Impact of imputation methods on the CPI and HICP in view of the COVID-19 crisis

Koen Link and Antonio Chessa[1]

April 20, 2020

Abstract

The SARS-CoV-2 virus and COVID-19 disease outbreak have led to a worldwide economic inactivity of sectors that are most affected by the imposed social distancing measures. Eurostat has recently set up guidelines, which should help national statistical agencies to continue publication for product aggregates for which sales have collapsed or price collection is temporarily impossible. This paper describes and examines the impact of three imputation methods on the overall Dutch CPI and HICP. The study is confined to seasonal items, specifically to flights and package holidays. This paper recommends imputing aggregate indices with their month on month index from the previous year. The results show that the imputed indices for flights and package holidays have a negligible effect on most all-items annual rates of change in the CPI and HICP. The recommended method requires little implementation effort, as it naturally fits with common imputation practices. This study also suggests a simple way of monitoring the performance of the method ahead of monthly publication.

Keywords: CPI, price imputation, seasonal items, market crash, COVID-19.

1 Introduction

Since its first outbreak in the Chinese city of Wuhan in December 2019, the SARS-CoV-2 or coronavirus 2 made its appearance in Europe with a number of positive test cases by the end of January 2020. After a slow increase in the number of cases in February, the numbers of infected persons started to increase in an uncontrolled way in the first week of March. These rapid developments led governments by mid March to impose strict measures on people's conduct and on the availability of public services and businesses.

In particular the social distancing measures have a huge impact on nations' economies. Restaurants and pubs have closed, services where clients get in close contact with service providers have been suspended, airlines have suspended their flights until further notice and travel agencies do not allow customers to book holidays for at least the next two months in the Netherlands.

[1]Statistics Netherlands, CPI department. E-mail: k.link@cbs.nl, ag.chessa@cbs.nl

The economic inactivity in many sectors poses National Statistical Institutes (NSIs) in front of huge challenges. Sales values will collapse to (nearly) zero in the aforementioned and other sectors in the coming months, which has direct implications for the compilation and publication of important statistics. In particular regularly published statistics will have to respond quickly to the abruptly changed situation and NSIs have to come up with acceptable solutions in order to meet publication deadlines. The CPI and HICP are an example of such statistics, which have to be compiled and published each month in most countries.

Expenditures will get close or equal to zero for a considerable number of COICOPs in at least the next two months. Such COICOPs can be found in different divisions, which are partly affected or even up to their full extent. An example of the latter is division 11 Restaurants and hotels, while divisions 07 Transport and 09 Recreation and culture contain the aggregates 0733 Passenger transport by air and 096 Package holidays respectively, which are challenging because of their seasonal nature and their weights in the CPI and HICP.

The sum of the weights of the COICOPs for which expenditures have collapsed lies between 10 and 15 per cent in the Dutch CPI. The high weight share and seasonal character of several aggregates are factors that may cause a severe impact on the overall CPI and HICP. Aggregates with (nearly) zero expenditures should not or hardly affect the all-items annual rates of change. From a transactions point of view, price indices with zero sales are undefined, as quantities sold and monthly expenditure shares of such items are equal to zero. However, the use of yearly fixed weights at COICOP level forces NSIs to compile indices for such aggregates, since the weights cannot be modified in the course of a year.

Eurostat has recently issued a guidance note, which contains a description of imputation methods that NSIs are allowed to use [1]. Seasonal items are given special attention in this note. Eurostat states that seasonal patterns should be preserved and advocates three imputation methods: (1) according to the annual rate of change of so-called "reliable indices", (2) carry forward with a seasonal correction factor, or (3) by taking the month on month index from the previous year of a seasonal item. The month on month index of all "reliable indices" in the current period will break seasonal patterns but is allowed for other items that are not or hardly transacted.

This paper describes and compares three imputation methods: aforementioned methods (1) and (3) and the month on month index of the current period. Although the latter method is not recommended for seasonal items in the Eurostat guidance note, we included this method in the present study in order to assess the performance of a method that does not continue seasonal patterns. The three methods were applied to passenger transport by air (flights) and package holidays. Dutch CPI and HICP data from 2018 and 2019 were used for this purpose.

Throughout this paper we will use a different terminology than in the Eurostat note. Eurostat defines an aggregate as "reliable" if it is imputed for less than 50 per cent of its weight ([1], p. 5). Since the aggregates that are severely affected by the COVID-19 crisis constitute a large share of the CPI and HICP, we decided to use the simplified and more direct distinction between aggregates that are "economically active" (instead of

"reliable"), where transactions still take place on a large scale, and aggregates with low or zero economic activity. Aggregates that are economically active or have a low or zero activity should be identified case by case by NSIs.

This paper is organised as follows. The three imputation methods are described in Section 2. The index formulas are given for each method and the expressions for the all-items indices and year on year indices are derived and compared. Section 3 presents the results of a case study, in which the three imputation methods are applied to flights and package holidays in the Dutch CPI and HICP. The analytical and empirical findings are used in Section 4.2 to formulate a recommended approach for imputing prices and indices for aggregates that are affected during the coronavirus crisis. Section 5 concludes and mentions possible directions for further research.

2 Imputation methods

This section describes three imputation methods that are also recommended in the Eurostat note [1]. Index formulas are derived for each method, followed by an analytical comparison of the three methods. The following notation is introduced for this purpose.

The set of aggregates for which markets are economically active is denoted by A^+. The set A^- consists of aggregates for which markets have collapsed, in the sense that economic activity is limited or even zero for some time. The set of all aggregates that fully covers the CPI or HICP is denoted by A, so it is the union of A^+ and A^-. The price index $P^a_{y,m}$ is the index of aggregate a in month m of year y. The base period of this index (i.e., a month or year) is typically established at a rebasing of the CPI (e.g., the year 2015 in the Dutch CPI). The annual fixed weight of aggregate a in year y is denoted as w^a_y. The weights are normalised in this study, such that $w^a_y \in [0, 1]$ for all $a \in A$ and $\sum_{a \in A} w^a_y = 1$.

