Seasonal adjustment by frequency determined filter procedures

For the purposes of monthly time series analysis the original data are assumed to be additively composed of a trend-cycle, a seasonal, a calendar-effect, and a residual component, the last of which may include a few extreme values. The first mentioned two components are estimated by moving filter applications derived from approximating functions by a regression approach. The appropriateness of the filters is judged and controlled by their transforms into the frequency domain. While model 3 of this procedure was designed series-specifically, the improved model 4 provides adaptive and end-stable results by a uniform set of filters to both components where general optimization aspects had already been included. The mathematical models used, their specifications and development, and some features of the new model like practical simplicity and definiteness are indicated. A hint is given to the integrated identification of extreme values and estimation of calendar-effects which in future will both be performed before the eventual seasonal adjustment.

1. Introduction

We have to distinguish between several *versions* of seasonal adjustment by Frequency Determined Filter Procedures (FDF) used in the Federal Statistical Office of the Federal Republic of Germany. Since 1975 version 3 has been applied for decomposition of monthly time series. This version includes filter sequences for estimation which are designed series-specifically [2]. The improved version 4, which has just been introduced, tries a uniform approach to every statistical series with monthly data [3] (or resp. quarterly data). In the following short description reference will be made to FDF 3 or FDF 4 where necessary.

2. Models of analysis

The originally observed values of a (monthly) statistical time series are assumed to be *additively composed* of a smooth component *T* (for trend-cycle movement), a seasonal components *S* (for seasonal movement), and a residual component *R* (for short-term irregular movements). The residual component may still include extreme values or outliers which have to be eliminated before estimating *S* (in FDF 3 this is performed through conditional prediction by a Wiener filter). In some sets of series the values are affected by irregularities of calendar; week-day effects are separable from *R* by regression estimation (FDF 3).

The estimation of *T* and *S*, respectively, is performed in appropriately defined time intervals moving over the whole time series where the possible course of the component values is approximated by a suitable set of mathematical functions. Here, methods of *regression* estimation are applied in FDF 3 whereas FDF 4 uses weighted least squares, their weights having a two-part linear distribution with (highest) value 1 at the month of estimation and smallest values at both the ends of the estimation interval.

The functions of approximation are called *basic functions* [1,4]. In order to estimate *S* of a monthly series they are composed of the 11 trigonometric functions $\cos j(\pi/6)t$ ($j = 1,2,...,6$) and $\sin j(\pi/6)t$ ($j = 1,2,...,5$) where *t* denotes the current number of the month. Ordinary polynomials up to the third degree are used to estimate *T*, some of which might be combined with up to two trigonometric functions of different wave lengths of more than two years (FDF 3). Instead, FDF 4 is confined to a third-degree polynomial which proves flexible enough because of weighting the least sqares with window effect.

The respective partial intervals of the time series subjected to the approximation are called *reference ranges*. With monthly data in FDF 3 they may vary between 36 and 59 months for seasonal estimation or between 25 and 39 months for the estimation of *T*. In FDF 4 these figures are 38 to 51 and 25 to 30 months, respectively.

The estimation of a component is not performed for all the months of the relevant reference range but only for the best suited month within this range and its relative position is called the *estimation point*. For the estimation of the following month, the whole reference range is moved by one month in order to get the estimation point unchanged. This principle of moving estimation can be realized only for the middle part of a time series and has to be modified when the reference range has reached the end of the given time series. In this situation several new (less favourable) filters must be developed and applied one after another, where the estimation point shows a stepwise approach to the end of the reference range (and of the time series).

By means of regression methods, a set of elements or weighting factors can be computed for any set of basic functions and any length of reference range. The original values in the reference range are multiplied individually by these factors before aggregation in order to get the result of the relevant component at the estimation point. Each set of weighting factors is called a *filter*. The separate estimation of *T* or *S* emerges unbiased only if the sum of the filter elements is equal to 1 or 0, respectively, which is always automatically realized. While estimating successively *T* and *S*, the capability of the regression procedure is utilized to take into account the potential existence of the respective other (non-estimated) component in the sense of segregation. Therefore, potential biasing interactions are considerably reduced or even avoided at all by this approach.

