The BV4.1 Procedure
for Decomposing and Seasonally Adjusting
Economic Time Series

(Translation of ‘Methodenbericht, Volume 3’)
## Content

1. Introduction ................................................................................................................................. 1
2. The time series model .................................................................................................................. 3
3. The method of estimating the trend-cycle and seasonal components ...................................... 4
   3.1 Basic model ............................................................................................................................ 4
   3.2 Basic principle of estimation ................................................................................................. 5
   3.3 Estimating the trend-cycle component ............................................................................... 8
   3.4 Estimating the seasonal component .................................................................................. 12
4. Estimating outliers, calendar effects and influences of user-defined variables ..................... 15
   4.1 Outlier component ............................................................................................................... 15
   4.2 Calendar component .......................................................................................................... 19
   4.3 User component and level shifts .................................................................................... 20
   4.4 Estimation procedure ....................................................................................................... 21
5. Overview of the procedure ....................................................................................................... 23
6. The PC software for BV4.1 .................................................................................................... 24
7. Literature .................................................................................................................................. 26

Annex: Example of applying the BV4.1 software ....................................................................... 27
The BV4.1 Procedure for Decomposing and Seasonally adjusting Economic Time Series

1 Introduction

To examine current economic activities, the Federal Statistical Office calculates the data of a multitude of short-term indicators at monthly or quarterly intervals. However, what is really at the centre of interest for purposes of analysing economic trends is the development of indicators over time, rather than the individual data themselves. A major pillar of analysing economic trends therefore is the analysis of time series, that is of chronological data of short-term indicators.

On the one hand, current analyses of economic trends require the short-term – i.e. infra-annual – monitoring of economic variables and, on the other hand, this creates one of the main problems. The goal of such analyses is to derive from the data information on the phase of the economic cycle the overall economy or a specific economic branch currently is in. However, for infra-annual time series, the cycles that are relevant for the economic trends and which by definition have periodicities longer than one year, are generally affected by considerable short-term fluctuations. As a consequence, analysing the current business-cycle phase is quite difficult. To improve the accuracy of the economic trend analyses, it is common to apply mathematical filtering techniques to remove such fluctuations from the time series.

Developing mathematical adjustment methods always requires defining models which adequately describe the time series. A widespread modelling approach is based on the empirical finding that for infra-annual economic time series there are specific relations – usually of a periodical type – between adjacent observations. For most of those series, the relations are similar, which is due to the fact that the same or similar factors exert direct or indirect influence on most of the economic variables. The typical series structures originating from those relations are grouped under the following terms:

- trend,
- economic fluctuations,
- seasonal fluctuations,
- calendar fluctuations,
- residual or irregular fluctuations.

What is meant by the trend of a time series is its basic tendency. It reflects the influence of factors having a long-term impact such as technological progress or population growth. Economic fluctuations comprise the impact of all factors that have a medium-term influence on the series level (e.g. changes in consumption behaviour or investment behaviour caused by political actions such as tax policy and government expenditure). What is
grouped under the term “seasonal fluctuations” is the impact of factors taking effect at annual intervals (e.g. climatic conditions). Consequently, seasonal fluctuations are variations with a more or less annual periodicity. The term “calendar fluctuations” relates to the influences exerted on time series observations by different numbers of the various weekdays and public holidays in the underlying reference periods (months, quarters). Therefore, calendar fluctuations are important especially for time series of cumulative variables (e.g. production). The residual fluctuations cover the effects of any other factors. Generally, these are effects of a multitude of conditions having only a short-term impact on the values of the time series (e.g. strikes, errors in data collection, etc.), so that the residual fluctuations are also referred to as irregular fluctuations.

Against that background, it is an obvious choice to imagine economic time series as being composed of those fluctuations in a suitable way and, consequently, to use the terms trend, business cycle, seasonal, calendar, and irregular components. Because of definitional problems of delimitation, the trend and the cycle components are usually combined to form the trend-cycle component.

The first four of the above components are also referred to as systematic components. It is, however, definitely possible that even the residual or irregular component of a time series includes elements that must be considered as systematic. This is always the case where specific factors exert permanent influence on the values of a series in a way causing fluctuations which cannot be assigned to the fluctuations of the systematic components that are typical of economic time series.

The importance of the above-explained approach to modelling economic time series by means of a small number of movement components arises from the fact that it is the basis of the so-called decomposition procedures – a group of mathematical methods to facilitate carrying out economic trend analyses that is well established worldwide and, not least of all, in official statistics. Its purpose is to decompose economic time series into their various components. After decomposition, the economic analyst generally has two tools at his disposal to assess the economic situation. The first is seasonally adjusted series. These are produced by removing from the time series the seasonal and calendar components obtained through a decomposition procedure, so that the trend-cycle movements are revealed more clearly. Second, decomposition procedures directly provide estimates of the trend-cycle components of the time series. (For the benefits and shortcomings of using seasonally adjusted series and of using trend-cycle components cf. Schmidt (1991), Speth (1994) and Höpfner (1998).)

The BV procedure, which has been applied for more than 30 years at the Federal Statistical Office, is also one of the decomposition procedures. BV stands for Berliner Verfahren (Berlin procedure). The mathematical bases of that method, which is suitable for analysing monthly and quarterly economic series, were developed in the late 1960s in Berlin, that is at the Technical University and at the German Institute for Economic Research (Nullau, Heiler et al. (1969)). From 1983, the version BV4 – an own further development (Nourney (1983 and 1984)) – was applied at the Federal Statistical Office, which has now been replaced by BV4.1. The latter differs from BV4 by methodical improvements in the treatment of calendar effects and so-called outliers – which refers to time series values differing substantially from the general structure of the series values because of unusually large irregular influences. Also, BV4.1 now allows the user to separate from the residual component effects of known influential variables by explicitly take account of them as independent time series components, thus further improving the
efficiency of the method. The detailed presentation of those new features is the subject of this methodology report. However, the description of BV4.1 is not limited to the new elements of the procedure. For the sake of completeness, those parts of BV4 that have been taken over without changes are also documented.

The improved BV4.1 procedure, just like the former version BV4, is particularly well suited to be used at statistical offices and other institutions where current analysis results have to be produced regularly for a large number of time series. Here are its major benefits:

- Good cost-benefit ratio because high-quality analysis results can be achieved without expert knowledge, costly user training or long experience in applying the method being required.
- Analysis results are basically independent of the users, so that it is ensured not only that they can be computational reproduced but also that they can be understood in their entirety.
- High efficiency of seasonal adjustment even for quickly changing seasonal fluctuations.
- The trend-cycle movements of the time series are depicted plausibly from an economic point of view.
- Components/seasonally adjusted series of sub-series can be summed up to form the relevant component of the aggregate series (provided that the same time series model is used for all series). This means that there is generally no difference between the results of so-called indirect and direct analyses of aggregate series.

(As a special user service, for major short-term indicators the Federal Statistical Office has been publishing since 2001 not only decomposition results of BV analyses but also seasonally adjusted series according to the X12-ARIMA method of the US Bureau of the Census (http://www.census.gov/).)

2 The time series model

The core of BV4.1 is the modelling of time series through the two systematic trend-cycle and seasonal components and an irregular component; for the latter it is assumed, contrary to the above general definition, that it consists only of random fluctuations. That decomposition approach, which in the following will be referred to as the basic model, therefore requires generally that the time series to be analysed are adjusted in advance for all those systematic influences which – such as calendar effects – cannot be assigned to the trend-cycle or the seasonal component. Thus it is useful for BV4.1 analyses – wherever possible – to integrate into the general time series model such effects in the form of one or more separate components, in order to be able to eliminate those systematic influences from the time series before applying the basic model. Furthermore, it is generally useful, before determining the trend-cycle and the seasonal component, to adjust the time series for effects of extreme random fluctuations which produce so-called outliers. The mathematical background here is that non-eliminated outliers will more or less strongly impair the efficiency of the estimation methods applied in BV4.1. Therefore, in BV4.1 an outlier component is separated from the residual component.
Assuming that there are \( n \) chronological observations \( x_t, t = 1, \ldots, n \), of a monthly or quarterly series. Then the general formula of the BV4.1 time series model is:

\[
x_t = f(m_t, s_t, c_t, a_t, e_t, u_t), \quad t = 1, \ldots, n.
\]  

(2.1)

This means that for the individual observations \( x_t \), it is assumed that there is a specific functional relation \( f \) with the relative values of the trend-cycle component \( m_t \), the seasonal component \( s_t \), the calendar component \( c_t \), the component of other systematic explanatory variables \( a_t \), the outlier component \( e_t \) and the component of random fluctuations \( u_t \).

