

Identifying and evaluating COVID-19 effects on short-term statistics

Arnout van Delden*, Martijn de Winter and Koert van Bommel
Statistics Netherlands, The Netherlands

Abstract. The economic downturn due to lockdown measures at the beginning of the COVID-19 crisis raised the question whether any adaptations to the short-term statistics (STS) were needed to ensure accurate and relevant output. We limit ourselves to STS on turnover and related variables like volume of production. We looked into the different stages of the production process – from data collection to output – and anticipated a number of potential lockdown effects. With respect to output relevance, there was an increased interest in faster and specific output. With respect to the output accuracy, we took measures to check whether the anticipated effects really occurred and measures to mitigate the consequences. Examples of such measures are the calculation of an additional editing score function, alternative imputations and extensions of the regular analysis step. In this paper we give an overview of the anticipated effects, the subsequent measures that we took, we evaluate to what extent the anticipated effects occurred in practice and we mention some unforeseen effects. We end this paper by discussing to what extent the developed measures are also useful to keep after the economy has recovered.

Keywords: Analysis measures, economic downturn, robustness of statistical estimates

1. Introduction

The situation around COVID-19 has led to various governmental measures in the Netherlands with the purpose to slow down the spread of the virus, see RIVM [1] for a detailed timeline. As part of those measures there were two lockdown periods in the Netherlands that seriously affected the economy: mid-March till end-June 2020 and mid-December 2020 till end-June 2021. During the second lockdown, the measures were gradually released from the beginning of April till end of June; the date and speed with which measures were released varied with industry. There was also a partial lockdown: from mid-October till mid-December 2020, but its effect on the economy was limited. Apart from the lockdowns, also some other measures were taken that impacted the economy in 2020–2021. During both lockdowns, enterprises of different kinds of economic activities experienced a serious drop in their

turnover. For some enterprises there was a complete stand-still. In a few economic activities enterprises had an increased turnover on average. In this period, having reliable statistical information on short-term business statistics (STS) was very important for policy makers: it provided them with information whether and where supportive measures were needed.

An economic crisis may potentially affect different stages of the statistical value chain, see for instance Brand [2] and Simkins et al. [3]. The enormous drop in activity of enterprises during lockdown raised the question at Statistics Netherlands (CBS) as to whether CBS should adapt any processing steps to ensure a good quality of STS output. Different teams were launched that looked further into the potential consequences on the STS of the economic situation in response to COVID-19 measures, especially consequences during and shortly after a lockdown. In the context of this paper, we refer to this as ‘COVID-19 effects’. The teams consisted of people working at organizational units responsible for producing the STS and people working in the Department of Methodology and Process Development. The authors of this paper participated in one such team, and

*Corresponding author: Arnout van Delden, Statistics Netherlands, P.O. Box 24500, 2490 HA Den Haag, The Netherlands. E-mail: a.vandelden@cbs.nl.

in this contribution we will describe the results. The scope of our contribution is a limited part of the STS output; it does not cover price statistics for example. The focus is mainly on turnover and related variables.

Our team looked into different production stages of the statistics – from data collection to output – and identified a number of issues that could potentially lead to biased results. Most of those issues referred to the lockdowns, some referred to long-term effects. The team members developed monitoring instruments to verify whether the expected effects actually occurred and if so to what extent. The team was also confronted with some unforeseen effects. Subsequently, we introduced some changes into the different processing steps of the statistics to adjust for the effects of the lockdowns.

All monitoring instruments mentioned in this paper were computed as well as the effects of all of the changes. This was done in small pieces of computer code (scripts) which were running outside the standard IT production system. Some of those scripts are now implemented into the production system, some will be implemented later and others will not be implemented; that depended on their effectiveness and on the extent that the lockdown affected the corresponding parts of the production process. The new editing score function (Section 3.3.2) has been implemented in the standard IT-system. Some of the additional analysis checks (Section 3.8) are also automated and available for the regular production and can be used during a crisis situation. In the near future, we intend to make the alternative nowcasting method (Section 3.5.4) available in a standard IT-system. Other measures may follow later, see also the discussion (Section 4).

The objectives of the current paper are to describe the COVID-19 effects that we anticipated, to describe the checks and changes and to evaluate how effective those measures were. Many of the quantitative examples that we give in the current paper refer to the second quarter (Q2) of 2020 because it was the first lockdown period on which we could evaluate our measures. The long-term objective of this paper is that it can be used as a reference for adaptations to the STS in case of a future economic downturn. The remainder of this paper is organized as follows. First we introduce the STS (Section 2). In Section 3 we describe the anticipated COVID-19 effects, the measures we took and the evaluation. Finally, in Section 4 we summarise our main findings and discuss which of those adaptations are worth keeping.

2. Introduction to the short-term statistics

The STS are produced for a large part of the 21 different economic sectors. By an economic sector we mean the first-level category of the NACE Rev 2 code classification (Eurostat [4]). Within a sector, the statistical output is estimated for different industries (smaller groupings of NACE codes). CBS uses two main production systems for the STS. First, there is a production system which is solely based on sample survey data (SSD). This SSD system is used for industries whose statistical output is published on a monthly basis, except for a large part of motor and car trade whose output is published quarterly and whose turnover figures were produced with the SSD system until 2021. Input for the SSD system is a collection of (short) online surveys, but in the remainder of this paper we briefly use the term ‘STS survey’. Depending on the industry, the STS data collected using the survey concerns a number of variables: domestic, non-domestic and total (= non-domestic plus domestic) turnover, and further the variables inventories, orders and hours worked. For all variables a number of common processing steps are used within the SSD system: automatic editing, imputation of nonresponse and index computation. However, in practice only the turnover data are analysed and interactively edited within the SSD system to be used to produce statistical output. The microdata on the variables orders and hours worked are exported to the production system for the ‘production index’, and within the latter system those variables are interactively edited, processed and used to produce statistical output. The microdata on inventories are processed in a separate system to produce output on inventories.

The second main production system for turnover uses Value Added Tax (VAT) data for the small and medium sized enterprises and survey data for the largest and most complex units. This results in observations for almost all units in the population, except for the early flash releases. We will refer to this production system as the census combined data (CCD) system. The CCD system is used to publish STS output for industries whose output is required on a quarterly basis and for which only turnover variables are needed. Note that in the Netherlands on a quarterly basis nearly all turnover of the target population is available of which a selective part of the population reports VAT monthly. Finally, from an output quality perspective it would be useful for CBS to estimate the quarterly output from the CCD system also for industries whose output is published on the basis of the SSD system and then to benchmark the

Table 1

Overview of output (periodicity and economic activities) of the two STS production systems and industries severely negatively affected by COVID-19 measures (see text)

STS system	Publ. period	Economic activities processed by the STS system (sorted by sector)	Industries per sector with a year-on-year turnover growth in 2020 of -50% or less due to COVID-19 crisis
CCD (VAT and survey)	Quarterly (*)	Motor and Car trade (G 45 except 45111) {from 2021}	None
		Whole sale trade (G 46)	None
		Transportation (H)	491: Passenger rail transport (no tram or metro), 51: Air transport, 5223: Support activities for air transport
		Accommodation and food service activities (I)	551: Hotels and similar accommodation, 56101: Restaurants, 562: Canteens and catering, 563: Bars
		Information and communication (J)	5914: Cinemas
		Professional, scientific and technical activities (M) (**)	None
		Administration and support service activities (N) (**)	7911: Travel agencies, 7912: Tour operators
SSD (survey)	Quarterly	Motor and Car trade (G 45, except 45111) {up to 2020}	None
		Mining and quarrying (B)	None
	Monthly	Manufacturing (C)	15. Manufacture of leather (products) of leather and footwear, 19. Manufacture of coke and refined petroleum products, 29. Manufacture of motor vehicles, (semi-)trailers
		Electricity gas, energy, steam and air conditioning supply (D)	None
		Water collection, treatment and supply (E 36),	None
		Materials recovery (E 383)	
		Construction (F)	None
		Retail trade (G 47)	47710: Shops selling clothes (accessories), 47782: Shops selling optical articles
		Car import (G 45111)	45111: Import of new passenger cars and light motor vehicles

(*) Except for the industries in sector S, all other industries have to be produced on a monthly basis in the near future, see the main text. (**) A few NACE codes are not included.

monthly SSD output to the quarterly CCD output. This is currently not done, because the quarterly turnover pattern from VAT differs from that of the survey data (Van Delden et al. [5]) and because it is complex to incorporate the extra analysis and the benchmarking step into the production process, because it has a tight schedule.

Due to the new Commission Implementing Regulation (EU) No 2020/1197 [6] all EU member states are required to provide monthly turnover figures where quarterly figures were required in the past for many industries in NACE codes G45, G46 from 2021 onwards and for sections H, I, J, L, M, N from 2024 onwards. These new monthly index series have to go back to January 2021. CBS obtained a derogation from the European Commission: all our new monthly index series only have to be provided from 2024 onwards and only have to go back to January 2023. Since 2021, part of the survey data of NACE code G45 and of sections H, I, J, M and N are already collected by CBS on a monthly basis. An STS system to produce the turnover figures for these industries on a monthly basis, is currently un-

der development. Also monthly figures on volume of production will be published.

The STS output is produced for different releases. For most industries there are three releases: an early, an intermediate and a final estimate. The exact timing depends on the industry, but the early estimate is usually available around 30 days after the end of the period. Obviously, for early releases there is less response than for later ones. Imputation is used to estimate a value for all non-responding survey units for the SSD and CCD systems and in the case of the CCD system, imputation is also used for the units without survey for which the VAT data are not available yet.

Table 1 gives an overview for which industries the output is produced, by which system and at which periodicity. The last column in Table 1 shows which industries had a year-on-year turnover growth rate in 2020 of -50% or less (the period in the previous year being pre-COVID-19) in at least one month for industries with monthly output and in at least one quarter for industries with quarterly output. We selected 2020 for this analysis since most drops in turnover due to lockdown measures were in 2020.

Table 2
Anticipated effects of COVID-19 on STS

Production stage	Anticipated effect
1 Sampling and data collection	Response rates to a survey may be lower and businesses may delay reporting VAT.
2 Editing	The score function for selective editing may be affected because the function takes the turnover of a previous period as reference values and large changes are seen as suspicious.
3 Imputation	The quality of the imputations for missing observations may be affected due to <ul style="list-style-type: none"> – increased heterogeneity of response values – increased selectivity of respondents
4 Nowcasting	Standard nowcasting methods with autoregressive terms may lead to a bias since time patterns in the past are no longer good predictors of the present situation.
5 Index computation	It is uncertain whether the STS-index series will return to the correct level after a great dip.
6 Seasonal adjustment	The seasonal adjustment procedure is no longer valid, because the regular seasonal pattern is affected by a lockdown.
7 Regular analysis	Since more extreme values and growth rates are expected, possibly combined with reduced response rates, more research capacity will be needed to validate the results.
8 Create output	There will be an increased demand from society for more, more timely and more detailed output to monitor COVID-19 impact.

3. COVID-19 effects at different productions stages

3.1. Overview

For the STS we identified several potential risks to the STS output quality. Most of them concerned a risk of biased output, some a risk of increased variance, and the remaining risks referred to other quality aspects, see Table 2. We limited ourselves to those risks that we expected to be most influential for output quality. The kind of risks that we identified were very similar to the ones described a decade ago for the 2008 economic downturn by Brand [2] and Simkins et al. [3]. We have ordered them by the different STS processing steps and refer to them as ‘anticipated effects’. Although it was uncertain whether the anticipated effects would really occur in practice, for most of them we developed analysis tools to monitor them or measures to mitigate them.

3.2. Sampling and data collection

3.2.1. Anticipated effects

We expected that the survey response rates and the VAT reporting rates might be reduced during periods with COVID-19 measures since entrepreneurs were expected to have other priorities than reporting data to external agencies.

3.2.2. Measures

The CBS data collection department (DVZ) has a standard procedure to get high response rates for the STS. Enterprises that have to provide quarterly survey data receive at the beginning of each quarter registered

letters and emails with access codes to online questionnaires. They have approximately one month to fill in those questionnaires about data of the previous quarter. The enterprises will receive a number of reminders via letters or electronically. From approximately 40 days after the date on which the registered letter/email was sent, DVZ can, besides sending reminder letters or emails, also call enterprises to increase response rates. Here, a top-down approach is used to ensure that the most influential enterprises are approached first. If enterprises still do not respond, an enforcement policy will be in place in which companies could potentially get fined for not sending data. For enterprises that have to provide monthly survey data, a similar approach is followed. They will receive registered letters/emails at the end of each month and a last reminder will be sent approximately 20 days after the date on which the registered letter/email was sent.

During the lockdowns, for STS, contacting by email and by phone was preferred over sending letters since DVZ anticipated that entrepreneurs whose enterprises were inactive or whose shops were closed would miss letters sent to them. Furthermore, the enforcement policy was put on hold.

