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

- Complexity of the Service
- Heterogeneity of Service Products
- Concentration of the Industry
- Number of Observations
- Share of Turnover
- Sample Size
- Weight of the Service
- Burden on Respondents

© Federal Statistical Office of Germany, Prices for services
Influencing factors

- Heterogeneity of Service Products
- Concentration of the Industry
- Share of Turnover
- Burden on Respondents

Sample Size
## Starting position

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

- Neyman – Allocation

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>maximizes survey precision given a fixed sample size</td>
<td>strata by turnover don't account for the variable to be measured&lt;br&gt;sample is greater the larger the standard deviation is</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>follows the principle of price statistics: number of units / observations proportional to weight</td>
<td>given a fixed sample size some industries would be insufficiently supplied with data</td>
</tr>
<tr>
<td></td>
<td>doesn't take into account the number of products</td>
</tr>
<tr>
<td></td>
<td>doesn't take into account the concentration of the industry</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>covers a specified portion of all sales</td>
<td>sample size is not fixed for slightly concentrated industries the effort is inadequate</td>
</tr>
</tbody>
</table>
Random sampling vs census

Companies that are not required for SPPI

Companies for random sampling

Companies below threshold

Market with high concentration of turnover

Market with low or medium concentration of turnover

© Federal Statistical Office of Germany, Prices for services
## Sample size after reallocation

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

© Federal Statistical Office of Germany, Prices for services
Influencing factors

- Complexity of the Service
- Heterogeneity of Service Products
- Number of Observations
- Share of Turnover
- Weight of the Service
- Burden on Respondents
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

© Federal Statistical Office of Germany, Prices for services