Another fundamental assumption of BV4.1 refers to the form of \( f \): An additive combination of the components forming the time series is assumed. So model (2.1) can be written as:

\[
x_t = m_t + s_t + c_t + a_t + e_t + u_t, \quad t = 1, \ldots, n.
\]  

(2.2)

This modelling means that no systematic (e.g. proportional) relations between the components are formulated; this does however not mean that they are ruled out.

As the basic model for estimating the trend-cycle and the seasonal component is used in a special way also for the necessary prior adjustments, it is the first item presented in the following. It is identical to that of BV4. What has also been taken over into BV4.1 is the BV4 approach to estimate the trend-cycle and the seasonal components. For the purpose of this methodology report, it will therefore be sufficient to present both of them in a rather concise form. For details see: Nullau, Heiler et al. (1969), pp. 19 et seqq., and Nourney (1983 and 1984).

3 The method of estimating the trend-cycle and seasonal components

3.1 Basic model

The methodical approaches of BV are based on the assumption that a monthly or quarterly economic series with \( n \) observations is the realisation of a discrete stochastic process \( \{ \xi_t | t = 1, \ldots, n \} \) which, according to the general decomposition approach, is composed additively of the stochastic sub-processes \( \{ \mu_t | t = 1, \ldots, n \} \), \( \{ \nu_t | t = 1, \ldots, n \} \) and \( \{ \epsilon_t | t = 1, \ldots, n \} \), which describe the trend-cycle fluctuations, the seasonal fluctuations and the random fluctuations of the time series. This means that an observation \( x_t \) is the realisation of the relevant random variable \( \xi_t \), which is itself composed of the realisations \( m_t \), \( s_t \) and \( u_t \) of the random variables \( \mu_t \), \( \nu_t \) and \( \epsilon_t \).
For the trend-cycle process \( \{ \mu_t \mid t = 1, \ldots, n \} \) it is assumed that it is a non-deterministic process with a high positive autocorrelation of the process variables, so that realisations are so “smooth” that they can be approximated with sufficient accuracy through low-order polynomials \( p \) with regard to shorter time periods. So, the functional formulation for modelling the trend-cycle component is:

\[
m_t = \sum_{j=0}^{p} \alpha_j t^j. \tag{3.1.1}
\]

For the seasonal process \( \{ \nu_t \mid t = 1, \ldots, n \} \) it is assumed that it is weakly stationary and that the annual periodical fluctuations of its realisations can change only slowly, so that for shorter time periods they can be represented with sufficient accuracy by finite Fourier series, that is:

\[
s_t = \sum_{j=1}^{l} \left( \beta_j \cos \lambda_j t + \gamma_j \sin \lambda_j t \right). \tag{3.1.2}
\]

Here, \( \beta_j \) and \( \gamma_j \) are the Fourier coefficients, \( \lambda_j = 2\pi / P \) is the angular frequency belonging to the annual period (with \( P \) being the number of observations of the time series within a year, i.e. \( P = 12 \) for monthly series and \( P = 4 \) for quarterly series), and \( \lambda_j = \lambda_1 \cdot j \), for \( j = 1, 2, \ldots \), with \( \lambda_1 \leq \pi \) are the harmonic waves for \( \lambda_1 \), that is infra-annual oscillations, which also have the period \( P \) (i.e. in (3.1.2) there is \( l = 6 \) for monthly series and \( l = 2 \) for quarterly series).

For the random process \( \{ \varepsilon_t \mid t = 1, \ldots, n \} \), white noise (i.e. mutually independent and identically distributed process variables \( \varepsilon_t \)) with an expected value of zero is assumed.

This gives the following so-called basic model of BV4.1:

\[
x_t = m_t + s_t + \varepsilon_t = \sum_{j=0}^{p} \alpha_j t^j + \sum_{j=1}^{l} \left( \beta_j \cos \lambda_j t + \gamma_j \sin \lambda_j t \right) + \varepsilon_t^*. \tag{3.1.3}
\]

Here \( \varepsilon_t^* \) generally differs from \( \varepsilon_t \) by the approximation error because of the chosen so-called basic functions \( \sum_{j=0}^{p} \alpha_j t^j \) and \( \sum_{j=1}^{l} \left( \beta_j \cos \lambda_j t + \gamma_j \sin \lambda_j t \right) \). According to the above assumptions, however, that error is considered negligible for sufficiently short periods, so that it will not be taken into account in the following.

### 3.2 Basic principle of estimation

For longer time series, the basic model is generally not suitable for the estimation of the trend-cycle and the seasonal components. Trend-cycle polynomials of low order generally cannot take sufficient account of the movements of the trend-cycle development in economic time series, and the approximation of the seasonal
component, too, is less and less meeting the requirements the longer is the time series because changing seasonalities cannot be covered by the basic function (3.1.2). Therefore, by means of the basic model, a moving estimation of the components is performed in BV4.1.

If $t_i$ refers to the period of the $i^{th}$ observation of a time series and $U = [t_1, \ldots, t_n]$ to the entire observation period, then the principle of moving estimation means that the estimation of the coefficients $\alpha_j$, $\beta_j$, and $\gamma_j$ of the trend-cycle polynomial and the Fourier series of the basic model (3.1.3) is first made on the basis of the first $k$ ($<n$) observations of a time series, that is the observations within the time interval (estimation range) $U^*_1 = [t_1, \ldots, t_{k*}]$, then to the observations of the estimation range $U^*_2 = [t_{k*}, \ldots, t_{k*+1}]$ etc., up to the observations of the estimation range $U^*_n = [t_{n-k*}, \ldots, t_n]$. In doing so, in BV4.1 the coefficients for the individual estimation ranges are determined by means of a least-squares criterion.

That way, for one and the same observation $x_i$ one firstly obtains several estimated values $\hat{m}_i$, and $\hat{s}_i$, according to the number of the different estimation ranges $U^*_l$, $l = 1, \ldots, n - k + 1$, which cover $t_i$. Results of the spectral analysis show, however, that the estimated values $\hat{m}_i$, and $\hat{s}_i$, do not show any leads or lags – so-called phase shifts – compared with the behaviour of the relevant components in the time series only if they are based on estimation ranges where the time point $t_i$ is exactly in the middle. Therefore, in BV4.1, the length $k$ of the estimation ranges $U^*_l$ is chosen odd-numbered and only estimated values for the trend-cycle component and the seasonal component concerning the middle time points of the $U^*_l$’s are used. (The one time point of an estimation range whose estimate is used for the BV4.1 analysis result will in the following be referred to as the “estimation point”. Furthermore, to avoid misunderstanding, the term “estimation time” will be used only for a time point regarding the entire analysis period of the time series.)

As of every estimation range $U^*_l$ only the estimated values for the trend-cycle component and the seasonal component for the relevant central estimation point are used, it is an obvious choice to minimise a weighted least-squares criterion for the (estimation range related) estimation of the coefficients of the basic model, with the weights being defined in a way that such time series values tend to have more influence the closer they are to the estimation point. The concentration aspect of the moving estimation method on time series values that are neighbouring the estimation point is thus logically continued and intensified. Additionally, the importance of the concrete formulation of the basic model (3.1.3) for the estimation results – which is relativised anyway by using a moving estimation approach – is further reduced. Despite using one and the same – and even rather simply structured – basic model for all time series, it is thus possible to depict the variety of trend-cycle and seasonal developments in economic time.

In concrete terms, the estimates $\hat{m}_l$ and $\hat{s}_l$ based on the estimation range $U^*_l$ are determined according to the following weighted least-squares criterion:

$$\sum_{l=1}^{n} w_i \left( x_i - \hat{m}_l - \hat{s}_l \right)^2 = \text{Minimum}, \quad (3.2.1)$$
where

\[ w_i = 1 - \frac{|t_i - t^+|}{D + 1}, \quad t_i \in U. \] (3.2.2)

Here, \( t^+ \) is the time point within the estimation range \( U \) to which is assigned the largest weight \( (w_i = 1) \) and \( D \) is the longer of the two distances (expressed in months or quarters) between \( t^+ \) and the first and \( t^+ \) and the last time point of \( U \).

The basic principle of estimation in BV4.1 provides that \( t^+ \) is to be set to the central estimation points of an estimation range. As a result, the weights, starting from \( t^+ \), will fall towards zero in a symmetrical and linear manner towards the ends of the estimation range.