Moreover, for industries in retail trade the date of starting with processing the response was advanced, and the output figures were published two weeks earlier than normal. Enterprises in those industries were actively contacted by phone and were asked to deliver their STS data earlier than normal. Some of the employees of DVZ that could not do their regular work due to COVID-19 situation were deployed to call those enterprises.

Table 3

Turnover response rates of early estimates for sectors with monthly turnover figures (SSD system). First lockdown period was from mid-March till end of June 2020

NACE code	2019						2020					
	Jan	Feb	Mar	Apr	May	Jun	Jan	Feb	Mar	Apr	May	Jun
(B)	99.8	79.5	99.4	99.3	86.8	97.7	96.7	98.9	97.1	96.3	98.0	97.7
(C)	90.2	91.6	94.2	94.8	96.6	96.1	92.1	92.6	94.1	92.4	95.0	93.7
(D)	93.0	91.6	97.4	93.1	95.8	99.4	86.5	92.0	91.7	95.5	96.7	95.8
(E 36), (E 383)	86.0	91.1	84.9	91.1	83.2	90.9	85.8	92.6	91.8	86.3	94.2	93.6
(F)	93.7	96.4	96.1	95.4	94.6	96.2	94.4	94.0	95.9	96.0	95.4	95.8
(G 47)	79.3	91.7	89.9	89.8	89.3	89.8	90.3	90.5	91.2	90.6	91.5	92.5

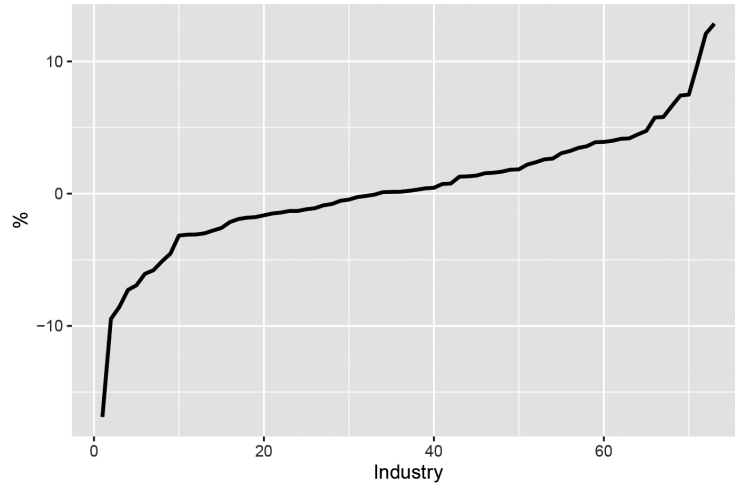


Fig. 1. Difference in turnover response rates per industry for the early estimates of the monthly STS statistics (SSD system) between the second quarter of 2020 (first lockdown) and the second quarter of 2019 (no lockdown). The 73 industries are sorted from small to large values.

3.2.3. Results

We monitored whether the survey response rates were indeed lower than normal. We computed the turnover response rates for period t and industry k , denoted by \mathcal{R}_k^t , given by:

$$\mathcal{R}_k^t = \frac{\sum_{i=1}^{n_k^t} a_{ki}^t d_{ki}^t \dot{Y}_{ki}^t}{\sum_{i=1}^{n_k^t} a_{ki}^t d_{ki}^t \dot{Y}_{ki}^t + (1 - a_{ki}^t) d_{ki}^t \tilde{Y}_{ki}^t}, \quad (1)$$

where a_{ki}^t denotes whether unit i of industry k in period t has responded ($a_{ki}^t = 1$) or not ($a_{ki}^t = 0$), \dot{Y}_{ki}^t denotes an observed turnover, \tilde{Y}_{ki}^t denotes an imputed turnover, d_{ki}^t denotes the weight of the unit which is usually the inverse of the inclusion probability but it may be adjusted in the case of an outlier, and n_k^t denotes the sample size. The response rates also depended on the actual release, but to simplify the notation we did not explicitly include the release in the notation of \mathcal{R}_k^t in Eq. (1).

The turnover response rates for the early estimates (published around 30 days after the end of the period) of the STS statistics of different economic sectors pro-

duced with the SSD system in pre-COVID-19 months were 86% or higher (see Table 3). During the first lockdown period (March-June 2020), the response rate for the early estimates remained at a high level and in some sectors it even slightly increased, see Table 3.

Figure 1 shows the difference in turnover response rate per industry for early estimates of the monthly turnover statistics for the second quarter of 2020 (lockdown) and the second quarter of 2019 (no lockdown). For most industries the difference in response rates was at most 5 percentage points. Only a few industries have a decline in turnover response rate larger than 5 percentage points. One industry had an increased response rate of 13 percentage points during the first lockdown. Note that some of the industries are small and their turnover response rates varied considerably from quarter to quarter depending on whether the larger enterprises already responded or not.

The turnover response rates for the quarterly estimates produced by the CCD system were also analysed. In the CCD system turnover values were imputed for all units with missing response. Hence, for the CCD

system, the weights d_{ki}^t in Eq. (1) are equal to 1 and n_k^t stands for all units of the population. Whereas the turnover response rates for the early estimates remained at the same level for the monthly survey data (SSD system), a decline in response rates for the early estimates was seen for both the VAT-data and the quarterly survey data (CCD system). The decline for the first quarter of 2020 as compared to the first quarter of 2019 was of the order of 10% over all industries. Note that whereas only a small part of the first quarter of 2020 coincides with the first lockdown period, the data collection period for data on the first quarter of 2020 was during the middle of the first lockdown period. A smaller decline in response rates was seen for the second quarter of 2020 (compared to the second quarter of 2019). The 10% decline in response rate for the early estimates in the first quarter of 2020 did not cause a problem with the total amount of response because, averaged over all industries in the CCD system, still 60% of the total turnover was based on response. Furthermore, before publishing the Quarterly National Accounts that were based on the early STS-estimates, CBS checked that the updated CCD-estimates from a couple of days later (based on far more response), did not differ too much from the early estimates. Although the size of the response for the early estimates was sufficient for both the SSD and CCD system, we were aware of the risk that the early response during periods of a lockdown may be selective; that issue is addressed in Section 3.4.

In conclusion, it turned out that the turnover response rates for the early estimates of the SSD system were at similar or even slightly higher levels compared to non-COVID-19 periods; also for industries affected by the COVID-19 crisis (see Table 1 for those industries). CBS mentioned the importance of figures for society in the first letter (email) that they sent to the enterprises. Perhaps this has promoted them to sent in a response. However, a decline in response rate was seen for both the VAT data and the quarterly survey data for the CCD system. It is unclear to us why response rates of survey data of monthly published industries in the SSD systems remained high while response rates of survey data used in the CCD system for quarterly published surveys were somewhat reduced. What is clear is that additional efforts for the retail trade statistics worked out well and allowed us to accelerate the output.

Statistics Portugal (Moreira et al. [7]) who did not report additional measures with respect to the data collection, found a drop in their response rates on monthly business surveys of approximately 10% in April and May 2020 as opposed to 2019. Statistics Slovenia (Šuš-

tar Kožuh [8]) gradually shifted to electronic communication with businesses, hardly used any reminders but extended the deadlines for data transmission for their STS. They found the monthly STS response in 2020 to be close to the response rates in 2019.

3.3. Editing

3.3.1. Anticipated effects

Before explaining the anticipated effects, we first briefly introduce the usual editing score function. CBS uses different score functions to identify enterprises within the STS whose turnover values are influential and possibly incorrect. The most important score function, denoted by $A_{ki}^{t,r}$, is based on influential growth rates. For the SSD system this score function is defined by:

$$A_{ki}^{t,r} = \frac{d_{ki}^t Y_{ki}^t - d_{ki}^r Y_{ki}^r}{Y_k^r}, \text{ with} \quad (2)$$

$$Y_k^r \equiv \sum_{i=1}^{n_k^r} d_{ki}^r Y_{ki}^r$$

where Y_{ki}^t and Y_{ki}^r denote the reported or imputed value for the target variable (mostly total turnover) of enterprise i in industry k for the current period t or reference period r respectively, d_{ki}^t and d_{ki}^r are the corresponding weights in the current and reference period respectively, and n_k^r stands for the number of sampling units in reference period r and industry k . Furthermore, Y_k^r stands for the weighted summed industry turnover. For the CCD system Y_k^r refers to the summation over all units in the population and the weights d_{ki}^t and d_{ki}^r are equal to 1.

Enterprises that are entering and those that are leaving industry k have a zero turnover value for the period that they are not present in industry k . To better understand $A_{ki}^{t,r}$ as an editing score function we can rewrite Eq. (2) as (assuming $d_{ki}^r Y_{ki}^r \neq 0$):

$$A_{ki}^{t,r} = \left(\frac{d_{ki}^r Y_{ki}^r}{Y_k^r} \right) \left(\frac{d_{ki}^t Y_{ki}^t - d_{ki}^r Y_{ki}^r}{d_{ki}^r Y_{ki}^r} \right)$$

$$= \omega_{ki}^r (\check{G}_{ki}^{t,r} - 1), \text{ with} \quad (3)$$

$$\check{G}_{ki}^{t,r} \equiv \frac{d_{ki}^t Y_{ki}^t}{d_{ki}^r Y_{ki}^r} \text{ and } \omega_{ki}^r \equiv \left(\frac{d_{ki}^r Y_{ki}^r}{Y_k^r} \right).$$

The symbol ω_{ki}^r stands for the proportion of the weighted turnover of enterprise i to the weighted summed industry turnover, which is the *influence* component of the score function. Further, $\check{G}_{ki}^{t,r}$ is the weighted growth factor. The term $(\check{G}_{ki}^{t,r} - 1)$ stands for the *risk* component of the score function, the extent to which the unit is outlying. Note that the sum

of $A_{ki}^{t,r}$ over all enterprises within the industry yields $\sum_{i=1}^{n_k^{t,r}} \frac{d_{ki}^t Y_{ki}^t - d_{ki}^r Y_{ki}^r}{Y_k^r} = G_k^{t,r} - 1$ with $G_k^{t,r} \equiv \frac{Y_k^t}{Y_k^r}$, $Y_k^t \equiv \sum_{i=1}^{n_k^t} d_{ki}^t Y_{ki}^t$, n_k^t stands for the sample size in period t and $n_k^{t,r}$ stands for the size of the union of units in the sample for period t and r .

In practice, the previous period was used as reference period: $r = t - 1$. Normally, an extreme absolute value of $A_{ki}^{t,r}$ indicates that the corresponding unit is an outlier and influential. However, a large absolute value of $A_{ki}^{t,r}$ is not necessary a indication for an outlier during lockdown. We anticipated that the absolute values of the score function $A_{ki}^{t,r}$ might increase greatly for many enterprises during lockdown since many enterprises will show a large turnover drop. We therefore expected $A_{ki}^{t,r}$ to be less suitable for finding outliers during lockdown.

3.3.2. Measures

Since the score function $A_{ki}^{t,r}$ was less suitable during the lockdown periods, analysts put more emphasis on scatterplots rather than on $A_{ki}^{t,r}$ to identify outliers for the monthly industries that were seriously affected by the lockdown. In the scatterplots Y_{ki}^r was plotted against Y_{ki}^t with $r = t - 1$ or $r = t - 12$. For the estimates made during 2020 this was the main measure taken, due to time limitations.

By the time the data pertaining to all months of 2020 were available in SSD, we had calculated a new editing score function to identify influential outliers that replaced $A_{ki}^{t,r}$. This new editing score function was used in production from the spring of 2021 onwards. The first estimates for which it was used, were the final estimates over the reporting year 2020.

This alternative score function, denoted by $B_{ki}^{t,r}$, aims to capture whether the growth of an enterprise is extreme relative to the average growth of the enterprises in the same industry. This score function is defined as

$$B_{ki}^{t,r} = \frac{d_{ki}^t Y_{ki}^t - d_{ki}^r \hat{Y}_{ki}^t}{Y_k^r}, \quad (4)$$

with $\hat{Y}_{ki}^t \equiv Y_{ki}^r G_k^{t,r}$ and the growth factor $G_k^{t,r} \equiv Y_k^t / Y_k^r$. We can rewrite Eq. (3) (assuming $d_{ki}^r Y_{ki}^r \neq 0$) as:

$$B_{ki}^{t,r} = \left(\frac{d_{ki}^r Y_{ki}^r}{Y_k^r} \right) \left(\frac{d_{ki}^t Y_{ki}^t - d_{ki}^r \hat{Y}_{ki}^t}{d_{ki}^r Y_{ki}^r} \right), \quad (5)$$

$$= \omega_{ki}^r (\check{G}_{ki}^{t,r} - G_k^{t,r}),$$

where $\check{G}_{ki}^{t,r}$ and ω_{ki}^r are defined in the second line of Eq. (3). Compared to Eq. (3) the term $(\check{G}_{ki}^{t,r} - 1)$ is replaced by $(\check{G}_{ki}^{t,r} - G_k^{t,r})$. The difference $G_{ki}^{t,r} - G_k^{t,r}$ stands for the (new) risk component. Note that $\sum_{i=1}^{n_k^{t,r}} B_{ki}^{t,r} = 0$.

