Discussion Paper

Consistent Multivariate Seasonal Adjustment for Gross Domestic Product and its Breakdown in Expenditures

2017 | 06

Reinier Bikker, Jan van den Brakel, Sabine Krieg, Pim Ouwehand, Ronald van der Stegen

Content

Summary
Seasonally adjusted series of Gross Domestic Product (GDP) and its breakdown in underlying categories or domains are generally not consistent with each other. Statistical differences between the total GDP and the sum over the underlying domains arise for two reasons. If series are expressed in constant prices, differences arise due to the process of chain linking. These differences increase if in addition a univariate seasonal adjustment, with for instance X-13ARIMA-SEATS, is applied to each series separately. In this paper, it is proposed to model the series for total GDP and its breakdown in underlying domains in one multivariate structural time series model with the restriction that the sum over the different time series components for the domains are equal to the corresponding values for the total GDP. In the proposed procedure this approach is applied as a pre-treatment to remove outliers, level shifts, seasonal breaks and calendar effects, while obeying the aforementioned consistency restrictions. Subsequently, X-13ARIMA-SEATS is used for seasonal adjustment. This reduces inconsistencies remarkably. Remaining inconsistencies due to seasonal and calendar adjustment are removed with a benchmarking procedure.

Keywords
Seasonal adjustment, discrepancies, Kalman filter, Multivariate structural time series models, X-13-ARIMA-SEATS, benchmarking

Acknowledgement
The views expressed in this paper are those of the authors and do not necessarily reflect the policy of Statistics Netherlands. The authors are grateful to Harm Jan Boonstra and Jan van Dalen for reading and commenting on a former draft of this paper.

1. Introduction

Adjustment for seasonal and calendar[1] effects is common practise amongst national statistical institutes (NSI's). The aim of these adjustments is to separate long term variation from seasonal fluctuations to make different reporting periods comparable. Most official statistics publish figures at an aggregated level and breakdowns in $K \geq 2$ domains, for instance the Gross Domestic Product (GDP) divided over industries or over expenditures. Seasonal and calendar adjustment procedures are generally based on univariate methods applied to the series of each publication domain separately. A consequence of such approaches is that adjusted figures at the aggregated level are not consistent with the sum over the adjusted figures of the underlying breakdown in K publication domains. In this paper, we focus on the Gross Domestic Product (GDP) and a breakdown in different expenditures. The proposed method is general and can be applied in any situation where consistent seasonal and calendar adjustment is required. This is a well-known problem and the status quo is that no adequate solution exists.

Eurostat's ESS guidelines on seasonal adjustment (Eurostat, 2015) suggests to compute the adjusted series at the aggregated level as sum of the adjusted underlying domains, which is often referred to as the indirect approach. A drawback of this approach is that the most reliable estimates at the aggregated level are disregarded. Alternatively, if the discrepancies are small enough, they can be distributed by means of multivariate benchmarking techniques, like the extended multivariate Denton method described in Bikker *et al* (2013). In the Netherlands the changes in the discrepancies between the quarterly figures of the directly adjusted GDP and the sum of the adjusted series of its expenditures, is often larger than the growth rate of GDP itself. This fact alone renders both suggested remedies unsuitable.

Another discrepancy is introduced by chain linking (see Bloem *et al*, 2001), when GDP and its expenditures are calculated as chain volumes. In chain linking, series of a constant price level are constructed by "chaining" volume growth rates. These volume growth rates are calculated by dividing the nominal growth rate by a price factor. As each of the series in the breakdown of GDP has its own price factor, discrepancies arise between the sum of the expenditures and GDP itself.

In the Dutch case, the discrepancies from chain linking typically are smaller than the discrepancies introduced by the adjustments for seasonal and calendar effects. Moreover, we noted that the typical size of the discrepancies due to seasonal and calendar adjustment grew larger in the period 2009–2013. This period is characterized by rapid changes in seasonal patterns following the financial crisis in 2007/2008.

[1] Calendar effects are variations in a time series that can be explained from variations in the calendar. This includes working day patterns and national holidays.

The increasing size of the discrepancies eventually lead to complaints from users. The discrepancies were noted in the press, and also professional users asked how to interpret our published results. By 2013 Statistics Netherlands developed an urgent wish to reduce or eliminate the discrepancies arising from the adjustment of seasonal and calendar effects under the constraints that the size of the revisions may not increase and the quality of the seasonal adjustment may not decrease on average.

When new data points become available and are added to the series, better estimates of the trend, the seasonal and calendar effect of all previous quarters can be made. Therefore revisions are inherent to seasonal and calendar adjustment.

Traditionally the quality of seasonal and calendar adjustment is assessed using a well-defined set of criteria. In the case of X-13ARIMA-SEATS (US Census Bureau, 2015), the method used at Statistics Netherlands, these are the Q and M-diagnostics. They are numerical scores given to properties like the amount of seasonality compared to noise and the rate at which the seasonal component changes in time. These criteria are optimised for each time series individually. After performing seasonal and calendar adjustment, the resulting discrepancies are calculated and only when these are very large, the seasonal and calendar adjustment may be changed. Revisions are monitored, but never lead to changes in the setup of seasonal and calendar adjustment. So the quality criteria we traditionally applied are (in order of decreasing importance):

1. Optimal quality diagnostics (specifically X-13ARIMA-SEATS's Q and M-values)
2. Minimal statistical discrepancies between GDP and the sum of expenditures
3. Minimal revisions of the seasonal effect under addition of new data points.

This is under the assumption that all criteria are within acceptable boundaries. In the Netherlands, this was not the case after the crisis of 2007/2008. Therefore it is the primary objective of this research to reduce the statistical discrepancy. This is achieved by introducing a multivariate approach. The consequence from the shift of an optimal univariate solution to a multivariate solution is that the seasonal and calendar adjustment of one series is influenced by another. Therefore some aspects of the multivariate adjustment can be less optimal, when compared to the univariate case. However, a slightly lower quality (according to the Q and M-diagnostics) can be just as acceptable for the users, as long as no residual seasonal effect can be found in a corrected series. Therefore our goal is that, on average, the revisions and the quality diagnostics should not deteriorate.

In this paper we describe two alternative approaches for adjustment for seasonal and calendar effects. The first approach employs a multivariate structural time series model to an aggregated series and its breakdown in K subseries. The model estimates all components subject to the constraint that the sum over the subseries is equal to the components of the aggregated series. Unfortunately, the results under this approach are not satisfactory due to numerical problems. Furthermore, the estimates for the seasonal components are considered to be too volatile. Therefore, a second approach is developed, which is based on a combination of a multivariate structural time series model and routines of X-13ARIMA-SEATS. Under this approach, the discrepancies are sufficiently reduced whereas the size of the revisions is in the same order as before.

In Section 2 we will first define the problem in a more precise way. Section 3 discusses four intuitive alternative approaches to this problem, which in the end did not result in a satisfactory solution of the problem. In Section 4 we present the method that did meet all our criteria. Special attention is given to outlier detection in Section 4.2, as this has a crucial influence on the quality of the results and the stability of the estimates. Section 5 discusses the results and finally, Section 6 concludes.

2. Problem definition

The statistical discrepancies that we discuss in this paper are those between an aggregate series and the sum of its breakdown in $K \geq 2$ publication subseries. Statistics Netherlands publishes quarterly figures for GDP with both the final expenditures and the value added by industry as domains. These breakdowns are called the expenditure approach and the production approach. Both breakdowns are computed in constant prices (chain linked volumes) and in current prices. In this paper we will focus on the expenditure approach in constant prices.
The aggregate B1G, the GDP, consists of its subseries[2] P7 (imports), P31S1A (consumption households), P31S13 (consumption government), P51G (gross fixed capital formation), P5M (changes in stocks and inventories), P6 (exports) and SD (statistical discrepancy due to chain linking), i.e.

$$B1G = -P7 + P31_{S1A} + P31_{S13} + P51G + P5M + P6 + SD. \qquad (1)$$

The way we will handle the discrepancies arising from chain linking is by considering them as an extra subseries of the aggregate. It is a series that must be adjusted for seasonal and calendar effects, together with the other subseries of the aggregate. The method we developed is therefore also suitable for series where no chain linking takes place, such as current price data and any other set of series where preserving additivity or at least reducing discrepancies due to adjustments for seasonal and calendar effects is required.

As we will apply our model to chain volumes of GDP and its expenditures, the total statistical discrepancy after adjustments for seasonal and calendar effects can be divided in two parts, each with its own origin.

[2] Note that CBS actually publishes a more detailed breakdown into expenditures than the one with seven expenditures in equation (1), see appendix 1. The production approach is not considered in this paper. Any breakdown of GDP has the same problems with additivity, so for brevity, we only use the breakdown above in this article.

2.1 Discrepancies arising from chain linking

The first part of the total discrepancies already exists before making adjustments for seasonal and calendar effects. These discrepancies are introduced by chain linking of constant price data. In each linking step the aggregate and each of its subseries are multiplied by a different price factor, leading to the loss of additivity.

The statistical discrepancies due to chain linking typically have a slow moving long term trend-cycle, combined with a strong short term pattern. The short term pattern has a seasonal and an irregular component. Figure 2.1 shows a typical example for total GDP.

The statistical discrepancies due to chain linking can be interpreted as the consequence of changes in relative prices of subseries of the aggregate. One can show that the sum of the chain linked expenditures is chain volume with different weights. In each link step, the chain volume of GDP is weighted with the relative values of its expenditures in the previous year, valued at prices of the previous year, whereas the sum of expenditures is weighted with the relative values of the previous year, in prices of the reference year. So the statistical discrepancy due to chain linking is the difference between the value of the aggregate valued in previous year's prices and the value of the aggregate valued in prices of the reference year. The discrepancies due to chain linking are therefore zero in the first year after the reference year and tend to be bigger the further away from the reference year.

