УДК 311.13:519.246.8

O. Vasyechko,
Doctor of Sciences (Economics),
Université Paris 1 Panthéon-Sorbonne;
M. Grun-Rehomme,
Doctor of Sciences (Mathematics, Economics),
Université Panthéon-Assas Paris 2

Endpoints Issue in Time Series Adjusting

Methods based on the moving averages are largely used for times series filters construction. Yet, the times series endpoints are too sensitive to the type of the method used. A preferable method is supposed to be chosen by a statistician according to his/her experience and competency. The article aims to study the issue of choice of a method of time series decomposition by comparing the existing approaches. The issue of endpoints adjusting by use of Henderson smoother is discussed, and a new smoothing based on Epanechnikov kernel is presented.

Key words: *endpoints of times series, kernel smoothing, Henderson filters, symmetrical moving average, asymmetrical moving average.*

Time series analysis aims to reduce the effects of random variations in order to extract meaningful statistics from the data. Time series analysis takes into account the fact that data points taken over time may have an internal structure consisting of different systematic pattern components (trend, cycle and seasonality, Han, Kamber, 2001) and of a random error component that should be accounted for. The trend and the cyclical component, usually referred to as trend-cycle, are estimated jointly. To estimate a seasonally-adjusted time series or one without seasonality, the most common tool is moving averages.

It is possible to build a symmetrical moving average series, with desirable characteristics in terms of trend saving, noise reduction, flexibility and/or non-phase effects that are well adapted to the goal of analysis. Different approaches can be used to estimate a trend-cycle component (Grun-Rehomme, Ladiray, 1994; Gray, Thomson, 1996; Guggemos et al., 2012). However, with symmetrical moving averages, the treatment of the endpoints remains an issue. A moving average of $2p+1$ order, for example, does not allow smoothing of the p endpoints of the time series, or the p beginning points (although beginning points may be less important in a long data series because they are the oldest).

We can mention the following techniques which are currently used to smooth time series endpoints (Bianconcini, Quenneville, 2010; Proietti, Luati, 2008):

1. The symmetric and asymmetric filters associated with the Henderson smoother. Following the tradition already established in the related literature, both symmetric and asymmetric filters are called Henderson filters. Henderson kernel filters of any length can be constructed, including infinite ones. (Dagum, Bianconcini, 2006).

2. Within the family of Henderson filters, we can distinguish the asymmetric Henderson smoothers developed by Musgrave (1964a and 1964b). They are based on the minimization of the 15 mean squared revisions between the final estimates (obtained by application of the symmetric filter) and the preliminary estimates (obtained by application of an asymmetric filter) subject to the constraint that the sum of the weights is equal to one (Laniel, 1985; Doherty. 1992; Dagum, Bianconcini, 2006). The same technique can be applied to the starting observations of a time-series. This technique is implemented in X11, for example (Doherty, 2001; Ladiray, Quenneville, 2001; Quenneville et al., 2003). However, while the Musgrave filter is optimal for revision (with knowledge of new data records) it does not use the Henderson optimization criteria.

3. Time series forecasting by the ARIMA model followed by use of the symmetrical moving average, as performed by X11-ARIMA (Dagum, 1982). The major innovation introduced by Dagum consists in extending the series with forecasts to lessen the use of X11's asymmetric filters. For this purpose, she proposed to measure the theoretical reduction in revisions.

This paper analyzes the adjusting of endpoints of time series following the Henderson technique and proposes a new smoothing technique to treat endpoints. We use the same length for the asymmetric moving averages to keep the same level of smoothing. This principle is based on the Loess estimator proposed by Cleveland (1979) and further developed by Cleveland and Devlin (1988), which is also known as locally-weighted polynomial regression. However, the Loess estimator approach (without moving averages) cannot be compared with the one proposed by Dagum and Bianconcini (2008; 2012), since they are completely different.

Research issue

We consider a monthly seasonally-adjusted time series x_t (the technique is similar for quarterly time series) which is additive and can be decomposed into a trend-cycle component (denoted by tc_t) and a random error component ε_t, called noise:

$$x_t = tc_t + \varepsilon_t \,. \qquad (1)$$

The random error component in the decomposition model is often presented as white noise with variance σ_t^2.

The goal of this study is to estimate a globally smooth trend-cycle component that is assumed locally to follow a polynomial. There exist different approaches, such as local polynomial regression, graduation theory, and so on. Here we are interested in a kernel smoothing method (based on weighted moving averages).

Let p and f be two non-negative integers. The value of the initial time series at time t is replaced by a weighted average (with θ coefficients) of p "past" values of the series, the current value and f "future" values of the series. The quantity $p+f+1$ is called the moving average order. When p is equal to f, the moving average is said to be centered. If, in addition, we have $\theta_{-t} = \theta_t$ for every t, the centered moving average is said to be symmetric.

