A NEW TECHNIQUE OF SEASONAL ADJUSTMENT WITH AN APPLICATION TO EPIDEMIOLOGIC DATA

Paolo d'Alessandro

Istituto di Elettrotecnica
Facoltà di Ingegneria
Università di L'Aquila
Monteluco, 67100 L'Aquila
(ITALY)

Eugenia De Luca d'Alessandro

Istituto di Igiene
Università di Roma
P.le Moro, 5
00185 Roma
(ITALY)

Keywords: Time series, Seasonal adjustment , Convex analysis.

ABSTRACT

A new method of seasonal adjustment is introduced. This is based on the technique optimal smoothing, which has been implemented via a fast projection algorithm. As example of application, numerical results are given for some epidemiologic data.

1. INTRODUCTION

Various methods are available to face the problem of seasonal adjustment. A survey may be found in [1] . All most common methods lean on the use of moving averages to smooth the time series, often repeated in several stages.

The method presented in this paper makes instead use of optimal smoothing technique introduced in [2] , and implementated by means of the fast projection method described in

[3].

The procedure consists of two main steps, one to separate the irregular component the other to separate the seasonal component. Other variants of the method might well be devised, for example to cope with different data structures.

To give an example of the performance of the procedure a numerical example for some epidemiologic data is given in the last section. Of course the method applies to data of any other kind as well. The particular example choosen is motivated by the fact that it was the first at hand, and from the interest of studying seasonal factors of epidemies, particularly for modeling purposes.

Further tests, particularly in the economic field, and comparisons with other methods is in the plan for subsequent research, along with other applications of the optimal smoothing technique. In the couse of this activity other variants of the method might be experimented.

2. OPTIMAL SMOOTHING

As anticipated earlier the optimal smoothing technique is the basic ingredient of the method.

The purpose of this section is to briefly recall such technique, whereas details may be found in [2] .

Consider a signal or time series, which is represented as an element s of the Euclidean space R^N, where N is the number of samples. Assume it is required to smooth such signal. Instead of adopting an heuristic approach like the popular mooving average technique, the problem may be addressed with a rigorous optimization formulation as follows. Define on R^N a smoothness (or variation) functional, say p, which will be assumed to be a norm (or pseudonorm) and suppose that p(s) must be reduced to r= c p(s) for some $c \in (0,1)$. Then the natural solution is to find $\hat{s}$ so as to minimize $\|s-\hat{s}\|$, subject to $\hat{s} \in S_p^r$, where S_p^r is the closed sphere of p about the origin and with radius r.

It is well-known that the solution exists, is unique and belongs to the boundary of S_p^r (so that actually $p(\hat{s}) = c\, p(s)$), in view of some basic results of convex analysis. Thus the solution defines a projection operator P of R^N onto S_p^r.

It might be useful to insist on the motivation of this approach. If any other technique for filtering the signal were adopted, then it would anyway result in some smoother signal. With the present method one can pick (adjusting c) the signal having the same degree of smoothness, that is the most faithful reproduction of the starting signal, thus obtaining a surely better result.

A major advantage of this "optimal smoothing" approach is just that it leaves a free parameter, namely c, to adjust the level of smoothness of the filtered signal $\hat{s}$, until it is best suited to some relevant specific purposes. This is essential in applications because it allows to tune the filter to satisfy suitable optimality criteria. The method presented in the next section is an example of this concept.

Another important degree of freedom is obviously the choice of the functional p. This point is not discussed here, and poses largely open question related, in the first place, to the performance (both in deterministic and stochastic terms), of the filter. Here it is choosen the norm w defined by

$$w(f) = |f(1)| + \sum_{i=1}^{N-1} |f(i+1) - f(i)| \qquad f \in R^N$$

The motivation for this choise leans mainly on technical convenience, for it allows to use the fast projection technique described in [3], and on the fact that in practice it looks quite satisfactory. Notice that the use of efficient numerical algorithms is essential in the present case, where hundreds of runs of the filter may be required for each run of seasonal adjustment.

3. THE SEASONAL ADJUSTING METHOD

To introduce the method, the time series considered is more conveniently represented by the usual model (see e.g. [1])

$$O_{ij} = T_{ij} + S_{ij} + R_{ij}$$

where O_{ij} is the original time series, T_{ij} S_{ij} and R_{ij} are the trend-cycle, seasonal and random or irregular component respectively, the index j varies over months (i.e. $j=1,...12$) and i over the years. In other words data are arranged in a matrix, whose rows yield values pertaining to each fixed year and columns values pertaining to each fixed month.

