

**Expert Meeting on Statistics for SDGs**

31 March - 1 April 2022, Geneva

Session 2: Data from non-traditional sources (including non-official statistics and their quality)

**A new design for statistics on household consumption – Statistics Sweden**

Prepared by Thomas Laitila and Marie Hollertz, Statistics Sweden

**I. INTRODUCTION**

1. The household consumption is important for consumer price index (CPI) and national accounts (NA). In CPI household consumption is used to weigh the prices and in the national accounts it is important for calculation of the gross domestic product (GDP) from the user side. Historically has data from the household budget survey (HBS) been used. But this survey has for many years been struggling with quality problems. There is low response rate and large partial non-response. Statistics Sweden decided for two years ago to start a project for household consumption with a new design.

2. The project is working to collect turnover on detailed product groups from enterprises. The approach is to split the turnover in households and others based on information from financial institutions. The turnover collection from enterprises will be built on cash-register data (EAN-code) from larger enterprises and on sample surveys for smaller ones. The financial institutions are sending transaction data which can be divided on activities and on households and others. The goal is to implement the new survey design in 2024.

**II. IDENTIFICATION AND INVESTIGATION OF POSSIBLE DATA SOURCES**

3. The new design requires information about: What is bought (in Sweden) and how much? Has the purchase been made by an individual or company?

**What is bought and how much?**

4. Statistics Sweden has continued to work with scanner data from large companies and chains. Scanner data from grocery chains has been used for measuring prizes for the CPI for about ten years. The development is on-going as we expand with more companies and industries. In this work, industries that consists mainly of large companies and chains are prioritized.

5. Statistics Sweden also purchases data from a data analytics company whose business concept is to collect and analyze scanner data. This delivery mainly concerns home electronics and white goods and has now, since 2020, been expanded.

6. In a store with a cash register system, the tax agency is collecting samples of transactions for revision purposes. The collected information is by legislation standardized, but the specific format in which it is stored is not standardized. There is presently work done to standardize the format as well, and by 2027, all cash register systems are to have a standardized format for exported files. The legislation on standardized receipt format/information do not apply for on-

line stores. This exemption is under investigation. Some similar regulations for on-line stores can be expected in the future.

7. The proposed upcoming standardized format can allow the collection of cash registers from small and middle-size companies. However, since each company sets its own product code and the level of detail may vary, AI-technology will be required for coding into COICOP.

8. Examples of model calculations that can be used are those for housing costs, energy consumption and health care. Other sources can be, for example, information on newly registered cars from the Swedish Transport Administration and the Swedish Gaming Inspectorate's has information on gambling and betting activities. These sources and models already exist and work today.

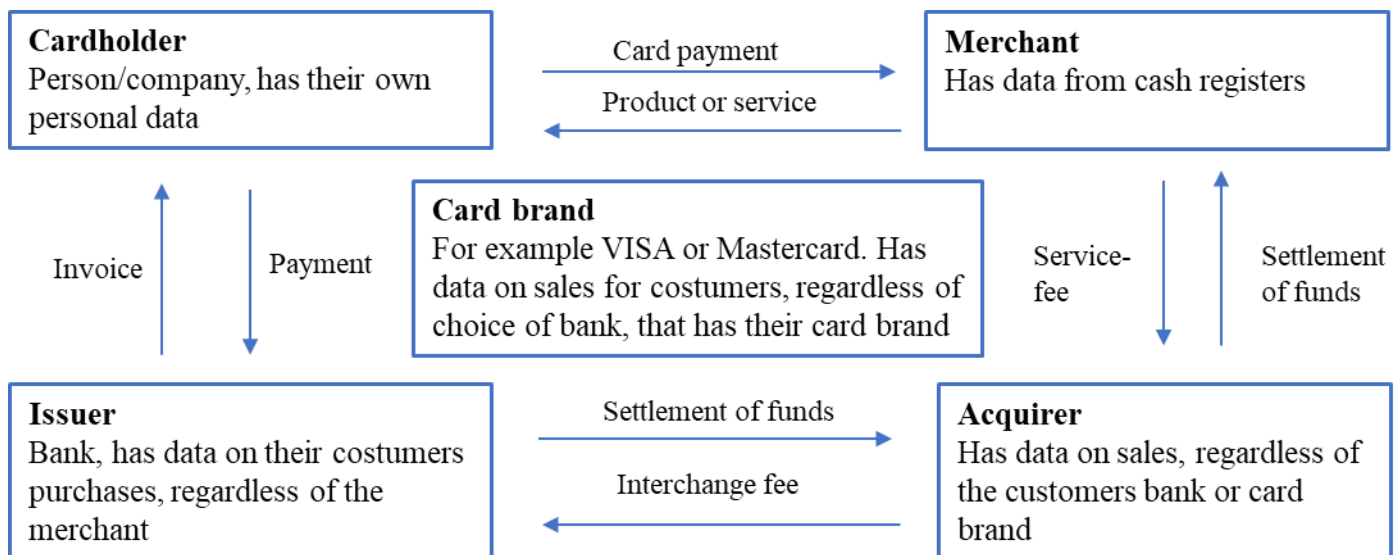
**Has the purchase been made by an individual or company?**