Due to that symmetry of weights, it is also ensured that the estimated components are free from phase shifts.

Another measure taken to improve the estimation quality is to apply the above approach to the estimation of the trend-cycle component and the seasonal component under different framework conditions that are component-specifically optimised. What is important here is especially the length of the estimation ranges used. This is because for the estimation of the seasonal component it is necessary to consider time series values of several years to have an information basis to produce accurate results. On the other hand, the estimation ranges necessary to achieve good seasonal estimates are generally too long to represent through a low-order polynomial all ups and downs of the trend-cycle development of a series.

Therefore, what is done in BV4.1 is to determine first the trend-cycle component of a series under framework conditions which are tailored to the goal of estimating the trend-cycle component. The estimatee of the seasonal component, which is simultaneously produced here, is not used, so that the approach is referred to as the partial estimation of the trend-cycle component. After this the seasonal component is estimated. For that purpose, the time series is adjusted for the estimated trend-cycle component (i.e. \( x_t - \hat{m}_t, \quad t = 1, \ldots, n \)) and then again analysed according to the basic principle, but based on conditions which are favourable for estimating the seasonal component. What is then used is only the estimate of the seasonal component (partial estimation of their seasonal component).

In both estimation steps, the (complete) basic model is used to avoid biased estimates for the component one is actually interested in. This is especially important when estimating the trend-cycle component because the time series here still include all seasonal effects. When estimating the seasonal component on the basis of the trend-cycle-adjusted series, this approach avoids that shortcomings which may have occurred in estimating the trend-cycle component can bias the estimates of the seasonal component.

A consequence of the requirement for estimation points lying exactly in the middle of the estimation ranges is that the described basic principle of moving estimation cannot be applied to the entire time series. In order to
perform the decomposition and the seasonal adjustment also at the ends of a time series, the basic principle is abandoned, that is, the estimation ranges there are located asymmetrically around the estimation point. Also, there are modifications concerning the basic functions and the relative position of \( t^\ast \) with regard to the estimation point (cf. chapters 3.3 and 3.4).

### 3.3 Estimating the trend-cycle component

Using the basic model (which is linear in the model parameters) and estimating the model parameters by minimising the least-squares criterion has the following effect: The estimates of the trend-cycle components \( \hat{m}_t \) for all estimation times \( \hat{t} \) result from linear combinations of the time series values \( x_t \) (where, in the context of decomposition and seasonal adjustment of time series, such linear combinations are referred to as linear filters).

This means:

\[
\hat{m}_t = \sum_{\ell=1}^n \delta_\ell(\hat{t}) x_t, \quad \hat{t} = 1, \ldots, n. \tag{3.3.1}
\]

In BV4.1, for the estimation of \( m_t \), only those observations \( x_t \) are used which fall within the related estimation range \( U_t \), that is: \( \delta_\ell(\hat{t}) = 0 \) for \( t \notin U_t \). Thus (3.3.1) can be written as

\[
\hat{m}_{\hat{t}} = \sum_{t \in U_{\hat{t}}} \delta_\ell(\hat{t}) x_t, \quad \hat{t} = 1, \ldots, n. \tag{3.3.2}
\]

(In (3.3.1) and (3.3.2) the bracket \( (\hat{t}) \) is to illustrate that the filter weights \( \delta_\ell(\hat{t}) \) are different for the various estimation times \( \hat{t} \)).

The filter weights \( \delta_\ell(\hat{t}) \), for \( t \in U_{\hat{t}} \), depend on, and are uniquely determined by, the following conditions of the estimation approach:

- length \( k_t \) of the estimation range \( U_t \) (number of months or quarters),
- position \( h_t \) of \( \hat{t} \) within \( U_t \),
- position \( q_t \) of \( t^\ast \) within \( U_t \),
- degree \( p_t \) of the trend-cycle polynomial of the basic model used for the estimation within \( U_t \).

For all \( \hat{t} \), they do not depend on the observations of the time series. Thus, a BV4.1 filter \( M_{t} \) for the estimation of the trend-cycle component for the estimation time \( \hat{t} \) can be described by \( M_t(k_t, h_t, q_t, p_t) \).
When developing BV4, comprehensive spectral analysis studies were performed in order to find combinations of the above conditions which result in filters that, owing to their properties in the frequency domain, are adequate for any monthly or quarterly economic time series.

For the trend-cycle estimation in monthly time series, the filter \( M_{(27,14,14,3)} \) was identified as most suitable. From the aspect of spectral analysis, the filter stands out by the fact that a trend-cycle component estimated with its help does not involve any phase shifts and represents all long-term and medium-term fluctuations of the filtered series nearly unchanged and also nearly unimpaired by variations of the time series with annual or higher frequency. It is applied to the middle part of the analysis period, which – because of the length of its the estimation range – extends from \( \hat{t} = 14 \) to \( \hat{t} = n - 13 \). \( M_{(27,14,14,3)} \) is therefore referred to as the central filter.

For estimations regarding the first and the last 13 months of the analysis period, subsequent filters had to be defined. When concretely defining those filters, the main focus was on ensuring that their properties are as similar as possible to that of the central filter. The reasons are, first, that the good filter effect of the central filter should be preserved as far as possible towards the ends of the time series and, second, that it should be avoided that the trend-cycle component (being composed of results of various filters) looked at as a whole shows level shifts or other discontinuities. However, it was of course not possible to avoid increasing quality losses (e.g. phase shifts) towards the ends of the analysis period, which was especially because of the necessity to use asymmetrical filters.

Based on the studies, the following subsequent filters were defined:
\[
\begin{align*}
M_{(28,16,16,3)}, M_{(29,18,18,3)}, M_{(30,20,20,3)}, M_{(30,21,21,3)}, M_{(29,21,21,3)}, \\
M_{(28,21,21,3)}, M_{(26,20,20,3)}, M_{(25,20,20,3)}, M_{(25,21,21,3)}, M_{(25,22,22,3)}, \\
M_{(25,23,23,3)}, M_{(26,25,25,3)} \text{ and } M_{(27,27,27,3)}.
\end{align*}
\]

To improve the estimate, however, determining the trend-cycle component for the last 6 estimation times is also based on the filters \( M_{(25,20,25,1)}, M_{(24,20,24,1)}, M_{(23,20,23,1)}, M_{(22,20,22,1)}, M_{(21,20,21,1)} \) and \( M_{(20,20,20,1)} \), where the trend-cycle component is modelled through a straight line. The final estimate is then calculated as the weighted mean of the two individual estimates, with – towards the end of the series – the weight of the filter with polynomial degree 1 gradually growing from \( 1/12 \) to \( 6/12 \) and the weight of the filter with polynomial degree 3 falling correspondingly from \( 11/12 \) to \( 6/12 \).

At the beginning of the analysis period, the analogue filters – which are mirror-inverted with regard to the estimation time \( \hat{t} \) – are used, that is:
\[
\hat{m}_{\hat{t}} = \sum_{\hat{t} = 1}^{\hat{t}} \delta_{\hat{t}-\hat{t}+1}(n-\hat{t}+1)x, \quad \hat{t} = 1,\ldots,13. \quad (3.3.3)
\]
Another consequence of the necessity of using subsequent filters is that former estimates of the trend-cycle component are revised when new analyses are performed including additional time series values. However, larger revisions that could be relevant for business cycle analysis generally occur only for the time points $n-1$ and $n-2$.

For the analysis of the trend-cycle component of quarterly series, the following filters or filter combinations were chosen: the central filter $M_L(9,5,5,3)$, valid for $\hat{t} = 5$ to $\hat{t} = n-4$, the subsequent filters $1/2 \cdot M_{n-3}(11,8,8,3) + 1/2 \cdot M_{n-3}(10,7,7,3)$, $1/2 \cdot M_{n-3}(10,8,8,3) + 1/2 \cdot M_{n-3}(9,7,7,3)$, $M_{n-1}(9,8,8,3)$ and $1/2 \cdot M_{n}(10,10,10,3) + 1/2 \cdot M_{n}(8,8,8,1)$ and the corresponding mirror-inverted filters for the first 4 time points of the analysis period.