2.1 Imputation with all-items year on year index

The first imputation method estimates an index for an aggregate $a \in A^-$, for which sales have dropped to (nearly) zero, by linking the year on year index of all economically active aggregates in A^+ to the index of 12 months ago. The index in month m of year y that corresponds with A^+ is denoted as $P^+_{y,m}$. This index of the set of economically active aggregates is sometimes also referred to as "all-items index", since the aim of using these aggregates is to neutralise the effect of low activity aggregates on the overall CPI and HICP. This imputation method yields an index in month m of year y as follows for an aggregate $a \in A^-$:

$$\hat{P}^a_{y,m} = P^a_{y-1,m} \cdot \frac{P^+_{y,m}}{P^+_{y-1,m}} \tag{1}$$

Throughout this paper we take the situation where month m of year y is the first month in which indices are imputed. With regard to the current crisis, this period corresponds with April 2020. Prices could still be collected in March, as the government measures entered into force from the second half of March. However, the formulas presented in this paper equally apply to successive months as well. The only difference is that some

notation would have to be adapted. For instance, if an index has to be imputed in May 2020, then the notation for the index of some aggregate a in the preceding month should be changed to $\hat{P}^a_{y,m-1}$. Furthermore, we assume that imputations will be needed only for the current year y. In case they would have to be done for an extended period, then also the notation of the indices in the previous year should be adjusted, as some of these indices are imputed.

Imputation with an all-items index could be considered in situations where the parent or nearest aggregate index does not have a strong relation with the aggregate that has to be imputed. For example, the parent of 096 Package holidays is 09 Recreation and culture. The aggregates in the parent do not have items that are related to package holidays. The same applies to 0733 Passenger transport by air (international flights), where the parent aggregate 073 Transport services consists of railway, road and waterway transport as the other main aggregates.

An additional motivation for choosing the year on year index of the economically active aggregates to impute indices of aggregates for which sales have collapsed is that these aggregates will hardly affect the year on year indices of the overall CPI or HICP. Flights and package holidays have been drastically reduced or suspended during the COVID-19 crisis, so that the expenditures of these items are close or equal to zero. Hardly any contribution from these items to the overall CPI should therefore be expected, which is mimicked by imputing with the year on year index for all other aggregates where transactions do take place. On the other hand, the index of the parent aggregate is likely to be affected, so the idea is to give precedence to the more important annual rate of change based on the overall CPI.

Let $P_{y,m}$ denote the index for the overall CPI (i.e., for the set A) in month m of year y and let $\hat{P}_{y,m}$ denote the index after imputation. Price indices for COICOPs are calculated by generating so-called "short" annual series, with December of the previous year as base month. These index series are combined with the annual weights w^a_y to obtain series for higher aggregates. The short series are linked to the index in the base month to yield "long" index series.

If we write $w^+_y = \sum_{a\in A^+} w^a_y$ for the sum of the weights of the aggregates for which transactions still take place on a large scale, then we can write the expression for the overall CPI in month m of year y as follows when imputing according to (1):

$$\hat{P}_{y,m} = P_{y-1,12} \sum_{a\in A} w^a_y \frac{P^a_{y,m}}{P^a_{y-1,12}} \tag{2}$$

$$= P_{y-1,12}\left(\sum_{a\in A^+} w^a_y \frac{P^a_{y,m}}{P^a_{y-1,12}} + \sum_{a\in A^-} w^a_y \frac{\hat{P}^a_{y,m}}{P^a_{y-1,12}} \right) \tag{3}$$

$$= P_{y-1,12}\left(w^+_y \frac{P^+_{y,m}}{P^+_{y-1,12}} + \sum_{a\in A^-} w^a_y \frac{P^a_{y-1,m}}{P^a_{y-1,12}} \frac{P^+_{y,m}}{P^+_{y-1,m}} \right) \tag{4}$$

Note that substitution of (1) in (3) yields expression (4).

The year on year index of the overall CPI can be related to the year on year index for

the set of economically active aggregates A^+ by slightly rewriting expression (4):

$$\frac{\hat{P}_{y,m}}{P_{y-1,m}} = \frac{P_{y-1,12}}{P_{y-1,m}}\left(w_y^+ \frac{P^+_{y-1,m}}{P^+_{y-1,12}} + \sum_{a\in A^-} w_y^a \frac{P^a_{y-1,m}}{P^a_{y-1,12}}\right)\frac{P^+_{y,m}}{P^+_{y-1,m}} \tag{5}$$

A number of observations can be made with regard to this expression. First, the year on year index for the overall CPI is equal to a factor multiplied by the year on year index for the aggregates in A^+, where the latter is used to impute indices according to (1) for aggregates with collapsed sales. The year on year indices for each of these aggregates are thus equal to $P^+_{y,m}/P^+_{y-1,m}$, but this does not hold for the year on year index $\hat{P}_{y,m}/P_{y-1,m}$ of the overall CPI.

Second, the term within brackets in (5) is the sum of the weights of all aggregates in the CPI, which are price backdated to month m of year $y-1$. In general, the ratio $P_{y-1,12}/P_{y-1,m}$ will not be equal to the term within brackets, so that $\hat{P}_{y,m}/P_{y-1,m}$ will differ from $P^+_{y,m}/P^+_{y-1,m}$. We expect this difference to be small in practice and that it will be driven by the differences between the weights of the aggregates in $y-1$ and y. We will investigate this in Section 3.

We conclude this subsection by listing some pros and cons of the imputation method:

- A major advantage of the method is that the impact of the imputed indices on the overall annual rate of inflation can be controlled and is expected to be limited.
- A second advantage is that the method links on regularly calculated indices, that is, on indices that do not suffer from the estimation problems caused by the current crisis.
- The imputation methods used for the Dutch CPI make use of month on month indices. Imputing according to expression (1) would require manual interventions, which makes the system sensitive to errors.

We will take these points into account in our recommendations for an approach to be followed in the next months in the Dutch CPI and HICP (Section 4.2).

2.2 Imputation with monthly change from previous year

With regard to seasonal items, the Eurostat guidance note states that "NSIs should ensure that the imputations do not break the seasonal pattern of the series" (see [1], p. 6). The method described in the previous subsection is expected to preserve seasonal patterns, since the year on year index $P^+_{y,m}/P^+_{y-1,m}$ used for the imputations is usually rather stable. Seasonal patterns will also be preserved by the method described in the present subsection.