The appropriateness of filters is judged by their *transfer functions* TF (i.e. by their Fourier transform into the frequency domain with λ varying from 0 to π; in squared form over the real and the imaginary part). The TF of a filter for estimating *T* should possess values near 1 (i.e. unchanged pass-through of the respective waves) for the lower frequency band ($\lambda < \pi/6$) and values near 0 (i.e. extinction) for the remaining frequency regions (see Figure 1). The TF of a filter for estimating *S* should have the value 1 at each of the six seasonal frequencies $\lambda = j\pi/6$, j = 1,2,...,6, and values near 1 in some narrow neighbourhood of these, and values near 0 for all the remaining frequency regions (see Figure 2). The last mentioned condition TF = 1 at $\lambda = j\pi/6$ is attained if the estimation points within the reference range in estimating monthly seasonals are located at the 12th or 24th or 36th ... month. The TFS of every possible filter show certain deviations of different extent from the ideal form for most of the remaining frequencies where $\lambda \neq 0$ and $\lambda \neq j\pi/6$. The combined effect of the estimation of first *T* and then *S* is decisive for the validity of seasonal adjustment.

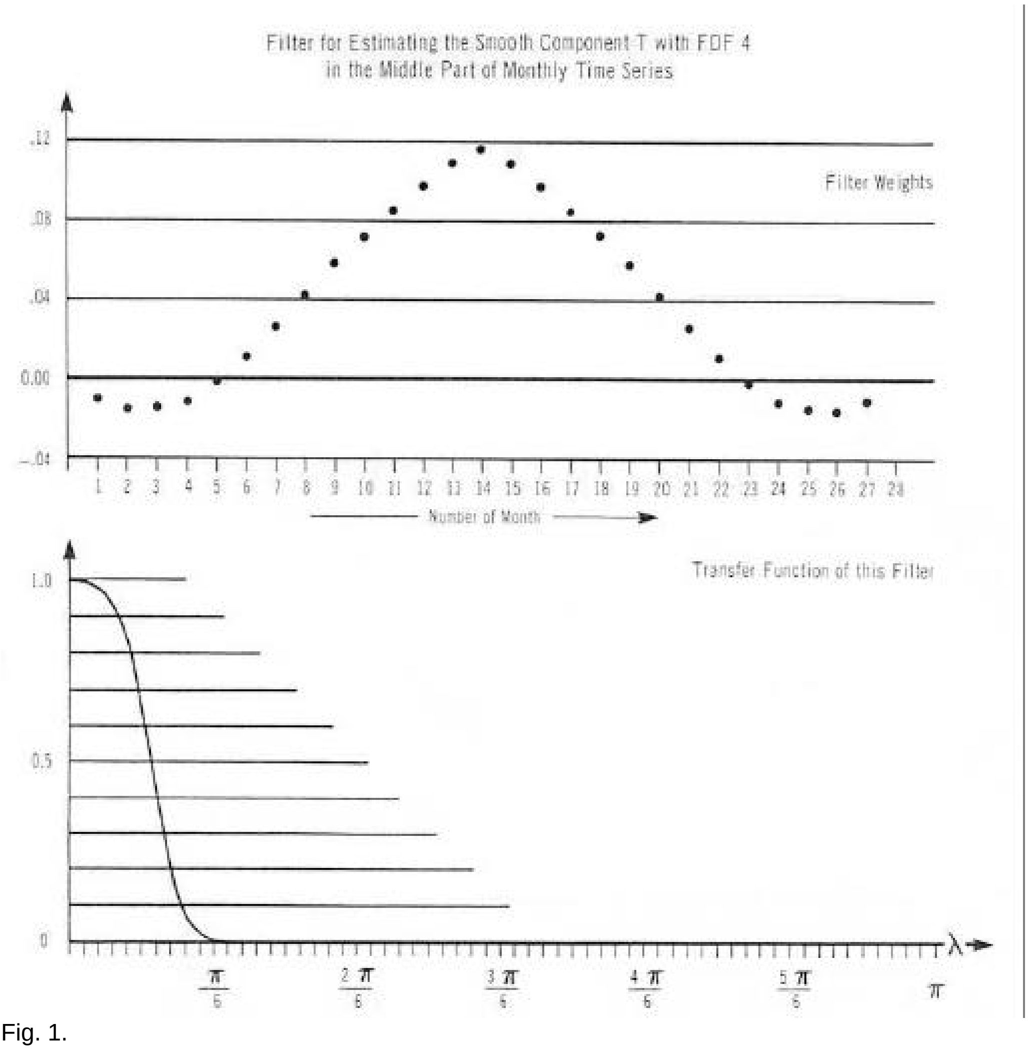

Fig. 1.

Time series analysis was performed in FDF 3 along the following *estimation process*: first, possibly existent extreme values are identified by means of short-term projections (conditional expectations) and eliminated. Then *T* is estimated by the moving application of filters and is subtracted from the original values of the series. Based on these figures *S* is estimated by the use of moving filters. The seasonally adjusted series is then calculated by eliminating *S* from the original data. Out of *R* (to be presented as a difference using *T* and *S*), a component of week-day effects is deduced by moving regression estimation and included into the seasonal adjustment. The FDF 4 differs from FDF 3 in starring with a combined extreme value and week-day effect identification related to figures which on both sides of the regression equation are transformed to unified month lengths as well as provisionally detrended and deseasonalized [3].

