CONFERENCE OF EUROPEAN STATISTICIANS

Workshop on Statistical Data Collection 'New Sources and New Technologies' Session 1

14-16 October 2019, Geneva, Switzerland 07 October 2019

New survey design

Sara Westling and Chandra Adolfsson (Statistics Sweden)

sara.westling@scb.se; chandra.adolfsson@scb.se

Abstract and Paper

In consequence of declining response rates in the Household Budget Survey (HBS) and the survey of Housing Costs, Statistics Sweden is redesigning the two surveys. The HBS has been especially problematic, mainly because of the high response burden. The main features of the new survey, to be launched in 2020, are introduction of web mode (not used in the HBS to date) and a new survey design. The new design consists of three modules, combined in a panel design where each household takes one module each year for three years. The main purpose of the panel design is to reduce the (perceived) response burden. It will also give an opportunity to estimate changes in consumption patterns over time, while meeting Eurostat requirements on micro data delivery. Web will be used as an alternative mode, to complement telephone interviews. Web has not been used previously in the HBS, but has been used successfully in the Housing Costs survey.

At the same time, new possibilities of using a simplified data collection combined with cash register data from Sweden's major grocery store chains are being investigated. If successful, this may replace the current collection of data on expenditures for groceries in the future.

We describe the new, combined, survey and the consequences of coordinating the two surveys, as well as the development of Statistics Sweden's data collection systems to facilitate increased use of mixed mode designs in social surveys and use of new data sources.

Keywords

[Keywords]

Chandra Adolfsson
Sara Westling

New survey design

Background

In consequence of declining response rates in both the Household Budget Survey (HBS) and the survey of Housing Costs, Statistics Sweden is currently redesigning the two surveys. The HBS has been especially problematic, mainly because of the high response burden.

In an attempt to counter the increasing nonresponse rates, the HBS data collection was adjusted in 2016. The most important change was to reduce the length of the reference period that households are asked to keep a diary of all their expenses, from two to one week. Also the telephone interviews, where data on less frequent purchases such as furniture or cars are collected, were simplified. This, however, did not lead to an improvement in response rates. When the survey was conducted according to the new design, the data collection was terminated prematurely because of the low response rate, about 20 percent. No data were published.

It became clear that the prerequisites to conduct a traditional Household Budget Survey with sufficient quality are no longer fulfilled. Thus, a new project was launched in 2017, aiming to find a new design that is both practically feasible and of sufficient quality. The starting point for this work was a thorough investigation of national user needs, as well as Eurostat requirements. So far, two alternative designs are proposed; merging of HBS and Housing Costs, and statistics based on alternative data sources.

We describe the new, combined, survey and the consequences of coordinating the two surveys, as well as the development of Statistics Sweden's data collection systems to facilitate increased use of mixed mode designs and use of new data sources.

Current designs

The design in the survey of Housing Costs has undergone major revisions in recent years. Before 2015, the survey covered both household incomes and household living costs, as well as other expenditures. In 2016, a new Swedish register on dwellings made it possible to split the survey in two. One is completely register based, using income variables from tax registers and household composition from the dwelling register. The other became the survey of Housing costs, in which data on housing costs are collected from a sample of around 17 000 households. The survey is conducted every two years. In

2015 and 2017, web mode was used as an alternative to the telephone interview. The response rate in 2017 was just over 40 percent, about half of which were web responses.

The HBS has run every four years, with a sample size of 7 500 households. Each household is randomly allocated to one week during the reference year and data is collected in a telephone interview and a self-administered diary. Collected variables include expenditures on a number of detailed consumption item groups. Some variables, such as income and education, are collected from registers.