3.3.3. Results

The alternative scores $B_{ki}^{t,r}$ (see Section 3.3.2), with $r = t - 1$, were used for editing of the data of Manufacturing and Retail Trade within the SSD system. When reviewing the STS for the reporting year 2020 (in the spring of 2021), a list of alternative scores was sent to analysts and validators of Manufacturing and Retail trade. They found that these scores were particularly interesting if turnover changes were mainly caused by a homogeneous group of units. Normally, all units with the usual score function $|A_{ki}^{t,r}| > 0.01$ were checked: in the second quarter of 2020 this concerned 313 units for Manufacturing and 438 for Retail trade, see Fig. 2. The alternative score function $|B_{ki}^{t,r}| > 0.01$ concerned 283 enterprises for Manufacturing and 372 for Retail trade, so those numbers were somewhat smaller. Some units were outlying according to $|A_{ki}^{t,r}| > 0.01$ but not according to $|B_{ki}^{t,r}| > 0.01$ and vice versa, see Fig. 2.

At the time of writing this paper, CBS is still in the process of evaluating and perhaps improving the editing score function.

3.4. Imputation

3.4.1. Anticipated effects

Missing turnover values in the STS are imputed by using a ratio imputation of the form:

$$\tilde{Y}_i^t = X_i \left(\frac{Y_d^t}{X_d} \right) \quad (6)$$

where \tilde{Y}_i^t denotes the imputed turnover value of enterprise i for period t , X stands for an auxiliary variable and Y_d^t / X_d is the ratio of the weighted total of Y for domain d in period t and the corresponding domain total for X . The domain d stands for a group of enterprises that are similar to unit i (based on the values of background variables such as size class and industry) and that hopefully have a similar Y/X ratio to unit i (on average over imputations). In most cases the chosen auxiliary variable X is the turnover in a reference period r in which case the equation is equivalent to $\tilde{Y}_i^t = Y_i^r (Y_d^t / Y_d^r)$. For new enterprises (births) the number of employees in period t is often used as auxiliary variable.

There are three points of attention with respect to the imputation (Van Bommel and Goorden [9]):

- In the SSD system enterprises with a net turnover of 0 are given the status of ‘inactive’ and are imputed with a value of 0 in subsequent months as long as no new response is received. In the CCD system, which was developed earlier than the SSD system, this rule was not included.

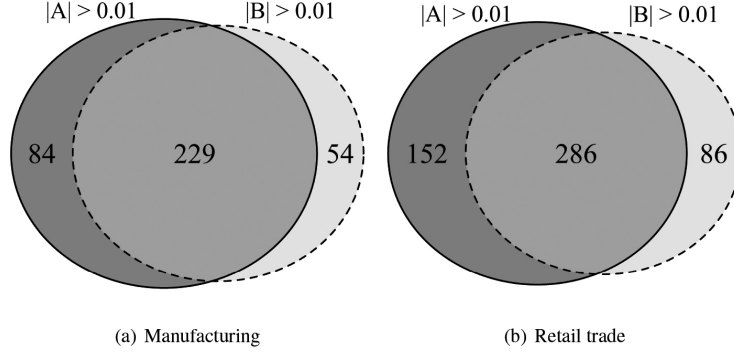


Fig. 2. Number of units with $|A_{ki}^{t,r}|$ and/or $|B_{ki}^{t,r}| > 0.01$ (score functions abbreviated as A and B in the figure) summed over April, May and June 2020, using the previous month as reference period, for Manufacturing and Retail trade.

- For most imputations the ratio Y_d^t/X_d concerns a year-on-year ratio.
- The domain d is chosen such that it contains at least 20 respondents. In SSD, but not in CCD, there is the additional constraint that in the domain at least 8 respondents have a turnover larger than 0.

We anticipated that the setting and cancelling of COVID-19 measures would result in large changes of growth rates in STS that may in turn impact the quality of the imputations of STS:

- Domains d may become heterogeneous in terms of the Y_d^t/X_d ratio: some enterprises are heavily affected by the lockdown measures while others are not. This could lead to a higher variance of the ratio Y_d^t/X_d , and a risk of a bias when heavily affected enterprises respond earlier or later on average.
- Imputation of units can go wrong when a value of 0 (in period t or a previous period) plays a role. In the SSD system units may incorrectly still have the status inactive after cancelling of the COVID-19 measures and are incorrectly imputed with a 0. In the CCD system units that are still closed due to the COVID-19 measures may incorrectly not be imputed with a 0 because they do not have status inactive and are for instance imputed with a year-on-year growth factor that is not 0 due to other respondents that are not closed.

We took the following measures:

- Count the number of units with turnover value 0 and the number of inactive units (see Section 3.8).
- Monitor whether there are domains whose heterogeneity has increased in the lockdown period (see Section 3.4.2).
- Prepare an alternative imputation method based on new auxiliary data (see Section 3.4.3).

3.4.2. A measure for heterogeneity change

As a starting point for developing a measure for heterogeneity change, we used some results from Van Delden and Scholtus [10] who made a regression of survey turnover, denoted by y_i to VAT turnover, denoted by x_i . They found that the variance of the residuals of y_i increased with the value of x_i : $\text{Var}(\epsilon_i) = \sigma^2(x_i)^\lambda$, where ϵ_i stands for the residual and σ^2 stands for a constant, see also Kutner et al. [11, pp. 424–431] on how to model heteroscedasticity. Van Delden and Scholtus [10] estimated λ and found $\hat{\lambda} = 1$ to be a good estimate for different economic sectors, see also their Figs 3–6. Given this model for the heteroscedasticity, we assumed that the ratio of the standard deviation of the turnover values Y_{di}^t in a domain d for period t over its mean (to the power $\alpha = \lambda/2$, with $\hat{\lambda} = 1$), $\text{Std}(Y_{di}^t)/(\bar{Y}_d^t)^\alpha$, with $\bar{Y}_d^t \equiv (1/n_d^t) \sum_{i=1}^{n_d^t} Y_{di}^t$, is relatively constant over time for a given domain. We refer to this ratio as the relative standard deviation of turnover of domain d and period t , denoted by $\text{RSD}(Y_{di}^t)$. For different populations (domains) however we expected different values for $\text{RSD}(Y_{di}^t)$.

Similarly, we defined the RSD of a growth factor of period t to r of unit i in domain d as:

$$\text{RSD}(G_{di}^{t,r}) = \frac{\text{Std}(G_{di}^{t,r})}{(\bar{G}_d^{t,r})^\alpha}, \text{ with} \quad (7)$$

$$\bar{G}_d^{t,r} \equiv \frac{1}{n_d^{t,r}} \sum_{i=1}^{n_d^{t,r}} G_{di}^{t,r}$$

and $n_d^{t,r}$ stands for the size of the union of units in domain d in the sample for period t and r . Since we were aware that the distribution of turnover tends to be very skewed and also that growth rates can have outliers, we defined a robust version of the RSD for the growth factor $G_{di}^{t,r}$ which is given by:

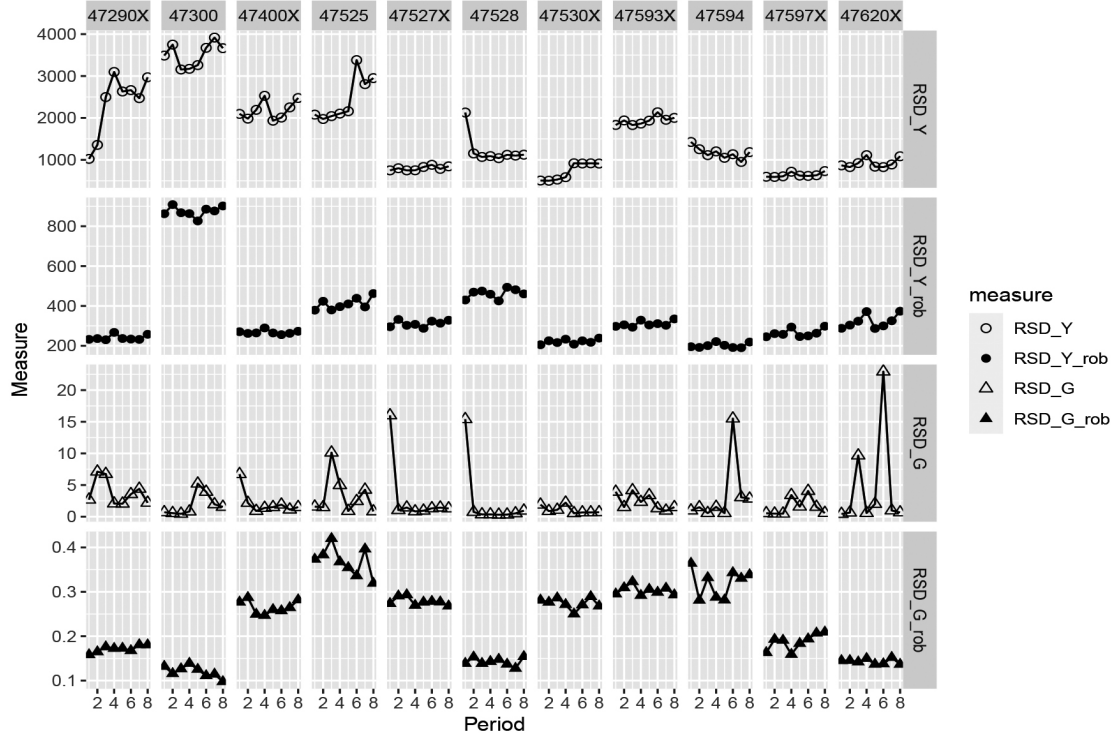


Fig. 3. Relative Standard Deviation of eleven NACE codes in Retail trade from the first quarter of 2015 (period 1) till the fourth quarter of 2016 (period 8). Growth factors are year-on-year. NACE codes are ending with ‘X’ when several underlying 5-digit codes are combined.

$$\text{RSD}(G_{di}^{t,r})(\text{rob}) = \frac{\text{Std}(G_{di}^{t,r})(\text{rob})}{[\text{median}(G_{di}^{t,r})]^\alpha}, \quad (8)$$

where the mean of the growth factor $G_{di}^{t,r}$ is replaced by its median and the robust standard deviation (Rousseeuw [12]) is given by

$$\begin{aligned} & \text{Std}(G_{di}^{t,r})(\text{rob}) \\ &= \frac{1}{\gamma} \text{median}|G_{di}^{t,r} - \text{median}(G_{di}^{t,r})|. \end{aligned} \quad (9)$$

with $\gamma = 1/0.6745$, which is a correction factor such that in case of a normal distribution, in expectation, the robust standard deviation equals the (classical) standard deviation. The robust RSD for the turnover Y_{di}^t is defined as in Eqs (8) and (9), but with $G_{di}^{t,r}$ replaced by Y_{di}^t . Cases in Eqs (7) and (8) where $Y_{di}^t = 0$ or $Y_{di}^r = 0$ so that $G_{di}^{t,r}$ is 0 or undefined, were treated separately. Furthermore, we tracked domains with many 0 values and verified the imputations that were based on them.

In order to verify whether it is reasonable to assume that the RSD per domain is constant over time but differs among domains, we computed the $\text{RSD}(Y_{di}^t)$ and $\text{RSD}(G_{di}^{t,t-4})$, and their robust version, for industries in Manufacturing, Retail trade and Employment services using historical VAT data of 2014–2016 for small

and medium-sized enterprises on a quarterly basis. We only included records with turnover values larger than zero for both quarters of the year-on-year factor. As an example we plotted the results for some industries in Retail trade, see Fig. 3. We found that for some industries $\text{RSD}(Y_{di}^t)$ and $\text{RSD}(G_{di}^{t,t-4})$ varied considerably over time, especially $\text{RSD}(G_{di}^{t,t-4})$. Their robust counterparts were more stable over time. For the imputations (see next section) often domains are used that are more detailed than the domains that we used in Fig. 3. We did some further analyses on these data and found that domains should have 30–40 units for the $\text{RSD}(\text{rob})$ to be stable over time.