For a more detailed illustration of the properties of the filters, chart 1 and chart 2 show the gain functions (cf. e.g. Leiner (1976)) of the central filter and of the three subsequent filters for $\hat{t} = n$, $n-1$ and $n-2$ (or $\hat{t} = 1$, 2 and 3) for the trend-cycle estimation for monthly and quarterly time series, respectively. The value of the gain function for a frequency $2\pi/P$ indicates the factor with which the amplitude of the relevant oscillation in the time series to be filtered has to be multiplied in order to arrive at the amplitude of the oscillation in the filtered series. Consequently, the relevant filter does not have any impact on the amplitudes of oscillations with frequencies for which the gain function has the value 1. They are included in the filtered series without changes in this respect. Where the values of the gain function are equal to 0, this means that the relevant oscillations are completely eliminated by the filter. They are no longer contained in the filtered series. Values of the gain function between 0 and 1 indicate that the amplitudes of the relevant oscillations are reduced by the filter process, while values larger than 1 indicate that such amplitudes are reinforced.
Chart 1: Gain functions of three subsequent filters ($\hat{f} = n$, $n-1$ and $n-2$) and of the central filter (-----) for the trend-cycle estimation for monthly series.

Chart 2: Gain functions of three subsequent filters ($\hat{f} = n$, $n-1$ and $n-2$) and of the central filter (-----) for the trend-cycle estimation for quarterly series.
3.4 Estimating the seasonal component

The situation for estimating the seasonal component is generally the same as that for estimating the trend-cycle component. Again, the estimated values of the seasonal component can be represented as results based on linear filters:

\[
\hat{S}_t = \sum_{i=1}^{q} \vartheta_i \hat{t} x_i, \quad \hat{t} = 1, \ldots, n.
\]  

(3.4.1)

with: \( \vartheta_i (\hat{t}) = 0 \) for \( t \notin U_\hat{t} \). The filters – now denoted as \( S_\hat{t} (k_t, h_t, q_t, p_t) \) – are also uniquely determined by the specification of \( k_t, h_t, q_t \), and \( p_t \), and the filter weights \( \vartheta_i (\hat{t}) \), \( t \in U_\hat{t} \), again do not depend on the observations of the time series to be analysed.

Therefore, when defining the filters for the estimation of the seasonal component, it was possible again to use spectral analyses in order to identify universally suitable filters. In contrast to the situation when looking for suitable trend-cycle filters, an additional condition that had to be taken into account here was that unbiased seasonal component estimates can be guaranteed only by those symmetrical filters for which either the position \( h_\hat{t} \) of the estimation point \( \hat{t} \) within the estimation range \( U_\hat{t} \) is a whole-number multiple of the annual period \( P \), i.e. \( h_\hat{t} = j \cdot P \), or where \( h_\hat{t} = k_\hat{t} - (j \cdot P) + 1 \), \( j \in \mathbb{N}_+ \).

For the analysis of monthly series, the central filter chosen for the seasonal estimation is the filter combination \( 6/7 \cdot S_{\hat{t}} (47,24,24,3) + 1/7 \cdot S_{\hat{t}} (47,24,24,1) \), applicable for \( \hat{t} = 24 \) to \( \hat{t} = n - 23 \). Contrary to the basic principle, a combination filter was chosen here even for the central filter to improve the filter effects. The filter has the following benefits: It does not cause any phase shifts, the seasonal effects entirely ‘get through’ for all series, and even stronger annual changes in the seasonal effects can be represented by the estimated seasonal component largely without problems. Due to that kind of flexibility of the central filter, which is a highly important property especially of filters applied universally, it may happen – depending on the situation of the individual series – that the seasonal component estimate is influenced by parts of the short-term fluctuations in the time series. That situation should however not be considered as a shortcoming with regard to the purpose of determining seasonally adjusted figures, that is determining figures which reveal more clearly the trend-cycle movements of the time series examined.

The required 23 subsequent filters are:

\[
\begin{align*}
6/7 \cdot S_{n-23} (46,24,24,3) &+ 4/5 \cdot S_{n-21} (45,24,24,3) + 1/5 \cdot S_{n-21} (45,24,23,1), \\
4/5 \cdot S_{n-20} (44,24,24,3) &+ 1/5 \cdot S_{n-20} (44,24,23,1), \\
2/3 \cdot S_{n-19} (43,24,24,3) &+ 1/3 \cdot S_{n-19} (43,24,22,1), \\
2/3 \cdot S_{n-18} (42,24,24,3) &+ 1/3 \cdot S_{n-18} (42,24,22,1), \\
1/2 \cdot S_{n-17} (41,24,24,3) &+ 1/2 \cdot S_{n-17} (41,24,21,1), \\
1/2 \cdot S_{n-16} (40,24,24,3) &+ 1/2 \cdot S_{n-16} (40,24,21,1), \\
1/2 \cdot S_{n-15} (39,24,24,3) &+ 1/2 \cdot S_{n-15} (39,24,20,1), \\
1/2 \cdot S_{n-14} (38,24,24,3) &+ 1/2 \cdot S_{n-14} (38,24,20,1), \\
2/9 \cdot S_{n-13} (37,24,24,3) &+ 1/9 \cdot S_{n-13} (39,24,36,3) + 6/9 \cdot S_{n-13} (37,24,19,1). 
\end{align*}
\]
Those subsequent filters, too, were determined from the aspect that their properties should be as similar as possible to that of the central filter. Time lags – especially for the estimated values of the last year of the analysis – and weaker performance with regard to an adequate representation of changing seasonalities must, however, be put up with.

For quarterly series, the filter sequence is as follows:

1/2 \( S_{n} \) (15,8,8,1) + 1/2 \( S_{2n} \) (15,8,8,3) for \( \tilde{t} = 8 \) bis \( \tilde{t} = n - 7 \) (central filter),

1/4 \( S_{3n} \) (14,8,7,1) + 1/4 \( S_{4n} \) (14,8,7,3),

1/2 \( S_{6n} \) (13,8,7,1) + 1/2 \( S_{7n} \) (13,8,8,3),

5/10 \( S_{9n} \) (12,8,7,1) + 3/10 \( S_{10n} \) (16,12,11,3) + 2/10 \( S_{11n} \) (12,8,8,3),

4/5 \( S_{13n} \) (11,8,6,1) + 1/5 \( S_{14n} \) (15,12,11,3),

2/5 \( S_{15n} \) (13,12,10,3) + 1/5 \( S_{16n} \) (17,16,15,3),

5/10 \( S_{17n} \) (8,8,7,1) + 2/10 \( S_{18n} \) (12,12,9,3) + 3/10 \( S_{19n} \) (16,16,14,3).

At the beginning of the analysis period, the corresponding mirror-inverted filters are used both for monthly and quarterly series.

Spectral analyses performed as part of developing the BV4.1 procedure have shown that a pre-adjustment of the time series for the trend-cycle components estimated with the above-described filters is altogether more advantageous for the estimation of the seasonal component than the direct application of the seasonal filters \( S_{j}(k_{j},q_{j},p_{j}) \) to the original time series. Therefore, in BV4.1, the seasonal component is not estimated according to (3.4.1), but rather by means of

\[
\hat{s}_{x} = \sum_{t=1}^{n} \delta_{t}(\tilde{t})(x_{t} - \tilde{m}_{t}) = \sum_{t=1}^{n} \delta_{t}(\tilde{t})(x_{t} - \hat{m}_{t}) = \sum_{t=1}^{n} \hat{\delta}_{t}(\tilde{t})x_{t}, \quad \tilde{t} = 1,\ldots,n. \tag{3.4.2}
\]

To assess the BV4.1 seasonal component, it is therefore not sufficient to just take account of the properties of the filters \( S_{j}(k_{j},q_{j},p_{j}) \). It is rather necessary to examine the – again linear – filters which are obtained through
the combination of the methodological steps of trend-cycle estimation, trend-cycle adjustment, and seasonal estimation, and which are denoted by $\mathcal{G}(\tilde{t})$, $t' \in U'$. 

In charts 3 and 4, the curves of the gain functions are shown for some selected combination filters for seasonal adjustment $x_i - \hat{s}_i$. 

Chart 3: Gain functions of three subsequent filters ($\tilde{t} = n$, $n - 1$ and $n - 2$) and of the central filter (—) for the seasonal adjustment of monthly series.
4 Estimating outliers, calendar effects and influences of user-defined variables

Where explanatory variables are not taken into account in linear regression models, this will lead to systematically biased least-squares estimates because the effects of the lacking variables are in part taken over by the others. Therefore it is generally useful to add further explanatory variables to the basic model.