Since the seasonal component should be, by its own nature, fairly constant for each fixed month over the years, it appears reasonable to filter first the time series represented by the columns in order to remove the irregular component. Each of these 12 steps is optimized adjusting the filter in such a way to make the residual (i.e. the difference between the filtered signal and the original signal) as incorrelated as possible. As usual, to test incorrelation, use is made of the so-called Von Neumann ratio (see e.g. [1]). In other words the smoothing parameter is adjusted to minimize the absolute value of the difference between Von Neumann ratio relative to the residual signal and the optimum Von Neumann ratio (that is 2.1).

The time series obtained in this way is assumed to represent the component $T_{ij} + S_{ij}$, whereas the residue gives the component R_{ij}.

To remove the seasonal component S_{ij} the filter is applied again, this time on each year. However after some experimentation to improve the performance of the filter, it was found out that the best results were obtained using a window technique, in the sense that the year under consideration is viewed as a window on a time series starting six months ahead and ending six months later than the given year. After filtering the whole time series only data pertaining to the year are of course considered. In this case the criterium for choosing the parameter c is essentially that of performing the minimum amount of smoothing, compatible with the condition that the filtered signal, that should give the trend-cycle time series T_{ij}, is free of seasonal components. This is easily translated in an optimality condition (which for semplicity is illustrated in the case of higher values during the good season; any other case requires obvious adaptations only). For as c increases, if a seasonal factor is present within the year, a more and more simplified seasonal patter is observed in the filtered signal, in the sense that the number of points, at which the signal varies, decreases. This phenomenon is fairly clear from the theory of fast projections [3] and to be more precise points of variation correspond, in the case of functional w, to vertexes of the polytopes onto which the projection is made. From a certain value of c on it is not possible anymore to individuate two months, one in spring the other in autumn in such a way that the average value in the interval from the first to the second month is greater than those from January to the first and from the second to december. It is thus choosen as optimum the minimum value of c such that this condition holds.

In respect to the choice of c in each step of the method, it is useful to recall that the filtered signal depends continuously on c. In practice exploiting this fact a cut and try procedure has been followed to find each value of c. In the numerical example that follows, it was decided to use for c a precision of three significant digits after decimal point, since the effort for a greater resolution is not worth the small refinement of the result.

Thanks to the fast projection even this cut and try procedure is rather handy. However an automatical procedure is fairly feasible, and software oriented to this purpose might well be developed.

It might be worth noticing that in the specific example the Von Neuman ratios obtained turned out to be optimum in all cases without exceptions.

4. NUMERICAL RESULTS

The numerical results presented regard the time series of cases/100.000 inhabitants of typhoid fever in Italy in the years from 1960 to 1976. The data were derived from those on cases and population issued by ISTAT (which is the National Istitute of Statistics) [5]. For semplicity they were approximated to the third significant digit.

Original data are given in Table 1, whereas the three component produced by the method are given in Tables 2, 3 and 4. Table 5 gives Von Neuman ratios of irregular components of each month and the corresponding variation cuts used to separate such components. Finally table 6 gives the variation cuts used to separate the trend-cycle component. In this respect it must be adviced that convention has been adopted of computing the percent of variation on the only variation part of the functional w. Therefore values greater than 100% are allowed.