Merging of the two surveys

The plan is to merge the designs and data collection for the two surveys, Housing Costs and the HBS, while still publishing separate results. The surveys are similar in many ways. The target populations and domains of interest only differ slightly, while the observation variables in the expenditure categories that overlap can be used in both surveys. However, the reference periods differ. The survey on Housing Costs refer to calendar year, while the HBS cover the reference week or the previous 12 months. It is suggested that the HBS reference period is used for most variables. One of the consequences for the Housing Costs statistics is that the relationship between household income and housing cost will be less clear, since income variables from registers typically refer to calendar year.

To reduce response burden, a new survey design is proposed. A split questionnaire design is to run over three consecutive years. Also, a panel approach is considered, where sampled households respond to one of the three blocks each year so that all three blocks are covered, each for a third of the sample. Each block from the questionnaire will cover different consumption item groups. The main purpose of the split questionnaire design is to reduce the response burden. Combined with the panel design, it will also give an opportunity to estimate changes in consumption patterns over time, while meeting Eurostat requirements on micro data delivery.

Web will be used as a mode for data collection, as a complement to telephone interviews. Web has not been used previously in the HBS, but has been used successfully in the Housing Costs survey. The aim is to develop a web version of the diary, facilitating respondents' task. To simplify further, it is also possible for respondents to send in all their receipts from purchases during the reference week, instead of filling in the diary. The receipts are scanned and coded at Statistics Sweden.

In 2019, an experiment is performed to test different alternatives for collecting data on expenditures for food, beverages, clothes and restaurants, which is one of the blocks in the proposed split questionnaire design. One of the alternatives is "traditional" data collection, with a telephone interview and diary. This is to be compared with a collection of much less detailed data, where households estimate their total expenditures and the proportion spent on different consumption item groups for the reference week retrospectively.

Results from the experiment will be available at the beginning of 2020. The complete split questionnaire design will be tested in 2020-2022.

Despite all efforts, the response burden will still be considerable, and it has also become increasingly difficult to make contact with selected households. Therefore, there is a need to find alternatives, using data already available from other sources.

Alternative data sources

New possibilities of using a simplified data collection combined with cash registry data from Sweden's major grocery store chains are being investigated. If successful, this may replace the current collection of data on food expenditures in the future. The principle may be applied for other consumption item groups in the future, if information from other data sources can be found.

The desired information is average household expenditures on different food item groups within household types, as well as associated margins. Instead of collecting data from households to obtain direct estimates of table cells, data sources on aggregate data across consumption item groups is available. Modelling the distribution across different household types of aggregate expenditures may lead to estimates of sufficient detail and quality. The main data source on aggregate expenditure for food item groups is cash registry data, already provided to Statistics Sweden, although for other purposes. Information on expenditure among different types of households may be obtained from a simplified version of a traditional household expenditure survey, similar to the one being tested in the 2019 experiment.

The proposed estimation approach involves modelling, and methods for reducing the risk of bias in estimates due to misspecified models need to be worked out. One method is calibration of estimates on known or unbiasedly estimated totals. The design and methods need to be further developed and evaluated before it can be used for the production of official statistics.

General efforts to reduce response burden

Statistics Sweden is working in many areas to reduce response burden, while maintaining high quality in survey results. There is a long history of using many types of administrative data sources and in recent years this work has been intensified due to the decreasing response rates and a governmental directive to reduce the response burden on enterprises. Statistics Sweden is currently investigating the possibilities of sharing more data in smart and safe ways. An example of this is cash registry data to be used as auxiliary information throughout the production process and from 2019 has replaced collection of price data for food and beverages in the production of the Swedish Consumer Price Index, CPI. Other related examples are PAYE tax reports, submitted monthly to the Swedish Tax Agency, and digital annual reports from enterprises that hopefully can be used in both social surveys and business surveys.

Strategies involve, except the use of new data sources and redesign of surveys and of data collection, increasing the awareness of Statistics

Sweden in order to build a strong trademark, cooperation between governmental agencies and software suppliers and giving the respondents the possibility to choose from more alternatives regarding reporting mode, login method, device (smartphone/tablet) etc.