

Sampling, Other Selection Methods and Important Aspects of the Swedish CPI

Kristina Strandberg

Methodologist

Method Unit, Enterprises, Stockholm

Department for Development of Processes and Methods

Statistics Sweden, PMU/MFS



Swedish CPI



Use and Relevance of Swedish CPI

- Measure of compensation
 - Measure the effect of price changes on household cost of living
 - Approximation of Cost of Living Index, COLI
- Fixed basket of goods can be more or less relevant
 - ‘Plutocratic’ basket of goods: weighting factor based on the total expenditure per purpose within the total private consumption – good for economic analyses
 - ‘Demographic’ approach: average of all households’ individuals’ price – more relevant for compensation?



Use and Relevance of Swedish CPI

- Measure of inflation
- Swedish Riksbank (central bank) has inflation goal of 2% according to the CPIF (fixed interest rate) change over the past 12 months
 - Switch from CPI to CPIF
- Indexation
 - Contracts and agreements
- Price base
 - Projected forward annually with CPI inflation in June
 - Taxes and public expenditures
- Economic analyses



Use and Relevance of Swedish CPI

- Volume calculations (deflation) of private consumption in the National Accounts
- Detailed levels of CPI
 - Quality on lower aggregates is important

Different Measures

- CPI-CT – constant taxes
- CPIF – fixed interest rate
- HICP – harmonized for comparisons between EU countries
- CPIF-CT – constant interest rate and constant taxes
- CPIF-XE – ... excluding energy

Different measures → different delimitations of the basket of goods

CPI Board

- Members representing main users
 - Ex. Sveriges Riksbank, the Ministry of Health and Social Affairs, the Ministry of Finance
- Experts
 - Ex. Statistics Norway, university professors, consultants
- Two sessions a year
- Until 2016 the CPI board was decisive
- From 2016 and forward the CPI board is advisory



Contents of the Statistics



**The CPI shall measure the average price trends
for all private consumption in Sweden**

Units and Population

- **Population of interest** consists of *all transactions* in private consumption in Sweden in the period to which the price index refers
 - **Target population** = goods offered for sale to private consumers
 - Target population includes goods that may not be sold during the reference period
 - Representative products
 - Target population includes consumption by foreigners in Sweden
 - ... does not include consumption of Swedish people abroad
(Domestic approach vs. National approach)

Target Population

- A population of retail outlets
 - Unit = shops, service outlets, websites etc.
- A population of products
 - Unit = goods and services
- **Product offer**
 - The combination of retail outlet and product
 - An observable good or service that is offered for sale at a given price in a certain retail outlet, on certain terms and conditions and at a certain point of time
- Primary **variable** is the price paid by the consumer – or rather, the offer price



Samples

- KPI consists of around 360 product groups divided into around 60 different surveys
 - different frames and different sampling methods
- Large parts of KPI are based on samples in two steps
 1. Sample of outlets (stores)
 2. Sample of products in each outlet (goods and services)
- Frame → Stratification → Probability Sample
 - Sometimes simple (Swedish Business Register)
 - Sometimes hard (no frames for products)

Main "Systems"

- Daily necessities – scanner data
- Local price collection – all surveys where data is collected locally
 - By visit, phone call or via the internet
 - Clothing and footwear
- Central price collection – prices are collected centrally at Statistics Sweden
 - By phone calls, via the internet or through questionnaires

Order Sampling

Order sampling is a class of sampling schemes with *unequal inclusion probabilities* that provides interesting properties such as

- **Simple sample selection**
- **Good estimation accuracy**
- **Simple variance estimation**
- Simple procedures for **sample coordination** over times and between surveys (by permanent random numbers)

Why Use Order Sampling?

Good for sampling from a frame where we know we have over coverage.

