

SAMPLE SIZE AND NUMBER OF OBSERVATIONS

A balance between quality and producibility

Introduction

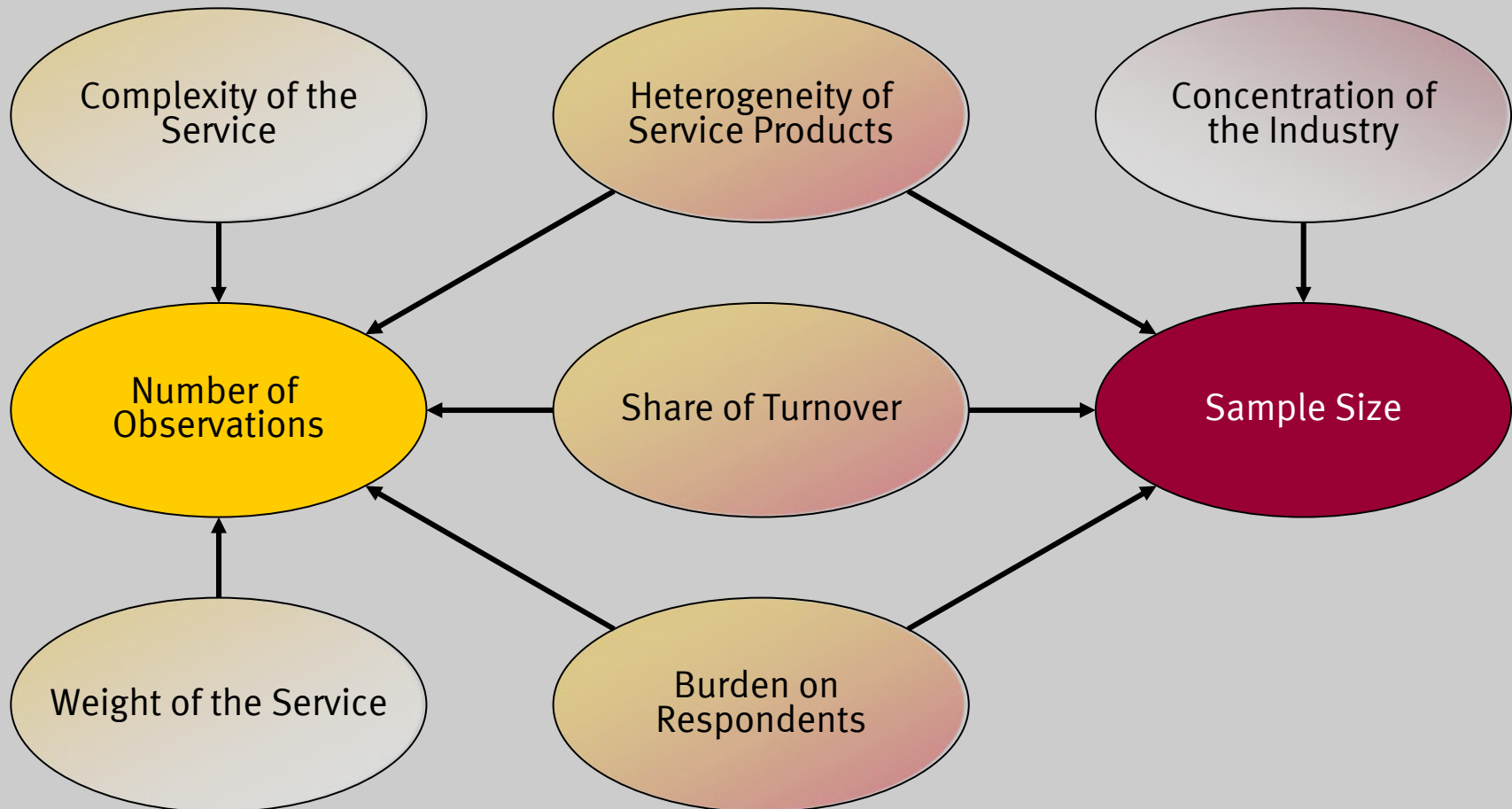
Key figures

- Number of indices: SPPIs for 21 industries
- covered turnover: 477 bn € in 2010
- number of respondents: 2.800
- number of collected prices: ca. 15.000