4.1 Outlier component

Outliers are individual and sporadically occurring observations of a time series which, for various and generally unknown reasons, differ to an unusually large extent from the general structure of the series values. If they are not taken into account in the time series model, the estimate of the trend-cycle component will be biased for that periods (months or quarters) in which the outliers occur and also for some directly neighbouring periods. Especially, however, the seasonal estimation will be adversely affected because the biases of the seasonal component will no longer occur limited to periods in the neighbourhood of outlying observations, but to a considerable extent also to the corresponding periods in neighbouring years.
The following approach to identify outliers has been taken largely unchanged from BV4 to BV4.1.

The method is based on the assumption that, with the exception of a small number of observations, the time series \( \{x_t, \ t = 1, \ldots, n\} \) to be analysed is the realisation of a discrete stationary stochastic normal process \( \{\xi_t | \ t \in \mathbb{Z}\} \) with expected value \( E(\xi_t) = \nu \) and autocovariances \( \gamma_{\eta} = E((\xi_{\eta} - \nu)(\xi_{\nu} - \nu)) \), \( \eta = 0, \pm 1, \ldots \). On that basis, and drawing on the observations within time intervals of fixed length \( M \) (the so-called reference ranges) moving over the analysis period, the conditional expected values \( E(\xi_t | x_{t-M}, x_{t-2M}, \ldots, x_{t-M}) \) are determined for \( t = M + 1, \ldots, n \). Here, it is also assumed that there are no outliers among the first \( M \) observations of the time series.

A time series value \( x_t, M < t \leq n \), is regarded as an outlier if the probability is low that the value has been created by the same process as the preceding \( M \) values. This is considered to be the case if the deviation of the observation from the relevant conditional expected value of the process variable \( \xi_t \) is more than the product of a suitable (confidence) factor \( \tau \) and the process variable’s conditional standard deviation \( \sigma_t \), that is, \( x_t \not\in [l_t, u_t] \), where \( l_t = E(\xi_t | x_{t-M}, x_{t-2M}, \ldots, x_{t-M}) - \tau \sigma_t \) and \( u_t = E(\xi_t | x_{t-M}, x_{t-2M}, \ldots, x_{t-M}) + \tau \sigma_t \).

If for a time point a time series value is thus identified as an outlier, it is not appropriate to continue the procedure by using that value in the moving calculation of the conditional expected values of the subsequent time series values. For the outlier identification procedure it is therefore replaced by a rough estimate of the outlier-adjusted value, that is – depending on the direction of the deviation – by either the upper or the lower limit of the relevant tolerance range \( [l_t, u_t] \).

\( M \) and \( \tau \) can be defined by the BV4.1 user. In order to keep the number of outliers small, the standard setting should be \( \tau = 3 \). What should also be chosen by default is \( M = 24 \) for monthly series and \( M = 8 \) for quarterly series because, due to the seasonal effects in economic time series, it is useful to use the observations of at least two complete years to estimate the conditional expected values.

Here are more details: If

\[
\Gamma_{M+1}^{-1} = \begin{pmatrix}
\gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_M \\
\gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{M-1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\gamma_M & \gamma_{M-1} & \gamma_{M-2} & \cdots & \gamma_0
\end{pmatrix}
\]

is the autocovariance matrix of the above normal process \( \{\xi_t\} \) with regard to a reference range of length \( M \) and if \( \Gamma_{M+1}^{-1} = (\kappa_{ij}) \), \( i, j = 0, 1, \ldots, M \), is the relevant inverse, then the general formula for the required conditional expected values is

\[
E(\xi_t | x_{t-M}, x_{t-2M}, \ldots, x_{t-M}) = \nu - \frac{1}{\kappa_{00}} \sum_{j=1}^{M} \kappa_{0j} (x_{t-j} - \nu)
\]  

(4.1.1)
and the formula for the conditional variances is

\[
V(\xi_t | x_{t-1}, x_{t-2}, \ldots, x_{t-M}) = \frac{1}{\kappa_{00}} = \sigma_t^2, \quad t = M + 1, \ldots, n.
\] (4.1.2)

To estimate \( E(\xi_t | x_{t-1}, x_{t-2}, \ldots, x_{t-M}) \), first of all the expected value \( \nu \) is estimated by

\[
\hat{\nu} = \frac{1}{n} \sum_{i=1}^{n} x_i.
\] (4.1.3)

Estimated values for \( \kappa_{ij} \), \( i, j = 0, 1, \ldots, M \), are obtained by \( \hat{\kappa}_{ij} = (\hat{\Gamma}_{M+1})^{-1} \), where to estimate the autocovariances the formula

\[
\hat{\gamma}_\eta = \frac{1}{n_{\eta+1}} \sum_{i=n-\eta}^{n} (x_i - \hat{\nu})(x_{i-\eta} - \hat{\nu}), \quad \eta = 0, 1, \ldots, M
\] (4.1.4)

is applied. The estimated values \( \hat{\gamma}_\eta \) are biased because in (4.1.3) no division by \( n - \eta \) is performed, but that shortcoming is put up with because this is the only way to ensure that \( \hat{\Gamma}_{M+1} \) is positively definite and thus invertible (cf. e.g. Parzen (1961)). Also, this generally reduces the variances of \( \hat{\gamma}_\eta \).

Thus, the estimation formula for (4.1.1) is this:

\[
\hat{E}(\xi_t | x_{t-1}, x_{t-2}, \ldots, x_{t-M}) = \hat{\nu} - \frac{1}{\kappa_{00}} \sum_{i=1}^{M} \hat{\kappa}_{ij}(x_{t-i} - \hat{\nu}) = \hat{x}_t, \quad t = M + 1, \ldots, n.
\] (4.1.5)

According to (4.1.2), for estimating the conditional variances or standard deviations, the element \( \kappa_{00} \) from \( (\hat{\Gamma}_{M+1})^{-1} \) could be used. However, for reasons of programme efficiency, the following estimates are performed:

\[
\hat{\sigma}_t^2 = \frac{1}{n-M} \sum_{i=M+1}^{n} (x_i - \hat{x}_i)^2, \quad t = M + 1, \ldots, n.
\] (4.1.6)

The advantage here is that, if outliers are identified, it is not necessary to perform the inversion of the autocovariance matrix reestimated on the basis of the updated outlier-adjusted values.

By means of those formulae, and in a step-by-step process, for \( t = M + 1, \ldots, n \) those observations \( x_t \) are identified as outliers for which there is:
\[
X_i \notin \left[ \hat{X}_i - r \sqrt{\sigma_i^2}, \hat{X}_i + r \sqrt{\sigma_i^2} \right]. \tag{4.1.7}
\]

However, as soon as the first outlier has been identified, for example for period \( t^e \), further outliers will be identified by using the following estimation formulae instead of (4.1.5) and (4.1.6):

\[
\hat{E} \left( \xi \mid X_{t-1}, X_{t-2}, \ldots, X_{t-M} \right) = \hat{\nu} - \frac{1}{M} \sum_{j=1}^{M} \hat{X}_j \left[ (X_{t-j} - \hat{\nu}) \cdot 1_{t-j \leq t^e} + (\hat{X}_{t-j} - \hat{\nu}) \cdot 1_{t-j > t^e} \right] \quad \hat{X}_i \quad \left[ (X_{t-j} - \hat{\nu}) \cdot 1_{t-j > t^e} + (\hat{X}_{t-j} - \hat{\nu}) \cdot 1_{t-j \leq t^e} \right] \quad \hat{\nu} 
\]

\[
\hat{\sigma}_i^2 = \frac{1}{n-M} \sum_{j=1}^{t^e} (X_i - \hat{X}_i)^2 + (\hat{X}_i - \hat{\nu})^2 + \sum_{j=t^e+1}^{t} (\hat{X}_j - \hat{X}_i)^2 + \sum_{j=M}^{n} (X_j - \hat{X}_i)^2 \quad \hat{\sigma}_i 
\]