- ❑ "Gross sample" → clear from over coverage → "net sample"
- ❑ Ex. We have a hard time to find stores that sells strollers and sample from industry "other". Pareto πps-sampling allows us to skip outlets in the gross sample that does not carry strollers.
- ❑ We want to coordinate samples over time and between surveys

Coordinated sampling (1)

Co-ordinated sampling for business surveys has 3 objectives:

- To obtain comparable statistics by promoting the use of the same version of the business register for many surveys
- To obtain positive co-ordination of samples over time for the same survey
- To obtain samples that are coordinated - negatively or positively - among different surveys

SAMU The system for co-ordination of frame populations and samples from the Business Register at Statistics Sweden, Background Facts on Economic Statistics, 2003:3 Inquiries: annika.lindblom@scb.se

Coordinated sampling (2)

- The basic idea is to associate a *permanent, independent and unique* random number
PRN $\sim U(0,1)$
with every unit in the register.
- To coordinate:
 - Stat Sweden groups all units in the register into 5 rotation groups
 - The random numbers for the units are then shifted by 0.10 in one rotation group each year.

Coordinated Sampling (5)

It has been decided that the objective should be to keep the selected units in the sample for 5 years. There are at least two possible methods for rotation:

- Shift the starting points
- Shift the random numbers
- Statistics Sweden groups all units in the register into 5 rotation groups. The random numbers for the units are then shifted by 0.10 in one rotation group each year.

Daily Necessities - Scannerdata



Daily Necessities and Scannerdata

- 2012: One big chain of stores for *daily necessities*
 - Food, except perishable fruits, vegetables and meat
 - Non-alcoholic beverages
 - Personal hygiene
 - House upkeep
 - Tobacco
 - Non-food
- 2013: Two more chains for daily necessities join
 - Three chains cover ~ 80% of the market for daily necessities
 - Daily necessities account for ~16% of the Swedish CPI
 - Price collection in some stores from other chains
- 2019
 - New chains → no more price collection?

Choice of Outlets

- We used to draw a rotated sample of stores
- Rotation is a disadvantage
- Now a Sequential Poisson sample without rotation
- We keep the same stores as long as they exists
- Maybe it's time to rotate?

Frames of Products

- Frames from the three chains contain 50 000 – 100 000 products each
- Classifications differ between Statistics Sweden and each of the chains
- Coding of frame into product groups is done manually
 - Key between GTIN-code and product groups from Food Statistics
 - Sales numbers, good base for weights
 - Very good for food
 - Not as good for other

Frame Coverage

- For the sample for year y we use sales statistics for year $y - 2$
 - under coverage and over coverage
- Update the frame with information from CPI Scanner Data, $y - 1$
 - Over coverage of around 25 %, value of sale
 - > Products leave the market
 - > Products change GTIN-code
 - Under coverage is smaller
 - > Products new to the market
 - > We add the most important (highest sales)

Allocation

- Neyman allocation based on variance estimates over six years
- For each product group (ex. rice, coffee) the sample is allocated proportionally to weight and the square root of the variance
- Assumption: cost of price collection is the same across product groups
- Ca. 800 products per chain
- For each strata, the sample is allocated proportionally to sales

Sample

- Stratified Pareto $\pi p s$ sample of products with sample probabilities proportional to turnover
- One sample per chain
- No sample rotation → keep the sample as long as possible
- As little overlap between the three samples as possible → possible by using permanent random numbers from SAMU

2-Dimensional Sampling/Cross- Classified Sampling

Our population of interest is observed bi-dimensionally and can be represented by a rectangular array.
Neither of the dimensions is first or second.

	Prod 1	Prod 2	Prod 3	Prod 4
Shop A				
Shop B				
Shop C				
Shop D				
Shop E				
Shop F				
Shop G				
Shop H				
Shop I				
Shop J				
Shop K				
Shop L				
Shop M				

Local Price Collection



Frame for Outlets

- Swedish Business Register
 - SAMU frame, “frozen” four times a year
 - Clear the frame from entries that do not offer sales to consumers
- Stratification by industry
 - Some industries are joined
 - A few big retail chains for clothing have been given their own strata
- No stratification by size or geography
- Internet and mail order is most likely underrepresented

Allocation Between Strata (Industries)

- Based on jack-knife* estimates of between outlet variances
- Assumption: the sample sizes for products are fixed
 - Variance between products shows as variance between outlets
 - Smaller sample of outlets and bigger sample of products
- Allocation per industry (h) is proportional to the square root of the variance over the cost
 - We assume the same cost of price collection over all strata



* Ex. Särndal, C-E., Swensson B. and Wretman J. (1992) *Model Assisted Survey Sampling*, Springer-Verlag, New York, 1992.