9. For whether the purchase been made by an individual or company, we believe that sources that holds information on different kinds of payments could be part of the future design.

**Card payment terminals and banks**

10. Figure 1 illustrates involved parties in a bank/credit card payment. The card issuer (a bank) has access to its customers' purchases. The card brand has data on its customers' purchases for all their card issuers. These data sources are both limited in the sense they cover purchases made by their own customers. There are many card issuers, and they hold information on all card payments made by a customer.

Figure 1  
**Four-part model over card payment transactions**



11. An interesting data source is the acquirers. They capture all payments made irrespective of the card issuer and has information on whether it is a privately owned card or a business card as well as the business industry of the purchase. The number of acquirers is also small.

12. Statistics Sweden has received data from one out of three of the Swedish card payment terminal players in Sweden. The data material covers 50 - 60% of card purchases in Sweden. The different acquirers have different profiles and cover different parts of the market. At present, we have good/full coverage in some industries but hardly anything at all in others, so we continue to work to get data from additional acquirers.

13. From meetings and dialogue with Swedish banks, we know that also the card issuers, i.e. banks, have the data needed to distinguish private consumption from corporate consumption.

### **Payment service providers**

14. There are a few Payment service providers (PSP:s) in Sweden. For example, the banks offer to pay purchases over the mobile phone with an app called Swish. Swish payments are not included in the terminal data and swish is increasingly used.

### **Consumer loans**

15. Consumer loans were initially most common in online shopping but have now spread to retail in physical stores. Here, we have so far only begun the process of understanding what type of data these actors may have.

### **Cash payments**

16. Of course, customers can normally still pay cash. Cash withdrawals correspond to about 10% of card terminal payments. Adding swish, invoice payments and online shopping, cash payments can be considered of less importance overall, but it can be of a larger proportion in some industries.

## **III. DIFFICULTIES ENCOUNTERED**

17. A prerequisite for the new design is to utilize new data sources to largest extent possible. With respect to the difficulties in collecting data directly from households and limited resources, it is a valid starting point.

18. We have found that in some cases, data holders are very willing to share both information in meetings and most importantly, their data with Statistics Sweden. In other cases, getting the data we want and need to test our theory, has been more challenging. Data holders' systems might not allow for extractions that we have asked for. There have also been cases where data holders have been unwilling to share data because of privacy reasons. The dialogues and relations building process takes a lot of time and effort.

19. We have concluded that we must find ways to secure deliveries of data to build a reliable method. To make sure that our access to data does not depend on the willingness of one or two specific persons at a company.

20. Development in the payment area is moving fast. New ways of paying, both on-line and in physical stores, are continuously emerging and taking shares of the market. The design must be flexible.

#### **IV. STATISTICAL VALIDATION**

21. Several different data sources will be compiled in the calculations of CPI weights and levels of household consumption on product levels. Most of them will not cover the whole population for which they are used. Therefore, there is a risk of systematic errors in the statistics calculated because of selectivity bias.

22. Different specific analyses of survey data are to be conducted measuring effects of such systematic errors. The specific information to be extracted are decided on by evaluations of the sensitivity to errors in different sources.

23. Two survey data sources available are the SBS and the new, simplified HBS launched in 2021. A few additional sample and register surveys may also be necessary.

24. Apart from data sources it is expected the calculation of statistics and the validation part will require development of new estimators for statistical inference. One example is estimating households shares of SBS enterprises total sales on product level. There is an estimate given by the enterprises of households share of total sales over all products. Work on projecting this “over all” share on to the different products is ongoing.

#### **V. THE REMAINDER OF THE WORK**

25. By the end 2022 enough data will have been attained to fully test and evaluate the new design. This includes point of sales data that covers some of the industries in which we have lacked information so far. It also includes enough data about payments to separate household consumption from consumption by companies, at least in some of the industries. By the end of 2022 the access to data will not be by any means perfect, but it will hopefully be enough to decide on the new design.

\*\*\*