The second imputation method calculates an index for an aggregate $a \in A^-$ in month m of year y by linking the month on month index from the previous year of the same aggregate to the index in the previous month of the current year:

$$\hat{P}^a_{y,m} = P^a_{y,m-1} \cdot \frac{P^a_{y-1,m}}{P^a_{y-1,m-1}} \tag{6}$$

This imputation method gives rise to the following year on year index of aggregate a:

$$\frac{\hat{P}^a_{y,m}}{P^a_{y-1,m}} = \frac{P^a_{y,m-1}}{P^a_{y-1,m-1}} \tag{7}$$

The year on year index in month m is thus equal to the year on year index in the preceding month, which means that this imputation method is equivalent with carrying forward the year on year index of an aggregate.

We now derive an expression for the overall index in month m of year y. Substituting expression (6) in (3) gives:

$$\hat{P}_{y,m} = P_{y-1,12}\left(w^+_y \frac{P^+_{y,m}}{P^+_{y-1,12}} + \sum_{a \in A^-} w^a_y \frac{P^a_{y-1,m}}{P^a_{y-1,12}} \frac{P^a_{y,m-1}}{P^a_{y-1,m-1}} \right) \tag{8}$$

The year on year index that follows from (8) can be rewritten in a form similar to (5). However, it is more instructive to relate the year on year indices that result from the two imputation methods. To this end, we denote the overall index for the method presented in Section 2.1 by $\hat{P}^{(1)}_{y,m}$ and the overall index given by (8) by $\hat{P}^{(2)}_{y,m}$. The difference between the year on year indices for the two imputation methods can be expressed as follows:

$$\frac{\hat{P}^{(1)}_{y,m}}{P_{y-1,m}} - \frac{\hat{P}^{(2)}_{y,m}}{P_{y-1,m}} = \frac{P_{y-1,12}}{P_{y-1,m}} \sum_{a \in A^-} w^a_y \frac{P^a_{y-1,m}}{P^a_{y-1,12}} \left(\frac{P^+_{y,m}}{P^+_{y-1,m}} - \frac{P^a_{y,m-1}}{P^a_{y-1,m-1}} \right) \tag{9}$$

It may be clear that the difference between the two year on year indices only contains a summation over the aggregates that are imputed.

It is interesting to note that the term within brackets is a difference between two year on year indices. The first ratio is the year on year index that corresponds with the economically active aggregates, while the second one is the year on year index of aggregate a in the previous month, which follows from expression (6). Year on year indices can be expected to be more stable than month on month indices for seasonal items. As a consequence, the difference between the overall year on year indices for the two imputation methods could be small. This will be investigated in Section 3.

The imputation method described in the present subsection has a number of additional properties, which are summarised below as pros and cons of the method:

- In contrast with the imputation method in Section 2.1, the method defined by expression (6) results in imputed indices that only make use of prices and indices of the aggregate that is imputed.

- A potential drawback of the method is that it uses the previous month as linking month in order to derive an imputed index. The index in the previous month may be imputed or based on a sparse sample, which consequently influences the reliability of the imputed index in the current month.

- In addition to the previous point, note that the right-hand side of expression (9) only contains one index in the current period. Moreover, the index $P^+_{y,m}$ is expected

to be more stable than the index for the overall CPI, as seasonal items like flights and package holidays are excluded. By making an expert guess for $P^+_{y,m}$ prior to publication, we can use expression (9) to derive an estimate of the difference between the year on year indices for the imputation methods in sections 2.1 and 2.2. This gives NSIs the possibility for timely tracking the validity of the second method, if imputation by the all-items year on year index would be used as a benchmark.

The last point is one of the useful insights obtained from this study. If NSIs would choose the imputation method described in this subsection, then they have the possibility of monitoring its behaviour compared to the first method ahead of monthly publication. We will take this additional feature into account in our final recommendations (Section 4.2).

2.3 Imputation with all-items month on month index

The third method described in this paper is based on one of the standard imputation methods used in the Dutch CPI. If an index of some aggregate has to be imputed in the current month, then the nearest aggregate is usually selected and its month on month index is used to derive an index for the aggregate to be imputed. In the exceptional situation where sales have collapsed, it can be argued to take the month on month index of all economically active aggregates in order to impute an index. The underlying thought is similar to the one behind the method described in Section 2.1, in the sense that the influence on the overall CPI from aggregates without sales is neutralised. Instead of choosing the year on year index, the focus now is on the month on month index.

An imputed index for an aggregate $a \in A^-$ in month m of year y is obtained as follows:

$$\hat{P}^a_{y,m} = P^a_{y,m-1} \cdot \frac{P^+_{y,m}}{P^+_{y,m-1}} \tag{10}$$

Substitution of expression (10) in (3) gives the following expression for the all-items index:

$$\hat{P}_{y,m} = P_{y-1,12}\left(w^+_y \frac{P^+_{y,m}}{P^+_{y-1,12}} + \sum_{a \in A^-} w^a_y \frac{P^a_{y,m-1}}{P^a_{y-1,12}} \frac{P^+_{y,m}}{P^+_{y,m-1}} \right) \tag{11}$$

It can be easily verified that the month on month index that follows from this expression is equal to $P^+_{y,m}/P^+_{y,m-1}$. This means that imputation according to expression (10) will fully neutralise the impact of aggregates with collapsed sales on the month on month index of the total CPI. This can be expected since this imputation method merely uses prices and indices from the current year. We are therefore not left with a factor like in expression (5) when using the year on year index for imputations.

Like was done in Section 2.2, it is instructive to relate the year on year index for the overall CPI to one of the other two methods. A suitable choice is the method described in Section 2.2. We denote the overall index given by expression (11) by $\hat{P}^{(3)}_{y,m}$. Since both methods link on the index of the previous month, we need to distinguish between the imputed indices in the previous month for the two methods from the second month of imputation. We therefore write the imputed indices for an aggregate $a \in A^-$ as $\hat{P}^{a\,(2)}_{y,m-1}$

and $\hat{P}^{a\,(3)}_{y,m-1}$ for the two methods that are compared. The difference between the year on year indices for the two methods can be written as follows:

$$\frac{\hat{P}^{(3)}_{y,m}}{P_{y-1,m}} - \frac{\hat{P}^{(2)}_{y,m}}{P_{y-1,m}} = \frac{P_{y-1,12}}{P_{y-1,m}} \sum_{a \in A^-} w^a_y \left(\frac{\hat{P}^{a\,(3)}_{y,m-1}}{P^a_{y-1,12}} \frac{P^+_{y,m}}{P^+_{y,m-1}} - \frac{\hat{P}^{a\,(2)}_{y,m-1}}{P^a_{y-1,12}} \frac{P^a_{y-1,m}}{P^a_{y-1,m-1}} \right) \tag{12}$$