Cut Off of Municipalities

- In 2015 the CPI board agreed to a cut off sample of outlets that covers 90% of total private consumption in Sweden
 - Combination of total trade and home location of price collectors
 - Cut off of the 7.5%
 - We adjust for the cut off within 6 geographical regions

Sample of Outlets

- Sample in two phases
 1. Geographical sample
 - Zip codes ~10,000 in Sweden
 - Systematic π ps sample with the frame sorted by municipality code and zip code
 2. Sample of outlets
 - Sequential Poisson sample with inclusion probabilities proportional to store size measure multiplied with the inverted inclusion probability from phase 1
 - Gross sample \rightarrow net sample

Frame and Sample for Products

- Except for daily necessities, we have no frames over products in the outlets
 - Clothes, shoes, household goods, sporting goods, radio & TV, computers, et. c.
- The sample is chosen subjectively by the price collector assigned to the store
- Broader product definition than for daily necessities

... and the Design?

- A stratification of the markets by the representative items, which covers all or a large part of the market. A random (arbitrary) sample of product offers in one single stage, independent between outlets
- Ex. Clothing store → strata: dresses, pants, jackets...
- This design allows us to calculate variances

Central Price Collection

- All surveys where prices are collected from the central office at Statistics Sweden
- Prices are collected monthly by questionnaire, internet searches or direct contact with informant
- Often no frames → create frames based on statistics on turnover
 - Ex. Electricity consumption, statistics from us and other sources
- In general: outlet – product → price offer

Variance Estimation

Variance

- Statistical measure of the variation of an estimate due to sampling
- Here: the estimate is *Inflation*
- Inflation is calculated using a sample of outlets, items and product offers for two years and with a chain index with annual links that are mostly new, but overlapping samples
- Assumption: We have statistical samples, i.e. samples that have been selected using random methods and with known inclusion probabilities
- Only true for the outlet and product samples in daily necessities
- Our variance estimates are calculated using replication techniques

Bias Study



Bias Study

- Update of the bias study done for the SOU 1999:124, by Dalén
- Based on the Boskin report (1996, 1998)
- Martin Ribe, Statistics Sweden
- Bias can be defined as the expected difference between the estimate and the statistical value it is set to measure
- Swedish CPI is a cost-of-living index (coli)
 - Should follow the cost for a consumer to keep a constant standard of consumption (SOU 1999:124) or constant standard of living (ILO, 2003)

New

- Increased internet trade
- Scannerdata as the source of data
- Increased rate of change in products
- Quality assessment methods
- Digitalization
 - Changes the behavior of shops (pricing)
 - Changes the behavior of consumers

Conclusions

- Most important bias sources
 - New or improved products on the market
 - Quality adjustment methods
 - Weights
- Sampling and non-sampling inaccuracies described in the quality declaration
 - Available at www.scb.se → CPI

On the Horizon for the Swedish CPI



More Scannerdata!

- Transaction data
- Hopefully coming:
 - Charter travel
 - Dental care
- Looking at:
 - Clothing
- New methods for index calculation!
 - How handle “*all*” the data?
 - How handle replacements?
 - Multilateral methods



Hedonic Models for Quality Adjustments for Home Electronics

- Historically done by price collectors
- Now expert groups at Statistics Sweden
- Web scraping → cheap data → hedonic models
 - TV
 - Coffee maker

Challenge due to a Changing Market

The Digital Economy

- New products and new sales channels
 - > Web stores
 - > International trade
- Substitutions are easy between physical outlets and web stores
 - > Ex. clothes
- Do our sample designs mimic consumer behavior? Is the outlet dimension necessary?



Plans

- New products and new sales channels:
 - Map up problematic areas of consumption in the Swedish KPI/HIKP
 - Try solutions that have been recommended internationally (if possible)
 - Eurostat recommendations and regulations
- New sample designs:
 - Clothes: sample brands instead of stores

Obrigado!

Kristina Strandberg | Metodstatistiker

Metodenheten företag, Stockholm

Avdelningen för process- och metodutveckling

Statistics Sweden

Box 24300, 104 51 STOCKHOLM

Besöksadress: Karlavägen 100

010-4794034

www.scb.se