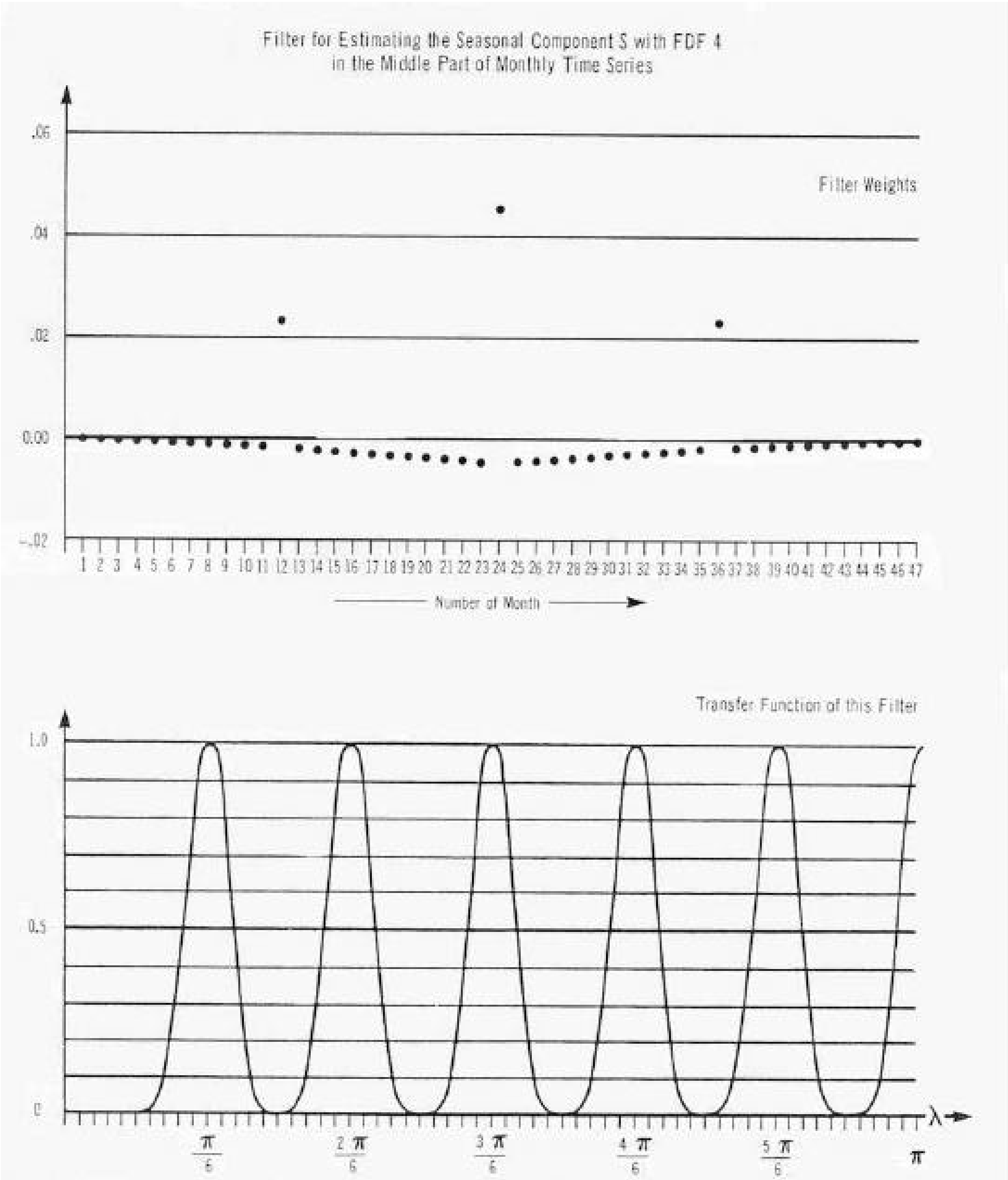

Fig. 2.

3. Specifications of the models

The essential methodological problem was to select filters which would enable the user to reliably analyse the series. Looking at the TFs of the combined effect of the estimation of both the components, several decisions in FDF 3 could be made in advance so that the multitude of possible parameter combinations would be reduced [2]. So it was seen, for instance, that the most suitable reference ranges for seasonal estimates in the middle part of a series should cover just 47 or 59 months (the corresponding estimation points being fixed at the 24th or 36th month, respectively). In addition it was found that a variation of the wave lengths of the superimposed trigonometric functions included in the set of basic functions for *T* had very little effect, so that those wave lengths could be prefixed to 36 and/or 60 months.

Another limitation of the eventual set of filters is provided by the fact that reference ranges, having reached the end of the series after moving over the middle part, might vary their lengths during the following estimation steps but should cover always the last part of the series. In the course of this procedure, the estimation point approaches the end of the reference range more and more so that the relation between the length of the reference range and the position of the estimation point is currently fixed and only one of them is still open for optimization. This principle, however, cannot be realized perfectly in seasonal estimation because the estimation point is not permitted to vary continuously but works without bias only at multiples of the basic period of 12 months (for the first 11 months of a series, biased results seemed tolerable). When, for example, a reference range of 59 months with estimation points 36 in FDF 3 has reached the end of the series, it