TABLE 1 - ORIGINAL DATA

	JAN.	FEB.	MARCH	APRIL	MAY	JUNE
1960	2.74000	1.67000	2.29000	2.06000	3.37000	4.54000
1961	2.17000	1.76000	1.62000	1.94000	2.47000	2.78000
1962	2.22000	1.71000	2.01000	2.54000	3.91000	4.14000
1963	2.10000	1.46000	1.68000	1.81000	3.39000	3.16000
1964	1.83000	1.01000	1.14000	1.80000	1.95000	1.95000
1965	2.15000	1.36000	1.25000	1.49000	2.37000	2.44000
1966	1.14000	1.72000	1.65000	1.48000	1.88000	1.89000
1967	6.70000E-1	1.10000	1.42000	1.40000	1.57000	1.63000
1968	1.55000	1.59000	1.53000	2.11000	3.88000	2.29000
1969	2.18000	1.78000	1.51000	1.61000	1.90000	2.18000
1970	2.13000	1.82000	1.21000	1.59000	1.72000	2.12000
1971	1.95000	1.64000	1.24000	1.47000	1.87000	1.90000
1972	1.58000	1.12000	1.43000	1.44000	1.58000	1.86000
1973	1.99000	1.38000	1.44000	1.47000	2.03000	1.97000
1974	7.00000E-1	6.90000E-1	7.60000E-1	8.30000E-1	8.80000E-1	1.11000
1975	9.90000E-1	8.50000E-1	9.80000E-1	1.27000	1.70000	1.54000
1976	1.82000	1.27000	1.30000	1.02000	1.22000	1.43000

TABLE 1 - CONTINUED

	JULY	AUG.	SEPT.	OCT.	NOV.	DEC.
1960	3.70000	3.60000	3.80000	2.48000	2.20000	1.66000
1961	2.85000	2.71000	3.42000	2.68000	1.68000	8.90000E-1
1962	4.33000	4.39000	4.42000	2.97000	2.52000	2.04000
1963	3.27000	3.15000	3.20000	2.38000	1.71000	1.36000
1964	2.47000	2.62000	2.35000	1.85000	1.41000	1.17000
1965	2.98000	2.96000	2.54000	2.14000	1.90000	1.73000
1966	2.20000	1.99000	2.64000	2.45000	1.69000	1.82000
1967	1.96000	2.10000	2.42000	2.17000	1.64000	1.83000
1968	2.28000	2.58000	2.23000	1.75000	1.66000	1.65000
1969	2.10000	2.46000	2.37000	1.94000	1.69000	1.18000
1970	2.82000	2.95000	2.76000	2.00000	1.58000	1.40000
1971	2.21000	2.83000	2.41000	1.63000	1.25000	1.01000
1972	1.91000	2.08000	1.92000	1.62000	1.49000	7.50000E-1
1973	2.24000	2.67000	3.01000	1.30000	7.50000E-1	1.03000
1974	1.44000	1.99000	1.90000	1.27000	8.10000E-1	5.50000E-1
1975	1.97000	2.11000	3.57000	2.76000	1.80000	1.41000
1976	1.49000	1.62000	1.66000	1.17000	7.50000E-1	7.20000E-1

TABLE 2 - IRREGULAR COMPONENT

	JAN.	FEB.	MARCH	APRIL	MAY	JUNE
1960	4.21556E-1	1.20787E-1	4.88460E-1	2.14720E-1	8.44991E-1	1.38267
1961	-2.49999E-2	2.10787E-1	-1.81540E-1	9.47201E-2	-5.50086E-2	-3.77334E-1
1962	2.50001E-2	1.60787E-1	2.08460E-1	6.94720E-1	1.38499	9.82666E-1
1963	2.38419E-7	-1.19209E-7	-1.19209E-7	2.38419E-7	8.64991E-1	2.66600E-3
1964	-1.60000E-1	-4.10591E-1	-2.76667E-1	3.57628E-7	-3.80000E-1	-2.45000E-1
1965	1.60000E-1	-6.05907E-2	-1.66667E-1	-1.30000E-1	4.00002E-2	2.45000E-1
1966	1.19209E-7	2.99409E-1	2.33333E-1	-1.40000E-1	-4.50000E-1	-1.32000E-1
1967	-4.21555E-1	-3.20591E-1	3.33309E-3	-2.20000E-1	-7.60000E-1	-3.92000E-1
1968	1.19209E-7	5.59020E-3	1.13333E-1	4.90000E-1	1.55000	2.68000E-1
1969	2.33852E-1	1.95590E-1	9.33331E-2	2.38419E-7	-2.38419E-7	1.58000
1970	1.83852E-1	2.35590E-1	-1.20000E-1	3.57628E-7	-7.99999E-2	9.80000E-2
1971	3.85225E-3	5.55903E-2	-9.00002E-2	3.57628E-7	7.00002E-2	-1.00001E-2
1972	-2.05000E-1	-1.30000E-1	9.99999E-2	-1.49996E-2	-2.20000E-1	-5.00001E-2
1973	2.05000E-1	1.30000E-1	1.10000E-1	1.50005E-2	2.30000E-1	5.99999E-2
1974	-3.55778E-1	-3.26181E-1	-3.39230E-1	-3.77360E-1	-8.93327E-1	-5.81778E-1
1975	-6.57778E-2	-1.66181E-1	-1.19230E-1	6.26402E-2	-7.33272E-2	-1.51778E-1
1976	2.10778E-1	2.46181E-1	2.00770E-1	-1.87360E-1	-5.53327E-1	-2.61778E-1