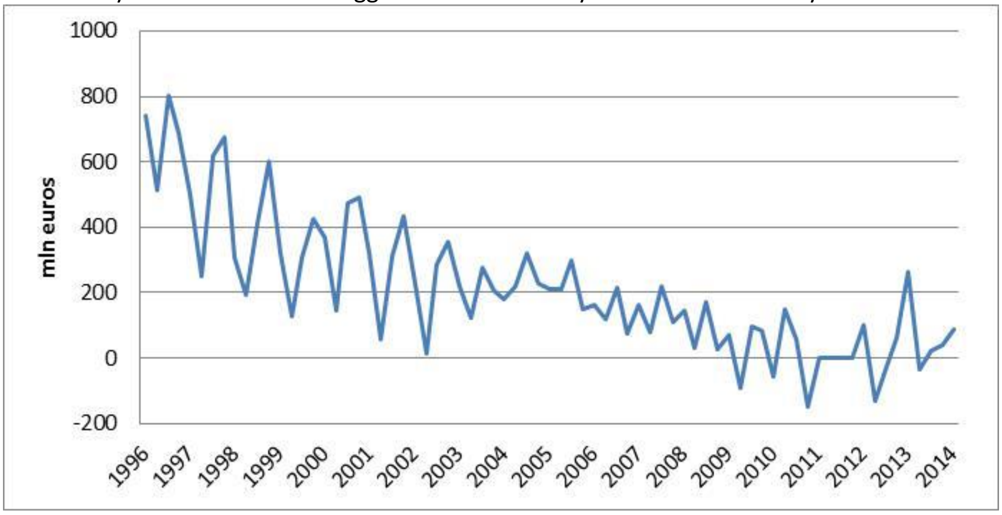

Figure 2.1. Statistical discrepancies due to chain linking between Dutch GDP and the sum of the final expenditures (before adjustments for seasonal and calendar effects), reference year = 2010.

The statistical discrepancies due to chain linking are inherent to the way they are defined and should not be corrected, as this would harm the essence of a chain linked volume.

Just like the discrepancies themselves, the seasonal pattern of these discrepancies is the difference between the pattern of the aggregate valued in previous year's prices and the pattern of the aggregate valued in prices of the reference year. This seasonal pattern can be removed. In theory, the discrepancies due to chain linking could also show calendar effects with seasonal adjustment. However, in the case of the Dutch economic series, they are negligibly small and we choose to ignore them.

The quarter-to-quarter changes of GDP are the single most important result from the economic analysis. Therefore it makes sense to also calculate the quarter-to-quarter changes of the statistical discrepancy and compare them to GDP as follows:

$$\%SD_t = (SD_t - SD_{t-1})/B1G_{t-1} * 100\%. \qquad (2)$$

When these quarter-to-quarter changes are of similar magnitude or larger than the changes of GDP itself, the analysis of the latter by breaking it down in components is severely hampered.

Especially when looking at quarter-to-quarter changes, removing the seasonal pattern can lead to a large reduction in the size of the discrepancies from chain linking. This is shown in Figure 2.2 and 2.3, where the quarter-to-quarter changes of the discrepancies in percentages of GDP have been adjusted for seasonal effects (i.e. the remaining series represents trend-cycle + irregular). The value of this series is mainly between –0.1% and 0.1%. To put this into perspective, the majority of GDP growth rates is between –0.5% and 0.5%.

As can be seen, the seasonal component is by far the largest component of the statistical discrepancy arising from chain linking. Therefore, with ideal seasonal and calendar adjustment, the adjusted statistical discrepancy should have a small influence on the interpretation of the economic growth and its components.

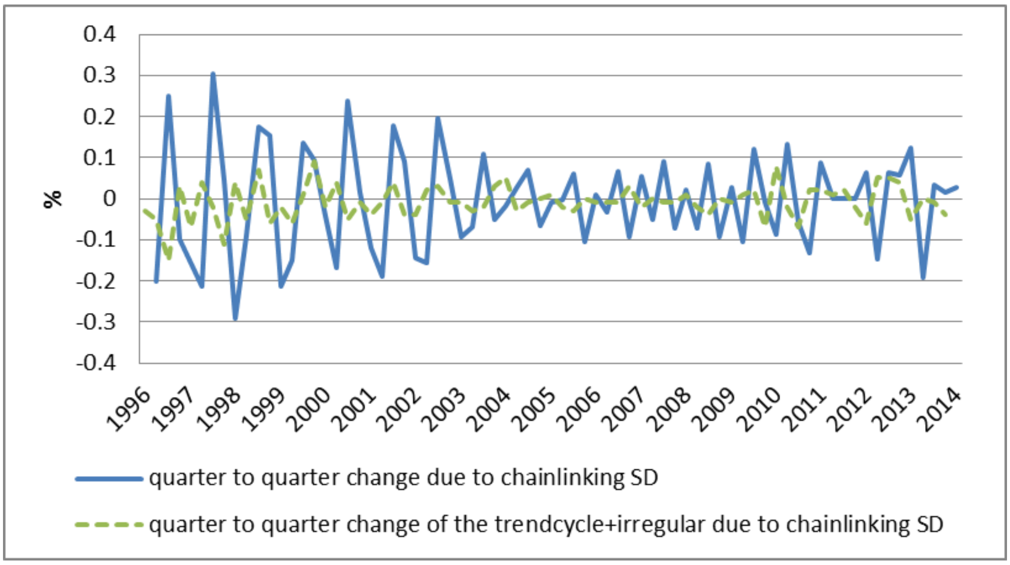

Figure 2.2. Univariate seasonal correction of the discrepancies arising from chain linking.

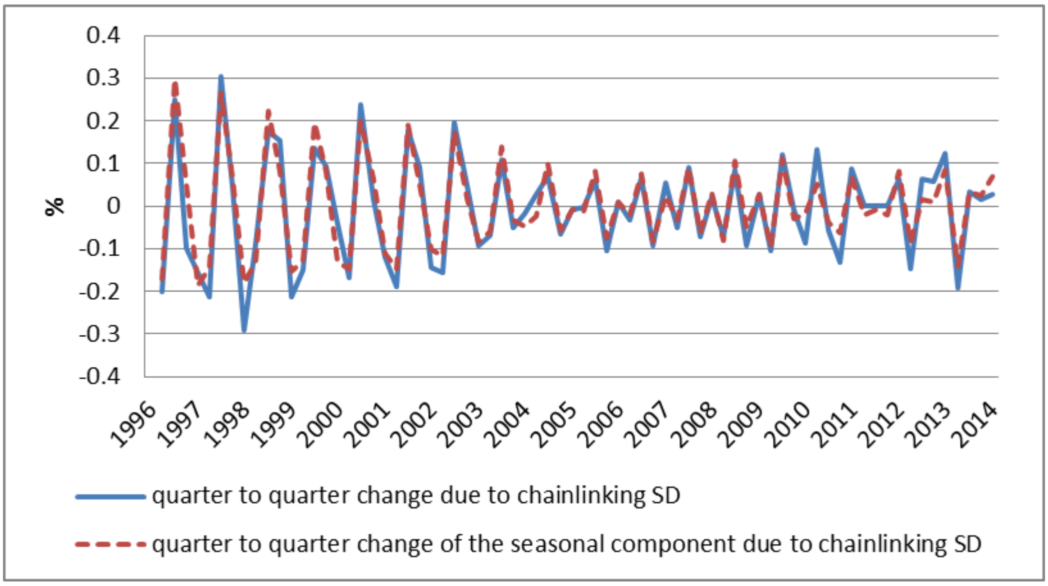

Figure 2.3. Seasonal component of the discrepancies arising from chain linking.

2.2 Discrepancies from adjustments for seasonal and calendar effects

The second part of the total statistical discrepancies is introduced by the estimation of seasonal and calendar effects. Seasonal and calendar adjustment assumes the following decomposition:

$$y_{tk} = L_{tk} + S_{tk} + \beta_k' x_{tk} + OL_{tk} + SB_{tk} + I_{tk}, \text{ t=1,...,T.} \quad (3)$$

Where y_{tk} stands for any of the k-th series appearing in equation (1), L the trend-cycle, S the seasonal, $\beta' x$ the regression component with x an auxiliary variable and β the regression coefficient, OL additive outliers and level shifts and SB seasonal breaks. Finally I is an irregular for the unexplained variation. The regression component is used in this application to adjust for calendar effects. In general, other regression effects can also be included, but this is not applied here. In a fully consistent adjustment, equation (1) holds for each of the components in equation (3). However, when these relations are not explicitly enforced, discrepancies will arise.

The process of seasonal and calendar adjustment consists of a pre-treatment phase and the actual seasonal adjustment. In the pre-treatment phase it is chosen between multiplicative or additive adjustment. In the first case, the original series are logarithmically transformed before the decomposition (3) is computed. The other parts of the pre-treatment phase are adjustments for calendar effects and other regression effects, removal of additive outliers, level shifts and seasonal breaks and extrapolation of the series in order to apply symmetric filters. The actual seasonal adjustment consists in the application of seasonal and trend filters. In this phase, again outlier detection takes place. After seasonal adjustment, additive outliers and level shifts are reintroduced into the series. The final adjusted series is therefore equal to:

$$y_{tk}^{SA} = y_{tk} - S_{tk} - \beta_k' x_{tk} - SB_{tk} = L_{tk} + OL_{tk} + I_{tk}. \qquad (4)$$

All steps of the process may cause discrepancies:

Logarithmic transformation: Usually this is done when this yields a better model fit, as for instance is current practice in X-13ARIMA-SEATS, see US Census Bureau (2015). For some of our series, this would indeed be the preferred option. However, when multiplicative adjustment is chosen in at least one series, the logarithmic transformation can cause additional discrepancies.

Outlier detection: when each time series is analysed separately for significant outliers, discrepancies may arise when an outlier is significant in one series but not significant or even detectable in another. These may lead to relatively large incidental discrepancies. The outlier detection in both the pre-treatment phase and the filtering phase can generate discrepancies. Here, the general term outlier is used for the combination of additive outliers, level shifts and seasonal breaks.