(M < \( t^e \), \( t \leq n \), where

\[
\hat{X}_i := \begin{cases} 
\hat{X}_i + r \sqrt{\sigma_i^2} & \text{if } X_i > \hat{X}_i + r \sqrt{\sigma_i^2} \\
\hat{X}_i - r \sqrt{\sigma_i^2} & \text{if } X_i < \hat{X}_i - r \sqrt{\sigma_i^2}
\end{cases} \quad \hat{X}_i \quad \left[ (X_{t-j} - \hat{\nu}) \cdot 1_{t-j > t^e} + (\hat{X}_{t-j} - \hat{\nu}) \cdot 1_{t-j \leq t^e} \right] \quad \hat{\nu} \tag{4.1.10}
\]

and

\[
\hat{X}_i := \begin{cases} 
X_i & \text{if } \hat{X}_i - r \sqrt{\sigma_i^2} < X_i < \hat{X}_i + r \sqrt{\sigma_i^2} \\
\hat{X}_i + r \sqrt{\sigma_i^2} & \text{if } X_i > \hat{X}_i + r \sqrt{\sigma_i^2} \\
\hat{X}_i - r \sqrt{\sigma_i^2} & \text{if } X_i < \hat{X}_i - r \sqrt{\sigma_i^2}
\end{cases} \quad \hat{X}_i \quad \left[ (X_{t-j} - \hat{\nu}) \cdot 1_{t-j > t^e} + (\hat{X}_{t-j} - \hat{\nu}) \cdot 1_{t-j \leq t^e} \right] \quad \hat{\nu} \tag{4.1.11}
\]

As an extension, compared with BV4, in BV4.1 that outlier identification procedure is also applied backwards. Thus it is possible, in particular, that outliers among the first \( M \) series values can be detected, too. Also, the formula (4.1.9) constitutes a modification compared with the corresponding BV4 formula. In contrast to BV4, all outliers, so far as identified, are now replaced by their tolerance limits (\( l \) or \( u \)) and all new estimates \( \hat{X}_i \) are also included.

After, for example, \( m \) periods \( t^e \) with probably outlying observations have been identified, the following outlier component is added to the time series model:

\[
e_t = \sum_{i=1}^{m} \nu_{i,t} e_{i,t} \quad t = 1, \ldots, n \quad \left[ (X_{t-j} - \hat{\nu}) \cdot 1_{t-j > t^e} + (\hat{X}_{t-j} - \hat{\nu}) \cdot 1_{t-j \leq t^e} \right] \quad \hat{\nu} \tag{4.1.12}
\]

where the \( e_{i,t} \) s denote the corresponding outlier dummy variables, that is,
\begin{equation}
\begin{aligned}
\gamma_{i,t} = \begin{cases} 
1 & \text{if } t = t^f_i, \\
0 & \text{otherwise}, 
\end{cases} 
\end{aligned}
\quad i = 1, \ldots, m,
\end{equation}

and where the \( \gamma_{i,t} \) s denote the coefficients to be estimated.

### 4.2 Calendar component

The observations of most economic time series are influenced by calendar effects. That is always the case, where the differences between periods (months or quarters) with regard to the number of days, the number of working days, or the number of individual weekdays and public holidays have an impact on the economic activities observed. Therefore, an obvious approach to modelling calendar effects is the following weekdays-based approach:

\begin{equation}
c_t^i = \sum_{i=1}^{8} \gamma_{i,t} d_{i,t}, \quad t = 1, \ldots, n.
\end{equation}

where, for \( i = 1 \) to 6, the \( d_{i,t} \) s denote the numbers of occurrences of Mondays to Saturdays in period \( t \) which are not public holidays, for \( i = 7 \) the number of Sundays and for \( i = 8 \) the number of public holidays not falling on a Sunday, and where the \( \gamma_{i,t} \) s denote the relevant coefficients.

If, for \( i = 1 \) to 8, the \( d_{i,t} \) s denote the mean values of the numbers of the “weekdays” \( i \) (according to the definition above) over all periods of the same “name” as \( t \), and if the \( d_t^i \) s are the average numbers of the “weekdays” \( i \) per period, then (4.2.1) can be written as:

\begin{equation}
c_t^i = \sum_{i=1}^{8} \gamma_{i,t} d_t^i + \sum_{i=1}^{8} \gamma_{i,t} \left( d_{i,t} - d_t^i \right) + \sum_{i=1}^{8} \gamma_{i,t} \left( d_{i,t} - d_{i,t}^i \right), \quad t = 1, \ldots, n.
\end{equation}

Here, \( c_t^i \) is a constant. In BV4.1, that part of the calendar effect is considered an element of the trend-cycle component. \( c_t^i \) denotes the length-of-period effect. Because of its seasonal character, it is considered an element of the seasonal component. Therefore, in BV4.1, what is included in the calendar component is only the effects of the deviation of the actual weekday and holiday structure from the mean period-specific structure:

\begin{equation}
c_t^i = c_3 = \sum_{i=1}^{8} \gamma_{i,t} \left( d_{i,t} - d_{i,t}^p \right), \quad t = 1, \ldots, n.
\end{equation}

What can be also used in BV4.1 to estimate the calendar effects, as alternatives to this weekdays-based calendar component, are calendar components based on the numbers of the working days not counting or counting Saturdays. Here, it is assumed that Saturdays, Sundays and holidays or Sundays and Holydays, respectively, do
not have any effect on the time series values and that the other weekdays do not differ in their effect, so that it is sufficient to use just their total number:

\[ c^A_t = \nu^A_t \left( d^A_t - \bar{d}^A_t \right), \quad t = 1, \ldots, n, \tag{4.2.4} \]

where

- \( d^A_t = \) number of the weekdays Mondays to Fridays in period \( t \) which are not holidays,
- \( \bar{d}^A_t = \) mean value of the numbers of the weekdays Mondays to Fridays which are not holidays over all periods of the same “name” as \( t \).

or

\[ c^W_t = \nu^W_t \left( d^W_t - \bar{d}^W_t \right), \quad t = 1, \ldots, n, \tag{4.2.5} \]

where

- \( d^W_t = \) number of the weekdays Mondays to Saturdays in period \( t \) which are not holidays,
- \( \bar{d}^W_t = \) mean value of the numbers of the weekdays Mondays to Saturdays which are not holidays over all periods of the same “name” as \( t \).

For the sake of simplification, no distinction will be drawn below between these three variants of the BV4.1 calendar component; the general formula

\[ c^A_t = \sum_{i=1}^{n} \nu^A_i k^A_{i,t}, \quad t = 1, \ldots, n, \tag{4.2.6} \]

will be used.

### 4.3 User component and level shifts

If there are available, say \( l \), suitable time series \( a_{i,t}, i = 1, \ldots, l, t = 1, \ldots, n \), on systematic explanatory variables which cannot be subsumed in the standard components of time series decomposition, it is possible to include that information by adding to the basic model a component of so-called user-defined explanatory variables. The general linear model for the user component is:

\[ a_t = \sum_{i=1}^{l} \nu p^i a^i_t, \quad t = 1, \ldots, n, \tag{4.3.1} \]
where the $\nu_{\lambda i}$ s denote the coefficients to be estimated.

What is referred to as level shift is an abrupt and permanent change of the mean level of a time series (generally caused by changes in the definition of the item covered by the time series values). If the user has information on level shifts in a time series which is to be included in the analysis, this is particularly important for the BV4.1 process (cf. chapter 5). The reason is that level shifts are incompatible with the stationarity assumption required for outlier identification. Therefore it is useful here to decompose the user component $a_t$ into a level-shift component $a_{s,t}$ and a component $a_{u,t}$ for the other user-defined explanatory variables:

$$a_t = a_{s,t} + a_{u,t}. \quad (4.3.2)$$

The separate determination of a level-shift component is necessary also because, once determined, the level-shift component generally – and also in BV4.1 – is assigned to the trend-cycle component.

If, for example, $p$ periods $t^i_t$ with level shifts are known, then the level-shift component is:

$$a_{s,t} = \sum_{i=1}^{p} \nu_{s,i} a_{s,t}^i, \quad t = 1,...,n, \quad (4.3.3)$$

where the $a_{s,t}^i$ s denote the values of the level-shift dummy variables, that is

$$a_{s,t}^i = \begin{cases} 1 & \text{if } \left\{ \begin{array}{l} t \geq t^i_t \\ \text{otherwise} \end{array} \right. , \quad i = 1,...,p. \end{cases} \quad (4.3.4)$$

The general formula for the component $a_{u,t}$ is:

$$a_{u,t} = \sum_{l=1}^{q} \nu_{u,l} a_{u,t}^l, \quad t = 1,...,n, \quad (4.3.5)$$

where the $a_{u,t}^l$ s denote the user-defined explanatory variables which are not level shifts ($p + q = l$).