TABLE 2 - CONTINUED

	JULY	AUG.	SEPT.	OCT.	NOV.	DEC.
1960	7.03237E-1	6.26175E-1	8.61311E-1	1.52730E-1	5.98209E-1	3.85000E-1
1961	-1.46763E-1	-2.63825E-1	4.8 1311E-1	3.52730E-1	7.82093E-2	-3.85000E-1
1962	1.33324	1.41618	1.48131	6.42730E-1	9.18209E-1	7.29502E-1
1963	2.73237E-1	1.76175E-1	2.61311E-1	1.05273	1.08209E-1	4.95019E-2
1964	-2.55001E-1	-1.70000E-1	-2.13278E-1	-3.02500E-1	-1.91791E-1	-1.40498E-1
1965	2.54999E-1	1.70000E-1	-2.32782E-2	-1.25000E-2	2.98209E-1	4.19502E-1
1966	-7.20010E-2	-4.95000E-1	7.67217E-2	2.97500E-1	8.82093E-2	5.09502E-2
1967	-3.12001E-1	-3.85000E-1	-1.43278E-1	1.75002E-2	3.82093E-2	5.19502E-1
1968	7.99894E-3	9.49998E-2	-3.33278E-1	-1.46666E-1	5.82094E-2	3.39502E-1
1969	-1.72001E-1	-2.50001E-2	-1.93278E-1	4.33335E-2	8.82093E-2	-1.10000E-1
1970	5.47999E-1	4.65000E-1	1.96722E-1	1.03334E-1	1.19209E-7	1.10000E-1
1971	-7.15256E-7	3.45000E-1	-1.53278E-1	-1.788410E-1	-1.20000E-1	-1.03876E-1
1972	-1.65000E-1	-2.95000E-1	-6.43278E-1	-1.88410E-1	1.20000E-1	-3.63876E-1
1973	1.64999E-1	2.95000E-1	4.46722E-1	-5.08410E-1	-5.60261E-1	-8.38758E-2
1974	-5.53826E-1	-2.42450E-1	-6.63278E-1	-5.38410E-1	-5.00261E-1	-5.63876E-1
1975	-2.38255E-2	-1.22450E-1	1.00672	9.51590E-1	5.79739E-1	2.96124E-1
1976	-5.03826E-1	-6.12450E-1	-9.03278E-1	-6.38410E-1	-5.60261E-1	-3.93876E-1

TABLE 3 - TREND-CYCLE COMPONENT

	JAN.	FEB.	MARCH	APRIL	MAY	JUNE
1960	1.87862	1.87622	1.87862	1.87862	2.20694	2.20694
1961	2.13625	2.13625	2.13625	2.13625	2.20061	2.20061
1962	2.13625	2.13625	2.13625	2.13625	2.17761	2.17761
1963	2.10432	2.10432	2.10432	2.10432	2.10432	2.10432
1964	2.02461	2.02461	2.02461	2.02461	2.02461	2.02461
1965	1.95720	1.95720	1.95720	1.95720	1.95720	1.95720
1966	1.86119	1.86119	1.86119	1.86119	1.86119	1.86119
1967	1.79339	1.79339	1.79339	1.79339	1.84522	1.84522
1968	1.85367	1.85367	1.85367	1.85367	1.85367	1.85367
1969	1.86721	1.86721	1.86721	1.86721	1.86721	1.86721
1970	1.85043	1.85043	1.85043	1.85043	1.85043	1.85043
1971	1.83353	1.83353	1.83353	1.83353	1.83353	1.83353
1972	1.73141	1.73141	1.73141	1.73141	1.73141	1.73141
1973	1.70848	1.70848	1.70848	1.70848	1.70848	1.70848
1974	1.55788	1.55788	1.55788	1.55788	1.55788	1.55788
1975	1.53553	1.53553	1.53553	1.53553	1.53553	1.53553
1976	1.54501	1.54501	1.54501	1.54501	1.68725	1.68725