Calendar effects: Estimating the regression coefficients for the calendar effects for each series separately contributes to the discrepancies. The calendar effect in some series are not significantly different for zero (at a 5% significance level). Therefore it is not incorporated in the model of these series. This leads to relatively small discrepancies, evenly distributed along the length of the time series.

Extrapolations: The extrapolations are very sensitive to the model choice in X-13ARIMA-SEATS and to outliers at the beginning and end of the time series. This may lead to relatively large discrepancies at the beginning and end of the time series and it is a major source of revisions.

Seasonal and trend filters: when different time series are treated with filters of a different length, which can easily happen when optimal models are detected automatically and used for seasonal adjustment as a default, relatively large discrepancies will arise in the seasonal components along the full length of the series.

The discrepancies are larger and more volatile when multiplicative adjustment is used. Another source of discrepancies is the situation where the seasonal patterns change rapidly. In this case discrepancies may arise around the period where the rapid change occurs, because these periods are considered to be outliers.

The statistical discrepancy can be calculated by rewriting equation (1):

$$SD = B1G - (-P7 + P31_{S1A} + P31_{S13} + P51G + P5M + P6). \qquad (5)$$

The right hand side of this equation is called the indirect discrepancy, the left hand side the direct discrepancy. This equation holds not only for the series itself, but in an ideal world also for each of the components of equation (3). However due to the arguments mentioned above, this is not the case for the seasonal component, as is shown in Figure 2.4 for the period 1996 – 2014. The purple line is the indirect seasonal component from the chain linked index, calculated as the seasonal component of GDP minus the seasonal component of all other expenditures (the right hand side of equation (5)). The purple line is very different from the seasonal component in Figure 2.3, here repeated in red (equal to the left hand side of equation (5)). The result is an increase in quarter to quarter change of the discrepancy instead of an significant decrease. Therefore, the analysis of the economic growth and its components is severely hampered.

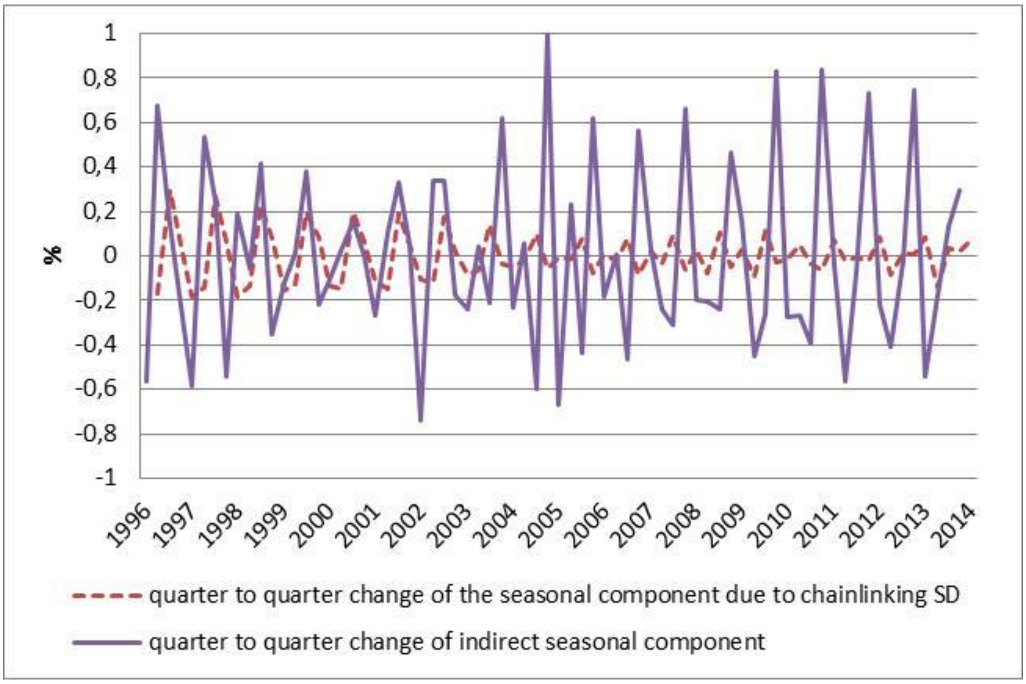

Figure 2.4: Seasonal component of the discrepancies arising from chain linking calculated directly (red) and indirectly (purple) with conventional univariate seasonal adjustment. The red line is the same as in Figure 2.3.

3. Alternative solutions

Before presenting the successful multivariate approach for pre-treatment in the next section, we first review several alternative solutions that we considered in order to eliminate or reduce discrepancies from seasonal and calendar adjustment. The first two alternatives are the indirect and benchmarking approach, which are both recommended by Eurostat (Eurostat, 2015) in order to avoid or remove discrepancies. The third approach is to optimize the adjustment of all series individually with the objective to minimize discrepancies. The fourth approach is to estimate a multivariate structural time series model that incorporates constraints that ensure the absence of discrepancies.

Each of the solutions has drawbacks in terms of the three objectives described in the introduction. However, they have given useful insights that we used in the multivariate approach for pre-treatment that is described in Section 4. In this section, we draw conclusions as to why each of the four alternative solutions does not give satisfactory results.

3.1 Indirect approach

In order to reduce statistical discrepancies, an obvious solution is to follow the indirect approach to seasonal and calendar adjustment. This approach first computes adjusted subseries for the most detailed breakdown in the publication domains, which are then aggregated in order to obtain an adjusted aggregate series. This in contrast to the direct approach, where the aggregate and subseries are adjusted independently. The advantage of the indirect approach is that it eliminates statistical discrepancies due to seasonal and calendar effects completely. This is, however, at the cost of reducing the univariate quality of the adjustment for the aggregate series. Firstly, the entire discrepancy due to chain linking and seasonal adjustment, is attributed to the aggregate series, while this choice bears no relation to the origin of the discrepancies. Secondly, the approach results in residual seasonality that can be detected in the aggregate. Since the growth rate of the GDP is one of the key economic indicators and the discrepancies are often larger than the growth rates of GDP itself, this is not a desirable approach. Apart from that, there are different approaches to GDP (such as expenditures approach and production approach), which would lead to different adjusted GDP series.

3.2 Benchmarking

A second approach to remove statistical discrepancies is by means of a multivariate benchmarking technique, like the extended multivariate Denton method described in Bikker et al (2013). Here, we first adjust all series individually. Next, the discrepancies are distributed among all series, so that they are effectively removed. The quality of this approach is only acceptable if the discrepancies are small enough. This is because this method relies on weights in order to distribute the discrepancies among the series. The usual choice of weights in a benchmarking procedure is a covariance matrix, but variances do not bear any relation to the sources of the discrepancies that we detailed at the end of Section 2.2. Furthermore, the procedure introduces residual seasonality in the adjusted series in the Dutch case, and therefore results in an unacceptably low univariate quality.

3.3 Improved univariate seasonal adjustment

The two approaches described above aim directly at removing the discrepancies, and therefore result in a lower statistical quality of seasonal and calendar adjustment for individual series and also in larger revisions. A third approach works the other way round. It aims at univariate adjustment of all series in such a way that the discrepancies are reduced, while maintaining acceptable quality for the aggregate series and all subseries, and thereby hopefully also avoiding large revisions.

Since X-13ARIMA-SEATS can only perform univariate seasonal adjustment, the magnitude of the discrepancies can only be established after all subseries are adjusted. The indirect relation between adjustment settings and the quality measure of interest makes it tedious to find a solution. We employed the following procedure.

For eight series of the expenditures approach (GDP and breakdown into 7 subseries) we did a detailed study in order to optimize individual adjustment settings. However, we concluded that a high quality (in terms of the M- and Q-diagnostics) adjustment for each series did not necessarily result in a smaller discrepancy. Based on this, we tried a large number of ideas in order to reduce the discrepancies while univariate quality is at least close to the optimal M- and Q-diagnostics:

- Outliers detected automatically in the subseries were also modelled in the aggregate (even if the outlier was not significant for the aggregate), in order to preserve additivity.
- All series were modelled using the same auxiliary variable (even if not significant). The value of the regression coefficient was estimated by X-13ARIMA-SEATS.
- All series were adjusted either additively or multiplicatively.
- The same ARIMA model was used for the extrapolation for all subseries as well as the aggregate.
- The same filter settings were used for all series.
- Seasonal and calendar adjustment was done separately for the period before and after 2008. This could be useful since the economic crisis had changed the seasonal pattern.
- The ARIMA model was estimated based on the period up to 2008, and kept fixed for the period thereafter.

We also examined partial aggregates in order to determine which series contributed most to the discrepancies. We found that especially the series with the largest outliers cause large discrepancies.

The extrapolations made by X-13ARIMA-SEATS for each series have a strong influence on the resulting adjustment. Since some of the series are very volatile the forecasts can be volatile as well. Since X-13ARIMA-SEATS uses symmetrical filters, the extrapolations are needed for adjustment at the end of the series, but can thus have consequences for the statistical discrepancies at the end of the series.

We found that:

- The largest part of the discrepancies is caused by the treatment of outliers.
- An additive adjustment approach for all series helps to reduce the discrepancies compared to a multiplicative adjustment. Although a multiplicative adjustment is preferred for the series of export, GDP and consumption of households, an additive adjustment is applied. Discrepancies are larger if a multiplicative adjustment is applied to all series instead of an additive adjustment.

Discrepancies are further increased if for some of the series an additive adjustment and for other series a multiplicative adjustment is chosen.
- If pre-treatment leads to small discrepancies, then most likely they are still small after seasonal adjustment itself.
- The best results could be obtained by an equivalent approach, i.e. using the same settings (ARIMA models, filters), modelling outliers in the same periods, and using the same set of auxiliary variables for all series. This results in a less optimal adjustment according to the Q and M-diagnostics of X-13ARIMA-SEATS.

Nevertheless, it is not possible to reduce the discrepancies sufficiently to apply a benchmark approach for eliminating them completely. On the contrary, they are still unacceptably large even under an optimally chosen equivalent approach. These findings have led to the conclusion that a multivariate approach is needed in order to reach all three objectives, i.e. reducing discrepancies while maintaining univariate quality and avoiding large revisions. In the next subsection, we present a multivariate structural time series approach.