If the current month m is the first month of imputation, then this expression can be written in the following simplified form:

$$\frac{\hat{P}^{(3)}_{y,m}}{P_{y-1,m}} - \frac{\hat{P}^{(2)}_{y,m}}{P_{y-1,m}} = \frac{P_{y-1,12}}{P_{y-1,m}} \sum_{a \in A^-} w^a_y \frac{P^a_{y,m-1}}{P^a_{y-1,12}} \left(\frac{P^+_{y,m}}{P^+_{y,m-1}} - \frac{P^a_{y-1,m}}{P^a_{y-1,m-1}} \right) \tag{13}$$

One of the most interesting terms in (13) is the term within brackets. It is the difference between two month on month indices: one of the indices applies to the set of economically active aggregates, for which transactions still take place, while the second one is the month on month index of a specific aggregate a in the previous year $y-1$ that has to be imputed in the current period. This difference can become quite large for aggregates with volatile indices, such as flight tickets and package holidays. We therefore expect bigger differences compared with the first two methods, as the difference term in expression (9) consists of year on year indices.

Also this section concludes with a list of pros and cons of the method just described:

- Imputing according to expression (10) has the advantage of neutralising the impact of aggregates with collapsed sales on the month on month index of the overall CPI.
- The method is also technically in line with common imputation practices in the Dutch CPI.
- Like the method in Section 2.2, a potential drawback is that it uses the previous month as linking month in order to derive an imputed index. The index in the previous month may be imputed or based on a sparse sample, which consequently influences the reliability of the imputed index in the current month.
- As was already pointed out, a second disadvantage of the method is that the difference terms in (12) and (13) may become quite large.
- As was already observed in Section 2.2, also expressions (12) and (13) offer the possibility of monitoring the behaviour of the index $\hat{P}^{(3)}_{y,m}$ and its year on year index, in this case with respect to the method presented in the previous subsection.

The three methods described in this paper have their pros and cons, and each method is expected to give different results. The methods in sections 2.1 and 2.2 preserve seasonal patterns, and the first method also directly controls the all-items year on year index. The method presented in this subsection neutralises the impact of aggregates with zero sales on the all-items month on month index, but will break seasonal patterns. In the next section we will investigate how these two groups of methods perform on data drawn from the Dutch CPI and HICP.

Beside methodological and empirical results, an additional aspect that has to be taken into account is the amount of work needed in order to implement and use a method in practice, as resources at statistical agencies are often limited and also time in this particular event.

3 Results

3.1 Impact on overall CPI and HICP

We have applied the three imputation methods to 0733 Passenger transport by air and 096 Package holidays by using data from 2018 and 2019 of the Dutch CPI and HICP. Imputed indices were calculated for these two aggregates from April 2019 until and including December 2019. The resulting year on year and month on month indices were compared with the overall CPI and HICP as published in 2019. We first applied the method that imputes with the year on year index for all other aggregates. In this particular case study, this set of aggregates not only consists of all economically active aggregates, but also of aggregates with low or zero activity during the crisis, like restaurants and hairdressers. The results for this method are summarised in Table 1.

CPI	**19.01**	**19.02**	**19.03**	**19.04**	**19.05**	**19.06**	**19.07**	**19.08**	**19.09**	**19.10**	**19.11**	**19.12**	**Average**
Published (P)	2.2	2.6	2.8	2.9	2.4	2.7	2.5	2.8	2.6	2.7	2.6	2.7	2.6
Method (M)	2.2	2.6	2.8	2.7	2.6	2.5	2.5	2.7	2.5	2.5	2.5	2.7	2.6
M minus P	0	0	0	-0.2	0.2	-0.2	0.0	-0.1	-0.1	-0.2	-0.1	0.0	0.0
HICP	**19.01**	**19.02**	**19.03**	**19.04**	**19.05**	**19.06**	**19.07**	**19.08**	**19.09**	**19.10**	**19.11**	**19.12**	**Average**
Published (P)	2.0	2.6	2.9	3.0	2.3	2.7	2.6	3.1	2.7	2.8	2.6	2.8	2.7
Method (M)	2.0	2.6	2.9	2.8	2.5	2.6	2.6	3.0	2.6	2.5	2.5	2.7	2.6
M minus P	0	0	0	-0.2	0.2	-0.1	0.0	-0.1	-0.1	-0.3	-0.1	-0.1	-0.1

Table 1: CPI and HICP annual rates of change (%) when imputing with the all-items year on year index.

Both the CPI and HICP show lower year on year indices in April 2019 compared to the published indices. The imputed year on year indices for flights and packages holidays (+2.7 per cent in both cases) are lower than the published CPI figures (+16.6 and +4.9 per cent respectively). Subsequent months show similar or smaller differences.

Expression (5) in Section 2.1 shows that the year on year index of the overall CPI is equal to a factor multiplied by the year on year index $P^+_{y,m}/P^+_{y-1,m}$ of the set of all other aggregates, which was used to impute the indices of flights and package holidays. It is reasonable to require a negligible or zero impact of these aggregates on the overall year on year index. We calculated the aforementioned factor in expression (5). The monthly values in the period April-December 2019 lie between 0 and 0.043, which means that the factor has no effect up to the first decimal of the overall annual rate of change. The imputed aggregates therefore hardly affect the year on year indices of the CPI and HICP when using this method. The imputation method also produces results that are very close to the published figures.

The second method described in Section 2 imputes indices of flights and package holidays by taking their month on month indices from the previous year. The results for

this method are shown in Table 2. The results hardly differ from those in Table 1. This confirms our expectation expressed in Section 2.2. The difference between the all-items year on year indices for the two imputation methods, as formalised by expression (9), contains a difference term that compares two year on year indices. The impact of these differences for flights and package holidays on the overall annual rates of change indeed turns out to be very small and is even negligible in most months.