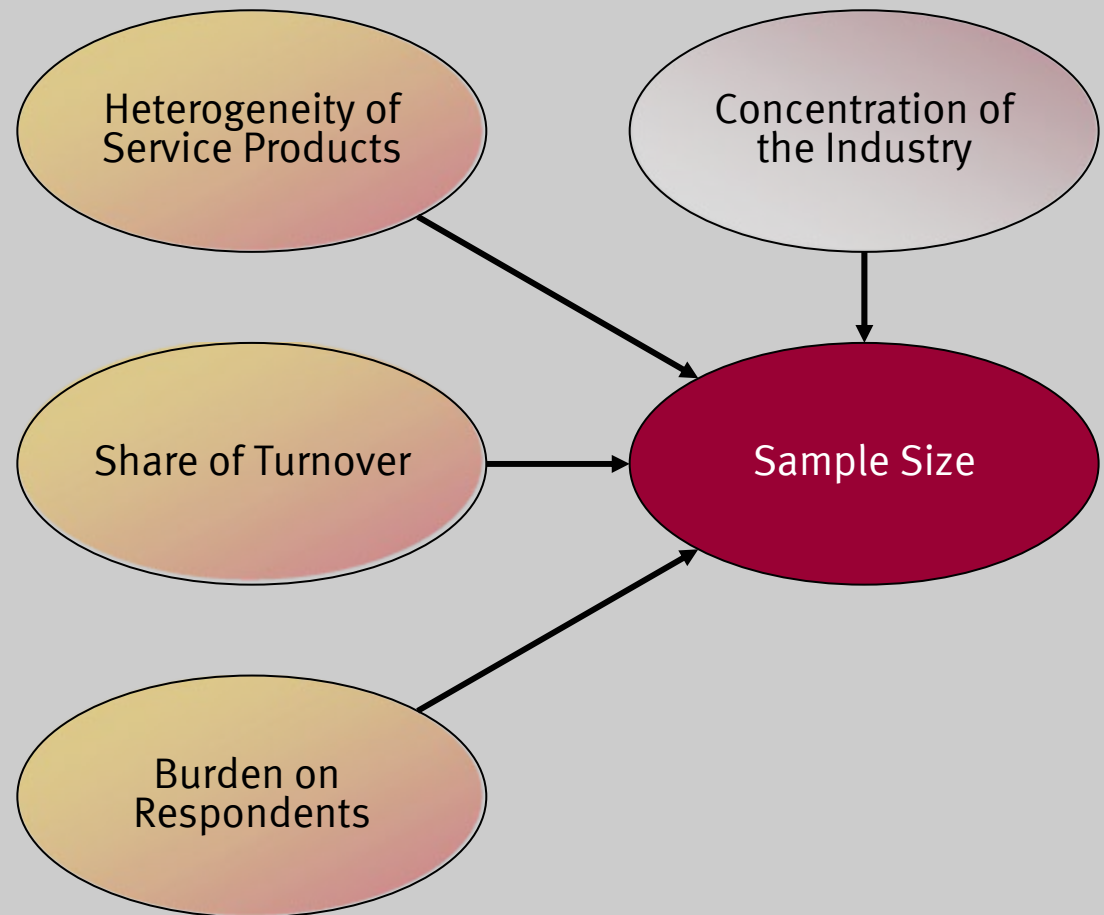
Aim of the presentation

- standardization of sample size and price obs. across indices
- focus on comparison of relative quantities / proportions
- decision support for rebasing
 - number of products to keep
 - number of prices to collect per elementary index

Influencing factors



Influencing factors



Starting position

ISIC	Title	Sample size in 2006	Sample size / Turnover (bn €)	Sample size / corrected TO (bn €)	Portion of population
4912	Freight rail transport	62	11.22	11.22	68.89 %
53	Post and courier services	350	13.60	13.60	2.80 %
61	Telecommunications	30	0.39	0.39	1.17 %
6910	Legal activities	360	21.51	76.32	0.70 %
6920	Accounting, bookkeeping, auditing, tax consultancy	150	7.29	23.44	3.92 %
80	Security and investigation activities	90	22.73	23.47	2.54 %

Approaches to determination of sample size

- **Neyman – Allocation**

Pros	Cons
<ul style="list-style-type: none"> ■ maximizes survey precision given a fixed sample size 	<ul style="list-style-type: none"> ■ strata by turnover don't account for the variable to be measured ■ sample is greater the larger the standard deviation is