We expected that in some domains the lockdown measures might have affected part of the units in a domain while other units were not or less affected. For instance because they were less dependent on a physical shop for their sales. In such conditions, it might well be that $\text{RSD}(G_{di}^{t,r})(\text{rob})$ is increased compared to a pre-COVID-19 period. Since the growth factor of a domain, $G_d^{t,r}$, is often used in the ratio imputation of the STS statistics, a larger spread of these growth factors may lead to imputations of a lower quality. To test whether a lockdown indeed increases the $\text{RSD}(G_{di}^{t,r})(\text{rob})$, we computed the $\text{RSD}(G_{di}^{t,r})(\text{rob})$ for the same industries

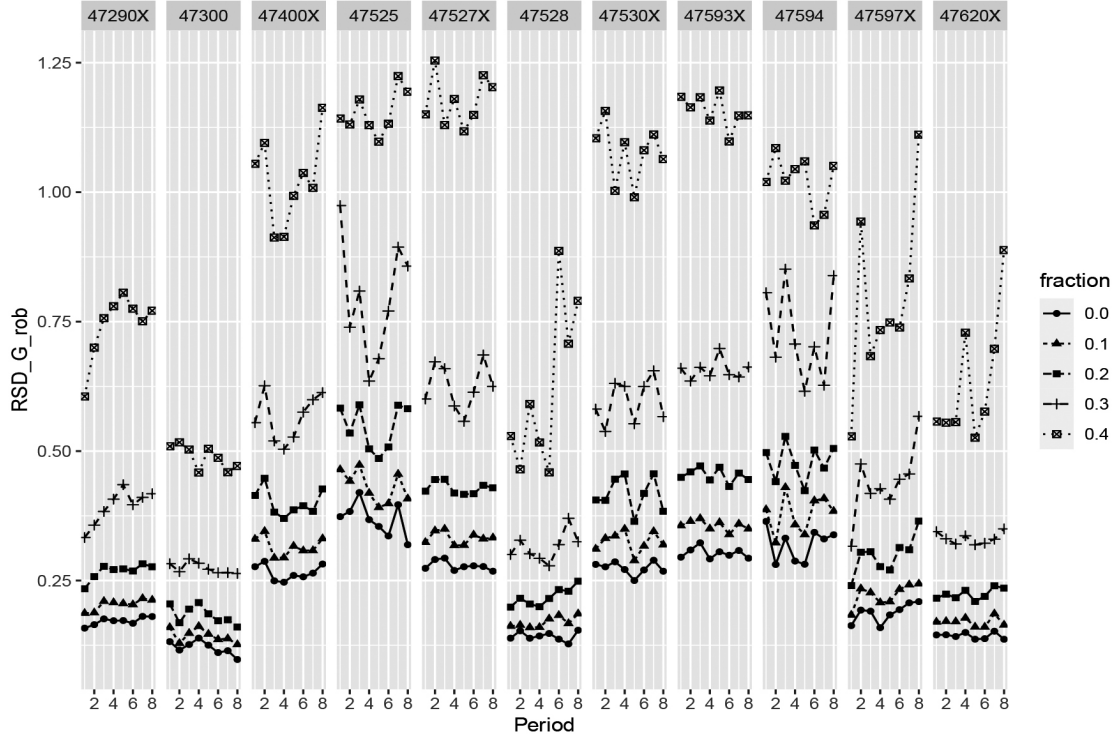


Fig. 4. Robust Relative Standard Deviation of the period-to-period growth factor as a function of the fraction of units with a very small growth factor, for eleven NACE codes in Retail trade from the first quarter of 2015 (period 1) till the fourth quarter of 2016 (period 8). NACE codes are ending with 'X' when several underlying 5-digit codes are combined.

and historical data as explained before, but this time we simulated that some of the units (randomly drawn) were affected by a lockdown. We tested fractions of 0.0, 0.1, 0.2, 0.3, 0.4, 0.45 and 0.5. For the affected units, we drew a growth factor at unit level (Y_{di}^t/Y_{di}^r) from a normal distribution with mean 0.05 and standard deviation 0.0025 while the other units kept their original values. Results for the fraction up to 0.4 are shown in Fig. 4 for the same industries and periods as Fig. 3. The $RSD(G_{di}^{t,r})$ clearly increased with an increased proportion of units that is affected by the lockdown. However, at proportions larger than 0.4 (not shown for clarity of the figure), the $RSD(G_{di}^{t,r})$ dropped down to a small positive values due to the use of the absolute median difference. For some industries this drop occurred already earlier, because in the original data were also units with very small growth factors.

To support the production of the STS, we did not directly look at the $RSD(rob)$ itself, but we computed the ratio $H_d^{t,r,j}$ (where H stands for heterogeneity measure) defined as:

$$H_d^{t,r,j} = \frac{RSD(G_{di}^{t,r})(rob)}{RSD(G_{di}^{t-j,r-j})(rob)}, \quad (10)$$

where t stands for a period with lockdown measures, reference period r is either one period earlier or the same period one year earlier and $t-j$ and $r-j$ stand for pre-COVID-19 periods.

The ratio $H_d^{t,r,j}$ was used to support the production of the STS from the first lockdown period onwards. For the results over years 2018–2020 we took j to be one year (Dutch: jaar), so $t-j$ and $r-j$ are one year before periods t and r respectively. For 2021 we took j to be two years to ensure that $t-j$ and $r-j$ stand for pre-COVID-19 periods. Figure 5 shows the values of $H_d^{t,t-12,j}$ for year-on-year growth factors of monthly turnover for domains processed by the SSD system. We found an increased number of domains with $H_d^{t,t-12,j} > 2$ in March–June 2020 and in January–May 2021, which concerned mainly domains in Retail trade (not shown). In these two periods there was a lockdown in which many shops were closed. From the beginning of April 2021 till end of June 2021 the lockdown measures were gradually released. In a very limited number of domains we found values of $H_d^{t,t-12,j}$ that were close to 0.

We aimed to compute a similar series of $H_d^{t,t-4,j}$ for year-on-year growth factors for quarterly turnover for

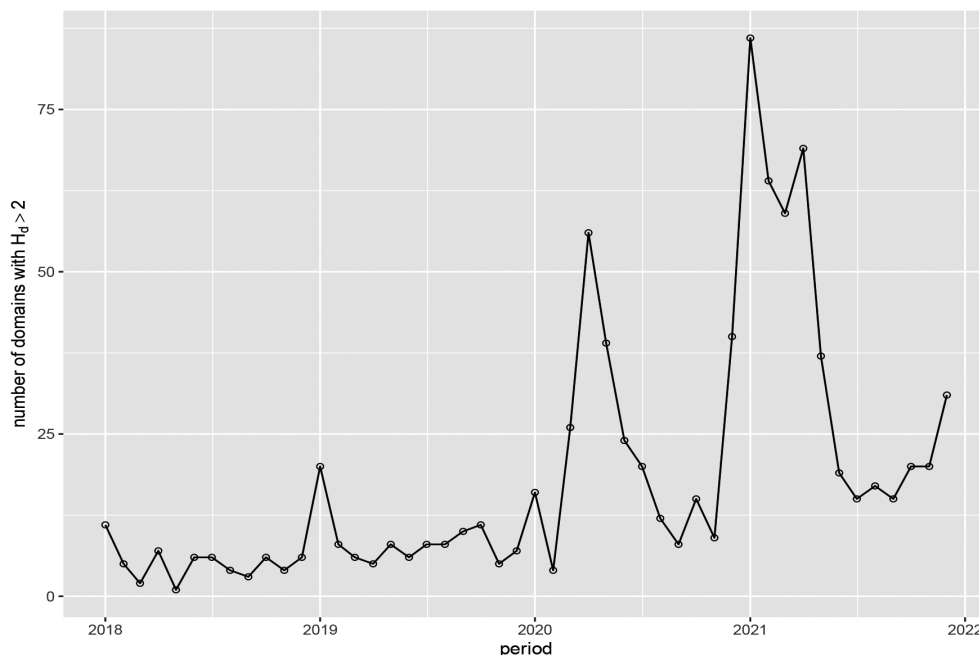


Fig. 5. For each month, the number of domains in the SSD system with the ratio $H_d^{t,t-12,j}$ larger than 2. The first plotted month is January 2018.

domains in the CCD system, but we could not easily produce such a time series. Still, results for the second quarter of 2020 have been calculated and stored. In the second quarter of 2020 there were a large number of domains with $H_d^{t,t-4,j} > 2$, for instance domains within NACE 5510 (Hotels and similar accommodation), 7912 (Tour operator activities) and 96021 (Hairdressing). We also found a considerable amount of domains where RSD (rob) dropped nearly to zero, also domains affected by the lockdown. This might be largely due to the aforementioned limitation of RSD (rob) in situations where a large proportion of the units in the domain has a turnover drop to almost zero.

For all domains with an increased heterogeneity, the imputed values were checked on plausibility. In case of the SSD system, for a few industries imputed values were adjusted. In case of the CCD system, first alternative imputations were computed for all domains, see Section 3.4.3, including domains with increased heterogeneity and domains where RSD (rob) dropped to small (positive) values. For a few industries the early estimates were adjusted, see also 3.4.3.

The heterogeneity score function has been a useful supportive means to check the quality of the imputations. Nonetheless, it would be good to try to improve the stability of this measure for smaller domain sizes since the domains that we use in practice often contain around 20 units. It would be good to test if using an

$\alpha = 1$ makes RSD (rob) more stable at smaller domain sizes. Furthermore, preferably a measure of heterogeneity does not suddenly drop down to small values when fractions larger than 0.4 of the population become inactive.

3.4.3. Alternative imputations: measures and results

Industries that were heavily affected by the lockdown measures and without high response rates during the early estimates, were mostly industries with quarterly statistics that are produced with the CCD system. During the production of the early quarterly estimates, used by National Accounts, many enterprises have not yet responded. Contributions of imputations to total turnover at industry level of more than 40% were no exception. We anticipated biased estimates due to an increased heterogeneity coupled with differential nonresponse within the industries. For example, if enterprises that were more affected by the lockdown responded earlier on average, then the early estimates would be biased downwards. Therefore, we took some measures for these quarterly statistics (CCD system). For the monthly statistics (SSD system) no further measures were needed, because the response rates turned out to be very high for the monthly industries that were affected by the lockdown measures, see Section 3.2.

As a potential measure to improve the imputations we analysed newly obtained auxiliary data, namely on

Table 4
Ratio of year-on-year turnover for different strata of a COVID-19 subsidy

Stratum	Applied	Received	Expected turnover drop	April		May	
				YOY-factor	Number	YOY-factor	Number
1	No	No	NA	0.92	5348	0.91	5372
2	Yes	No	20–100% (*)	1.03	933	1.13	26
3	Yes	Yes	20–30%	0.75	576	0.77	893
4	Yes	Yes	30–40%	0.62	952	0.68	1299
5	Yes	Yes	40–60%	0.50	844	0.62	1058
6	Yes	Yes	60–100%	0.46	893	0.63	990

(*) An enterprise could only apply for the subsidy when its expected turnover drop was at least 20%.

enterprises applying for a COVID-19 subsidy (compensation for the wages of the employees). Those enterprises had to report an ‘expected turnover drop’ compared to the pre-COVID-19 situation. We analysed these new auxiliary data in relation to available turnover data for industries with monthly statistics. By the time that we received those data, two months of actual turnover data were already available. Since we intended to use the auxiliary data for the early estimates of the quarterly statistics, there was no time to perform a good analysis with actual turnover information for industries with quarterly estimates. However, the analysis with turnover data for industries with monthly statistics gave us confidence that the auxiliary data could be used in an imputation model also for the industries with quarterly statistics.

We made strata based on whether the enterprise applied for the subsidy and actually received a subsidy. The enterprises that actually received a subsidy were further stratified according to the expected turnover drop (20–30%, 30–40%, 40–60%, 60–100%), see Table 4. Based on the monthly survey data of all industries available in SSD for April and May, we computed separate year-on-year growth rates for each of those strata, see Table 4. The actual observed turnover drop as averaged over industries was found to be small for enterprises that did not receive a subsidy, whereas it increased with an increase of the reported expected turnover drop for the enterprises that did receive a subsidy.

We concluded that the new auxiliary data could be useful to improve the imputations for the early estimates of the quarterly industries in the CCD system (although the previous analysis was necessarily carried out for the monthly industries). We wrote a script to compute the same ratio imputation as in CCD (see Section 3.4.1, also with the same auxiliary information), but with the domains further subdivided according to strata based on the new auxiliary data. We computed four alternative imputations, both a simple stratification (presence/absence in application) and extended stratifications based

on expected turnover drop categories. In a few cases no alternative imputations were calculated because the number of respondents was deemed too low (fewer than 10). Apart from these alternative imputations, the heterogeneity measures described in Section 3.4.2 were calculated as well.

In general, we found that the overall effect of the alternative imputations was limited. Depending on the exact method, the year-on-year growth rates with the alternative imputation method were 0.2–0.6 percentage points higher than with the normal imputation procedure (over all sectors with an early estimate from CCD). This indicated that early respondents on average had a lower turnover growth than late respondents within the original domains (without using the new auxiliary data). The effect was of the same order of magnitude as the known bias of the early CCD estimates in pre-COVID-19 periods, see also Section 3.5.1.