3.4 Multivariate structural time series models

With a structural time series model (STM) a series is decomposed in a trend component, seasonal component, regression components and an irregular component. The model can be extended with other components as cyclic components, or with ARMA components in order to model autocorrelation beyond these structural components, but this is not applied in the present paper. For each component an appropriate stochastic model is assumed which allows the trend, seasonal, and regression coefficients to be time dependent. See Harvey (1989) and Durbin and Koopman (2012) for an extensive treatment of structural time series modelling. In multivariate STM's, two or more series are modelled simultaneously, which allows to model cross-sectional dependency between these series.

So for each of the series (both GDP and its breakdown in seven subseries) a STM is assumed, which decomposes the series into a trend component, a seasonal component, a regression component, and an irregular component. The first part of the regression component models the calendar effects, the second part models the outliers. For the trend, the seasonal and the regression component of the calendar effects, a stochastic model is assumed. The outliers in the model are additive outliers, level shifts and seasonal breaks which are assumed to be constant over time.

In order to avoid that discrepancies are increased with the seasonal and calendar adjustment of each of the series, several constraints are imposed on the time series components. These constraints ensure that for each of these components at each point in time, the value for the aggregate series is exactly equal to the sum over the values of the underlying subseries. The full model consists of six constraints that must hold for each time period of the observed series, for the following time series components:

4. The trend components
5. The regression coefficients for the calendar effects

6. The seasonal components
7. Additive outliers
8. Level shifts
9. Seasonal breaks

The model is described in detail in Appendix 2. Equations (A.2) – (A.8) describe how the components are modelled. A mathematical description of the constraints is given in (A.11) – (A.16).

Here we focus on several preliminary implementations of that model that we examined before formulating our final implementation, which is treated in Section 4.

We developed a multivariate STM for quarterly GDP, broken down into a hierarchy according to either the expenditures approach or production approach. Either hierarchy contains multiple levels. At every level there is a statistical discrepancy (due to chain linking) before pre-treatment and seasonal and calendar adjustment (but only if measured in constant prices). The first breakdown of GDP into 7 subseries (including the statistical discrepancy) was described in Section 2, and here we focus only on this first breakdown.

From early testing, it appeared that computation times grow exponentially with the number of series and constraints. Therefore we tested several alternative formulations, with only a subset of the constraints or a different estimation approach. These alternatives are described below. All models are variations of the model described in Appendix 2. The model in the Appendix is the approach described under Alternative 3 but using a slightly different method to impose the required constraints. Another differences between all three alternatives and the model in the Appendix is that seasonal breaks are not included in models of the following three alternatives.

Alternative 1

A multivariate STM using 6 expenditure series, the statistical discrepancy from chain linking, and total GDP, where constraints are imposed on the trend components and the seasonal components for all series that make up GDP. No correlations between disturbances of the trend and seasonal components are included in the model.

Alternative 2

A multivariate STM using 6 expenditure series, the statistical discrepancy from chain linking, but no GDP series. The seasonal pattern for GDP is derived from the sum over all 7 underlying seasonal patterns. At first, models were tried where no correlations between disturbances of the trend and seasonal components were allowed. In an extension to the model, the effect of adding these correlations was studied. This approach resembles the indirect approach. An important extension is modelling the correlation between the times series components. This improves the accuracy of the model estimates and partially compensates for the loss of accuracy that is caused by replacing a seasonal adjusted series derived from the aggregated series by the sum over the underlying seasonally adjusted series.

Alternative 3
In the first step a univariate model was fitted for total GDP. Based on this, a smoothed seasonal pattern for total GDP was obtained. In a second step, a multivariate model was fitted for the decomposition into seven series (but without GDP). In this model, a restriction is added that the sum of all seasonal patterns is equal to the smoothed seasonal pattern from step 1. This method was proposed by Doran (1992). We also investigated modelling correlations between disturbances of trend and seasonal components.

For all alternatives studied we found that the estimated components, especially the seasonal component, were volatile, and subject to large revisions when data points were added to the time series. From a practical perspective, this is not desirable.

After testing these three models, it was decided to simplify the approach and find a method that can be implemented quickly and solve the urgent problem of large discrepancies. Therefore the focus is on the components that contributed most to the discrepancies. This new approach consists of pre-treatment based on a multivariate STM that ensures that the components that contributed most to the discrepancies are consistent. The seasonal adjustment itself is carried out with X-13ARIMA-SEATS. This approach is described in the next section.

Nevertheless, it is possible that the approach which is fully based on a multivariate STM including GDP and the underlying breakdowns can give satisfactory results under the right model formulation. Especially formulations with seasonal breaks could be interesting. So far, different versions of this model are hampered due to numerical problems. Since a practical solution was needed in the production process, it was decided to implement the aforementioned pre-treatment of the series with a multivariate STM. A seasonal adjustment approach based on a multivariate STM is left as a topic for further research.

4. Multivariate pre-treatment

4.1 Multivariate approach

Based on the results of the previous section we developed the following procedure that combines the advantages of a multivariate approach and robustness of conventional univariate seasonal adjustment with X-13ARIMA-SEATS:

- *Step 1:* estimate the multivariate STM from Appendix 2, without the constraint on the additivity of seasonal patterns and trends, and without the correlations between the disturbances of the trend and the seasonal component. Remove the model estimates of all regression effects (additive outliers, level shifts, seasonal breaks, calendar effects) from the series.

– *Step 2:* use X-13ARIMA-SEATS for the extrapolation of the series and to obtain seasonally adjusted series by applying trend and seasonal filters. After that, reintroduce the desired regression effects back into the series. In this case the additive outliers and level shifts are reintroduced.
– *Step 3:* apply multivariate benchmarking in order to eliminate any remaining discrepancies.

The first step uses the model from Appendix 2, but with some modifications so that the resulting model is a special case of the full model, with only restrictions 2, 4, 5 and 6 as presented in the previous section. Effectively, this means that the multivariate model from Appendix 2 is only used for pre-treatment, comparable to this step in X-13ARIMA-SEATS (except for the extrapolation, which is done by X-13ARIMA-SEATS in our approach). However, after estimation all regression effects will be fully consistent between the subseries and the aggregate. This means that not only calendar effects will add up over all series, but also additive outliers, level shifts and seasonal breaks are consistent. To this end, we used the same regressors for all series, and outliers were modelled consistently, as described in the next subsection. This is based on the results of Section 3.3, where we learned that the largest part of the discrepancies is caused by the treatment of outliers. Also, we found that if pre-treatment yields small discrepancies then most likely the final seasonal adjustment will not increase these discrepancies much. By ensuring additivity for these components, we obtained a close to optimal result, which is more stable than the approach of Section 3.4, and has acceptable computation times.

After pre-treatment by the multivariate model, final seasonal adjustment is done by X-13ARIMA-SEATS using JDemetra+ (Grudkowska, 2015). Since the earlier study indicated that the best results could be obtained by a consistent approach, we use the same filter lengths for all series. We chose a short seasonal filter since the seasonal pattern changes quite rapidly. This had only a slight effect on the quality of univariate seasonal adjustment. The ARIMA model used for extrapolation (in order to be able to apply symmetric filters) was determined for the aggregate and applied for each series in the breakdown. This procedure results in seasonally adjusted series that have only small discrepancies. In order to remove these, a multivariate benchmarking procedure was applied (see Section 3.2).

4.2 Consistent outlier detection

Accurate outlier detection is a crucial step in achieving good quality seasonal adjustment. As outliers can be much larger than the seasonal and calendar effects, their influence can be very large. If the outlier detection is incomplete, the estimates of all other components will be biased. In some cases, more than one outlier is needed to model the economic events in a short period of time and it may be difficult to find the optimal combination of outliers and avoid overfitting of the series. We have observed that one of the consequences of not using the right combination of outliers is a serious deterioration of quality diagnostics of X-13ARIMA-SEATS. Another consequence is that the estimates of all state variables can become unstable.

In this context we use three types of outliers: additive outliers, level shifts and seasonal breaks. For every significant outlier in one of the series, there must a counterpart in one or more of the other series to achieve consistency. These counterparts are not necessarily significant, in fact they may not even be detectable. We therefore need additional information to make the decision how to model the outliers consistently. In our case the additional information is provided by the fact that our macro-economic series are based on detailed quarterly supply-use tables.

When we were able to find an economic or statistical explanation for an detected outlier, the type of the outliers, the timing, the counterparts and, in some cases, the size of the outliers are known. When the size of an outlier can be determined from statistical/economic analysis, the outlier can be manually removed from the series and does not need to be modelled in the multivariate STM. This results in a more parsimonious model, which is therefore preferred. When the size of the outliers cannot be determined, we model the outliers in the STM. In the cases where the type, the timing or the series in which the outlier occurs are not known from the statistical/economic analysis, all possible combinations (given the restrictions based on the economic or statistical explanations) of the outliers are tested and the solution with the optimal AIC is selected.

There are several methods to find outliers. The largest outliers are easily detectable using the automatic detection in the pre-treatment phase of X-13ARIMA-SEATS. In this method outliers are detected with a RegARIMA model, which models the series with a linear regression model and applies an ARIMA model to the error terms to capture remaining serial correlation. These outliers are removed in a multivariate STM. During this calculation, additional outliers can be detected by considering the residuals of the STM using a disturbance smoother (Harvey and Koopman, 1992). For all residuals with a t-value larger than 2.5, an outlier and its counterparts are added to the set.