can at first be shortened several times by one month each (leaving the estimation point unchanged). After an appropriate number of repetitions, the estimation point must be shifted from 36 to 48 consequently the length of the reference range is enlarged abruptly, for example, from 43 to 54 (or from 40 to 51) months before its step-by-step shortening can be continued. Seasonal filters of length 47 will usually need two shifts of their estimation point (from the first estimation point 24 to points 36 and 48).

There were no further definite preferences in the class of still disposable estimation filters because they differ mutually by the fact that rather favourable values of the TF in certain frequency regions are often accompanied by rather unfavourable values in other regions. Here, FDF 3 took reference to the fact that various time series themselves, subjected to seasonal adjustment, have different properties which can be made visible very distinctly by their spectral representations. The differently distributed spectral intensities of the individual time series let expect reasonably good results of the analysis by an appropriate allocation of filters to series.

The FDF 3 procedure was series-specific as it minimizes the effect of estimation bias of the desired filters on the relevant time series. That combination of filters for *T* and *S*, for which the average over the whole frequency range (except the six frequency lines at the seasonal part of the spectrum) of the squared deviations between the spectral intensities of the original series and those of the seasonally adjusted series is minimum, was considered as optimum. Using a search algorithm, the biasing variances of the TFs of all filters to be taken into consideration for a special time series position were calculated, and the filter which gave the minimum value was chosen for the respective series.

