Statistics Norway
Statistisk sentralbyrå

On optimal sampling designs for price index surveys

Susie Jentoft, Tora Löfgren, Anne Vedø and Li-Chun Zhang

Statistics Norway

Statistics Norway
Statistisk sentralbyrå

Background

- Strategy for methods 2014-2017 at Statistics Norway
- Sample coordination is a good reason for quality control and efficiency improvement
- Response burden can be reduced significantly if the best design methods are used in production

Background, continued…

- Standardised way of working with sample designs
- Quality control of sampling plans
- In all 100 plans for business surveys to go through

- Package in R for planning and allocation of samples
- Set of functions depending on statistics produced; level, ratio and difference statistics as well as price indexes

Optimal sampling design

- Minimise variance of estimator (given fixed sample size)

- Minimised sample size
  - Or a sample-size dependent cost function

- Wish to achieve a balance between
  - Production costs
  - Response burden
  - Accuracy of the output

Our work – Norwegian CPI data

- 2 major theoretical challenges planning price index surveys

  - The indirect sampling design
  - The lack of a finite-population sampling variance of the index

- Many data sources in CPI; here sample survey data is used

- 2 phase approach
- Minimise variance given total sample size (e.g. current)
- Minimise sample size subjected to efficiency loss (compared to phase 1)

Price index

All the goods are divided into elementary groups, g = 1, …, G

The survey CPI ($\widehat{P}$) is a weighted sum of the Jevons index ($\widehat{P}_g$), one for each elementary group given as

$$\widehat{P} = \sum_g w_g \widehat{P}_g$$

$$\widehat{P}_g = \left(\prod_{j=1}^{n_g} y_{gj}\right)^{\frac{1}{n_g}} \Big/ \left(\prod_{j=1}^{n_g} x_{gj}\right)^{\frac{1}{n_g}}$$

w_g is the weight for elementary group *g* which stands for the proportion of total expenditure for that group

x_{gj} is the base period price of item *j* in *g*

y_{gj} is the price of item *j* in the statistical period of interest

n_g is the number of price observations for items in group *g*

Statistics Norway
Statistisk sentralbyrå

Model-based variance for Jevons index

$$\widehat{Var}(\hat{P}) = \sum_{g} w_g^2 \widehat{Var}(\hat{P}_g)$$

$$\widehat{Var}(\widehat{P_g}) = \frac{\hat{\sigma}_g^2}{n_g a_g}$$

i.e. conditional on n_g where $\hat{\sigma}_g^2$ is the estimated variance of items in group g given by

$\hat{\sigma}_g^2 = \sum_j (z_{gj} - \bar{z}_g)^2/(n_g - 1)$ and $z_{gj} = \log(y_{gj}/x_{gj}) - \log \hat{P}_g$

a_g is the adjustment factor associated with a Jevons index

$$a_g = 1/\hat{P}_g^2$$

On sample allocation

- Two way classification; elementary group g and strata h
- Same group of goods may be found in businesses from different strata
- A business unit can provide prices to several elementary groups
- Wish to allocate the sample among the strata with highest number of price observations in groups with the highest variance
- we do not know how many price observations we will collect in each group, so n_g is a random variable

The anticipated variance

Matrix (b_{hg}) based on historic data, where each element gives us the average number of price observations in each stratum-group

$$b_{hg} = \frac{E(n_{hg})}{m_h}$$

n_{hg} is the number of price observations from the sample in stratum *h* of group *g*