In Table 5 we give illustrative examples for some selected industries. The results for the four alternative imputations are given and the last column shows the final proposed values for our early CCD estimates (that are subsequently used as input by National Accounts). For most industries the effect of the alternative imputations was negligible, even for industries that were seriously affected by the lockdown measures like NACE code 55100 (‘Hotels and similar accommodation’) and 56101 (‘Restaurants’). For some industries a larger relative effect was found, but these often turned out to be small industries, like NACE code 58200 (‘Software publishing’). For a small number of larger industries, such as NACE code 63100 (‘Data processing, hosting and related activities; web portals’) and 71120 (‘Engineering activities and related technical consultancy’), a moderate effect was seen. But overall, the impact of the alternative imputations was small. One should keep in mind that the early estimates are not published at this detailed level. They are used as input for the early National Accounts estimates which are published at a much less detailed level. The alternative imputations for the CCD system were used to adjust the estimates

Table 5
Net turnover for the second quarter of 2020 and 2019, corresponding original year-on-year growth rate (YOY), and difference between new and original YOY growth rates, for different alternative imputations

Industry	Turnover 2020Q2 (10 ⁶ euro)	Turnover 2019Q2 (10 ⁶ euro)	YOY (%)	$\Delta \text{ alt } m = \text{YOY (alt } m) - \text{YOY (\%)}^{(*)}$				
				$\Delta \text{ alt } 1$	$\Delta \text{ alt } 2$	$\Delta \text{ alt } 3$	$\Delta \text{ alt } 4$	$\Delta \text{ final}$
55100	372	1722	-78.4	0.0	0.0	0.0	0.0	0
56101	973	2357	-58.7	0.1	0.1	0.0	0.0	0
58200	17	16	6.6	6.1	6.1	6.1	6.1	6
63100	1816	1578	15.1	2.1	2.1	2.4	2.2	2
71120	4386	4483	-2.2	1.0	1.0	1.0	1.0	1

(*) The alternative methods used different combinations of the strata mentioned in Table 4: $m = 1$: {1, 2-6}, $m = 2$: {1, 2, 3-6}, $m = 3$: {1, 2, 3, 4, 5, 6}, $m = 4$: {1-2, 3, 4, 5, 6}, 'final' is the rounded average of $m = 1, \dots, 4$.

of a few of the industries that were sent to National Accounts.

3.5. Nowcasting

3.5.1. Anticipated effects

The CCD system for the quarterly estimates uses ratio imputation to account for missing turnover values, see Section 3.4. Unfortunately, we know that the early estimates are typically slightly biased downwards. Therefore, for a few years alternative nowcast estimates have been used to support the regular analysis. The CCD estimates are still the basis for the early estimates used by National Accounts, but the analysts can adjust these estimates. The nowcast estimates use ARIMAX and VAR models with autoregressive terms and, as explanatory auxiliary variable, a time series based on the historical turnover of the enterprises with early response for the current period, as explained in Schouten [13].

We anticipated that, during the COVID-19 crisis, the nowcasting method with autoregressive terms might lead to a bias since – during periods of fast economic changes – time patterns in the past are no longer good predictors of the present.

3.5.2. Measures

We advised the analysts not to use the quarterly nowcast estimates as a reference during the crisis, but only to use the early CCD estimates based on ratio imputation. The analysts have followed this advice and have not used the nowcast estimates as a reference ever since.

3.5.3. Results

Figure 6 shows year-on-year growth rates (%) for the sector “Hotels and similar accommodation” for quarters just before the crisis and during the beginning of the crisis. The dotted line shows the late estimates when

most enterprises have responded. These results are reliable and so early estimates close to the late estimates are considered to be good estimates. Both the early CCD estimates (solid line) and the quarterly nowcast estimates (long dashed line) are also shown in Fig. 6. For this industry a vector autoregression (VAR) model was used as the nowcast, see Schouten [13] for details on the model. In the VAR model dummy variables for the quarters are also included to account for seasonal effects.

In Fig. 6 a dramatic turnover drop of 79% is seen for the second quarter of 2020. The quarterly nowcast model also showed a large drop of 62%, but the estimate was 17 percentage points too high. The turnover drop started in the first quarter of 2020, which was expected since the first lockdown started mid-March 2020 in the Netherlands. For this quarter the nowcast year-on-year estimates were also far too high. Furthermore, the nowcast underestimated the partial recovery in the third quarter. The same pattern was observable for many other industries. At an aggregate level of all industries with quarterly turnover estimates in NACE sections I, J, M, N and S together, the turnover was estimated to drop by 19% in the second quarter according to the late CCD estimates and the early CCD estimates whereas the nowcast estimated a 11% drop. While the nowcast estimates were not accurate during the crisis, the regular early CCD estimates were rather accurate. The aforementioned bias in the early CCD estimates is of an order of magnitude of 0.5 percentage points, much smaller than the effects seen for the nowcast estimates during the economic downturn.

The bad performance of the quarterly nowcast estimates during the crisis is logical. Autoregression terms in ARIMAX and VAR tend to pick up sudden changes slowly. This might be desirable for early estimates during economically stable periods because a sudden change in turnover of a limited number of observed

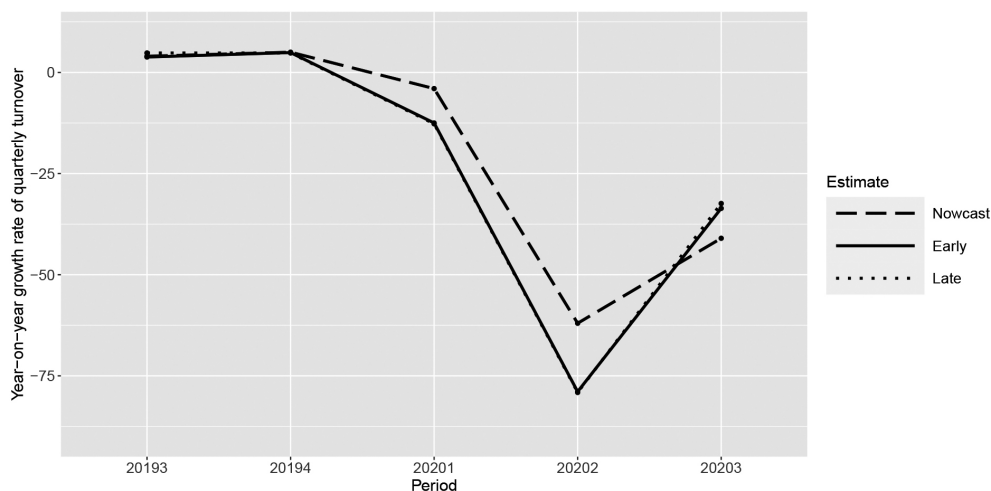


Fig. 6. Year-on-year growth rates of quarterly turnover of Hotels and similar accommodation, for the nowcast estimate, and early and late CCD estimates.

units might be due to noise. In case of a crisis however, sudden changes in observed turnover might be mainly caused by real economic changes and therefore should not be suppressed. Based on the reliable late estimates we concluded that the changes during the crisis were predominantly real and that the suppression of these changes was largely undesirable.

3.5.4. Implications for the monthly nowcast

The findings with these kind of autoregressive models during this crisis also impacts another development, namely the new monthly figures that have to be produced by the Netherlands from 2024 for NACE sections G45, G46, H, I, J, L, M, N, see our explanation in Section 2. CBS has decided to estimate these monthly figures for all these sections, except for G46 and L, using a nowcasting model with monthly VAT data from most enterprises and monthly questionnaires for a small group of large enterprises as input. Since most businesses declare VAT on a quarterly basis in the Netherlands, the monthly input data is incomplete and selective. The monthly nowcast method aims to correct for that selectivity, based on quarterly information from the past, using autoregressive terms.

Because such a standard nowcasting approach may lead to biased estimates during a crisis, alternatives are investigated at CBS. This is work in progress. One idea is to incorporate imputed data in the nowcasting models, since the ratio imputations turned out to react quickly to sudden changes in growth rates which may occur when the economic conditions changes. It will be interesting to compare different imputation methods on their reaction to sudden economic changes. Another idea is to

have a system that can accommodate two methods, the autoregressive nowcast model during “normal” periods and a method leaning heavier on recent response for periods with large economic changes. A critical question is how to determine when a crisis has started, see also the discussion in Smith and Lorenc [14]. In case of the COVID-19 crisis with accompanying lockdowns, this might be obvious but the situation might be more subtle for other crises. Furthermore, a crisis may not only affect the accuracy of nowcasts during a crisis, but due to its impact on model estimates (e.g. estimates of trends, seasonal patterns, autoregressive relations) it may also affect the accuracy of nowcasts after a crisis has passed. Therefore, a relevant question is also to ask, after a crisis, when to switch back to a nowcasting model that leans heavier on historical patterns. One could investigate this question by studying nowcasts using data of a real crisis that occurred in the past, which requires that such long microdata time series are available. To use the real data on the economic crisis of 2008, preferably microdata from about 2001 onwards should be available. Unfortunately it was not easy to construct such a microdata set due to changes in the IT systems used for the STS that occurred over the years. Instead, Zult et al. [15] have simulated three different crisis scenarios using the currently available microdata as a starting point and investigated their impact on different nowcast estimates.

3.6. Index computation

3.6.1. Anticipated effects

The estimated index of the target variables of the STS of industry k of the current period t (with base

period 0), denoted by $\hat{I}_k^{t,0}$, is computed as a chained index, i.e. as the index of the previous period times a period-on-period growth factor:

$$\hat{I}_k^{t,0} = \hat{I}_k^{t-1,0} \hat{G}_k^{t,t-1} \quad (11)$$

with

$$\hat{G}_k^{t,t-1} = \frac{\hat{Y}_k^{t(t,t-1)}}{\hat{Y}_k^{t-1(t,t-1)}} \quad (12)$$

and

$$\hat{Y}_k^{x(t,t-1)} = \sum_{i=1}^{n_k^{x(t,t-1)}} \delta_{ki}^{x(t,t-1)} d_{ki}^{x(t,t-1)} Y_{ki}^{x(t,t-1)},$$

$$\text{with } x = t \text{ or } x = t - 1 \quad (13)$$

where $\hat{G}_k^{t,t-1}$ stands for the estimated period-on-period growth factor of industry k . Throughout this paper a hat stands for an estimate. The weights $d_{ki}^{x(t,t-1)}$ are based on the sampling design of periods t and $t - 1$ and can be lowered in case of outliers; in case of the CCD system all weights $d_{ki}^{x(t,t-1)}$ are 1. In case of the SSD system, $n_k^{x(t,t-1)}$ refers to units in the sample, in case of the CCD system it involves all units in the population. The symbol $\delta_{ki}^{x(t,t-1)}$ indicates if unit i is included in the summation ($\delta_{ki}^{x(t,t-1)} = 1$) or not ($\delta_{ki}^{x(t,t-1)} = 0$). Sometimes units are excluded in situations of a restructuring. A restructuring concerns a merger, split and so on. In such a case, depending on certain conditions, either the unit is excluded for both periods ($t, t - 1$) or the turnover is estimated for both periods ($t, t - 1$) according to the same structure. Further, enterprises that change from industry are also excluded from both periods. For the various reasons just described, the turnover estimate of industry k for period t (denoted by $\hat{Y}_k^{t(t,t-1)}$) that is used to calculate the growth factor $\hat{G}_k^{t,t-1}$ can differ from the turnover estimate for period t (denoted by $\hat{Y}_k^{t(t+1,t)}$) for the subsequent growth factor. The estimates $\hat{Y}_k^{t(t,t-1)}$ and $\hat{Y}_k^{t-1(t,t-1)}$ are not intended to be accurate level estimates, the pair of estimates is used to estimate a period-on-period growth factor. Therefore, the index after a number of periods is not necessarily equal to the ratio of two directly estimated levels.

Due to an economic crisis, the accuracy of the index can be affected by a number of issues:

- a. Extreme growth factors: an economic downturn may result in very small growth factors while economic recovery may result in very large growth factors. Under these conditions the errors of estimation are larger. The compounding effect of repeatedly applying a factor with a large error may result in indices ending far up from their true value.

- b. Subpopulations within a stratum for which growth rates are computed may be affected differently by the lockdown measures, leading to an increased heterogeneity. This increased heterogeneity may lead to an increased variance of the growth rate, and thus of the index. In practice there is always a certain proportion of units that needs to be imputed. The respondents on which the imputation values are based may not be affected in the same way by the lockdown measures as the units to be imputed, which may lead to a bias of the index. See also Section 3.4.
- c. Outliers: economic downturn or recovery may lead to more outliers than in periods of economic stability and it is uncertain whether the treatment of outliers (reducing their weight) may result in a bias of the long term growth rate when a period of economic downturn is followed by an economic recovery.
- d. Secondary activities: our experience is that turnover of secondary economic activities is not always reported on the sample survey, probably because the respondent expects that CBS is only interested in turnover from the main activity. (For instance a supermarket only reports its customer sales, but not its income from rental activities). Turnover from secondary activities may become relatively more important during a lockdown.
- e. Dynamics in population composition: an economic downturn may lead to more bankruptcies. Furthermore, the rules to handle restructurings (mergers, splits and so on) may become inappropriate. When the rules are inappropriate, more manual corrections are needed.
- f. The effect of an economic downturn on different forms of turnover – domestic versus non-domestic; turnover on main activities versus turnover on secondary activities; turnover on trade – may be different. One must check whether the derivation rules to achieve consistency among the index of specifications of turnover with their total turnover index are still appropriate.