The set of outliers that is modelled explicitly in the pre-treatment phase is removed from the series and the pre-treated series then become the input for the seasonal adjustment phase. Nevertheless, the seasonal adjustment phase in X-13ARIMA-SEATS can detect additional outliers. This phase applies filters to extract trend and seasonal components from the series. In case the actual series is significantly different from the modelled series (i.e., outside the confidence interval around the modelled series), an observation is considered an outlier. The size of the interval is controlled by the parameter *sigmalim*, which consists of two values, representing an upper and lower limit. These values are multiplied by the standard deviation of the irregular (estimated by the filters). If the irregular component for a single observation is below/above the lower/higher boundary times the standard deviation, it gets the full/zero weight. If it is in between, it gets a proportion of the weight according to a linear scale. X-13ARIMA-SEATS thus gives these outliers a lower weight in filtering the trend and seasonal effect. As these decisions are made for each series separately, they will again lead to statistical discrepancies. Therefore we must model these events with additional outliers in the pre-treatment phase. In order to reduce the number of outliers detected in the seasonal adjustment phase, we increased the lower and upper limits of *sigmalim* from 1.5-2.5 to 2.5-3.5 (resulting in a wider boundary around the modelled series). All outliers that were detected above this level, were added to the set of outliers modelled in the pre-treatment phase. This process is iterated until no new outliers are detected.

The final seasonally adjusted series are the seasonally adjusted series plus the level shifts and additive outliers removed in the pre-treatment. Seasonal breaks are not added back to the series, because the purpose of seasonal adjustment is to remove seasonal patterns.

5. Results

In this section, we apply the seasonal and calendar adjustment approach obtained with standard X-13ARIMA-SEATS (old method) and the proposed (improved) method to the cycle of releases of the quarterly GDP and its components (as described in Section 2) in an annual estimation cycle and compare the results obtained under both approaches.

The quarterly GDP figures are produced twice: 45 days after the end of a quarter a ‘flash’ estimate is published. Then, 85 days after the end of the quarter a new so-called ‘regular’ estimate is published based on more complete data sources. When the regular estimate of the fourth quarter is published in March, the figures for the first three quarters are revised again.

The quarterly figures are revised three more times after that: for each new annual estimate the quarterly figures are adapted such that the four quarters add up to the new annual figure. This happens for the first time in June and for the second time one year later when the final annual figures are published. Finally, one year after this, the quarterly figures are revised one more time without changing the annual results. Every time a new quarter is added, the seasonal adjustment procedure is applied to the entire time series, potentially affecting all quarters. Normally, though, revisions of seasonal adjustments to earlier figures are small.

This means that once a year, at the time the regular estimate for the first quarter (1r) is made, large changes are made to the (unadjusted) time series. Therefore, it is necessary to derive new settings for the seasonal and calendar adjustment at this point in time every year. The annual estimation cycle starts with the regular estimate for the first quarter, and the derived settings are then used for all subsequent estimates of the annual estimation cycle. The 1r estimate is followed by the first (flash) estimate of the second quarter (2f) and the second estimate of the second quarter (2r). This scheme is continued till the second estimate of the fourth quarter (4r). The first estimate of the first quarter of the current year comes before the large updates of 1r and is therefore also part of the same cycle. It is called 5f, to emphasize that the settings of previous year are applied.

Below, the old and improved method are compared according to the three quality criteria described in Section 1:
- The statistical discrepancy between the seasonally adjusted GDP and its components.
- The standard quality diagnostics of X-13ARIMA-SEATS: M1 till M11 and Q.
- The revisions of the published results between the subsequent estimations.

5.1 The statistical discrepancy due to seasonal and calendar adjustment

In this subsection we discuss the discrepancies added by the seasonal and calendar adjustment process. This process estimates the seasonal components of each of the series. In both the old and improved approach, the estimated seasonal components of all subseries do not add up to the seasonal component of GDP, and result in a residual:

$$\Delta = S_{B1G} + S_{P7} - S_{P6} - S_{P31_{S1A}} - S_{P31_{S13}} - S_{P51G} - S_{P5M}$$

Before seasonal adjustment, we already had a statistical discrepancy SD, which also contains a seasonal component, S_{SD}. Therefore, Δ is the statistical difference due to seasonal adjustment and including S_{SD}. The difference between these two is the added statistical discrepancy due to seasonal adjustment.

In Table 5.1, we compare the added statistical discrepancy of the old and the improved method, by taking the relative added statistical discrepancy, computed as a percentage change from seasonally adjusted GDP:

$$\left|\frac{(S_{SD}-\Delta)}{B1G^{SA}}\right| * 100\%$$

Table 5.1 presents the average and maximum of this difference over the entire time series.

Table 5.1. Average and maximum absolute discrepancies due to seasonal adjustment.

	old method		improved method	
	Avg %	Max %	Avg %	Max %
1r	0.331	1.01	0.001	0.01
2f	0.329	1.00	0.001	0.01
2r	0.321	1.00	0.001	0.01
3f	0.321	0.99	0.001	0.01
3r	0.320	0.99	0.001	0.01
4f	0.322	0.98	0.001	0.01
4r	0.330	0.99	0.001	0.01
5f	0.313	0.99	0.000	0.01

Table 5.1 shows that a significant reduction in statistical discrepancy due to seasonal and calendar adjustment can be achieved by using the improved method. With the old method, interpretation of the GDP growth was on average hampered by the discrepancy by 0.3 %, with a maximum of 1 %, while with the new method the disturbance is negligible.

5.2 The standard diagnostics of X-13ARIMA-SEATS

Software of the X-11-family summarizes the quality of the seasonal and calendar adjustment with M1 till M11 and a Q-diagnostics. These diagnostics value different aspects of the seasonally adjusted series. For the meaning of the values, see Appendix 3. These statistics vary between 0 and 3 but only values smaller than 1 are acceptable. The lower the value, the better. Tables 5.2 and 5.3 present the diagnostics of the seasonal adjustment with the old method and the improved method of estimate 1r.

Table 5.2. Quality of seasonally adjusted estimate 1r with old method (missing values are 0, values >1 are bold).

	Import (P7)	Consumption HH ($P31_{S1A}$)	Consumption Govern ($P31_{S13}$)	GCF (P51G)	Stocks (P5M)	Export (P6)	GDP (B1G)
M1	0.49	0.42		0.09	0.33	0.10	0.05
M2	0.64			0.06	0.19	0.05	0.02
M3	*0.13*			*0.31*	0.39		
M4	0.18	**1.16**	0.95	1.05	0.84	0.84	0.84
M5	0.24	0.20	0.20	0.20	0.41	0.20	0.20
M6	*0.12*	*0.44*	**1.00**	*0.16*	0.64	0.52	*0.22*
M7	0.38	0.39	0.16	0.13	0.32	0.24	0.06
M8	0.91	0.71	0.51	0.53	**1.12**	0.60	0.17
M9	0.54	0.64	0.26	0.14	0.67	0.34	0.06
M10	**1.33**	**1.04**	0.38	0.54	**2.30**	0.98	0.17
M11	**1.30**	**1.02**	0.22	0.44	**2.30**	0.97	0.17
Q	0.46	0.45	0.22	0.29	0.64	0.34	0.16

Table 5.3. Quality of seasonally adjusted estimate 1r with improved method (missing values are 0, values >1 are bold).

	Import (P7)	Consumption HH ($P31_{S1A}$)	Consumption Govern ($P31_{S13}$)	GCF (P51G)	Stocks (P5M)	Export (P6)	GDP (B1G)
M1	0.33	0.61	0.02	0.27	0.73	0.18	0.04
M2	0.17	0.02	0.03	0.18	0.55	0.08	0.02
M3	*0.30*	*0.17*	*0.27*	*0.86*	**1.10**	*0.03*	
M4	0.95	0.51	0.51	0.40	0.18	**1.05**	0.40
M5	0.20	0.20	0.20	0.60	0.95	0.20	0.20
M6	*0.08*	**1.32**	*0.74*	**1.41**	*0.81*	*0.31*	*0.16*
M7	0.30	0.19	0.24	0.14	0.19	0.18	0.10

M8	0.82	0.72	0.51	0.68	**1.02**	0.60	0.33
M9	0.45	0.13	0.26	0.19	0.21	0.25	0.22
M10	0.64	0.52	0.36	0.78	**1.23**	0.38	0.22
M11	0.32	0.40	0.24	0.17	0.48	0.16	0.12
Q	0.40	0.30	0.24	0.39	0.61	0.28	0.14

In Table 5.2, eight diagnostics are between 1 and 2 and two of them are above 2. This shows that the quality of the seasonal and calendar adjustment is not always satisfactory, but further improvements are not possible by traditional methods. For the new method, the results have improved to six diagnostics above 1 and none above two. On the other hand, the new method has less quality diagnostics with very small values. On average, the quality improves slightly. This is due to the improved analysis of the outliers, the multivariate aspect of the new method causes the reduction in the number of quality diagnostics with very high and very low values.

The quality for GDP is almost the same as before despite the fact that in the multivariate approach the excellent univariate seasonal and calendar adjustment is slightly disturbed by the other series. The quality for the GCF (gross and capital formation, P51G) deteriorates because the M5 worsens due to of the larger level shift in 2009-Q1 (resulting in less trend) in the multivariate case compared to the univariate case. The M1 and M2 deteriorate because this series contains a larger irregular component because the values for *sigmalim* have been increased. This also affects the M6 and M8 resulting in larger arbitrary changes of the seasonal component. The results for M9 till M11 are strongly improved. This is caused by the modelling of the seasonal outliers. In both methods the sum over the four quarters of a seasonal outlier adds up to zero. However, in the improved method, a seasonal outlier has in every quarter a different magnitude, while in the old method the outlier is determined in one quarter and the three other quarters have a third of its opposite magnitude. A disadvantage of the improved method is that it uses three quarters in the time series to determine the outlier, while the old method only uses one degree of freedom.

Remarkably large differences in diagnostics are found for M4 for the import (P7) and the export (P6), both deteriorate. Further analysis showed that M4 could be improved by adding an extra seasonal break in 2003 which becomes more pronounced due to the seasonal break of 2008, however this was unknown during the implementation of the new method for the seasonal adjustment of the Dutch quarterly national accounts.