- **determination by share of turnover**

- **Census with threshold determined by coverage rate**

Approaches to determination of sample size

- Neyman – Allocation
- determination by share of turnover

Pros	Cons
<ul style="list-style-type: none"> ■ follows the principle of price statistics: number of units / observations proportional to weight 	<ul style="list-style-type: none"> ■ given a fixed sample size some industries would be insufficiently supplied with data ■ doesn't take into account the number of products ■ doesn't take into account the concentration of the industry

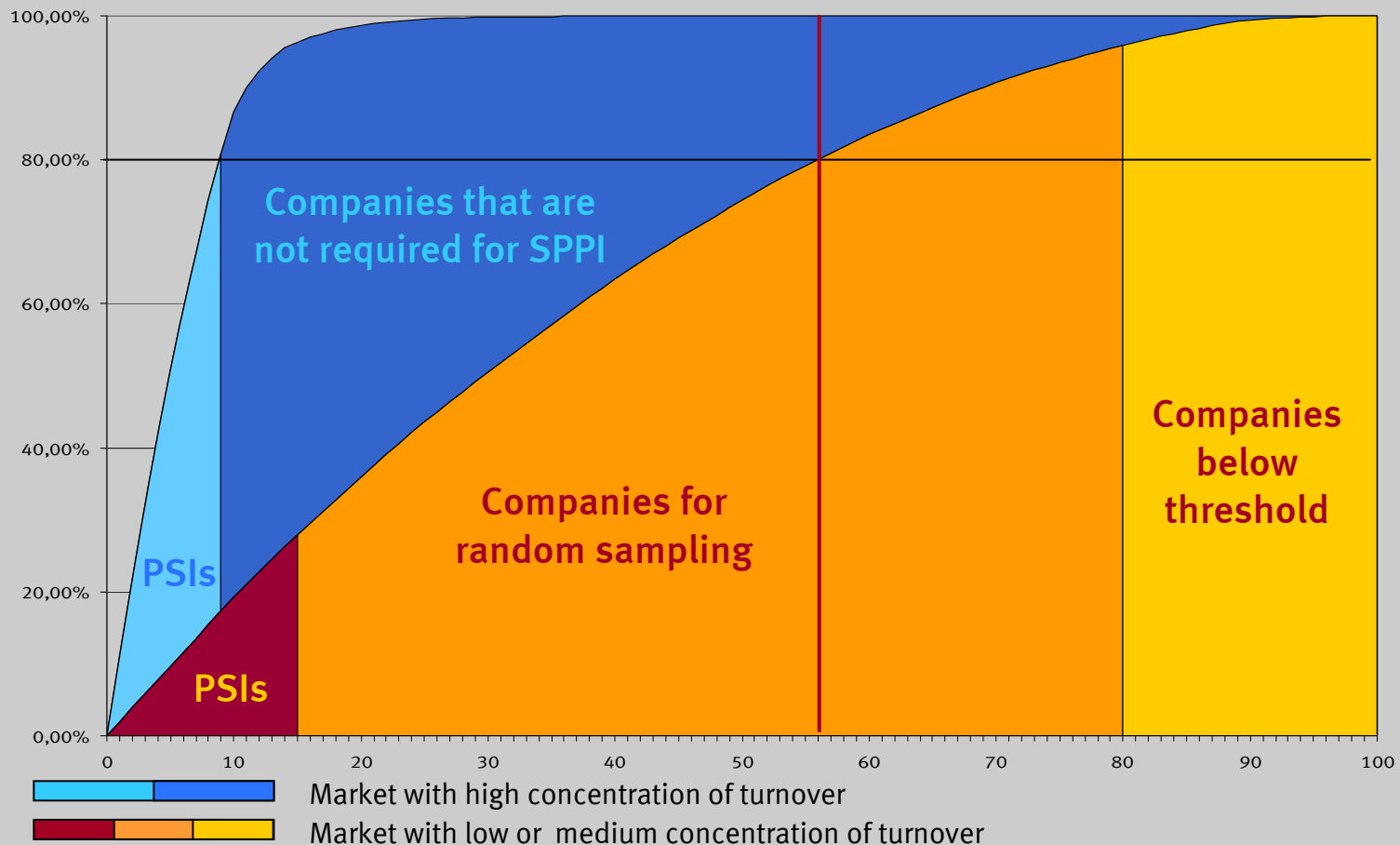
- Census with threshold determined by coverage rate

Approaches to determination of sample size

- Neyman – Allocation
- determination by share of turnover
- Census with threshold determined by coverage rate