The transformation of x_t using moving averages MA can be written as follows:

$$x_t^* = MA(x_t) = \sum_{-p}^{f} u_k x_{t+k} . \qquad (2)$$

It is easy to show that for a moving average MA to conserve a polynomial of a certain degree d, it is necessary and sufficient that the coefficients θ_t verify:

$$\sum_{-p}^{f}\theta_i = 1 \quad \text{and} \quad \forall k \in \{1,2,\ldots,d\}, \quad \sum_{-p}^{f} i^k \theta_i = 0 . \qquad (3)$$

The returned coefficients, when applied to data, perform a polynomial least-squares fit within the filter window. The symmetrical moving averages have some good properties (without phase shift), but they are not convenient for estimation of the time series endpoints.

Henderson moving averages are mostly used for time series smoothing. An estimate tc_t of the trend-cycle must be a smooth curve. Let us denote the Dirac time series $\delta_t^{t_0}$ by:

$$\delta_t^{t_0} = \begin{cases} 1 & \text{if } t = t_0 , \\ 0 & \text{if } t \neq t_0 . \end{cases}$$

The application of moving averages of order $p+f+1$ and coefficients $\{\theta_t\}$ transforms it into the following time series:

$$MA(\delta_t^{t_0}) = \begin{cases} \theta_{t_0 - t} & \text{if } t = t_0 , \\ 0 & \text{otherwise}. \end{cases}$$

It is sufficient to impose that the curve of the coefficients of moving averages is smooth. The Dirac time series is a base of the set of time series, since all series x_t can be written as $x_t = \sum_{t_0 \in Z} x_{t_0} \delta_t^{t_0}$. Henderson's initial requirement is that the filter should reproduce a cubic polynomial trend. He suggested using, as criterion of smoothness, the quantity: $\sum_{i=-\infty}^{i=+\infty} (\nabla^3 \theta_i)^2$, where ∇ represents the first-order difference operator ($\nabla X_t = X_t - X_{t-1}$). The lower this quantity, the more flexible are the transformed series. The symmetric Henderson filter is an unbiased estimator for polynomials of degree 3.

Henderson's weights (θ_t) are solutions of the following optimization problem:

$$\min_{\theta} \left\{ \sum_{-p}^{f} (\nabla^3 \theta_t)^2 \,/ \sum_{-p}^{f} \theta_t = 1, \; \sum_{-p}^{f} t\theta_t = 0, \; \sum_{-p}^{f} t^2 \theta_t = 0 \right\} . \qquad (4)$$

As we are interested in monthly time series, we will consider a Henderson moving average of order 13 ($p+f+1=13$) with p varying from 6 (centered average) to 12 (last known point). As we are interested only in the most recent values of the series, we assume that $f \le p$.

For asymmetric filters, we keep the same length of smoothing as for symmetric filters in order to maintain the same level of smoothness. For moving averages H-6-6 to H-9-3 (Table 1), we can see that the coefficient of the current value is larger than the other

Table 1

Coefficients of the Henderson filter according to moving average order

t	H-12-0	H-11-1	H-10-2	H-9-3	H-8-4	H-7-5	H-6-6
–12	0.08514						
–11	0.14861	0.04644					
–10	0.10217	0.07662	0.01625				
–9	–0.05239	0.04257	0.02167	–0.00542			
–8	–0.23577	–0.04912	0	–0.01625	–0.01858		
–7	–0.34294	–0.14736	–0.03930	–0.02554	–0.03715	–0.02322	
–6	–0.30007	–0.18933	–0.06877	–0.02292	–0.03406	–0.04102	–0.01935
–5	–0.10288	–0.13503	–0.06001	0	0	–0.02554	–0.02786
–4	0.17683	0.01072	–0	0.04501	0.05894	0.02947	–0
–3	0.41914	0.19647	0.10002	0.10502	0.12574	0.10806	0.06549
–2	**0.51083(*)**	0.34383	0.20630	0.16504	0.18004	0.18219	0.14736

Table 1 continued

–1	0.40867	**0.38313**	**0.27506**	0.20630	0.19647	0.22220	0.21434
0	0.18266	0.29334	0.27245	**0.21285**	**0.20576**	**0.22505**	**0.24006**
1		0.12771	0.19505	0.17879	0.15718	0.17683	0.21434
2			0.08127	0.11378	0.10217	0.10806	0.14736
3				0.04334	0.04954	0.04257	0.06549
4					0.01393	0.00232	0
5						–0.00697	–0.02786
6							–0.01935

* In this table, bold font indicates the highest value in each column (coefficients of moving average)

coefficients. This corresponds to the idea that the current value "dominates" the estimated value of the trend at the current date.

For the last three orders (i. e. from 10 to 12), the largest value of the coefficients remains around date t-2; this is not satisfactory because it indicates that the last observed values weigh less than those at t-2. Such an observation was already formulated by Dagum and Bianconcini (2008; 2012). Also note that the Henderson moving average, which only keeps the straight lines, has the same disadvantage.

A new approach

The kernel method is generally used to estimate the density of a probability distribution of a sample, taking into account the local character of this density

Briefly, let $x_1,\dots,x_n$ be a random sample drawn from an unknown continuous distribution with density function *f*. The kernel density estimator of *f* is:

$$f_n(x) = \frac{1}{n\lambda}\sum_{1}^{n} K\left(\frac{x - x_i}{\lambda}\right),$$

where *K*(.) is a kernel, an integrated function, positive (but not necessary positive when we are fitting a cubic polynomial locally) and λ > 0 is a smoothing parameter called the bandwidth. *K* is a probability density:

$\int_R K = 1$ with $\operatorname*{Max}_x K(x) = K(0)$.