3.6.2. Measures

As a first measure we did a number of specific analyses to address the issues a–f. These issues were analysed separately in the SSD and in the CCD system. Some of them were already part of the regular analysis, which is described in more detail in Section 3.8. In cases where an issue is expected to have a large impact on the results of an industry, we intended to do further analyses for

that industry, for example by comparing results of the SSD system with those of the CCD system.

SSD and CCD were not equally sensitive to each of the mentioned potential problems. While SSD is sensitive to all the issues a–f, the CCD system is less sensitive to the issues b, c, d and f. Issue b is less relevant for CCD because its final estimates are based on census data; only the early estimates are based on an incomplete set of data, in which case heterogeneity may have an impact on the imputations, see Section 3.4. The CCD system is also less sensitive to c and f because it is not based on a sample but on a census. In general estimates based on census data encounter fewer estimation problems. Further, CCD is less sensitive to issue d because we expect that reporting errors on secondary activities mainly occur in survey data which is used for only a part of the target population.

As a second measure we intended to look into differences between the SSD and CCD results (for sections whose regular output is produced with the SSD system). This evaluation refers to different aggregation levels, to verify whether there are small effects within each industry that accumulate in the same direction. We intended to investigate these differences not only for total turnover but also for domestic and non-domestic turnover. Because issues a–f did not cause serious effects, see Section 3.6.3, we did not use this second measure.

A third measure was to verify for all industries whether the index returned to the correct level after the economy had normalised. This was done as follows. The first step was to detect whether there were any industries affected by the lockdown measures in which the index series might not have returned at their correct level: potentially incorrect series. This was done by computing trend lines in the seasonal adjustment software JDemetra+, by critically evaluating the plausibility of computed index series and – in case of industries produced by the SSD-system – by comparing their year-on-year changes with year-on-year changes of SSD-panel data. In the second step, we recomputed the indices of industries whose index series were detected as potentially incorrect. We recomputed the indices in different ways, making use of alternative index computations, which are described in detail in Van Delden and Van Bommel [16]. An alternative index computation for Eq. (11) is given by:

$$\hat{I}_k^{t,0}(\text{alt}) = \hat{I}_k^{u,0} \times \hat{G}_k^{t,u}(\text{alt}), \quad (14)$$

where u is defined as a period before COVID-19 measures were taken, for instance January 2020, and t is the

period of interest where the economy is (more or less) normalised. $\hat{G}_k^{t,u}(\text{alt})$ stands for a growth factor between period t and u that is computed in an alternative way. A first alternative is:

$$\hat{G}_k^{t,u}(\text{alt1}) = \frac{\hat{Y}_k^{t(L)}}{\hat{Y}_k^{u(L)}}, \quad (15)$$

where $\hat{Y}_k^{t(L)}$ denotes a level estimate for period t and industry k and $\hat{Y}_k^{u(L)}$ the corresponding level estimate for period u . In the CCD system such a level estimate can be approximated. However, for the SSD system, it requires quite some effort to make an accurate level estimate because the SSD system has not been designed for that purpose. The main purpose of the STS is to estimate a growth factor that is corrected for restructurings and for units with a change in industry. We therefore also defined a second alternative estimator for the growth factor:

$$\hat{G}_k^{t,u}(\text{alt2}) = \frac{\hat{Y}_k^{t(t,u)}}{\hat{Y}_k^{u(t,u)}}. \quad (16)$$

This growth factor is computed similar to the computation of $\hat{G}_k^{t,t-1}$ in Eq. (12), but now for the periods t and u . In alt2 one uses pair-wise outliers and corrections for restructurings and for units that change from industry.

3.6.3. Results

With respect to the first measure separate analyses were performed for the SSD system concerning the aforementioned issues a–f and for the CCD system for issues a and e, leading to the following findings:

1. For nearly all of the industries, turnover figures in the SSD system for lockdown periods were not very small and therefore extreme growth rates did not occur.
2. In the CCD system very low index values were occurring for NACE 79120 (Tour operating agencies; index value 1.1 in the second quarter (Q2) 2020 and 9.6 in Q1 2021) and 59140 (Motion picture projection activities; 4.3 in Q1 2021). Moderately low index values were occurring for NACE 55100 (Hotels; 24.2 in Q1 2021) and 56300 (Beverage serving activities; 20.0 in Q1 2021). For those four industries alternative indices were computed (see details below).
3. The observed variation in growth rates within strata did not create substantive issues for the computation of index figures.
4. The number of automatically detected and man-

Table 6
Original and alternative quarterly index series for NACE code 79120

Series	20201	20202	20203	20204	20211	20212	20213
Original	60.1	1.1	50.4	14.2	9.6	39.3	224.6
Alternative	60.1	1.1	27.0	7.6	5.1	21.1	120.6

ually appointed outliers were not significantly higher during the lockdown compared to previous years. Furthermore, we did not find that the identified outliers were causing a bias.

5. Analysts found that the reported proportion of turnover in secondary economic activities during the lockdown hardly changed compared to the previous periods.
6. More bankruptcies were expected to occur during the COVID-19 crisis, but this was not found. Furthermore, the automatic rules to handle restructurings were found to be still appropriate.
7. We did a first check whether the different forms of turnover (domestic, nondomestic) were affected differently by COVID-19 by using trend lines in JDemetra+, in which X-13ARIMA-SEATS was used. We did not find such an effect. Therefore, we did not change the derivation rules to achieve consistency among the different indices.

In conclusion, the analysis of the specific issues indicated that most of the anticipated effects did not occur in practice. Still, we would do a similar analysis in a future economic downturn because it helped us to assure that the output was of a sufficient quality. Because no large effects were found for the SSD results, the second measure of comparing them with CCD results was not used.

With respect to the third measure, we did a preliminary check whether the index levels of affected industries in SSD and in CCD returned to their pre-COVID-19 levels by using trend lines in the seasonal adjustment software JDemetra+, in which the method X-13ARIMA-SEATS was used. Also, for SSD, year-on-year changes of the series were compared with year-on-year changes of panel data. These analyses did not yield any index series where adjustments were needed.

Next, validation of the times series up to Q3 2021 showed that the indices of NACE codes 79120, 59140, 55100 and 56300 had a very sharp drop during one or both of the lockdown periods and the indices might not have returned at the correct level after the measures were lifted. Therefore, indices were recomputed using different forms of Eq. (15) that varied in which set of two periods (t, u) was used. From that step it was concluded that only for NACE code 79120 (Tour operating

agencies) a revision of the published data was necessary, for the other NACE codes differences between the alternative indices and their original ones were limited. For the revision of the NACE 79120 indices, we used Eq. (16) with t equal to Q3 2020 and u equal to Q1 2020. We used Eq. (16) because in the STS one aims to correct for the effects of restructurings. The original versus the recomputed index series for NACE code 79120 is given in Table 6. Especially at period Q3 2021 there was a large downwards correction of the original published index. Later, when the results for all quarters of 2021 were finalised, we computed final corrections.

3.7. Seasonal adjustment

3.7.1. Anticipated effects

For the STS, index series are created which are calendar and seasonally adjusted. The first effects of the lockdown measures on the adjusted series were observed for the reporting period of March 2020. At that time, we anticipated the occurrence of a sharp decline or increase in the values of economic variables from March 2020 onwards. In crisis situations, the need for adjusted figures is even higher than normally because of the need to be able to separate the “crisis effect” from the calendar and seasonal effects. However, a crisis can destabilize the seasonal trends if extreme values are not accounted for in the (X-13 ARIMA-SEATS) models. In particular, the model estimates are very sensitive to the decision whether the most recent data points are deemed atypical or not. If crisis periods are not designated as outliers in the modelling, this could lead to unreliable figures and to substantive revisions for the already published figures (the whole series).

In seasonal adjustment, we consider four types of outliers:

1. An ‘additive outlier’ concerns an abnormal value for an isolated point of the series.
2. A ‘transitory change outlier’ concerns a point jump which is followed up by a smooth return to the original path of the times series.
3. A ‘level shift outlier’ increases or decreases the level of the time series by some constant amount.
4. A ‘seasonal outlier’ can be used when breaks occur in the seasonal pattern due to specific and unusual events.

Table 7

Reporting months that have been appointed as additive outliers (AO) for the volume of production figures of Manufacturing industries after the analysis of reporting month July 2020. When analyzing reporting month July 2020, data was added for reporting month July 2020 and was revised for reporting months April 2020–June 2020. Cells with white backgrounds refer to pre-specified outliers that have been appointed prior to the analysis of reporting month July 2020 (using a 5% threshold see text), shaded cells refer to new pre-specified outliers appointed after the analysis of July 2020 (using a 5% threshold see text), and underlined cells refers to previously pre-specified outliers which have been manually removed after the analysis of July 2020

Industry	Mar 2020	Apr 2020	May 2020	Jun 2020	Jul 2020
10 Manufacture of food products		AO	AO		
11 Manufacture of beverages	AO	AO	AO		
12 Manufacture of tobacco products					
13 Manufacture of textiles	AO	AO	AO	AO	
14 Manufacture of wearing apparel	AO	AO			
15 Manufacture of leather and footwear	AO	AO	AO	AO	
16 Manufacture of wood products					
17 Manufacture of paper		<u>AO</u>	<u>AO</u>	<u>AO</u>	
18 Printing and reproduction	AO	AO	AO	AO	
19 Manufacture of coke and petroleum			AO	AO	
20 Manufacture of chemicals					
21 Manufacture of pharmaceuticals					
22 Manufacture rubber, plastic products	<u>AO</u>	AO	AO	AO	
23 Manufacture of building materials			AO		
24 Manufacture of basic metals		AO	AO	AO	
25 Manufacture of metal products		AO	AO	AO	AO
26 Manufacture of electronic products					
27 Manufacture of electric equipment	AO				
28 Manufacture of machinery n.e.c.					
29 Manufacture of cars and trailers	AO	AO	AO		
30 Manufacture of other transport					
31 Manufacture of furniture		AO	AO		
32 Manufacture of other products					
33 Repair and installation of machinery		AO	AO	AO	AO

See for more detailed information JDemetra+ [17] and United Nations [18].

Eurostat [19] gave advice with regard to calendar and seasonally adjusted figures, which also follows the ESS Guidelines (Eurostat [20]). Two potential outlier approaches were advised:

- Model the latest (atypical) values of a times series as additive outliers, and when data for new periods become available change it into a level shift or into a transitory change outlier. Changing an outlier from an additive outlier to a level shift outlier or transitory change outlier will give rise to revisions.
- An alternative is to fix the seasonal factors, for example by using 2019 factors during 2020 ('projected seasonal factors'). In this way one achieves approximately the same solution as the one described above. One has to decide how to update the models in the next year(s).

3.7.2. Measures and results

For most industries that were affected by COVID-19, the first advised outlier approach of Eurostat (in Section 3.7.1) was used. During normal economic periods,

the estimate of a period was detected as an outlier by using a Student t -test at a 1% significance level. The outlier detection at this threshold can be done automatically in the software JDemetra+. From the reporting month March 2020 onwards, the estimate of a period was detected as a potential outlier by using a Student t -test and by comparing its result with a 5% critical value of the Student's t distribution, to be able to also appoint periods with less outlying values as outliers. A potential outlier was manually appointed to be a pre-specified, fixed, outlier in JDemetra+ when it was judged that the outlying effect was indeed due to the economic conditions. CBS made use of pre-specified outliers to ensure that the same periods are set as an outlier each time the time series is updated with new data. In contrast to the fixed, pre-specified, outliers, the periods that are automatically appointed as an outlier in JDemetra+ can change when the time series is updated.