CPI	**19.01**	**19.02**	**19.03**	**19.04**	**19.05**	**19.06**	**19.07**	**19.08**	**19.09**	**19.10**	**19.11**	**19.12**	**Average**
Published (P)	2.2	2.6	2.8	2.9	2.4	2.7	2.5	2.8	2.6	2.7	2.6	2.7	2.6
Method (M)	2.2	2.6	2.8	2.7	2.6	2.5	2.5	2.7	2.5	2.4	2.5	2.6	2.5
M minus P	0	0	0	-0.2	0.2	-0.2	0.0	-0.1	-0.1	-0.3	-0.1	-0.1	-0.1
HICP	**19.01**	**19.02**	**19.03**	**19.04**	**19.05**	**19.06**	**19.07**	**19.08**	**19.09**	**19.10**	**19.11**	**19.12**	**Average**
Published (P)	2.0	2.6	2.9	3.0	2.3	2.7	2.6	3.1	2.7	2.8	2.6	2.8	2.7
Method (M)	2.0	2.6	2.9	2.8	2.5	2.5	2.5	3.0	2.5	2.5	2.5	2.7	2.6
M minus P	0	0	0	-0.2	0.2	-0.2	-0.1	-0.1	-0.2	-0.3	-0.1	-0.1	-0.1

Table 2: Annual rates of change (%) when imputing with the aggregates' month on month indices from the previous year.

The third method imputes indices by taking the month on month index of all other aggregates in the current month, which is identical to the all-items month on month index. The results for this imputation method are shown in Table 3.

CPI	**19.01**	**19.02**	**19.03**	**19.04**	**19.05**	**19.06**	**19.07**	**19.08**	**19.09**	**19.10**	**19.11**	**19.12**	**Average**
Published (P)	2.2	2.6	2.8	2.9	2.4	2.7	2.5	2.8	2.6	2.7	2.6	2.7	2.6
Method (M)	2.2	2.6	2.8	2.6	2.4	2.5	1.9	2.2	2.5	2.4	2.5	2.6	2.4
M minus P	0	0	0	-0.3	0.0	-0.2	-0.6	-0.6	-0.1	-0.3	-0.1	-0.1	-0.2
HICP	**19.01**	**19.02**	**19.03**	**19.04**	**19.05**	**19.06**	**19.07**	**19.08**	**19.09**	**19.10**	**19.11**	**19.12**	**Average**
Published (P)	2.0	2.6	2.9	3.0	2.3	2.7	2.6	3.1	2.7	2.8	2.6	2.8	2.7
Method (M)	2.0	2.6	2.9	2.7	2.4	2.6	1.9	2.4	2.5	2.4	2.5	2.7	2.5
M minus P	0	0	0	-0.3	0.1	-0.1	-0.7	-0.7	-0.2	-0.4	-0.1	-0.1	-0.2

Table 3: Annual rates of change (%) when imputing with the all-items month on month index.

This imputation method shows the largest differences, both compared with the published figures and with the two other imputation methods. Imputing with the all-items month on month index breaks the seasonal pattern, which is the main reason for the large differences observed in July and August compared to the published figures.

The differences with the other two imputation methods can be traced back to expression (12) in Section 2.3. The difference term in that expression compares the month on month indices for flights and package holidays with the month on month index for all other aggregates. The differences between the aggregate and all-items month on month indices may become very large for seasonal items and may thus severely affect the all-items year on year indices. This was already noted in Section 2.3 and is confirmed by the results in Table 3.

The right-hand sides of expressions (9) and (12) contain the contribution of each aggregate to the difference between the all-items year on year indices for two imputation methods. These differences allow us to quantify the impact of each aggregate on the all-items year on year index. The results are shown in Figure 1. The graph on the left shows the contribution of each aggregate to the difference between the all-items year on year

indices for the method that uses the year on year index of all other aggregates and the method that takes the month on month index from the previous year for imputation. The first method directly controls and neutralises the impact on the all-items annual rate of change, which means that we can assess the capability of the other method in neutralising the impact of flights and package holidays on the all-items year on year index.

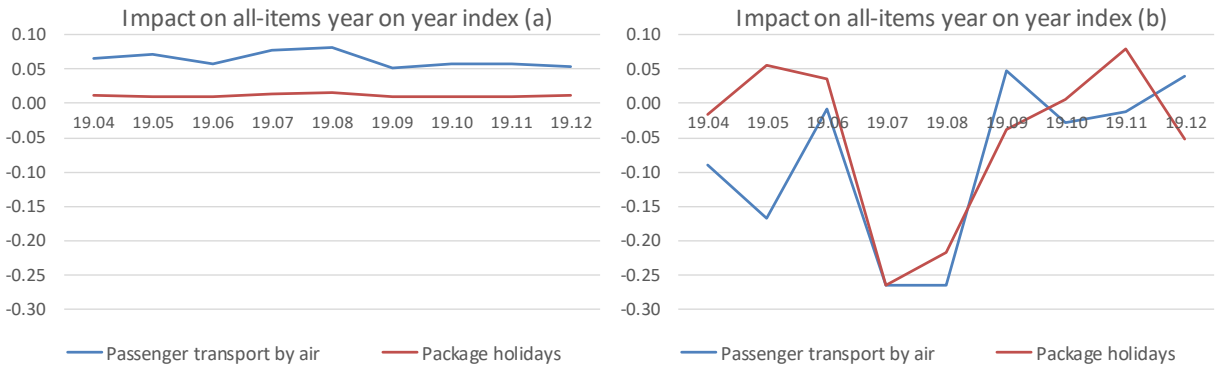

Figure 1: Contribution of flights and package holidays, in percentage points, to the difference between the all-items year on year indices for the CPI according to (a) expression (9) and (b) expression (12).

Figure 1 shows that the impact of package holidays is almost zero, while the impact of flights is slightly larger but still confined to several hundredths of a percentage point. This means that the method that imputes with the month on month index from the previous year shows excellent performance as well. The graph on the right quantifies in fact the performance of the third method, which imputes indices according to the all-items month on month index. The results show that this method has a significant impact on the all-items year on year index, and that the all-items index is affected by both aggregates. Adding the contributions for each aggregate gives the differences between the three imputation methods that can be derived from tables 1-3. The differences do not fully match in some months because of rounding effects. The results clearly show that breaking the seasonal pattern may adversely affect the all-items annual rate of change.

The month on month indices for the three imputation methods are shown in Table 4. Also these results show substantial differences between the published figures and the method that imputes according to the all-items month on month index. The implications of the break in the seasonal patterns can be noted in particular in July and September. The other two methods preserve seasonality and result in much smaller differences with the published CPI and HICP.

