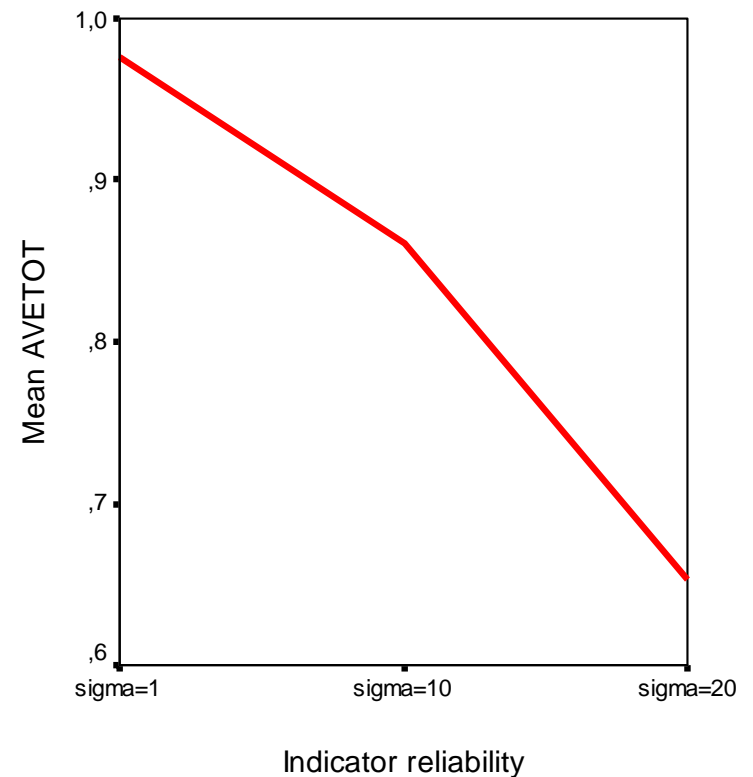
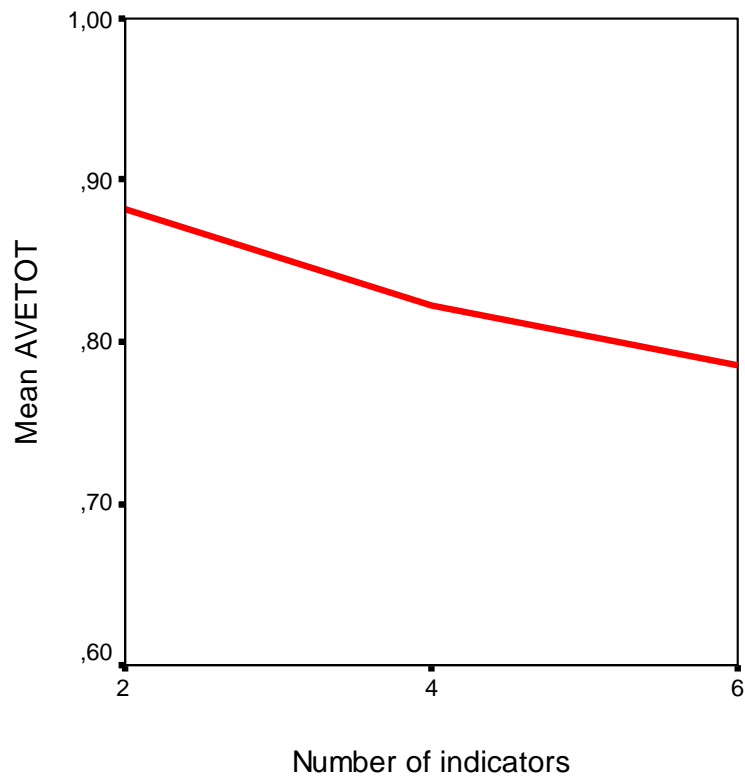
# Standard deviation of Gamma 21 as a function of sample size and number of indicators



# Degree of explanation.



# Average variance extracted.



## A couple of rough rules of thumb concerning the absolute bias of the indices

- Let  $\sigma$  be the standard deviation of the manifest variables,  $n$  the sample size, and  $p$  the number of indicators, then:
  - $\text{BIAS}(k\sigma, n, p) = k \text{BIAS}(\sigma, n, p)$ .
  - $\text{BIAS}(\sigma, kn, p) = (1/\sqrt{k}) \text{BIAS}(\sigma, n, p)$ .
  - $\text{BIAS}(\sigma, n, kp) = (1/k) \text{BIAS}(\sigma, n, p)$ .

## Conclusion to the simulation study

- **Basically our results support The Cassel, Hackl, Westlund results where comparable:**
  - *Misspecification* is in general a serious problem with severe parameter bias.
  - *Skewness* of distribution is of minor importance to the PLS estimates.
  - *Multicollinearity* between the latent variables is without importance for the estimated indices. It has a significant but small impact on the bias of the path coefficients. It has a significant effect on all standard deviations.
  - *Size of the sample* has no influence on the bias of the path coefficients. It has a large effect on all standard deviations.
- **In addition:**
  - *Indicator reliability* has an enormous influence on all measured responses, i.e. bias, standard deviation and fit measures. Furthermore several cases of two-factor interaction with both *multicollinearity*, *sample size*, and the *number of indicators* were found.
  - Likewise the *number of indicators* has a strong impact on all responses, and also a strong two-factor interaction with *sample size* and *reliability*.

## General conclusion

- PLS provides reasonably robust estimates of a customer satisfaction index in a usual practical setting where the sample size is  $n=250$ , the standard deviation around  $\sigma=20$ , and the average multicollinearity around  $\rho=.60$ .
- In a usual practical setting the bias of the indices is low and usually not larger than .50 (on a 100 point scale).
- The parameter estimates are in general biased. The bias can be both positive and negative depending on the model structure. The relative bias will in a usual practical setting be in the area of 10-20%.