One constantly monitored whether pre-specified outliers should be removed or added, whether other types of outliers were needed or whether one should define separate regressors (other auxiliary variables) within the seasonal adjustment to model the COVID-19 crisis. For series with too many appointed outliers (> 6 in 2020 for

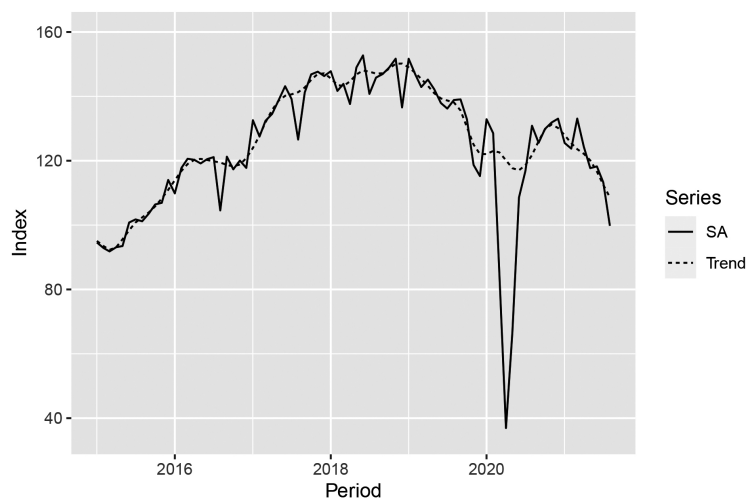


Fig. 7. Calendar and seasonally adjusted series (label: SA) of the volume of production and the trend line for NACE code 29 (Manufacture of motor vehicles and trailers).

monthly figures and > 2 in 2020 for quarterly figures), the second advised outlier approach of Eurostat was used.

CBS put clarifications in the explanatory texts that is provided when publishing figures, to inform users about the outlier approaches and their effects. The approach where outliers are monitored “along the way” during an economic downturn enabled CBS to also take into account the seasonal effects for reporting periods that were less or not affected by COVID-19. Better estimations of the seasonal effects can be made years after the crisis has passed, see also the discussion in Section 4.

In Table 7 an illustration of this approach is given for the time series up to July 2020 for the index of the volume of production for the Manufacturing industries. The time series of volume of production represents the short-term volume changes of value added. For most industries CBS uses the deflated turnover to calculate these changes. For some industries CBS uses other information from big enterprises like working hours, produced quantities and orders. Table 7 shows that most outliers occurred in the second quarter of 2020 due to COVID-19 effects. Furthermore, some of the previously pre-specified outliers were dropped after the first analysis was done for the reporting month July 2020.

In Fig. 7 an example is given for the industry “Manufacture of motor vehicles and trailers” (NACE code 29). At the end of March 2020 and in April 2020, several influential manufacturers in the industry temporarily closed their enterprises and therefore we saw extreme production declines in these particular months. In May 2020 the volume index figures recovered. The pre-

Table 8

Appointed additive outliers (AO) and level shift outliers (LS) for the time series of NACE code 29 (Manufacture of motor vehicles and trailers)

Outliers	Coefficients	T-stat	P $[T > t]$
AO (5-2020)	-51.0	-9.3	0.0000
AO (4-2020)	-84.3	-14.8	0.0000
AO (3-2020)	-41.4	-7.7	0.0000
LS (12-2008)	-54.1	-7.9	0.0000
AO (11-2008)	-24.0	-4.1	0.0001
AO (8-2004)	-21.9	-4.3	0.0000
AO (7-2004)	17.3	3.4	0.0007

specified outliers for NACE code 29 are presented in Table 8. In the future, we will probably use a transitory change outlier because the index series has returned to a similar level as in the pre-COVID-19 situation. Without identifying the outliers in 2020, the values in March-May 2020 would have caused large revisions of the calendar and seasonally adjusted figures for reporting periods prior to 2020. Furthermore, it would also create problems for 2021. If the abnormal low values in 2020 had been accepted rather than set as pre-specified outliers, then the Students t -test for the months March, April and May 2021 would have been significant at a 5% threshold (see Table 9). Also, the trend for the previous and coming years would be somewhat affected if the outlying values in the second quarter of 2020 are not accounted for (see Fig. 8).

In the future we aim to determine if the additive outliers for all the impacted industries can be replaced by transitory change outliers or level shift outliers.

One peculiarity we would like to mention is that the seasonal adjustment in our business cycle survey

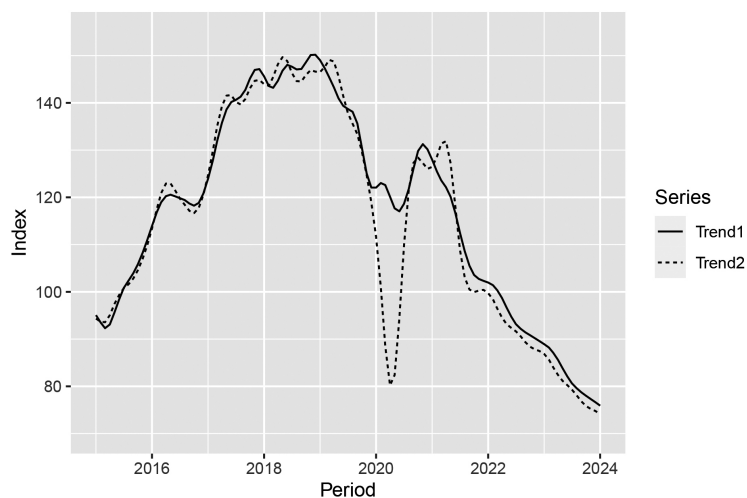


Fig. 8. The STS volume of production for NACE code 29 (Manufacture of motor vehicles and trailers), with two trend lines. For trend line 1 COVID-19 outliers are set for March, April and May 2020. Trend line 2 is without setting COVID-19 outliers for March, April and May 2020. The trend lines have been derived from the observed values from January 2000 up to and including August 2021 and predicted values from September 2021 up to and including August 2025. Results are shown up to the end of 2023.

Table 9

Additive outliers (AO) detected (using a 5% threshold see text) in 2021 for the time series of NACE code 29 (Manufacture of motor vehicles and trailers), when we do not treat any observations in 2020 as additive outliers

Outliers	Coefficients	T-stat	$P[T > t]$
AO (5-2021)	27.5	2.2	0.0341
AO (4-2021)	46.1	3.6	0.0005
AO (3-2021)	31.3	2.6	0.0104

resulted in a value that was out of range. In this survey, one of the questions is whether respondents expect turnover in the coming three months to increase, decrease, or stay approximately the same. The estimated net effect yields a value between -100% (all entrepreneurs expect a turnover decrease) and 100% (all entrepreneurs expect a turnover increase). However, the calendar and seasonally adjusted figure for the hotel industry in Q2 2020 was estimated to be $-112,6\%$ due to an unadjusted value that was close to -100% combined with a negative (additive) seasonal adjustment term. CBS decided to suppress the seasonally adjusted figures for the hotel industry for Q2 2020.

3.8. Regular analysis

The analyses on the effects of new or adapted methodologies as described in the previous sections were carried out by a team consisting of statistical researchers and methodologists. In this section, the focus lies on the regular analyses of the CCD and SSD results that are carried out by industry analysts. To monitor

the quality of the estimations and the reported data during lockdown periods, they used economic information on the lockdowns and extra indicators. The COVID-19 crisis already started to influence some statistics for the reporting period March 2020. Therefore, for the reporting period March, we already aimed at acquiring insights into the impacts of COVID-19 per industry by collecting information from additional sources. We collected information via financial newspapers, websites of several big enterprises, governmental websites about COVID-19-regulations and publicly available PIN debit transaction figures from banks for several economic activities to anticipate which industries needed more attention during future analyses. Besides the use of these additional sources, we put extra effort to the analysis by using a combination of normal and additional indicators. Overall, extra capacity was required compared to normal conditions, analyses started earlier than normal, and more attention was paid to specific industries. Furthermore, we extended our analysis by addressing lockdown effects. COVID-19-spreadsheets were created with special indicators for the CCD and SSD system. Some of these indicators are related to the specific issues mentioned in Section 3.6.1.

An analysis always consisted of an overview of the following five indicators listed below. These indicators were used in a top-down approach to select industries for which the underlying microdata are further inspected. Each of those indicators concerns the computation of a score that is a quantitative measure of quality (of some aspect) of the results of the in-

dustry. The threshold values of the indicators were set by researchers based on their experiences with earlier analyses.

1. **Indicator for population/sample dynamics.**

The combined influence of inflow (e.g. births) and outflow of enterprises (e.g. bankruptcies) on the month-on-month (or quarter-on-quarter) growth rates of industries. We used an indicator at industry level. If the dynamics caused a change of less than -2 percentage points or more than 2 percentage points on the month-on-month (or quarter-on-quarter) growth rates of an industry, analysts had to pay special attention to the inflow and outflow of enterprises.

2. **Indicator for influence of imputations.**

Analysts also used an imputation-indicator for enterprises that were present in the industry in the current period and the previous period. The imputation-indicator measures the relative difference between the month-on-month (or quarter-on-quarter) growth rate of enterprises for which we have received data in the previous and the current reporting periods to the month-on-month (or quarter-on-quarter) growth rate of all enterprises (responses and nonresponses) in the industry. This is in particular an important indicator for industries with low turnover response rates. If the indicator value was below -5 percentage points or above 5 percentage points, analysts had to pay special attention to the imputations.

3. **Indicator for period-on-period changes.**

This indicator measures for every industry the difference between the month-on-month (or quarter-on-quarter) growth rate of the current period and the month-on-month (or quarter-on-quarter) growth rate of the same period in the previous year. If the indicator value was below -2.5 percentage points or above 2.5 percentage points, an analyst had to pay special attention to detecting anomalies in the microdata within those industries. During the COVID-19 crisis, most industries needed special attention with regards to this indicator. The original score functions for the microdata described in Section 3.3 and scatter plots were used to detect and analyse extreme values in the data. Analysts also considered the differences between score functions in the current reporting period and the same reporting period in the previous year. In this way, analysts could investigate which enterprises showed a difference in influence on the month-on-month growth factors in the current re-

porting period versus those of the previous year (or the quarter-on-quarter changes) for certain industries. Some validators also used the outlier information from the seasonal analyses to validate the results (see Table 7 in Section 3.7)

4. **Influence of enterprise restructurings on short term changes.**

We compared the month-on-month (or quarter-on-quarter) growth rates of industries with and without enterprises that change from industry or are involved in restructurings like mergers, takeovers and reorganizations. If the difference between these two growth rates was smaller than -1 percentage point or larger than 1 percentage point, the relevant enterprises were checked by analysts.

5. **Revisions compared to a previous release.**

The revisions of month-on-month growth factors of the current release compared to the previous release were also analysed. If an industry had revisions below -2 percentage points or above 2 percentage points, analysts examined the underlying microdata more closely.

Table 10 shows a part of a spreadsheet for reporting period April 2020 for industries in the SSD system. Analysts examined the microdata in more depth when the values of the indicators exceeded their limits.

Besides the spreadsheets, we analysed 0 values reported by enterprises, outliers and the imputations for industries with a high heterogeneity. Analysts also examined in more depth industries where 0 or very low values were expected.

3.9. Output

A special COVID-19 dashboard was launched by CBS, for all kinds of monthly and quarterly statistics for which COVID-19 effects were interesting, see URL [21]. Furthermore, as explained in Section 3.2, retail figures were published two weeks earlier than before to inform people about the COVID-19 crisis. The press releases on STS and related statistics paid a special attention to COVID-19 effects. In the explanations of the figures it was underlined that results of industries severely affected by the lockdown might be less accurate than usual.

During the first months of the first lockdown a desire for additional new output arose; for instance, to deliver monthly rather than quarterly figures for NACE section I (“Accommodation and food service activities”). A first idea was to use the newly developed monthly nowcasting method, see Zult et al. [22]. However, because of

Table 10

Analysis-indicators (see text; in percentage points) of industries for reporting month April 2020. Indicator-values in bold have exceeded their threshold-value(s)

Industry	Population/sample dynamics	Influence imputations	Period-on-period changes	Restructurings	Revisions
29000	-0.1	0.1	-57.7	0.0	-0.3
47742	0.6	0.3	-48.9	0.2	0.0
45111	0.0	0.0	-48.3	0.0	-0.2
47782	0.0	0.2	-48.0	-0.5	0.0
47762	0.0	-6.1	42.8	0.0	0.2
09000	0.0	2.7	-33.9	0.0	0.1
47890X	0.0	28.8	-33.0	0.0	11.9
19000	0.0	0.0	-32.6	0.0	0.0
15000	0.0	-3.6	-31.1	0.0	1.1
18000	-0.3	0.9	-29.8	0.1	-1.0
32000	-0.8	10.8	-28.3	0.0	0.3
13000	0.2	-0.4	-28.2	0.0	1.0
38300	-0.2	-3.6	-27.8	0.0	0.5

the expected problems with the autoregressive terms, this idea was dropped. Instead, some monthly indicators were derived by estimating month-on-month growth rates from units that report VAT on a monthly basis. Each month-on-month growth rate was estimated by only selecting units that had responded on both months of the corresponding growth rate ('a response panel'). However, the results were considered unreliable and were therefore not published, because of the uncorrected selectivity of the monthly reporters.