Filter allocation in FDF 3 was performed according to certain rules. With respect to the particular importance of reliable results at the end of a series, the determination of filters started with the last month of each time series. The optimum parameter constellation found there (with regard to the kind of basic functions for *T*) was taken unchanged in all further filter determinations for that series in order to diminish the amount of search. Handling the second last till the sixth last month of a series, the lengths of reference ranges for *T* were observed not to vary too much; if necessary, near-optimum solutions were taken. The bridge between the sixth last month and the separately determined optimum solution for the middle part of the series was fixed in such a way as to get a steady connexion. In order to keep the number of such filter combinations within acceptable bounds, attempts were made to use the same combination for several time series of similar structure. By proceeding in this way, 47 filter combinations were developed for the more than 300 monthly analysed and published time series of this office.

The parameters in FDF 4 were fixed aiming at a uniform set of filters for all monthly time series (and another set for quarterly series) [3]. For estimating T in the middle part of a series, a symmetric filter with parameters 27 [14] was chosen, where the first figure refers to the length of the reference range and the second one (in brackets) to the relative position of the estimation point. Near the end of a series the next (unsymmetrical) filters 28(16), 29(18), 30(20), 30(21), 29(21), 28(21), 26(20) were determined so as to minimize changes in the real part of TF in the lower frequency band against the neighbouring ones (based always on polynomials of the third degree and on weighted least squares). Because of the increasing necessity of reducing the real-value part of TF at the frequencies around the ¾-year-wave while further approaching the end of a series (and also because of logical considerations), this kind of filter has been superimposed more and more by filters based on first-degree polynomials, the lengths of which are steadily decreasing from 25 to 20 months. This complete set of filters proved to yield flexible as well as end-stable results and is used in reverse order to estimate T for the first 13 months of a series.

The set of filters for estimating S was established in a similar manner in FDF 4. For the middle part of a series the symmetric filter 47(24) was chosen, being mainly based on polynomials of the third degree in the part of T and on weighted least squares. While approaching the end of a series this filter was shortened until 38(24) and afterwards superimposed by 49(36) down to 36(36) with slightly increasing participation weights. From 27(24) to 24(24) and 39(36) to 36(36) these filters were further stabilized by a component from 51(48) to 48(48), respectively. All these filters proved to be very flexible, comprising fairly wide frequency bands of TF at the six seasonals.

4. Main feature of the new model

There is some methodological guarantee that the FDF procedure does separate the smooth component *T* from the seasonal component *S* clearly without disturbing interactions (in contrast to the Census X 11-procedure where the 12-months-wave seems to be not completely separated from *T*).

In addition, FDF 4 tries to separate a calendar-specified component (and extreme values) out of a residual component *R* not being influenced by the former estimation of *T* and *S*. Thus, a comprehensive system of simultaneous estimation will be attained.

The application of FDF 4 is particularly simple by using only one set of universal filters for the estimation of T and S, respectively, without any iteration and without any reference to peculiarities of the various time series. Investigations in the search for optimization are already included in the filter construction and need not be repeated again for every time series.

The results of time series analysis by FDF 4 are reliable, definite, and comparable between all users of the procedure and of its results.

The aggregation of partial time series results from FDF 4 is directly possible and correct (because of additivity and definiteness) without further mutual adjustment. If extreme values occur, some precautions must be envisaged.

Later revisions of preliminary results at the end of a time series are considered to be necessary in view of the generally better properties of the so-called transversal filters (for estimation in the middle part of a series) in contrast to the extremely unsymmetrical recursive filters having been applied to the last months of a series.

References

[1] M. Nourney, Methode der Zeitreihenanalyse, Wirtschaft und Statistik 1 (1973) 11–17 (relates to FDF 2).
[2] M. Nourney, Weiterentwicklung des Verfahrens der Zeitreihenanalyse, Wirtschaft und Statistik 2 (1975) 96–101 (relates to FDF 3).
[3] M. Nourney, Umstellung der Zeitreihenanalyse, Wirtschaft und Statistik 11 (1983) 841–852 (relates to FDF 4).
[4] B. Nullau, S. Heiler, P. Wäsch, B. Meisner and D. Filip, Das Berliner Verfahren,ein Beitrag zur Zeitreihen-analyse, DIW-Beiträge zur Strukturforschung, Heft 7, Duncker & Humblot, Berlin, 1969.