Table 5.4 presents the difference in overall quality (as measured by the Q-diagnostic) between the two methods for all 8 estimates. Negative values (in bold) relate to an improvement by using the new method, positive to a deterioration. Both methods display an almost constant difference in quality during the annual cycle.

Table 5.4. Difference in Q-diagnostic between old and new method for seasonal adjustment (missing values are 0).

	Import (P7)	Consumption HH ($P31_{S1A}$)	Consumption Govern (P31S13)	Stocks (P5M)	GCF (P51G)	Export (P6)	GDP (B1G)
1r	**-0.06**	**-0.15**	0.02	0.10	**-0.02**	**-0.07**	**-0.01**
2f	**-0.08**	**-0.10**	0.03	0.10	**-0.07**	**-0.07**	**-0.02**
2r	**-0.07**	**-0.10**	0.03	0.10	**-0.07**	**-0.06**	**-0.01**
3f	**-0.08**	**-0.11**	0.02	0.13	**-0.06**	**-0.07**	**-0.01**
3r	**-0.08**	**-0.11**	0.02	0.13	**-0.06**	**-0.07**	**-0.01**
4f	**-0.07**	**-0.12**		0.16	**-0.04**	**-0.08**	**-0.01**
4r	**-0.06**	**-0.16**		0.09	**-0.03**	**-0.08**	**-0.02**
5f	**-0.05**	**-0.14**	0.01	0.12	**-0.03**	**-0.06**	**-0.02**

5.3 Revisions

In this section revisions of the quarter to quarter growth in %-point are investigated under the old and the improved method.

The quarter to quarter growth is defined as

$$\hat{\theta}_t = \frac{y_t^{SA} - y_{t-1}^{SA}}{y_{t-1}^{SA}} \cdot 100\%, \quad (6)$$

where y_t^{SA} denotes the seasonally adjusted figures. This is computed for the GDP and the variables of the breakdown.

The revisions are split into two types; the first are due to the updates from flash to regular estimate:

$$R_1 = \left(\sum_{t=2}^{4}\left|\hat{\theta}_{t|t}^{f} - \hat{\theta}_{t|t}^{r}\right|\right) \quad (7)$$

with $\hat{\theta}_{t|T}^{f}$ and $\hat{\theta}_{t|T}^{r}$ the flash and regular estimates of the growth rates, see formula (6), for quarter t based on the time series up to and including quarter T.

Note that the first quarter ($t = 1$) is excluded in R_1 since in the regular estimate of this quarter, the information on an annual basis is added which causes large revisions. The second type of revisions is due to adding the flash estimate of new quarter:

$$R_2 = (\sum_{t=1}^{4}\left|\hat{\theta}_{t|t}^{r} - \hat{\theta}_{t|t+1}^{f}\right|)$$

with again $\hat{\theta}_{t|T}^{f}$ and $\hat{\theta}_{t|T}^{r}$ the flash and regular estimates of the growth rates, similarly as in formula (7).

The average revision of the last quarter is therefore:

$\frac{1}{7}(R_1 + R_2)$

which is presented in Figure 5.1.

Similarly, Figure 5.2 presents the average absolute revisions of the quarter to quarter growth in %-point over 8 estimates (i.e. 7 differences) over the last year averaged per quarter:

$$\frac{1}{28}\sum_{j=0}^{3}(R_{1j} + R_{2j})$$

with $R_{1j} = \left(\sum_{t=2}^{4}\left|\hat{\theta}^{f}_{t-j|t} - \hat{\theta}^{r}_{t-j|t}\right|\right)$, j = 0, 1,2,3

and $R_{2j} = \left(\sum_{t=1}^{4}\left|\hat{\theta}^{r}_{t-j|t} - \hat{\theta}^{f}_{t-j|t+1}\right|\right)$, j = 0, 1,2,3

Series SD (statistical discrepancy due to chain linking) and P5M (changes in stocks) are both fluctuating around zero. Therefore, both can have huge growths in %-points because of small absolute values, resulting in huge revisions of the growth. As a consequence of that they are left out of the analysis. The figures show that the size of the revisions of the old method and the improved method are almost equal. A reduction of the revisions was not expected in advance, as adding or changing observations at the end of the series gives new information about trend-cycle and seasonal component. Revisions are therefore inherent to seasonal and calendar adjustment.

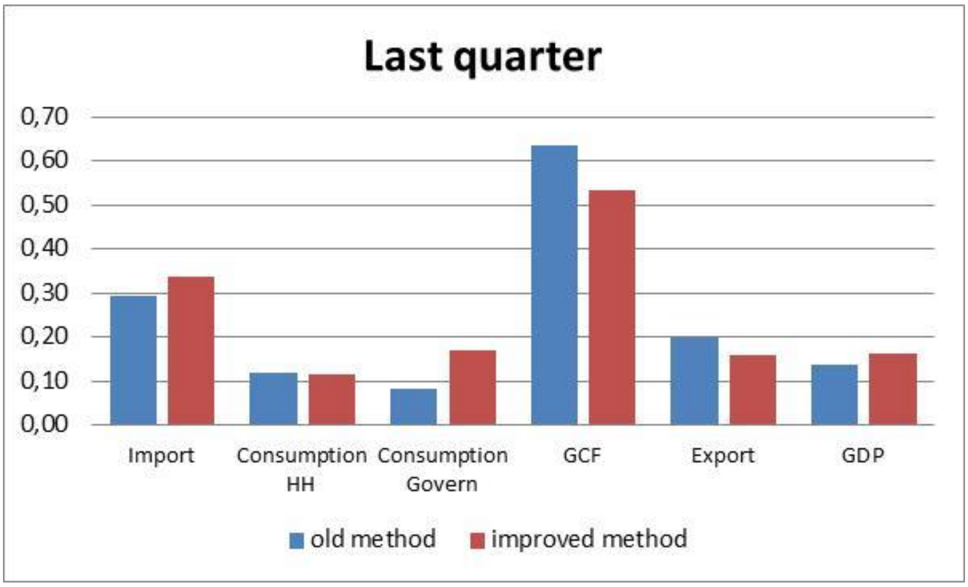

Figure 5.1: Revisions of last quarter.

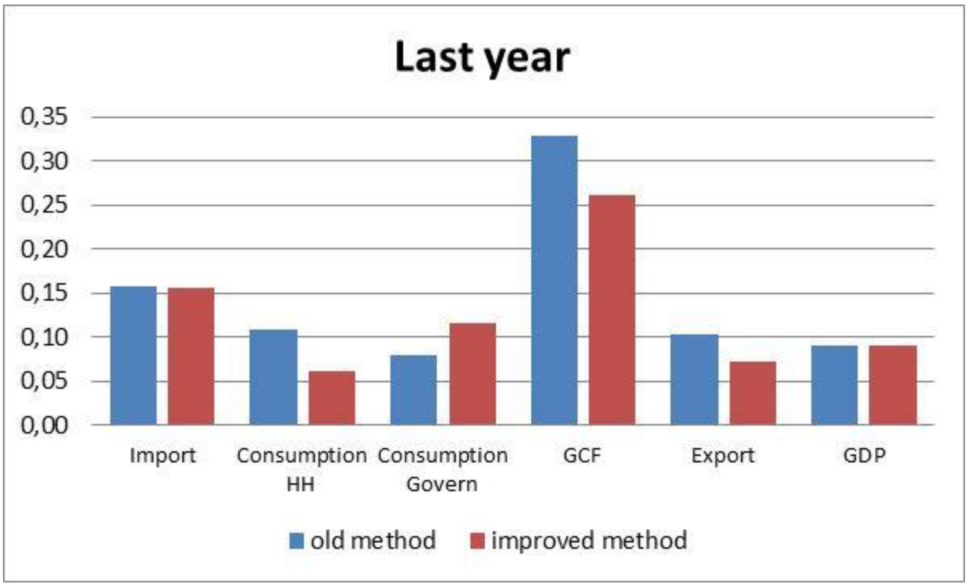

Figure 5.2: Revisions of last year.

6. Conclusion

Quarterly figures about GDP with a breakdown in K underlying subseries for e.g. expenditures or industries are produced by national statistical institutes to measure economic growth. Two factors are responsible for discrepancies between the sum over the underlying K subseries and the total GDP. The first factor arises due to the process of chain linking, which means that series of volume growth rates are expressed in constant price levels. Since the annual changes of these price levels differ between the series, statistical discrepancies between the sum over the underlying series and total GDP arise. The first factor does not arise if the estimate is in current values. The second factor arises after adjusting for seasonal and calendar effects using the standard approach based on X-13ARIMA-SEATS. Since 2009 the size of these discrepancies is often larger than the growth rates of GDP itself and hampers the interpretation of these figures.

Several intuitive approaches to avoid discrepancies are available in the literature, like the indirect approach and multivariate benchmarking. A major drawback of the first approach is that official figures about GDP are derived from the most detailed breakdown which contains the largest fluctuations while the most reliable estimates at the aggregated level are not used. Benchmarking is appropriate if the discrepancies are modest. In the Dutch application, the discrepancies are large, and benchmarking introduces a residual seasonal effect in the adjusted series.

In this paper an alternative approach based on a multivariate structural time series model is considered. The most intuitive approach is to construct a $K + 1$ dimensional structural time series model for GDP and its breakdown in K subseries. The model contains explicit constraints on the state variables to force that trend, seasonal effects, calendar effects and outliers in the GDP are equal to the sum over the *K* subseries of these components. In this way available series are consistently modelled and a two-stage approach, which was expected to be sub-optimal, is avoided. Nevertheless, the results obtained with this approach are not satisfactory since the estimated seasonal effects are too volatile. Furthermore, we observed numerical problems with the maximum likelihood procedure for the hyperparameters and there were too large revisions.