Pros	Cons
<ul style="list-style-type: none"> ■ covers a specified portion of all sales 	<ul style="list-style-type: none"> ■ sample size is not fixed ■ for slightly concentrated industries the effort is inadequate

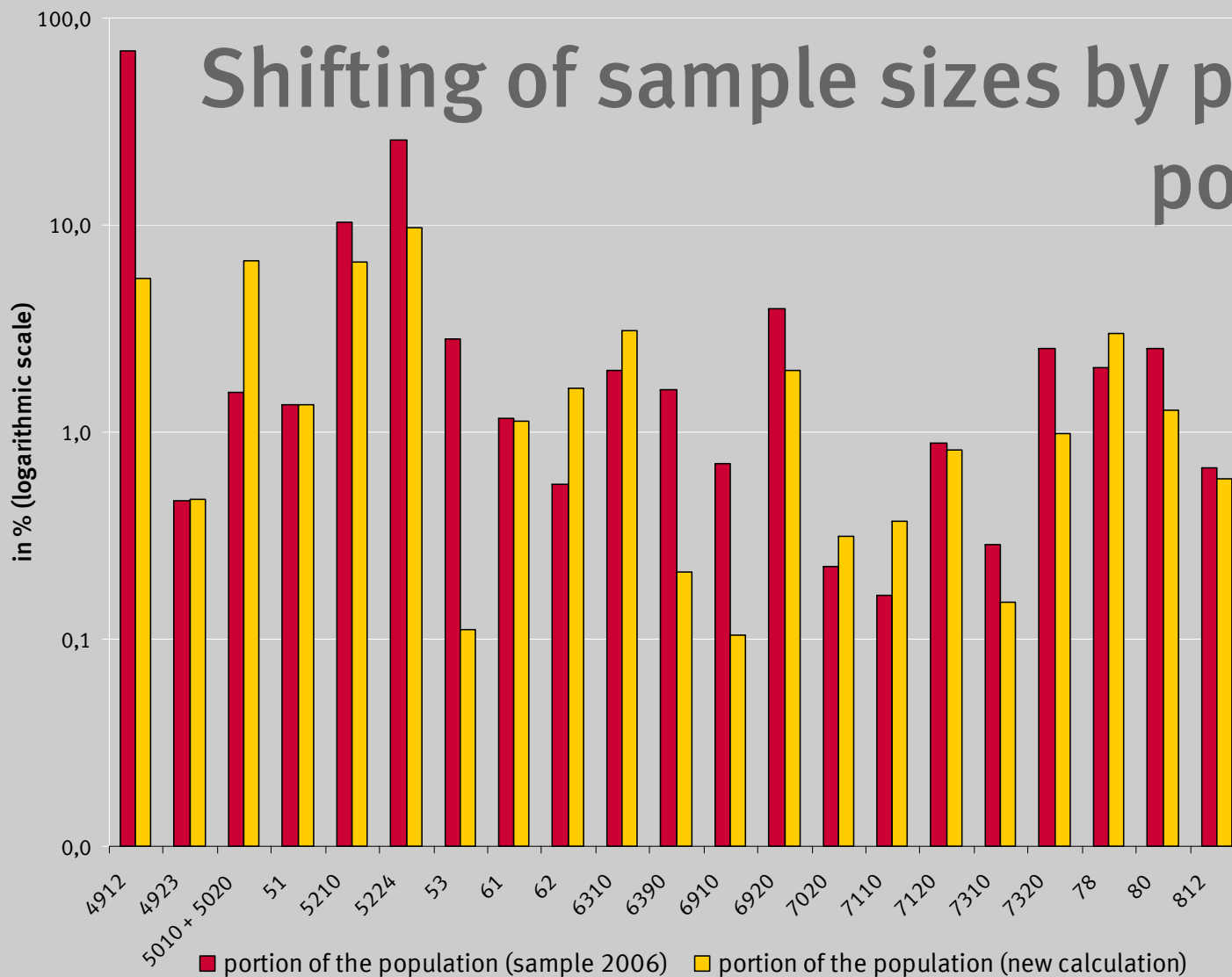
Random sampling vs census



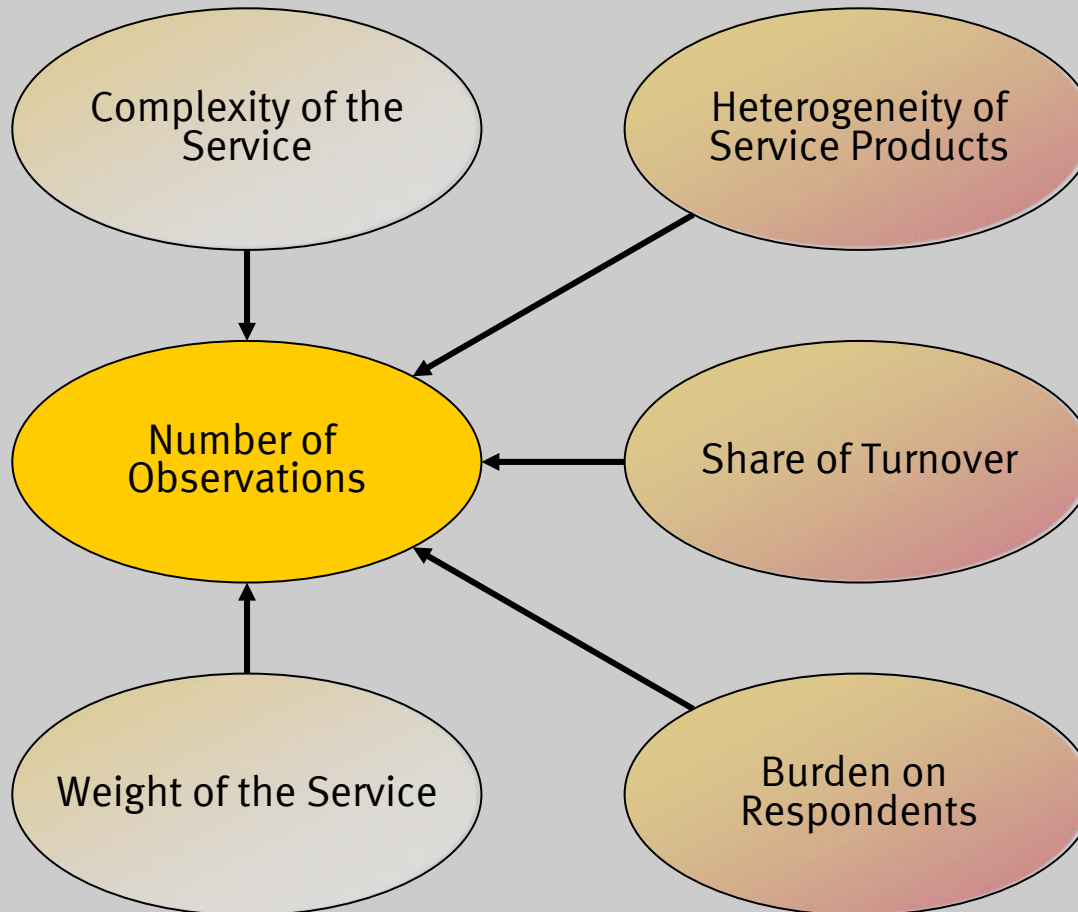
Sample size after reallocation

ISIC	Title	Sample size after reallocation (Sample size 2006)	Sample size / corrected TO in bn €	Portion of population
4912	Freight rail transport	5 (62)	0.83 (11.22)	5.56 % (68.89 %)
53	Post and courier services	14 (350)	0.51 (13.60)	0.11 % (2.80 %)
61	Telecommunications	29 (30)	0.40 (0.39)	1.13 % (1.17 %)
6910	Legal activities	54 (360)	10.68 (76.32)	0.10 % (0.70 %)
6920	Accounting, bookkeeping, auditing, tax consultancy	76 (150)	10.68 (23.44)	1.98 % (3.92 %)
80	Security and investigation activities	45 (90)	10.68 (23.47)	1.28 % (2.54 %)

Shifting of sample sizes by portion of population



Influencing factors



Share of turnover/ weight of the service

Historical role:

- Determining factor for
 - sample size and
 - number or price observations

Now:

- Indicator for 'importance' of sector or service category
- Upper limit for number of price observations

Complexity of the service

Determines the number of first aggregates/ elementary indices
⇒ receive homogeneous groups

How can we assess this complexity?