The comparisons with the published figures are not meant to express a preference in favour of a certain method or to disqualify other methods. The implications of using imputation methods that either preserve seasonal patterns or not are illustrated and discussed. The method that uses the all-items month on month index neutralises the effect of the imputed aggregates on the all-items month on month CPI and HICP, which is an additional desirable feature beside neutralising the effect on the all-items annual rates of change. Table 4 therefore also tells us to what extent the other two methods, denoted as 'All-items YoY' and 'MoM $y-1$' in the table, are capable of limiting the impact of imputations on the all-items month on month indices. Apart from July and September, the two methods perform very well in the other months.

CPI	**19.01**	**19.02**	**19.03**	**19.04**	**19.05**	**19.06**	**19.07**	**19.08**	**19.09**	**19.10**	**19.11**	**19.12**
Published	0.1	0.9	0.4	0.7	-0.1	-0.1	1.0	0.4	-0.6	0.4	-0.4	0.1
All-items YoY	0.1	0.9	0.4	0.5	0.2	-0.3	1.1	0.4	-0.7	0.3	-0.2	0.1
MoM $y-1$	0.1	0.9	0.4	0.5	0.2	-0.4	1.1	0.4	-0.7	0.3	-0.2	0.0
All-items MoM	0.1	0.9	0.4	0.4	0.2	-0.2	0.6	0.4	-0.2	0.3	-0.1	0.0
HICP	**19.01**	**19.02**	**19.03**	**19.04**	**19.05**	**19.06**	**19.07**	**19.08**	**19.09**	**19.10**	**19.11**	**19.12**
Published	-0.2	1.3	0.5	1.1	-0.3	-0.1	0.9	0.6	-1.1	0.6	-0.8	0.2
All-items YoY	-0.2	1.3	0.5	0.9	0.1	-0.4	1.1	0.5	-1.1	0.5	-0.5	0.2
MoM $y-1$	-0.2	1.3	0.5	0.8	0.1	-0.5	1.1	0.5	-1.2	0.5	-0.5	0.2
All-items MoM	-0.2	1.3	0.5	0.7	0.1	-0.3	0.4	0.6	-0.6	0.5	-0.4	0.1

Table 4: Monthly rates of change (%) for the three methods compared with the published figures.

3.2 Impact at division level

Although the methods described in this paper do not explicitly control the impact on the parent aggregates of flights and package holidays, it is nevertheless interesting to look at the indices produced by these methods at levels below the all-items level. To this end, we also calculated the indices at division level, that is, for 07 Transport and 09 Recreation and culture. The year on year indices for the two divisions that result from the three imputation methods are compared with the year on year indices after excluding flights and package holidays. The results are shown in Figure 2.

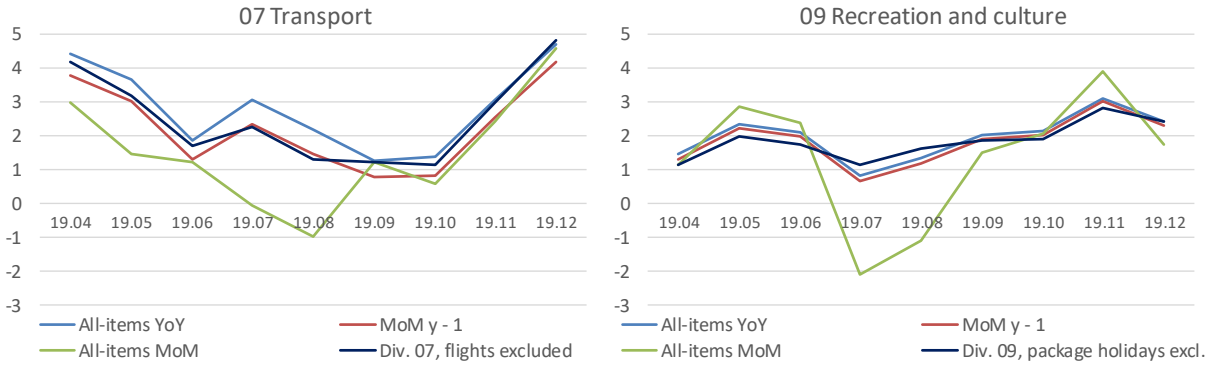

Figure 2: Year on year indices, in percentage points, for the three imputation methods at division level compared with the CPI division indices after excluding flights and package holidays.

The results show that the method that imputes with the year on year index for all other aggregates and the method that uses the month on month indices from the previous year of flights and package holidays compare very well with the division indices from which flights and package holidays are excluded. The method that imputes with the all-items month on month index leads to large deviations of up to three percentage points.

Two of the three methods give promising results, which demonstrate that these methods are not only able to limit or neutralise the impact of aggregates of low or zero economic activity on the all-items annual rates of change. The same effect should be required for parent aggregates. Although we did not have the time to explore ways of imposing this as an additional constraint on the methods, the two methods that give accurate results at overall CPI and HICP level are also able to produce satisfactory results at division level in an unconstrained way.

4 Discussion

4.1 Preliminary remarks

Ideally, a benchmark index would have to be defined in order to assess the performance of the three imputation methods and to make a well-informed choice. A benchmark index is hard to define, because we are trying to find solutions to a problem that introduces a highly dynamic situation regarding expenditure changes, while the CPI and HICP make use of yearly fixed weights for a fixed set of higher aggregates.

Different imputation methods may lead to different results. As the annual rate of change is the most important statistic in the CPI and HICP, the natural question arises what impact on the all-items year on year index could be expected from aggregates for which sales drop to zero. So far in this study we have insisted that aggregates with low or zero economic activity should hardly affect the all-items year on year index. We could gain more insight in this assertion by deriving an expression for the all-items rate of change.