### 4.4 Estimation procedure

The basis of estimating outliers, calendar effects and influences of user-defined variables is the comprehensive time series model (2.2), with the above-defined component models being used for the individual components:

$$x_t = m_t + S_t + c_t + a_t + e_t + u_t.$$
\[ m_t + s_t + \sum_{j=1}^{h} v_{k,j}^t \hat{K}_j^t + \sum_{j=1}^{l} v_{a,j}^t \hat{A}_j^t + \sum_{j=1}^{m} v_{e,j}^t \hat{E}_j^t + u_t, \quad t = 1, \ldots, n, \tag{4.4.1} \]

or using vectors:

\[ x = m + s + \sum_{j=1}^{h} v_{k,j}^t \hat{K}_j^t + \sum_{j=1}^{l} v_{a,j}^t \hat{A}_j^t + \sum_{j=1}^{m} v_{e,j}^t \hat{E}_j^t + u, \tag{4.4.2} \]

where \( x = (x_1, \ldots, x_n)' \), ..., \( u = (u_1, \ldots, u_n)' \).

As for the frame of that integrated estimation approach no suitable linear model components are available for the estimation of the trend-cycle and the seasonal component with regard to the entire time series, the system of equations (4.4.2) is first subjected to transformation with the linear filter \( F \) for trend-cycle and seasonal adjustment that is obtained from the BV4.1 filters for estimating the trend-cycle and seasonal components (described in chapters 3.3 and 3.4):

\[ F = (I - \hat{S})(I - \hat{M}), \tag{4.4.3} \]

where \( I \) is the \( n \times n \) unit matrix, \( \hat{M} = (\hat{M}_1, \ldots, \hat{M}_n)' \), \( \hat{S} = (\hat{S}_1, \ldots, \hat{S}_n)' \) and \( \hat{M}_t \) or \( \hat{S}_t \), \( t = 1, \ldots, n \), denoting the vectors of the weights of the relevant (combination) filters of BV4.1 for estimating the trend-cycle or seasonal component for the individual estimation times \( \hat{t} \).

This results in:

\[ F(x) = F(m + s) + \sum_{j=1}^{h} v_{k,j}^t F(k_j^t) + \sum_{j=1}^{l} v_{a,j}^t F(a_j^t) + \sum_{j=1}^{m} v_{e,j}^t F(e_j^t) + F(u). \tag{4.4.4} \]

As it can be assumed that \( F(m + s) = 0 \), the coefficients \( v \), that is to say the components \( c_t \), \( a_t \), and \( e_t \), are finally estimated by using the following linear regression model and the least-squares estimation method:

\[ F(x) = \sum_{i=1}^{h} v_{k,i} F(k_i) + \sum_{j=1}^{l} v_{a,j}^t F(a_j^t) + \sum_{j=1}^{m} v_{e,j}^t F(e_j^t) + \varepsilon, \tag{4.4.5} \]

where \( \varepsilon_i \) denotes the error term.

In unfavorable cases – especially if the confidence factor \( \tau \) of the outlier identification procedure is chosen too small–, inconsistencies in the form of different algebraic signs may occur between the extreme values according to the identification procedure on the one hand and according to the regression estimation on the other. Then the
relevant outlier dummy variables will automatically be deleted from the regression model (4.4.1) and the estimation is repeated.

5 Overview of the procedure

In what follows, the individual steps will be outlined by which the above-described elements are put together to form the BV4.1 procedure, depending on the procedural options defined by the user.

1. Setting of options concerning the extensions of the BV4.1 basic model (3.1.3) described in chapter 4:
   - Option “outlier component” $e_t$ (cf. (4.1.12)),
   - Option “calendar component” $c_t$ (cf. (4.2.6)) (with the variants “weekdays-based component” $c_t^w$ (cf. (4.2.3)), “working-day component” $c_t^d$ (cf. (4.2.4)) and “working-day component” $c_t^w$ (cf. (4.2.5))),
   - Option “level-shift component” $a_{s,t}$ (cf. (4.3.3)) and
   - Option “user component” $a_{u,t}$ (cf. (4.3.5)).

Note: To apply the options “user component” and “level-shift component”, the user has to supply BV4.1 with time series of the desired variables using specific input files. The same is true for the option “calendar component” if the user wants to use calendar variables other than the preinstalled ones, geared to the German public holiday regulation (cf. chapter 6).

2. If the option “outlier component” has been chosen: Identification of outliers according to chapter 4.1 and generation of the concrete outlier model.

3. If extension components have been chosen: Estimation of the relevant components ($\hat{e}_t$, $\hat{c}_t$, $\hat{a}_{s,t}$, and $\hat{a}_{u,t}$) (cf. chapter 4.4).

4. If both the option “outlier component” and the option “level-shift component” have been chosen:
   a) Performance of a provisional level-shift adjustment of the time series ($y_t = x_t - \hat{a}_{s,t}$).
   b) Repetition of the identification of outliers, which is now based on the provisionally level-shift adjusted time series $y_t$.
   c) Repetition of the estimation of the extension components, taking account of the revised model for the outlier component according to b) ($\hat{e}_t$, $\hat{c}_t$, $\hat{a}_{s,t}$, and $\hat{a}_{u,t}$).

5. If extension components have been chosen: Removal of all these components from the time series ($y'_t = x_t - \hat{e}_t - \hat{c}_t - \hat{a}_{s,t} - \hat{a}_{u,t}$ with $\hat{e}_t = \hat{e}_t$ or $\hat{e}_t'$, ...., $\hat{a}_{u,t} = \hat{a}_{u,t}$ or $\hat{a}_{u,t}'$).
6. Estimation of the trend-cycle component \( \hat{m}_t \) of the time series \( X_t \) according to chapter 3.3 (if extension components of the basic model have been chosen, the estimation is done on the basis of the time series \( y_t' \)).

7. Estimation of the seasonal component \( \hat{s}_t \) (cf. (3.4.2)) of the time series \( X_t \) according to chapter 3.4 (if extension components of the basic model have been chosen, the estimation is done on the basis of the time series adjusted not only for the trend-cycle component but also for the extension components, that is
\[
\hat{s}_t = \sum_{j=1}^n b_j (t)(y_t' - \hat{m}_t), \quad t = 1, \ldots, n.
\]

8. If the option “level-shift component” has been chosen, it will be added to the trend-cycle component \( \hat{m}_t \), so that the (final) trend-cycle component of BV4.1 is: \( \hat{m}_t' = \hat{m}_t + \hat{a}_{5,t} \).

9. Calculation of the seasonally adjusted series \( V_t = X_t - \hat{s}_t \). If the relevant components have been chosen, “seasonal adjustment” also means removal of the calendar and user components, that is:
\[
V_t = X_t - \hat{s}_t - \hat{r}_t - \hat{a}_{d,t}.
\]

6. **The PC software for BV4.1**

For the new version BV4.1 of the procedure, the Federal Statistical Office has developed a powerful and user-friendly software for Windows PCs (Windows NT 4.0/Windows 98 and later versions). For illustration, screenshots are given in the annex to demonstrate the process of a standard analysis.

The release candidate (RC 1) of the software may be obtained free of charge from the Federal Statistical Office at *bv4.1@destatis.de*.

The BV4.1 software has the following features:

- Performing BV4.1 analyses of monthly and quarterly time series. The maximum length of the series is limited to 360 observations. The minimum length is 60 observations for monthly series and 17 observations for quarterly series.

- Option to choose one of the 3 variants of calendar adjustment according to chapter 4.2. By default, the calendar regressors used are the deviations \( d_{t,j} - \overline{d}_{i,j} \), \( d_{A,t} - \overline{d}_{A,t} \) or \( d_{W,t} - \overline{d}_{W,t} \) based on the German public holiday situation (days which are public holidays only in part of Germany (Epiphany, Assumption Day, All Saints’ Day, Day of Repentance (since 1995) and Reformation Day) are not considered as holidays, Christmas Eve and New Year’s Eve together are counted as 1 holiday). Also, using the frequency distribution of the Easter date given in Ladiray and Quenneville (2001), the calculations of the mean values \( \overline{d}_{t,j} \), \( \overline{d}_{A,t} \) and \( \overline{d}_{W,t} \) are based on the entire cycle of the date of Easter (570000 years) and other related holidays.
If the user wants to base the calculation of the calendar component on a different than the German holiday situation described above, he may do so by replacing the content of specific input files. Of course, the user may also apply other calendar regressors by using the user component. (In this way, it is possible even to use entirely different (linear) concepts to define the calendar component.)

If the user applies the calendar option, he will also get information on the length-of-period effect. By default, the calendar adjustment is limited here to analysis periods within the period from 1949 to 2050.