CBS created additional output on short-term COVID-19 effects by asking questions on this topic in our business cycle survey that collects opinions of entrepreneurs with respect to their confidence in the economy. Within that survey there are three questions that CBS can change from time to time to give insight about current topics. These three questions were used to collect information about the type of COVID-19 measures that entrepreneurs applied and about the expected effects of the COVID-19 crisis on the turnover and on continuity of their enterprise, see URL [23].

The Dutch government implemented several support measures for enterprises that were impacted by COVID-19. At the request of the Ministry of Economic Affairs and Climate, CBS has, on a regular basis, reported how many enterprises participated and what the effects were of different support measures. The figures were stratified by several enterprise characteristics such as size, age and industry. Reports about the support measures are available on the CBS website, see URL [24].

4. Discussion

We identified several types of risks that might affect the quality of the estimated STS output as a result of

the economic impact of the COVID-19 crisis, mainly due to the lockdown measures. We developed a number of measures to analyse the occurrence of those anticipated effects and also some measures to mitigate them. For the industries whose monthly output estimates are produced using survey data (SSD system) we were positively surprised that response rates were not greatly affected. For the industries that are estimated with the combined survey and VAT data (CCD system), the response rates for the early estimates were lower than in the same quarter a year before, but still at an acceptable level. Early respondents were slightly selective, i.e. their growth factor was slightly smaller than that of late respondents, but this was also the case in pre-COVID-19 periods. Some of the domains which are used to compute growth factors for imputation had an increased heterogeneity, which is an indication that the corresponding units were affected differently by the lockdown measures. Such a differential effect is a risk for the quality of the imputations. The effect of alternative imputations (to replace imputations from heterogeneous domains and for industries strongly affected by the lockdown) using new auxiliary data was limited. The effect of the lockdown periods on outliers turned out to be very limited, and no adaptations were needed in the automatic rules to handle restructurings. Unfortunately, there was one industry with a very low index during the first lockdown period, where the index values were overestimated after this lockdown period. Those index values were revised afterwards, using an alternative index formula.

Two processing steps were clearly affected by the COVID-19 crisis. The first step concerned nowcasting methods based on autoregression terms with inaccurate results in case of an economic crisis. The estimates of the quarterly nowcasting model were no longer used

as a reference. We were fortunate that the newly developed monthly nowcasting model (to be used directly for estimation, not only as reference) had not yet been brought into production. We are currently working on developing a nowcasting approach that is accurate under a wider range of economic conditions. The second step concerned seasonal adjustment. We had to manually identify additional outliers to guarantee sensible seasonally adjusted results. As explained in Section 3.7, one can use different types of outlier adjustments in case of seasonal adjustment. A long period after an economic downturn, one can determine the correct seasonal adjustment more precisely and work out what model span to use and what type of outliers should be set. We considered four types of outliers during the lockdown periods. In the future we will investigate if *ramps*, another outlier type, can also be applied to several series. Ramps are used if one observes a smooth, linear or quadratic transition over a specified time interval, as explained in Eurostat [20]. Many different national statistical institutes are struggling to find the best way to achieve a seasonal adjustment during the COVID-19 crisis, see for instance Foley [25], Olaeta and Armendi [26] and Matthews [27].

In our description of COVID-19 effects on turnover-related STS we focused on the main effects and we do not claim that we have covered all effects. For instance, an internal reviewer remarked that CBS had to make adjustments to enterprises that report turnover on a four-week basis - which concerns mainly, but not only, some enterprises in Retail trade. Four-week values are transformed to corresponding monthly values in the SSD system using standard seasonal patterns. These seasonal patterns were adjusted since COVID-19 greatly affected Retail trade, especially during the first lockdown. The adjusted seasonal patterns were based on scanner data which are detailed data on the sales of goods, obtained by scanning the bar code of consumer goods. Another point with respect to COVID-19 effects that we have not described in the main text, is that CBS validated the values of the microdata of the structural business statistics over the year 2020 for enterprises with industries affected by lockdown measures. More specifically, this concerned enterprises who reported their results over a twelve month period which does not fully coincide with a calendar year. Before COVID-19 crisis, CBS assumed that the turnover in the months before the reporting year would approximately cancel out against the turnover of the months that are missing for the reporting year. That assumption was no longer valid during the COVID-19 crisis. Instead CBS computed a

correction, since the distribution of the values over the months of the year was affected by lockdown measures.

The economic crisis due to COVID-19 will not be the last crisis that can affect the quality of our business statistics estimates. Smith and Lorenc [14] remark that the current COVID-19 crisis may stimulate adaptations to the methodology of producing business statistics in order to make them more robust against sharp economic changes. They make a plea to share information among different national statistical institutes in what adaptations to the methodology are effective in making them more robust against economic downturns. The present paper and Jones et al. [28] are examples of documenting experiences with adaptations of the production of business statistics to an economic downturn.

The question arises: should we make permanent changes to the STS production systems to achieve accurate output (i.e. low bias and variance) under both steady economic conditions as well as situations of fast declines or sharp increases in growth rates? Furthermore, to what extent do those adaptations depend on the nature of the economic crisis? We believe that we should make a distinction among three groups of changes.

The first group concerns those indicators and monitoring instruments that are meant to detect economic changes which might have an impact on the assumptions used in the methodology (editing, imputation, estimation) to produce the STS statistics. Those indicators and monitoring instruments can be permanently built into the production systems and can be used when needed. An example is the heterogeneity measure. The indicators that are mentioned in the present paper are score functions that quantify some aspect of the quality of the results of industries or of individual records. Their use is only effective when it is clear what should be done when their limits are exceeded. It is not necessary to use all indicators and monitoring instruments during a stable economic situation.

The second group concerns adaptations that can be taken to make the methodology for STS statistics applicable for a wider range of economic conditions. The methods that are used rely on assumptions, which were not always explicitly identified. There is a risk that changing economic conditions causes some of the assumptions to become invalid. The adaptations in the second group provide alternative methods that are insensitive or less sensitive to effects of an economic downturn. An example concerns the nowcasting approach. CBS is investigating various alternative models that depend less on autoregression terms. A second, po-

tential point of permanent change is the use of the newly developed editing score function, in addition to the earlier editing score function. We are investigating which score function works best under which conditions.

Finally, the third group concerns measures which have a crisis-specific component. For instance, we used an auxiliary data source with respect to a subsidy regulation to improve our imputations; that regulation has been developed specifically to support entrepreneurs during the COVID-19 pandemic. This specific auxiliary data set is no longer available, and we do not know yet whether a similar regulation will be available during a next crisis affecting the economy.

Acknowledgments

We thank participants of the 2021 European Establishment Statistics Workshop for their useful comments during the workshop. Furthermore, we thank Jan de Haan, Ingrid Zum Vörde Sive Vörding and Derk van Wijk for reviewing an earlier version. We thank Leonoor Duijvenbode-van Well for her language corrections. Finally, we thank the anonymous reviewers who carefully read the submitted version and whose comments considerably improved the paper.

References

- [1] RIVM; 2021. Available from: <https://www.rivm.nl/gedragsonderzoek/tijdlijn-maatregelen-covid>.
- [2] Brand M. Methodological Issues Arising for the Office for National Statistics from the Recession. In: European Establishment Statistics Workshop, 7–9 September 2009, Stockholm, Sweden; 2009. Available from: <https://statswiki.unece.org/display/ENBES/Topical+Papers>.
- [3] Simkins A, Smith PA, Brand M. Financial crisis and recession: How ONS has addressed the statistical and analytical challenges. *Economic and Labour Market Review*. 2010; 4: 30–37.
- [4] Eurostat. NACE Rev 2. Statistical classification of economic activities in the European Community. Eurostat Methodologies and Working papers. 2008.
- [5] Van Delden A, Scholtus S, Ostlund N. Modelling measurement errors to enable consistency between monthly and quarterly turnover growth rates. In: Proceedings of Statistics Canada Symposium 2018: Combine to Conquer: Innovations in the Use of Multiple Sources of Data; 2018.
- [6] Commission Implementing Regulation (EU) No 2020/1197 of 30 July 2020 laying down technical specifications and arrangements pursuant to Regulation (EU) 2019/2152 of the European Parliament and of the Council on European business statistics repealing 10 legal acts in the field of business statistics. *Official Journal of the European Union*. 2020; L 271: 1–170.
- [7] Moreira A, Portugal A, Lima B, Poças J, Magalhães J, Cruz P, et al. The use of electronic invoice data in COVID time. In: European Establishment Statistics Workshop, 14–17 September 2021, online event; 2021.
- [8] Šuštar Kožuh L. Data collection in the light of COVID-19. In: European Establishment Statistics Workshop, 14–17 September 2021, online event; 2021.
- [9] Van Bommel KJH, Goorden M. Imputation of turnover for short term statistics in case of extreme growth rates due to the COVID-19 crisis. CBS report (Available upon request, in Dutch). The Hague/Heerlen: Statistics Netherlands; 2020.
- [10] Van Delden A, Scholtus S. Analysing response differences between sample survey and VAT turnover. CBS Discussion paper. The Hague/Heerlen: Statistics Netherlands; 2019. Available from: <https://www.cbs.nl/en-gb/background/2019/22/analysing-response-differences-in-vat>.
- [11] Kutner MH, Nachtsheim CJ, Neter J, Li W. *Applied Linear Statistical Models*. Fifth edition ed. New York: McGraw-Hill Education; 2005.
- [12] Rousseeuw P. Least median of squares regression. *Journal of the American Statistical Association*. 1984; 79: 871–880.
- [13] Schouten B. Fast nowcasts with incomplete response. CBS report (Available upon request, in Dutch). The Hague/Heerlen: Statistics Netherlands; 2018.
- [14] Smith PA, Lorenc B. Robust official business statistics methodology during COVID-19-related and other economic downturns. *Statistical Journal of the IAOS*. 2021; 37: 1079–1084.
- [15] Zult D, Krieg S, Schouten B, Ouwehand P, Van Den Brakel J. From quarterly to monthly turnover figures using nowcasting. *Journal of Official Statistics* (conditionally accepted). 2022.
- [16] Van Delden A, Van Bommel KJH. Indices as growth factors of the short term statistics in COVID-19 times. CBS report (Available upon request, in Dutch). The Hague/Heerlen: Statistics Netherlands; 2020.
- [17] JDemetra+. User-defined outliers (jdemetrdocumentation.github.io); 2021. Available from: <https://jdemetrdocumentation.github.io/JDemetra-documentation/pages/case-studies/detailedsa-outliers.html>.
- [18] United Nations. Practical Guide to Seasonal Adjustment with Demetra+: From source series to user communication; 2021. Available from: https://unece.org/fileadmin/DAM/stats/publications/Practical_Guide_to_Seasonal_Adjustment_final_web.pdf.
- [19] Eurostat. Guidance on treatment of COVID-19-crisis effects on data. Luxembourg; 2020. Available from: https://ec.europa.eu/eurostat/cros/system/files/treatment_of_covid19_in_seasonal_adjustment_methodological_note.pdf.
- [20] Eurostat. ESS guidelines on seasonal adjustment. Luxembourg; 2015. Available from: <https://ec.europa.eu/eurostat/documents/3859598/6830795/KS-GQ-15-001-EN-N.pdf>.
- [21] CBS; 2022. Available from: <https://www.cbs.nl/nl-nl/visualisaties/welvaart-in-coronatijd>.
- [22] Zult D, Krieg S, Schouten B, Ouwehand P, Van Den Brakel J. From quarterly to monthly turnover figures using nowcasting. CBS Discussion paper. The Hague/Heerlen: Statistics Netherlands; 2020. Available from: <https://www.cbs.nl/nl-nl/achtergrond/2020/14/maandcijfers-over-omzet-met-behulp-van-nowcasten>.
- [23] CBS; 2022. Available from: www.cbs.nl/coen.
- [24] CBS; 2022. Available from: <https://www.cbs.nl/nl-nl/maatwerk/2021/27/gebruik-van-steenmaatregelen-corona-per-31-mei>.
- [25] Foley P. Seasonal adjustment of Irish official statistics during the current COVID-19 crisis. In: European Establishment Statistics Workshop, 14–17 September 2021, online event; 2021.

- [26] Olaeta H, Armendi J. Adapting seasonal adjustment procedures in uncertain times. In: European Establishment Statistics Workshop, 14–17 September 2021, online event; 2021.
- [27] Matthews S. Seasonal adjustment during the covid-19 pandemic: Statistics Canada’s approach. In: European Establishment Statistics Workshop, 14–17 September 2021, online event; 2021.
- [28] Jones J, McLaren C, Elliott D, Mead D, Baird H, Sakshaug J, et al. Producing Official Statistics during the COVID-19 Pandemic. Chapter 11 in: Snijkers G, Bavdaž M, Bender S, Jones J, MacFeely S, Sakshaug J, et al., eds. *Advances in Business Statistics, Methods and Data Collection*. John Wiley & Sons, Inc.; 2023.