m_h is the number of sampled businesses in stratum *h*

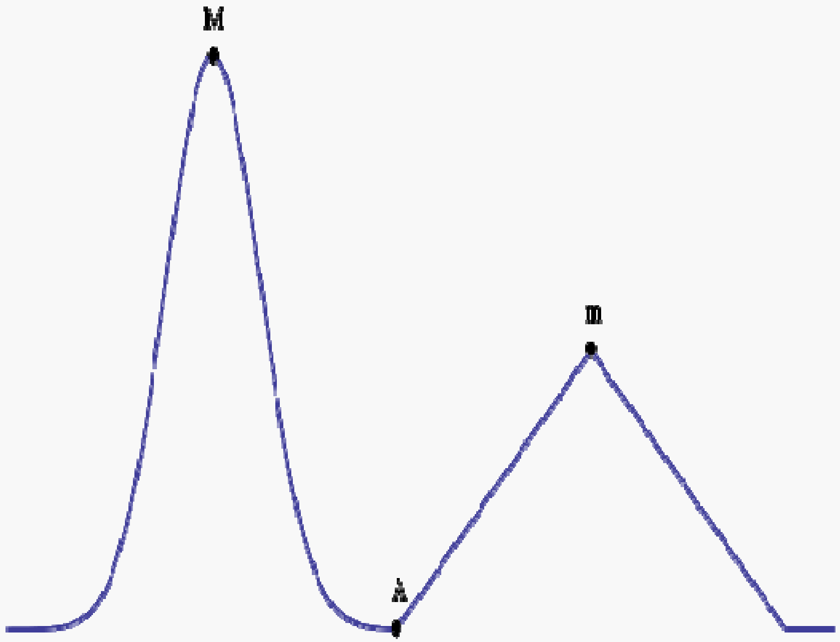
under s.r.s with stratum sample size m_h the expected number of price observations in stratum-group (hg) is given by

$$E(n_{gh}) = m_h b_{hg}$$

Substituting the resulting $E(n_g)$ into the model-based variance above gives us then an approximate anticipated variance

Allocation with greedy fill-up algorithm

- One business unit at a time…
- …in the stratum that results in the most decrease in variance
- Allows us to keep track of allocation and to control sample size
- Fill-up is good for imposing restrictions on stratum size



- Global maximum?

Statistics Norway
Statistisk sentralbyrå

Five different restrictions tested (fill-up)

- To prevent drastic changes
- Better selling point to person responsible for stat.

- Minimum restriction with one unit per strata
- Different ranges around todays sample size
- Range around proportional-to-size (pps, turnover in 2013) stratum sample size

Some results; fill-up with restrictions

Strata industry code	Restriction 1	Restriction 2	Restriction 3	Restriction 4	Restriction 5	Current allocation	Proportional allocation
4520	126	96	88	96	166	59	135
4532	1	39	47	20	20	78	39
4540	12	14	16	10	10	28	7
4711	40	96	96	319	319	193	638
4719	164	82	105	102	60	70	40
4724	15	21	38	4	14	42	9
4729	1	9	9	4	10	18	7
4730	89	72	119	99	146	87	198
4741	1	26	76	5	15	51	10
4742	1	26	26	3	10	51	6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Variance	1.10e-06	1.24e-06	1.61e-06	1.29e-06	1.64e-06	2.21e-06	2.34e-06

Allocated stratum sample size by fill-up algorithm with varying restrictions (1-5) and fixed total sample size (2127) for 10 strata. Anticipated index variance is given in the last row.

Allocation by swap algorithm

- Swap business units between two strata; tried different starting points
- Sample allocation updated by the move that leads to a smaller target variance

- Algorithm is terminated if a chosen amount of swaps do not yield an accumulated variance reduction larger than a threshold value

- Tried 100,1000 and 10000 swap attempts
- No guarantee for global optimum
- Results are similar to the fill-up allocation

Minimising sample size by down size algorithm

- Starting point result from fill-up
- Reduce sample one-by-one
- Choosing stratum with least variance increase in overall price index

The two phase approach for price index survyes

- Fill-up and swap
- Down-size and swap

Balance between

- production costs
- response burden
- statistical accuracy

To be improved

- R-package «AllocSN» to be improved! (Anyone interested in trying, please let me know; orl@ssb.no)
- Improve the allocation for the publishing level of CPI
- Different constraints give different solutions – why/how
- Global optimum vs. Local optimum

Thank you for listening!