- Possibility to include up to 15 user-defined explanatory variables.
- Graphical user interface (GUI).
- Support of the following input and output formats:
  - CSV,
  - EXCEL,
  - ACCESS,
  - SQL Server.
- Possibility of mass production of time series analyses. If linear dependencies occur between regressors of the extension components, the software will, if possible, automatically make changes as regards the selection of regressors to ensure smooth operation (e.g. by calculating a working-day component $c_t^d$ instead of a weekdays-based component $c_t^w$) and will inform the user about the modification.
- Possibility of performing successive analyses, that is of analyses where the analysis periods are successively extended by one additional observation. That option is particularly useful if revisions of analysis values are to be examined which are due to the filter sequence of BV4.1.
- Possibility of graphical presentation of analysis results.
- 2 variants to control the analyses:
  - metadata on the time series, series values and control parameters (options) are read from an input file,
  - metadata on the time series and the series values are read from an input file, while the control parameters are read from the GUI.
- For every analysed time series, a txt file is produced which contains comprehensive information on the analysis (information on the options chosen, the outliers identified, the regression coefficients and the test statistics on the outlier, calendar and user-defined regressors) and the complete results of the time series decomposition.

It should be noted here that the given numbers of degrees of freedom of the test statistics for assessing the significance of the coefficients and components estimated cannot be more than approximations. For simplification, and following the results of Lovell (1963), what is done here to take into account the trend-cycle and seasonal pre-adjustments by filter $F$ is just reducing the (nominal) numbers of degrees of freedom according to model (4.4.5) by 15 for monthly series and by 7 for quarterly series, because of the 4 parameters of the trend polynomials of the central filters (cf. chapter 3.3) and the 11 and 3 Fourier coefficients of the seasonal model (3.1.2) for monthly and quarterly series, respectively.

- 3 options for producing output files for the further processing of the analysis results:
  - Series-specific files with all components and the major adjustment results of an analysis.
- Series-specific files with a reduced number of components and adjustment results of an analysis.
- Component-specific files with all time series used in a program run or with the relevant analysis results for the trend-cycle components, the seasonally adjusted series or the calendar-adjusted series.

7 Literature


Annex: Example of applying the BV4.1 software

1 Front screen page

This is the front screen page of the BV4.1 application. The user can choose between an English and a German language version of the application by clicking on the button named "English" or "Deutsch", respectively. Clicking the "Quit" button closes the application.
2 Main menu screen page

Clicking one of the language buttons on the front screen page switches the graphical user interface (GUI) to the screen page of the main menu. The above screen shot shows the main menu with the defaults.

1 Possible formats of the file containing the time series to be analysed (⇒ section 3, point 2).

2 Field for the listing of the imported time series (⇒ section 4).

3 Options regarding the way to specify the options of the BV4.1 analyses:

- Using the option "Use of default options", all analyses are carried out with outlier adjustment (TAU=3.00 and M=24 for monthly time series, and M=8 for quarterly time series) as well as with calendar adjustment according to the trading day model. Level shifts and user-defined regressor variables won't be taken into account.

- Using the option "Use of input file options", all analyses are carried out according to the specification of the options given in the input file. Using this option, it is possible to analyses different time series with different BV4.1 options.
• Using the option "Inputation of options", the user has the possibility to specify the options by using the GUI (⇒ section 5). This option is useful to find out the best decomposition of a single time series.

4 Field to specify the name of the folder in which the output files of the program run will be stored.

5 Options regarding the output files of an analysis run:

• Option "integrated complete":
  For each of the analysed time series an output file is generated, which contains the complete set of the time series components as well as the important adjusted series.

• Option "integrated reduced":
  For each of the analysed time series an output file is generated, which contains a reduced set of components and adjusted series.

• Option "separated (reduced)"
  4 output files are generated, where all the time series analysed during the run, the corresponding trend-cycle components, the seasonally adjusted series and the calendar adjusted series are each stored separately.

The combination of different output variants is possible.

6 Option to carry out successive analyses.

7 Option to generate diagrams of the results (⇒ section 7).

To generate diagrams regarding results of past program runs it is sufficient to click on the "Continue" button (e.g. without reading in any time series). Then the GUI switches directly to the screen page "Generating graphs."

8 After filling in the main menu, the analysis run is started by clicking on the "Continue" button (⇒ section 6).
3 Specification of the decimal separator, the column separator for CSV files and the data path for the input and output files / Selection of the input file containing the time series to be analysed

Clicking on "Configuration" will open the "Options" window, where some default options may be changed.

For the decimal separator character, one can select between decimal point and decimal comma. For the column separator character for CSV files, one can select between semicolon and comma. The changes are valid only for the current BV4.1 run. If one restarts BV4.1, the defaults will be adopted again. The defaults for the German version are decimal comma and semicolon. For the English version, the defaults are decimal point and comma.
In the field “Data path” the path for the folder “Data” (containing the input and output files) can be determined.

To import time series first of all one has to specify the format of the file containing the desired series. A click on the corresponding format button (section 2, point 1) will open the “Load” window (for SQL Server: a window named “Selection”). All available files for the chosen format are shown in this window. By selecting the desired file and by clicking on the “Open” button, all existing time series in the file will be read in. If there are more than one sheet/table in an Excel- or Access- file another window will open to select the desired sheet/table. It is possible to read in time series from different sheets/tables by marking them.

4 Selection of the time series to be analysed and specification of the analysis periods

After selecting the input file containing the time series, all series from that file are listed in a table. Time series, which are to be analysed, have to be marked by a small hook in the column “Active”. As default, all imported time series are marked. For deactivating a time series, the user has to click on the corre-
sponding box in column "Active". By clicking on "Activate selected series", several time series can be simultaneously activated or deactivated. Before doing this, the desired time series have to be marked.

For the imported time series, the analysis periods are adopted from the input file. To change analysis periods the user has to click on the corresponding fields in the column "Analysis span". Then the dialog window named "Analysis span" will be opened where the changes can be made by using scroll down lists.

5 Manual specification of the analysis options by the GUI

Choosing the option "inputation of options" on the main menu screen page switches the GUI to the above screen page. It is possible now, to specify the analysis options independently of the specifications in the input file. Except for the presetting of the level shifts and the selection of the user-defined regressor variables, the specifications made will be adopted uniformly for all time series.

Field containing the options regarding the outlier adjustment.
2 Field containing the options regarding the calendar adjustment approaches.

3 Input fields for the presetting of level shifts.

4 Option to incorporate user-defined regressor variables in the time series analyses.

5 Field to specify the number of decimal places of the figures in the output files.

6 Result screen page

Clicking the "Continue" button on the main menu screen page opens the screen page shown above and the analyses will be carried out.

1 After the analyses are finished, the result for the first time series selected in the main menu is displayed. If so, the results for the other time series analysed during the analysis run can be displayed by clicking on the scroll down arrow in the box named "Time series".
The results may be printed out by clicking on the "Print" button.

2 For further processing of the analysis results, it is possible to export these results according to the output variants determined in the main menu. The export is carried out by clicking on the button of the desired file format.

3 Clicking the "Continue" button switches the GUI to the screen page for generating graphs (⇒ section 7).

7 Generating graphs

1 The box named "Code name of the desired analysis results" has to contain the code name of that result (⇒ section 2, point 4) which is to be plotted. By clicking on the scroll down arrow, a list of the currently held analysis results will be displayed.

2 These boxes have to contain the short names (as defined in the input file) of that time series for which graphs are desired. Per default, the short name of the first analysed series of the analysis run is dis-
played. Changes may be carried out by scroll down lists which contain the short names of all time series belonging to the chosen code name (⇒ point 1).

3 These boxes serve for choosing the components of the selected time series decompositions, which are to be plotted. Again, the selections are carried out by scroll down lists which open by clicking on the scroll down arrows of the boxes.

It is possible to plot up to 4 components.

4 If successive analyses were carried out, this option is activated, too. It is now possible to plot graphs, showing the results of successive analyses for the trend-cycle component, the seasonally adjusted series and the calendar adjusted series.

5 By clicking on the "Continue" button, the generated plot will be displayed (⇒ section 8).

8 Diagram screen page
In this field of the screen page, the ranges of the axes of the diagram shown can be changed. To change the value axis the desired numerical values have to be entered in the corresponding boxes. To change the time axis the beginning and the end of the intended interval have to be selected from the corresponding scroll down menus. Changes will be adopted by clicking on the “Show” button.

The diagram can be printed out by clicking on the “Print” button. By clicking on the “Back” button, the starting screen page for the generation of graphs will open again.