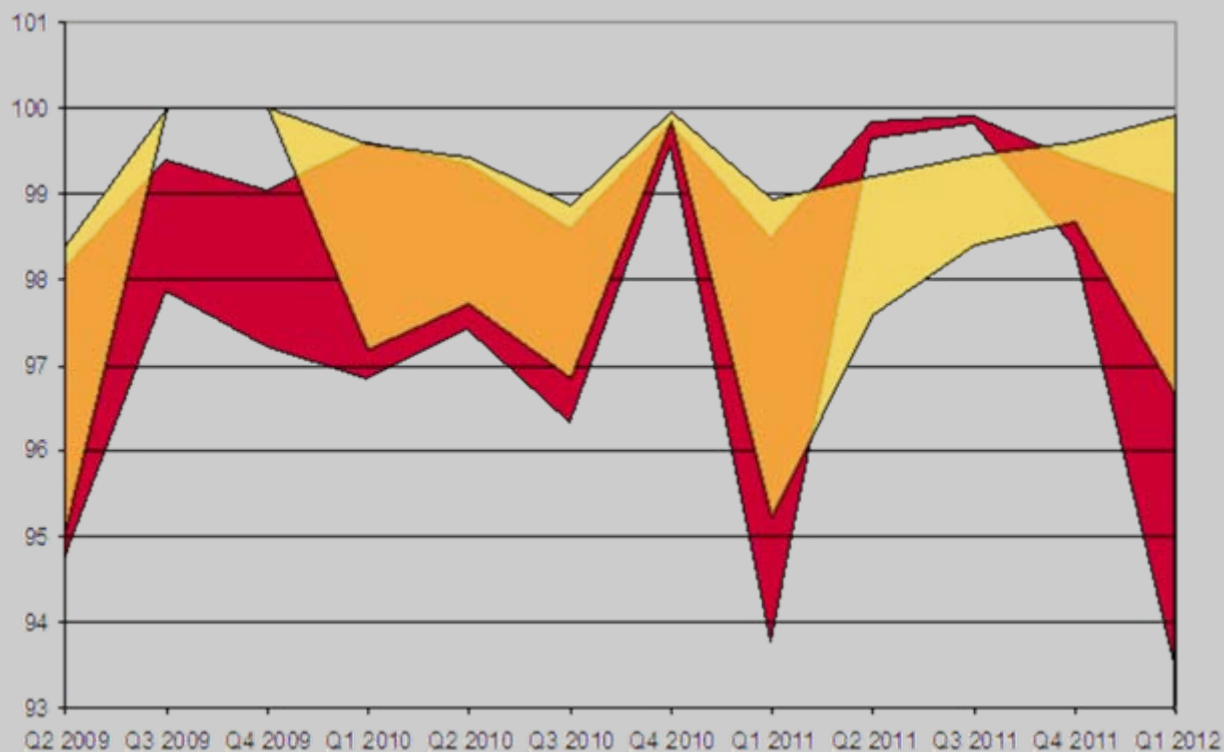
- Product catalogues
- Market expertise...

- Previous price information?
 - Caveat: ex-post analysis!
 - How to measure?

Complexity of the service II

Example from Cleaning Activities

95% Confidence Bands for Cleaning in Hotels ■
and Cleaning in Hospitals ■



Basis: price changes to previous quarter

Large overlap indicates higher inter-aggregate homogeneity

Standardized assessment of product detail by degree of heterogeneity possible

Heterogeneity of service products

**Determines the number of prices to collect
for each first aggregate/ elementary index**

⇒ minimize the effect of single price observations on index

**In case of heterogeneous price developments, we need more
observations per first aggregate**

**A standardization should take intra-aggregate heterogeneity
into account**

Burden on respondents

Depends on:

- Number of prices per sample size
 - In Germany, on average about 5 prices per respondent
- Kind of price collected
 - Contract prices more burdensome
 - In Germany, on average 6 prices per respondent
 - Hourly rates and transaction prices less so
 - In Germany, on average 4 prices per respondent

Limiting the burden on respondents both affects sample size and number of observations.

Outlook

Comparative calculations are starting point for further discussion

- **Sample size:**
 - Decisions on future sample sizes have to be taken
 - Decision making process has to be documented

- **Number of observations and first aggregates:**
 - Testing the results on heterogeneity / homogeneity by bootstrap analysis
 - Decisions on expansion or reduction of observations

YOU ARE WELCOME!



Dorothee Blang / Johanna von Borstel

Telephone: +49 (0) 611 75 -2319 or -4670

dorothee.blang@destatis.de

johanna.von-borstel@destatis.de

www.destatis.de