Impact of the COVID-19 virus on the data collection, data processing and index calculation of the Consumer Price Index

The coronavirus, COVID-19, has already restricted for a month and for some time will still restrict the activity of our society in an unprecedented way.

Several countries have had to start extensive restrictive measures to prevent the spread of the virus, covering every sector of society: citizens, enterprises, public administration and institutional population. Restrictive measures aim to limit interaction between people. Enterprises marketing consumer products and services have reacted to restrictions and possible falls in demand either by shortening their opening hours or closing down their activity for some time.

The effects of these restrictive measures will in time be visible in the statistics and the time series produced by them. Before that we need to estimate what impact the restrictions have on the collection of statistical data and the production of a reliable image for our society. In the following we will discuss how the changed situation is taken into consideration in data collection and processing of the Consumer Price Index so that these statistics will provide as correct as possible picture of the change in consumer prices also in the future.

Data collection

One of the most important factors influencing quality is the success of data collection in a similar manner from one month to the next. The collection of data has usually been made as a field collection around Finland and centralised at Statistics Finland. Some of the data are collected by interviewers in shops, some come as scanner data including sales information directly from enterprises' data systems to data processing and some of the prices are collected from the Internet or other sources.

Due to the restrictions on movement, traditional data collection cannot be carried out as a field collection anymore, but it must be replaced by alternative methods. Alternative collection methods are used so that the price collection data would not have too many missing prices.

In covering for missing data, the methods used are in the order of implementation as follows:

1. Prices of corresponding products or services are gathered from enterprises' Internet pages or by telephone collection
2. Missing prices are replaced by prices in the enterprise's scanner data with prices corresponding to the products in the collection
3. Missing prices are replaced with corresponding data from another outlet
4. A missing price is imputed[1] by the price change of other products or services in the same commodity group
5. The previous collected price data are taken as the price of the present month (so-called carry-forward)

Primarily the aim is to gather the prices from the enterprise's web pages so that no data are missing. If the enterprise does not have Internet pages or consumer prices

[1] Imputation refers to covering for missing observations with corresponding values.

are not on offer there, alternative data sources available in the statistics are considered.

If the first three alternative collection methods cannot be used, it is evaluated whether it is a question of temporarily missing data or a permanently unavailable commodity. It is also assessed whether it is a stable commodity whose prices seldom change or a volatile commodity whose price changes several times in a month (e.g. food and fuels). For stable commodities, in which prices can be assumed to be the same when the normal situation returns, the previous measured price can be used (alternative 5). For volatile commodities, alternative 4 is selected if data cannot be collected in other alternative ways. Imputation is further addressed in the next section “Treatment of missing data”.

Treatment of missing data

Missing prices of the commodities must be imputed especially when the product or service is limitedly available or is not in supply at all, as has now happened during the restrictive measures of the coronavirus. The Government’s decisions to close restaurants and restrict air and boat traffic to and from Finland are examples of reasons why the supply of products and services decreases and prices cannot be collected to the conventional extent anymore. Other currently challenging commodities are package holidays, cultural services, sports events, beauty care services such as hairdressers and barbers.

The most suitable method for imputing prices of temporarily missing products or services is the “nearest-neighbour-method”, where the price development of a similar available product or service (so-called substitution commodity) is utilised in estimating price development. From which level the nearest peer group is searched depends on the commodity and regional location of the missing prices. Alternative means are corresponding commodity groups of other areas or an adjacent commodity group of the same area.

Example:

> *Restaurant activities have clearly decreased during March and were totally discontinued at the beginning of April (excluding take-away). If prices of restaurant meals and drinks cannot be collected at all, the price development has to be estimated by means of the nearest neighbour, e.g. by the price change of subsidised staff meals or by calculating the price change from the nearest peer group takeaway+fast food+hamburger+cafe services, and by using this correction coefficient when estimating restaurant prices.*

If it should happen that a specific 5-digit level commodity cannot be collected at all, the EU instructs to use in imputation the price development of the nearest higher commodity group level.

Treatment of seasonal products changes

Seasonal products are products or services that are on offer in certain seasons of the year. In Finland seasonal products are such as seasonal clothes, men's and women's outdoor coats, winter coats and shoes and children's outerwear. In addition, utility goods connected to season, such as garden furniture, car summer and winter tyres, winter and summer sports equipment and spectator sports events such as football and ice hockey games are considered as seasonal products.

For seasonal products, the seasonal nature is taken into consideration when estimating the price development of the product or service. In practice, this means that when imputing missing prices, comparison data for imputation are searched

from the annual change of the same product or service in the previous year’s corresponding period or alternatively using the carry-forward method corrected with seasonal adjustment.

Example:

Ice hockey games were totally discontinued in mid-March. Prices could not be sufficiently collected in price collection, so the only alternative is to take the previous year's annual change and use it as a correction coefficient when estimating the current price for attending an ice hockey game.

Weight structure and commodities in index calculation

Weight structure

The weights of commodities are determined according to the previous year's individual consumption and they are kept the same throughout the year. According to the EU guidelines, commodity weights are not changed during the year, even if a distinct fall takes place in consumption for certain commodities, e.g. flights, restaurants, cafes.

Commodities in the Consumer Price Index

The commodity division of the Consumer Price Index is based on the international eCOICOP commodity classification (European Classification of Individual Consumption according to Purpose), in which products and services are divided into 12 main groups. The main groups are divided hierarchically into more detailed sub-groups so that the most detailed level separates products or services by means of the 7-digit commodity code.

Due to the restrictive measures connected to the coronavirus, some products or services are not available on the market or their availability is limited. Despite this, the commodity structure is kept unchanged in index calculation and when publishing statistical data; no commodity group is left out of reporting.

Imputation level in the Consumer Price Index

For March, the price collection was carried out more or less with the conventional model. For a few commodity groups the imputation level rose slightly compared to an average month.

More imputation of missing data will be done in April, because it is inevitable that the data collection cannot be carried out to the conventional extent. Information on these changes will be given in connection with the release for April.