In addition to the notation that was introduced in Section 2, let $P^-_{y,m}$ denote the price index of the set of aggregates with low or zero economic activity in month m of year y and let w^-_y denote the total weight of this set in year y. The all-items year on year index can be expressed as follows:

$$\frac{P_{y,m}}{P_{y-1,m}} = c\left(w^+_y + w^-_y \frac{P^-_{y,m}/P^+_{y,m}}{P^-_{y-1,12}/P^+_{y-1,12}}\right)\frac{P^+_{y,m}}{P^+_{y-1,m}} \tag{14}$$

in which the factor c is a function of price indices in $y-1$ and $y-2$ and weights in $y-1$:

$$c = \frac{1 + \frac{w^-_{y-1}}{w^+_{y-1}}\frac{P^-_{y-1,12}/P^+_{y-1,12}}{P^-_{y-2,12}/P^+_{y-2,12}}}{1 + \frac{w^-_{y-1}}{w^+_{y-1}}\frac{P^-_{y-1,m}/P^+_{y-1,m}}{P^-_{y-2,12}/P^+_{y-2,12}}} \tag{15}$$

The values of the factor c multiplied by the term within brackets in (14) oscillate around 1, as is shown in Figure 3 for the data used in this study.

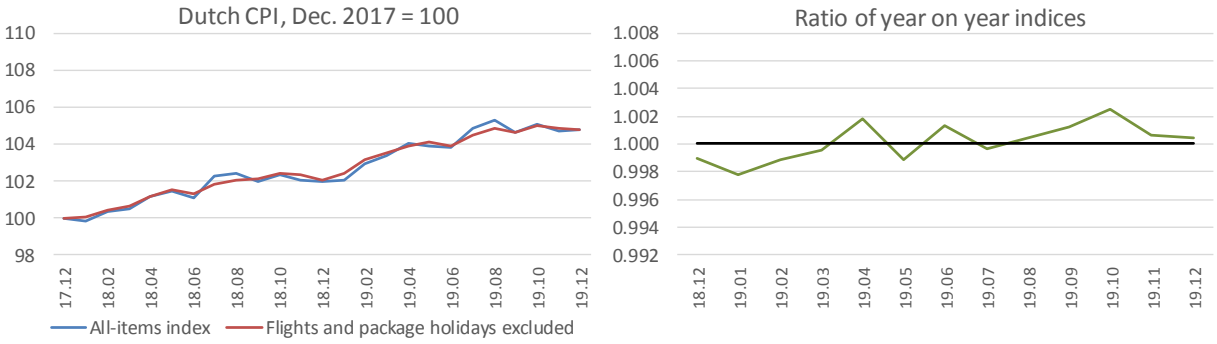

Figure 3: Left: all-items CPI and index for flights and package holidays excluded. Right: ratio of the two year on year indices.

The index $P^-_{y,m}$ within brackets in (14) does not exist when transactions do not take place and is imputed. There is no reason to expect that the term within brackets will be

systematically larger or smaller than 1. According to this reasoning, and supported by the data shown in Figure 3, the year on year index of the set of all economically active aggregates would thus serve as an adequate benchmark and choice for imputing indices of aggregates with collapsed sales.

Obviously, the same reasoning could be applied to all indices, irrespective of the comparison month. But, as we already stated, different imputation methods will usually lead to different results. A choice therefore has to be made for neutralising the impact of aggregates with low or zero sales on an index, and we prefer to choose the all-items year on year index since it is the most important statistic in the CPI and HICP.

The problem of defining and calculating index contributions from aggregates with sales that drop to (nearly) zero from one month to the next finds a natural solution with methods that make use of monthly varying weights, like multilateral methods. Baskets are then automatically adapted to the real situation in each month. Products that (temporarily) disappear may have an upward or downward impact on a price index, but also with this type of methods it is not reasonable to expect a systematically positive or negative contribution from disappearing products.

4.2 Suggested approach

From a methodological point of view, the preferred choice among the three investigated methods would be to impute with the year on year index of the set of economically active aggregates. However, the results in Section 3 have shown that the method that uses the month on month index from the previous year of each aggregate hardly differs from the method that uses the year on year index of economically active aggregates.

The results from Section 3 also show that imputing with the all-items month on month index may lead to considerable deviations from the all-items year on year indices. The differences increase to several percentage points at division level, as was shown in Section 3.2. Using a method that does not explicitly control the all-items annual rates of change and that does not respect seasonal patterns is therefore very risky and should be avoided.

The method that takes the month on month index from the previous year of each aggregate has several advantages. One of these is that the method naturally fits within the traditional system that carries out imputations. Time and resources are limited, and a quick decision has to be made in view of the monthly publication deadlines. We therefore recommend the use of the method described in Section 2.2. This recommendation applies to seasonal items, but could be extended to all aggregates that have to be imputed. A more comprehensive list of reasons that support our recommendation is given below:

- Aggregates for which products are (almost) not offered anymore hardly affect the overall annual rate of inflation, as would be reasonable to expect. This is supported by methodological arguments (expression (9) in Section 2.2) and empirical results (Section 3).

- An elegant feature of the method is that it only makes use of prices and indices of the aggregate that is imputed. This is not the case with the other two methods, which use an all-items index for imputation.

- As was stated near the end of Section 2.2, the performance of the recommended method can be easily monitored ahead of monthly publication. This can be achieved by making an 'expert guess' of the index $P^{+}_{y,m}$ for all economically active aggregates and by calculating expression (9) in Section 2.2. This allows to quantify the difference with the method that uses the year on year index of all economically active aggregates for imputation (Section 2.1) and to keep track of this difference over time.

- As was already stated at the beginning of this section, the recommended method fits within traditional methods for imputation that are used for the Dutch CPI. Time and resources are limited and any risk of errors should be minimised, given the number and total weight of the aggregates that are involved.

We conclude this subsection by restating the expression for the year on year index that results from our recommended method. In order to illustrate its character, we slightly rewrite expression (9) as follows:

$$\frac{\hat{P}^{(2)}_{y,m}}{P_{y-1,m}} = \frac{\hat{P}^{(1)}_{y,m}}{P_{y-1,m}} + \frac{P_{y-1,12}}{P_{y-1,m}} \sum_{a \in A^{-}} w^{a}_{y} \frac{P^{a}_{y-1,m}}{P^{a}_{y-1,12}} \left(\frac{P^{a}_{y,m-1}}{P^{a}_{y-1,m-1}} - \frac{P^{+}_{y,m}}{P^{+}_{y-1,m}} \right) \tag{16}$$

The overall year on year indices are thus equal to the year on year index that is obtained with the method that imputes with the index of all economically active aggregates, plus a sum of difference terms that involve the imputed aggregates. This illustrates the elegance of the recommended method: the main term $\hat{P}^{(1)}_{y,m}/P_{y-1,m}$ controls the overall annual rate of change, to which aggregate specific terms are added. The difference terms turn out to be small or even negligible in the examples of Section 3. In addition, imputing with aggregate specific indices should lead to better results than the two other methods for the parent aggregates, on which the impact of aggregates with (nearly) zero sales should be negligible as well. In this respect, also the results at division level in Section 3.2 look promising.