TABLE 3 - CONTINUED

	JULY	AUG.	SEPT.	OCT.	NOV.	DEC.
1960	2.20694	2.20694	2.20694	2.20694	2.20694	2.20694
1961	2.20061	2.20061	2.20061	2.20061	2.20061	2.20061
1962	2.17761	2.17761	2.17761	2.17761	2.17761	2.17761
1963	2.10432	2.10432	2.10432	2.10432	2.01838	2.01838
1964	2.02461	2.02461	2.02461	2.02461	1.89861	1.89861
1965	1.95720	1.95720	1.95720	1.95720	1.79712	1.79712
1966	1.86119	1.86119	1.86119	1.86119	1.76847	1.76847
1967	1.84522	1.84522	1.84522	1.84522	1.84522	1.84522
1968	1.85367	1.85367	1.85367	1.85367	1.82809	1.82809
1969	1.86721	1.86721	1.86721	1.86721	1.76743	1.76743
1970	1.85043	1.85043	1.85043	1.85043	1.73828	1.73828
1971	1.83353	1.83353	1.83353	1.77236	1.61664	1.61664
1972	1.73141	1.73141	1.73141	1.73141	1.63471	1.63471
1973	1.70848	1.70848	1.70848	1.70848	1.42407	1.42407
1974	1.55788	1.55788	1.55788	1.55788	1.45156	1.45156
1975	1.53553	1.53553	1.53553	1.53553	1.53006	1.53006
1976	1.68725	1.68725	1.68725	1.68725	1.68725	1.68725

TABLE 4 - SEASONAL COMPONENT

	JAN.	FEB.	MARCH	APRIL	MAY	JUNE
1960	4.39823E-1	-3.29408E-1	-7.70774E-2	-3.33374E-2	3.18074E-1	9.50394E-1
1961	5.87492E-2	-5.87041E-1	-3.34711E-1	-2.90971E-1	3.24403E-1	9.56723E-1
1962	5.87492E-2	-5.87041E-1	-3.34711E-1	-2.90971E-1	3.47405E-1	9.79725E-1
1963	-4.31919E-3	-6.44319E-1	-4.24319E-1	-2.94319E-1	4.20691E-1	1.05301
1964	-3.46100E-2	-6.04020E-1	-6.07940E-1	-2.24610E-1	3.05390E-1	1.70390E-1
1965	3.27982E-2	-5.36612E-1	-5.40532E-1	-3.37202E-1	3.72798E-1	2.37798E-1
1966	-7.21194E-1	-4.40604E-1	-4.44524E-1	-2.41194E-1	4.63306E-1	1.60806E-1
1967	-7.01829E-1	-3.72799E-1	-3.76719E-1	-1.73389E-1	4.84782E-1	1.76782E-1
1968	-3.03669E-1	-2.69259E-1	-4.36999E-1	-2.33669E-1	4.76330E-1	1.68330E-1
1969	7.89442E-2	-2.82796E-1	-4.50536E-1	-2.57026E-1	3.27941E-2	1.54794E-1
1970	9.57191E-2	-2.66021E-1	-5.20431E-1	-2.60431E-1	-5.04310E-2	1.71560E-1
1971	1.12620E-1	-2.49121E-1	-5.03530E-1	-3.63531E-1	-3.35306E-2	7.64695E-2
1972	5.35866E-2	-4.81413E-1	-4.01413E-1	-2.76413E-1	6.85865E-2	1.78587E-1
1973	7.65218E-2	-4.58478E-1	-3.78478E-1	-2.53478E-1	9.15216E-2	2.01522E-1
1974	-5.02097E-1	-5.41698E-1	-4.58647E-1	-3.50518E-1	2.15453E-1	1.33903E-1
1975	-4.79745E-1	-5.19345E-1	-4.36295E-1	-3.28165E-1	2.37805E-1	1.56255E-1
1976	6.42126E-2	-5.21187E-1	-4.45777E-1	-3.37647E-1	8.60759E-2	4.52590E-3