As an alternative, a *K* dimensional multivariate structural time series model with consistency restrictions on the additive outliers, level breaks, seasonal breaks and calendar effects (derived from a univariate model applied to the aggregated series) is only used to eliminate these effects from the observed series. Subsequently X-13ARIMA-SEATS is used for seasonal adjustment of all series. This reduces the inconsistencies remarkably. Finally a multivariate benchmarking is applied to restore consistency in the adjusted series and the additive outliers, level breaks, and calendar effects are added to the adjusted series. Although these pre-treatment approach appears to be suboptimal, a significant reduction of the statistical discrepancies is achieved whereas the quality of the adjustment in terms of the standard X-13ARIMA-SEATS quality measures is maintained or even improved for some series. In June 2015 this approach was implemented for production of the Dutch official statistics about economic growth.

The approach considered in this paper is generic and applies to many other applications at national statistical institutes. It is, therefore, worthwhile to further improve the *K + 1* dimensional structural time series model where consistent seasonal effects are directly estimated with the structural time series model.

7. References

Bikker, R., Daalmans J. and Mushkudiani, N. (2013), Benchmarking large accounting frameworks: a generalized multivariate model, Economic Systems Research, 25:4, 390-408, DOI: 10.1080/09535314.2013.801010. Available at: http://dx.doi.org/10.1080/09535314.2013.801010

Bloem, A., Dippelsman, R. and Maehle, N. (2001), Quarterly National Accounts Manual–Concepts, Data sources and Compilation, IMF, Washington D.C.

Doornik, J.A. (2009). An Object-oriented Matrix Programming Language Ox 6. London: Timberlake Consultants Press.

Doran, H.E. (1992). Constraining Kalman filter and smoothing estimates to satisfy time varying restrictions, Review in Economics and Statistics, 74, pp. 568-572.

Durbin, J., and Koopman, S.J. (2012). Time series analysis by state space methods, second edition. Oxford: Oxford University Press.

Eurostat (2015), ESS Guidelines on seasonal adjustment, 2015 edition, European Union, Luxembourg

Grudkowska, S. (2015). JDemetra+ User Guide. National Bank of Poland, Department of Statistics, Warsaw.

Harvey, A.C. (1989). Forecasting, structural time series models and the Kalman filter. Cambridge: Cambridge University Press.

Harvey, J.A. and Koopman, S.J. (1992). Diagnostic checking of unobserved components time series models, Journal of Business and Economic Statistics, 10, 377-389.

Koopman, S.J., N. Shephard, and J.A. Doornik (1999). Statistical algorithms for models in state space form using Ssfpack 2.2. Econometrics Journal, 2, pp. 113-166.

Koopman, S.J., N. Shephard, and J.A. Doornik (2008). SsfPack 3.0: Statistical algorithms for models in state space form. London: Timberlake Consultants Press.

Ladiray, D., Quenneville, B. (2001), Seasonal Adjustment with the X-11 method. Springer-Verlag, New York.

US Census Bureau (2015), X-13ARIMA-SEATS Reference Manual, Version 1.1 U.S. Census Bureau, Washington, DC.

Appendix 1: Detailed breakdown of GDP

The breakdown of GDP according to the expenditure approach that CBS publishes is actually more detailed than the breakdown we use in this article. The breakdown which we would like to consistently correct for seasonal and calendar effect consists of a tree structure with nine branches and is represented in Figure A.1 In each branch there is a time series marked SDx. These are the discrepancies arising from chain linking. They only occur in constant price data. We see that the breakdown comprises GDP itself and 20 subseries, complemented by nine different series for the discrepancies arising from chain linking (one for each branch of the tree).
A similar breakdown tree of GDP is used for value added by industry (the production approach). This tree consists of six branches, and comprises 20 subseries for branches of industry and of course six series for discrepancies arising from chain linking.

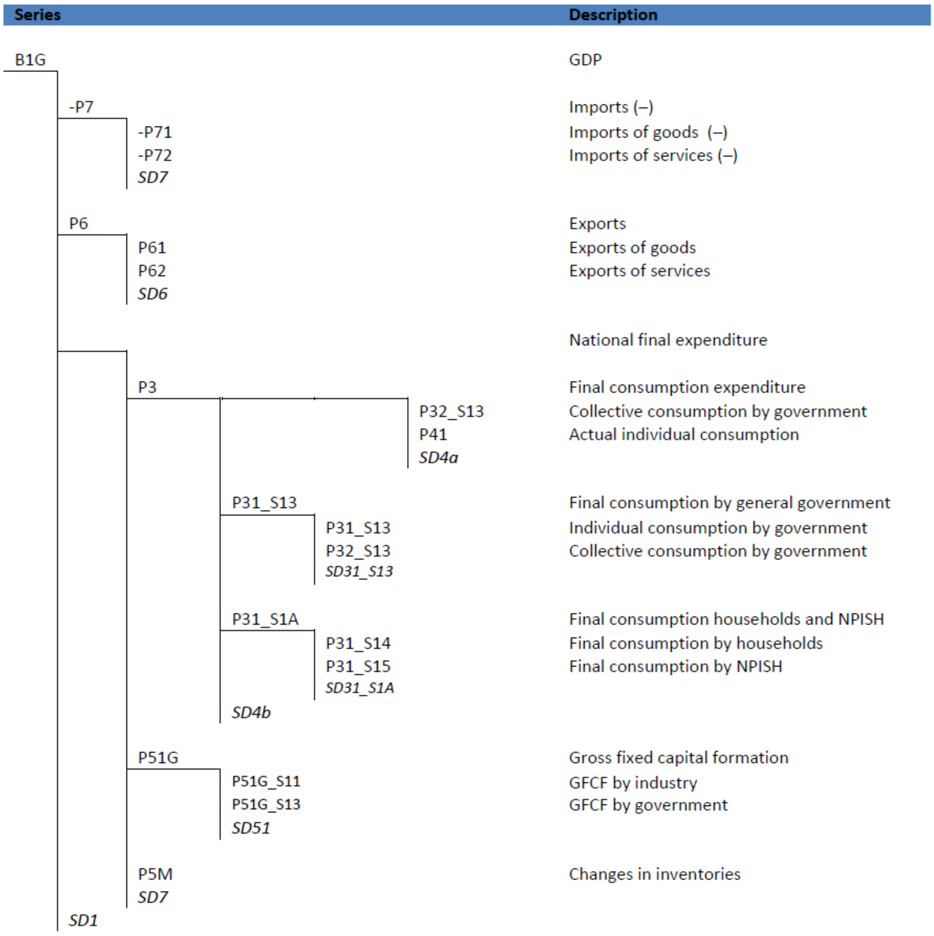

Figure A.1: Breakdown of GDP in components.

Appendix 2: Structural time series model

In this appendix, we describe a multivariate structural time series model (STM) for quarterly GDP, broken down into a hierarchy according to either the expenditures approach or production approach. The description of the model is for the general case. In this paper it is applied for the specific case of the expenditure approach with seven subseries.

Either hierarchy contains multiple levels. At every level there is a statistical discrepancy before seasonal and calendar adjustment (but only if measured in constant prices). In this paper, the number of working days is used as calendar effects. It is important that the consistency between all hierarchical levels of GDP is maintained. This is done by imposing restrictions that ensure that for every subsequent level, all time series components are benchmarked to estimates of the aggregate at the next level.

Let y_{t+} be the GDP as measured on a quarterly basis. In the first step, the following univariate STM is estimated:

$$y_{t+} = L_t + S_t + \alpha\delta_t + \beta_t\ x_t + \lambda\Delta_t^L + \gamma_t\ \Delta_t^S + I_t \tag{A.1}$$

In model (A.1), L_t is a time-dependent trend and S_t a time-dependent seasonal component. Furthermore δ_t is a dummy variable, indicating the period in which an additive outlier occurs, α the corresponding time-invariant regression coefficient measuring the magnitude of the outlier, x_t a variable containing auxiliary information about calendar effects in period t, β_t the corresponding time-dependent regression coefficient, Δ_t^L a dummy variable indicating the period in which a level shift occurs, λ a time invariant regression coefficient measuring the size of the level shift, Δ_t^S a dummy variable indicating the period in which a seasonal break occurs, γ_t a time invariant seasonal pattern measuring the size of the seasonal break, and I_t a disturbance term for any unexplained variations.

In the general case, multiple additive outliers, level shifts and seasonal breaks are possible, and multiple auxiliary variables may be useful. Then equation (A.1) can be adapted in a straightforward way.

The trend is modelled according to the so called smooth trend model (Durbin and Koopman, 2012) defined by:

$$\begin{aligned} L_t &= L_{t-1} + R_{t-1} \\ R_t &= R_{t-1} + \varepsilon_{tR} \end{aligned} \tag{A.2}$$

with ε_{tR} normally distributed disturbances with expectation 0 and variance σ_R^2. The seasonal pattern is modelled using a so-called trigonometric model, given by:

$$S_t = \sum_{h=1}^{2} S_{th}\ , \tag{A.3}$$

with

$$\begin{aligned} S_{th} &= \cos\left(\frac{\pi h}{2}\right) S_{t-1,h} + \sin\left(\frac{\pi h}{2}\right) S_{t-1,h}^* + \varepsilon_{th}\ , \\ S_{th}^* &= -\sin\left(\frac{\pi h}{2}\right) S_{t-1,h} + \cos\left(\frac{\pi h}{2}\right) S_{t-1,h}^* + \varepsilon_{th}^*\ , \text{ for h=1,2.} \end{aligned} \tag{A.4}$$

The disturbances ε_{th} en ε_{th}^* are normally distributed with expectation 0 and the same variance σ_S^2.

The additive outlier is modelled using a dummy variable:

$$\delta_t = \begin{cases} 1 & \text{for the period } t \text{ where an outlier occurs} \\ 0 & \text{for all other periods} \end{cases} \tag{A.5}$$

Where the magnitude of the outlier is measured by a time invariant regression coefficient α.