With regard to the Dutch CPI, the recommended method requires the month on month indices from the previous year to be derived for each aggregate, which can then be used to impute prices in the current month according to the standard procedure. CPI and HICP figures can be generated without any form of manual intervention. Price indices will automatically capture the price changes between the pre- and post-epidemic months. Specific attention in this respect will focus on the method for package holidays, in particular for travels that are temporarily unavailable and for which prices are carried forward.

4.3 Outlook on 2021

Given the big implications of the COVID-19 crisis on the CPI and HICP in 2020, it is important to think about the possible effects of the three imputation methods in 2021. Each method will have different effects on the annual rates of change. How large these effects will be depends on how the market behaves after the crisis. Generally speaking, the effects in 2021 will be opposite to those that will be observed in 2020. The size of the

effects will obviously also depend on the weights of each aggregate that will be established later this year for 2021. The weights will be affected by the very low or zero sales in the months during the crisis. It is therefore important that NSIs work towards a harmonised approach for determining the weights of affected aggregates in 2021.

The imputation method that uses the all-items year on year index will probably generate an upward effect on the overall annual inflation rate in April 2021, compared to normal conditions (no crisis). Imputing with the all-items year on year index will have a downward effect on the year on year indices of flights and package holidays in April 2020. (For example, the year on year index of airfares in March 2020 was 13.9 per cent in the Dutch CPI, while the annual rate of change for all aggregates, with flights and package holidays excluded, was 1.4 per cent.) On the other hand, only a part of Easter will enter the sample in April 2021, which will reduce the aforementioned upward effect. The impact on the overall year on year index will probably be smaller in subsequent months. A bigger impact is likely to occur again in the month when flights are resumed.

The method that imputes with the month on month index from the previous year will show a smaller effect on the overall annual rate of change in April 2021. This imputation method is equivalent with carrying forward the year on year index of the previous month, in this case March 2021. This expectation follows from the larger year on year index for airfares compared with the annual rate of change for all other items in March 2020, as was mentioned in the preceding paragraph. Since the differences with the method that imputes with the all-items year on year index were found to be small in Section 3, we expect that the differences between the two methods at all-items level will also be small in 2021.

The method that imputes with the all-items month on month index may deviate considerably from the other two methods. The use of the all-items month on month index eliminates the seasonal pattern from the indices of airfares and packages holidays. This is expected to generate a substantial impact on the overall annual rates of change in 2020 and 2021.

5 Conclusions

The wide-ranging implications of the COVID-19 outbreak and crisis have placed national statistical agencies in front of huge challenges. The crisis has a severe impact on national CPIs and HICPs, given the number of COICOPs that are involved and their corresponding weights. The rapid evolution of the virus outbreak also requires NSIs to react quickly in their quest for acceptable solutions for affected product aggregates.

Different statistical agencies and other international organisations started to set up impact analyses on imputation methods recommended in the recent Eurostat guidance note [1]. Three of these methods were investigated in the present study. Two imputation methods show promising results, in the sense that the impact of imputed indices of affected aggregates (flights and package holidays) on the overall annual rates of change is negligible in most cases. Also the results at division level are accurate.

The method that imputes with the all-items month on month index does not perform well, as it may deviate significantly from the year on year indices at all-item and division

level. This method does not explicitly control the year on year indices and does not respect seasonal patters, which, supported by the results, provide sufficient arguments to avoid this method.

This study recommends the use of the month on month index from the previous year of each aggregate to impute prices and indices during the crisis. Beside analytical arguments that are supported by empirical results, this method can be put into practice without much effort in the Dutch CPI and HICP. This is a big additional advantage given the limited amount of time that is left for implementation. A valuable insight that has emerged from this study is that the recommended method can be easily monitored. The performance of the method can be tracked over time, which allows users to verify to what extent the method neutralises the impact of affected aggregates on the overall annual rates of change. The monitoring can be done ahead of monthly compilation and publication. The recommended method has also found support in recent studies by the German NSI [2] and the ECB [3].

Time was simply too short to explore more refined methods, which is a topic that is consequently left for further research. In our opinion, expression (16) for the year on year index of the recommended method would be an excellent starting point for more sophisticated time series modelling (Section 4.2). The attention could then also shift towards the performance of different methods regarding their ability to neutralise the effect on the annual change rates of parent aggregates. The results at division level in Section 3.2 look very promising, although performance at lower aggregates was not the primary scope of this study. Embedding imputation methods within a constrained optimisation framework could be a possible direction worth of future exploration.

To conclude, the results of this study may have implications that stretch beyond the present scope that is dictated by the crisis. One of the topics in the recent debates on setting up recommendations for using multilateral methods in the CPI and HICP centres around the so-called "level of fixity". Expenditure patterns observed in transaction data can be highly dynamic. The level of fixity refers to the lowest level in the aggregation structures of the CPI and HICP at which annual fixed weights are defined. In theory, raising this level to some COICOP level implies introducing monthly weights for higher aggregates when using multilateral methods, which would eliminate problems that we are currently dealing with in a natural way.

On the other hand, we would end up in practice using annual weights for a part of the COICOPs and monthly weights for other aggregates. A more dynamic and realistic weighting scheme would be more time consuming and complex to work with and maintain in practice. The present study could reveal its usefulness also in this respect. Being able to neutralise the impact on the overall annual change rates from aggregates for which sales have drastically decreased, with a method that makes use of the traditional aggregation scheme of the CPI and HICP with annual weights, could therefore be extremely valuable in different respects.

References

[1] Eurostat (2020). *Methodological Guidance Note on the Compilation of the HICP in the Context of the COVID-19 Crisis.* Eurostat, 3 April 2020.

[2] FSO Germany (2020). *German Comment on the Treatment of Seasonal Products in CPI During Corona Crisis.* FSO Germany, Prices Division, 2 April 2020.

[3] Goldhammer, B. (2020). *Reply to Eurostat's "Methodological Guidance Note on the Compilation of the HICP in the Context of the COVID-19 Crisis".* European Central Bank, 30 March 2020.