TABLE 4 - CONTINUED

	JULY	AUG.	SEPT.	OCT.	NOV.	DEC.
1960	7.89824E-1	7.66884E-1	7.31754E-1	-2.15163E-2	-6.05146E-1	-9.31936E-1
1961	7.96153E-1	7.73213E-1	7.38083E-1	-1.51870E-2	-5.98817E-1	-9.25607E-1
1962	8.19155E-1	7.96215E-1	7.61085E-1	7.81465E-3	-5.75815E-1	-8.67105E-1
1963	8.92441E-1	8.69501E-1	8.34371E-1	8.11007E-2	-4.16585E-1	-7.07875E-1
1964	7.00390E-1	7.65390E-1	5.38670E-1	1.27890E-1	-2.96815E-1	-5.88105E-1
1965	7.67798E-1	8.32798E-1	6.06078E-1	1.95298E-1	-1.95334E-1	-4.86624E-1
1966	4.10806E-1	6.23806E-1	7.02086E-1	2.91306E-1	-1.66682E-1	-4.57972E-1
1967	4.26782E-1	6.39783E-1	7.18063E-1	3.07282E-1	-2.43427E-1	-5.34717E-1
1968	4.18330E-1	6.31331E-1	7.09611F-1	4.30006E-2	-2.26300E-1	-5.17590E-1
1969	4.04794E-1	6.17794E-1	6.96074E-1	2.94641E-2	-1.65637E-1	-4.77427E-1
1970	4.21569E-1	6.34569E-1	7.12849E-1	4.62390E-2	-1.58278E-1	-4.48278E-1
1971	3.76469E-1	6.51470E-1	7.29750E-1	0.00000	-2.46641E-1	-5.02761E-1
1972	3.43587E-1	6.43587E-1	8.31867E-1	4.09464E-2	-2.64707E-1	-5.20827E-1
1973	3.66522E-1	6.66522E-1	8.54802E-1	6.38815E-2	-1.13811E-1	-3.10191E-1
1974	4.35953E-1	6.74572E-1	1.00540	2.14482E-1	-1.41298E-1	-3.37678E-1
1975	4.58305E-1	6.96925E-1	1.02775	2.30834E-1	-2.19802E-1	-4.16182E-1
1976	3.06576E-1	5.45196E-1	8.76026E-1	8.51057E-2	-3.70994E-1	-5.73374E-1

TABLE 5

VON NEUMANN RATIOS AND VARIATION CUTS

MONTH	VARIATION CUT	VON NEUMANN RATIO
JAN.	59.9%	2.09909
FEB.	85.4%	2.09917
MARCH	95.3%	2.09857
APRIL	90.8%	2.10053
MAY	100.7%	2.10303
JUNE	99.1%	2.10086
JULY	97.0%	2.09526
AUG.	98.8%	2.10306
SEPT.	104.3%	2.100[illegible]3
OCT.	98.3%	2.10217
NOV.	104.4%	2.09745
DEC.	102.1%	2.09555

TABLE 6

VARIATION CUTS

YEAR	VARIATION CUT
1960	101.5%
1961	108.0%
1962	108.3%
1963	109.0%
1964	107.8%
1965	108.0%
1966	110.5%
1967	106.3%
1968	105.8%
1969	104.3%
1970	104.3%
1971	102.7%
1972	105.0%
1973	101.2%
1974	100.7%
1975	100.3%
1976	101.1%

REFERENCES

|1| M.M.G. FASE, J. KONING and A.F. VOLGENANT - "An experimental look at seasonal adjustment, a comparative study of nine alternative adjustment methods". European meeting of the Econometric Society, Oslo, Norway, August 8-31, 1973.

|2| P. d'ALESSANDRO - "Optimal smoothings" - Proceding of the 7th Symposium über Operations Research, St. Gallen, Aug. A982, Athenaüm Printers Königstein, Germany.

|3| P. d'ALESSANDRO and M. DALLA MORA - "Fast projection method for a special class of polytopes" - Research reports in Systems & Computers Science. Institute of Electrical Engineering, University of L'Aquila. Report No. 05-82, Feb. 1982.

|4| B. CVJETANOVIC, B. GRAB nad K. UEMURA - "DYnamics of acute bacterial deseases. Epidemiological models and their application in public helth". Supplement No.1 to vol.56 of the Bulletin of the world health organization, Geneve 1978.

|5| Annuario statistico italiano ISTAT (Istituto Centrale di Statistica). From 1960 to 1980.