A level shift is modelled using an intervention variable:

$$\Delta_t^L = \begin{cases} 0 & \text{for all } t \text{ before the period of a level shift} \\ 1 & \text{for all } t \text{ from (and including) the period of a level shift} \end{cases} \tag{A.6}$$

As mentioned, the magnitude of the level shift is measured by a time invariant regression coefficient λ. A break in the seasonal pattern is modelled with a similar intervention variable:

$$\Delta_t^S = \begin{cases} 0 & \text{for all } t \text{ before the period of a seasonal break} \\ 1 & \text{for all } t \text{ from (and including) the period of the seasonal break} \end{cases} \tag{A.7}$$

The magnitude of the seasonal break is measured by γ_t, which is defined by (A.3) and (A.4), but where the disturbances in (A.4) are equal to zero, making the pattern time invariant. This implies that all four quarters have their own break (adding up to zero) which remains the same through time.
Working day effects are measured using a time varying regression coefficient. This is done since GDP generally increases through time and therefore the size of the working day effect as well. The regression coefficient is modelled using a random walk:

$$\beta_t \; = \beta_{t-1} + \varepsilon_{t\beta}, \tag{A.8}$$

with $\varepsilon_{t\beta}$ a normally distributed disturbance term with expectation 0 and variance σ_β^2.

The GDP time series is broken down into its subseries in a hierarchical, nested way, at several levels. GDP is broken down into K domains. In constant prices one of the domains always represents the statistical discrepancy between the chained volumes for the respective hierarchical level. These K series can be represented by a K-dimensional vector $(y_{t1}, \dots, y_{tk}, \dots, y_{tK})'$. For every period, it holds that:

$$y_{t+} = \sum_{k=1}^{K} y_{tk}. \tag{A.9}$$

Due to the hierarchical structure, there are similar constraints for different subsets of series.

The K series can be modelled by a K-dimensional multivariate STM:

$$y_{tk} = L_{tk} + S_{tk} + \alpha_k \, \delta_t \; + \beta_{tk} x_t \; + \lambda_k \, \Delta_t^L + \gamma_{tk} \Delta_t^S + I_{tk} \;, k = 1 \dots K. \tag{A.10}$$

In model (A.10) we have

- L_{tk} a time varying trend for series k, defined according to (A.2), where the disturbances ε_{tRk} are normally distributed with
  $$Cov(\varepsilon_{tRk}, \varepsilon_{t'Rk'}) = \begin{cases} \sigma_{Rk}^2 & \text{if} \quad t = t' \text{ and } k = k' \\ \zeta_{kk'} & \text{if} \quad t = t' \text{ and } k \neq k' \\ & \text{if} \qquad t \neq t' \end{cases}$$

- S_{tk} a time-dependent seasonal component for series k, defined by (A.3) and (A.4) where the disturbances ε_{tkh} and ε^*_{tkh} are normally distributed with

$$Cov(\varepsilon_{thk}, \varepsilon_{t'h'k'}) = Cov(\varepsilon^*_{thk}, \varepsilon^*_{t'h'k'}) = \begin{cases} \sigma^2_{Sk} & \text{if} \quad h = h' \text{and } t = t' \text{ and } k = k' \\ \zeta_{kk'} & \text{if} \quad h = h' \text{and } t = t' \text{and } k \neq k' \\ & \text{if} \quad h \neq h' \text{ or } t \neq t' \end{cases}$$

- δ_t a dummy variable indicating the period where an additive outlier occurs in different series and α_k the corresponding time-invariant regression coefficient measuring the magnitude of the outlier in series k. α_k may be zero for some series k if the analysis has shown that the outlier does not occur in that specific series.
- x_t the calendar effect information for quarter t and β_{tk} the corresponding time-varying regression coefficient for series k, as defined by (A.8),
- Δ^L_t a dummy variable indicating the period in which a level shift occurs, as defined by (A.6), with λ_k a regression coefficient measuring the magnitude for series k. λ_k may be zero for some series d if the analysis has shown that the level shift does not occur in that specific series.
- Δ^S_t a dummy variable for the seasonal break, as defined by (A.7), and γ_{tk} a time-invariant seasonal pattern for series k. γ_{tk} may be zero for some series k if the analysis has shown that the seasonal break does not occur in that specific series.
- I_{tk} a disturbance term for the unexplained variance in series k, all mutually independent and normally distributed with every series its own variance.

Several constraints are imposed to the times series components of all series following an approach similar to the benchmark procedure proposed by Doran (1992). They ensure that the sum of the components over the subseries is equal to the same component of the aggregate, before and after removing all effects listed above. We have constraints for the following components:

- The trend components: $L_{t+} = \sum_{k=1}^{K} L_{tk}$ (A.11)
- The regression coefficients for the working day effects: $\beta_{t+} = \sum_{k=1}^{K} \beta_{tk}$, for all t (A.12)
- The seasonal components: $S_{t+} = \sum_{k=1}^{K} S_{tk}$ (A.13)
- Outliers: $\alpha_{+} = \sum_{k=1}^{K} \alpha_{tk}$ (A.14)
- Level shifts: $\lambda_{+} = \sum_{k=1}^{K} \lambda_{tk}$ (A.15)
- Seasonal breaks: $\gamma_{t+} = \sum_{k=1}^{K} \gamma_{tk}$ (A.16)

Similar constraints are added for different subsets of the series due to the hierarchical structure.

In the specific case considered in this paper, the subscript + refers to the GDP. In the general case it can also refer to the next aggregate.

In order to estimate the multivariate STM described above, it is written in state space form. Next, the Kalman filter is used to obtain optimal estimates for all state variables (see Durbin and Koopman, 2012, Harvey, 1989). The Kalman filter is a recursive procedure to obtain optimal estimates for the state vector at time t based in the data up to and including time period *t*, and are referred to as the filtered estimates. The filtered estimates of past state vectors can be updated, if new data become available. This procedure is referred to as smoothing. Several smoothing algorithms are available in the literature. In this paper, the Kalman filter estimates for the state variables are smoothed with the fixed interval smoother, which is a broadly applied smoothing algorithm, and these estimates are referred to as the smoothed estimates. The Kalman filter assumes that the hyperparameters are known, which is generally not the case. Therefore they are estimated with a maximum likelihood procedure. Finally, the Kalman filter is initialised by assuming diffuse priors for all the state variables.

These models are analysed with a program that was developed in Oxmetrics (Doornik, 2009), using the procedures of Ssfpack 3.0 (Koopman et al., 1999, 2008). Ssfpack is a library of subroutines developed for analysing (multivariate) STM's.

Several checks on model fit of the STM are performed. An important goodness-of-fit criterion is the AIC, which is used to choose between alternative models. A higher value is better. Next, model assumptions have to be verified. This is done using the residuals (the standardised prediction errors to be precise), which should be (in order of importance):
- Independent, i.e. show no autocorrelation.
- Homoscedastic: constant variance throughout the series.
- Normally distributed.

Each of these assumptions is verified using a statistical test. In our case, the results indicated no strong deviation from the model assumptions.

Appendix 3: quality diagnostics

X-13Arima-SEATS and comparable seasonal adjustment programs compute eleven quality diagnostics and a weighted average (Q) of these eleven values. In this appendix we present a short summary (Ladiray and Quenneville (2001, pp.176-182)).

M1: contribution of the irregular component to fluctuations
M1 diagnostics the relative contribution of the irregular component to the changes in the series. If this contribution is large, this means that the irregular component causes many more fluctuations than the seasonal component in the series. In such a case, a successful distinction cannot be made between the seasonal component and the irregular component.

M2: contribution of the irregular component in the stationary series
Just like M1, M2 also diagnostics the contribution of the irregular component to the total variance in the series. If M2 is large, the irregular component is also relatively large. However, the calculation of M2 differs from the calculation of M1.

M3: ratio of the irregular component to the trend
To properly determine the seasonal decomposition, it is important that the fluctuations in the irregular component are not too large compared to the

fluctuations in the trend. M3 diagnostics the ratio between the fluctuations of these two components.

M4: connection in the irregular component
One of the most important assumptions for the irregular component is that there is no connection between successive observations. If there is a strong connection between successive observations, this component is indeed not so irregular. M4 therefore diagnostics the connection in the irregular component.

M5: number of months for cyclical dominance (MCD)
Just like M3, M5 also diagnostics the changes in the irregular component compared to the changes in the trend cycle.

M6: ratio of the irregular component to the season
During the first two calculation rounds, a 3x5 filter is used for the calculation of the seasonal component. M6 checks whether the 3x5 filter is suitable for the given series. A large value of M6 means that the ratio of the irregular component and the seasonal component is either too small or too large for the filter.

M7: identifiability of the seasonal pattern
As we saw in table 1, M7 is the most important quality measure for the seasonal adjustment. If M7 is larger than 1, we may not, in principle, accept the seasonal adjustment as such. M7 indicates the extent to which the seasonal effect in the series is identifiable. If the seasonal effect is poorly identifiable, the absolute error in the ultimate seasonal component is large.

M8 to M11: change in the seasonal pattern over the years
M8 to M11 measure the extent to which the seasonal pattern in the series is subject to change. If the seasonal pattern changes strongly, the seasonal filters of X-13ARIMA_SEATS are not able to accurately estimate the seasonal pattern, and the error in the estimations is large. In particular, if the seasonal pattern in the end years changes strongly, the problem is large, because this means that the error in the most recent estimations is large. And it is exactly the most recent estimations that users of the series are interested in. The seasonal pattern can change in two different ways. First, more or less arbitrary fluctuations can occur in the seasonal pattern. Second, there can be a systematic increase or decrease. M8 and M10 measure the arbitrary fluctuations in the seasonal pattern. M9 and M11 measure the systematic increase or decrease in the seasonal pattern. In this context, M8 and M9 are calculated over the entire series. M10 and M11 are calculated based on the most recent years.

Explanation of symbols

Empty cell	Figure not applicable
.	Figure is unknown, insufficiently reliable or confidential
*	Provisional figure
**	Revised provisional figure
2015–2016	2015 to 2016 inclusive
2015/2016	Average for 2015 to 2016 inclusive
2015/’16	Crop year, financial year, school year, etc., beginning in 2015 and ending in 2016
2013/’14–2015/’16	Crop year, financial year, etc., 2013/’14 to 2015/’16 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures.

Colofon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Studio BCO

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contactform: www.cbsl.nl/information