For example: $K(x) = I_{[-1/2,1/2]}(x)$ where *I* is the indicator function, or $K(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$ (Gaussian kernel).

We can show (Weyman, Wright, 1983) that in the class of even, positive kernels verifying $\int_R x^2 K(x)\,dx = 1$, the minimum of the Mean Integrated Square Error (MISE): $\operatorname*{Min}_K \int_R E\left[f(x) - f_n(x)\right]^2 dx$ is reached for

$$K(t) = \frac{3}{4\sqrt{5}}\left(1 - \frac{t^2}{5}\right) I_{[-\sqrt{5},\sqrt{5}]}(t).$$

The Epanechnikov kernel is defined as

$$K_e(t) = \frac{3}{4}(1 - t^2)\, I_{[-1,1]}(t).$$

Hereafter we use the term "Epanechnikov kernel" for *K*(*t*), because we are dealing with probability densities.

The kernel method can also be used to define the coefficients θ_t of the moving averages, since we can write $\theta_t = \frac{K(t/\lambda)}{\sum_{-p}^{f} K(t/\lambda)}$, where λ is the bandwidth parameter chosen in order to ensure a filter length equal to $p+f+1$. Intuitively, one wants to choose as small as the data allow; however, there is always a trade-off between the bias of the estimator and its variance.

This moving average remains constant because $\sum_{-p}^{f} \theta_t = 1$. For example, for $\lambda=1$, we obtain $\theta_t = \frac{1}{2p+1}$.

Let us suppose that we wish to smooth a time series x_t, decomposed into a trend-cycle (noted tc_t) and a residual component ε_t (white noise):

$$x_t = tc_t + \varepsilon_t .$$

It can be noted that θ_0 (weight of the current date for the smoothing) will be always larger than the weights attributed to the previous or later dates. Moreover:

- For a symmetric moving average of order $2p+1$, it is necessary to choose $\lambda = \frac{p}{\sqrt{5}}$ to remain with in an interval $[-p, p]$, and $\theta_t = \frac{3p}{4p^2 - 1}\left(1 - \frac{t^2}{p^2}\right)$.

- For asymmetric moving averages, with $0 \le f \le p$, it is necessary to choose $\lambda \ge \frac{p}{\sqrt{5}}$, for example: $\lambda = \frac{p}{\sqrt{5}}$ and in this case

$$\theta_t = \frac{(1-\frac{t^2}{p^2})}{(p+f+1-\frac{S_2}{p^2})}, \text{ where } S_2 = \sum_{-p}^{f} t^2.$$

The main advantage of this kernel approach is that the current value has a larger weight, while the weights of past and future values decrease as we move away from the present value. This property also remains valid for asymmetric moving averages.

The largest weight at the current date permits to show the most recent variations (Table 2).

Table 2

Coefficients of moving averages according to the Epanechnikov kernel

t	N-12-0	N-11-1	N-10-2	N-9-3	N-8-4	N-7-5	N-6-6
–12	0						
–11	0.018821	0					
–10	0.036006	0.019700	0				
–9	0.051555	0.037523	0.020879	0			
–8	0.065466	0.053471	0.039560	0.022546	0		
–7	0.077741	0.067542	0.056044	0.042440	0.025084	0	
–6	0.088380	0.079737	0.070330	0.059681	0.046823	0.029412	0
–5	0.097381	0.090056	0.082418	0.074271	0.065217	0.054299	0.038461
–4	0.104746	0.098499	0.092308	0.086210	0.080267	0.074661	0.069930
–3	0.110475	0.105065	0.1	0.095490	0.091973	0.090498	0.094406
–2	0.114566	0.10975	0.105494	0.102122	0.100334	0.101810	0.111888
–1	0.117021	0.112570	0.108791	0.10610	0.105351	0.108597	0.122377
0	**0.117840***	**0.113508**	**0.109890**	**0.107427**	**0.107023**	**0.110859**	**0.125874**
1		0.112570	0.108791	0.106100	0.105351	0.108597	0.122377
2			0.105494	0.102122	0.100334	0.101810	0.111888
3				0.095491	0.091973	0.090498	0.094406
4					0.080267	0.074661	0.069930
5						0.054299	0.038461
6							0

* In this table, bold font indicates the highest value in each column (coefficients of a moving average)

The noise is transformed by the moving average into a sequence of random variables of constant variance: $\sigma^2(\sum_{-p}^{f}\theta_t^2)$. Reducing the irregular component, and therefore its variance, amounts to reducing the quantity $\sum_{-p}^{f}\theta_t^2$. The output signal is assumed to be 'as close as possible' to the input signal when noise components are removed; that is why we will call this criterion the 'fidelity' criterion. For this moving average, we obtain around 90% noise reductions (variance) for each order. Furthermore, it preserves constants, but not straight lines and parabolas.

In this part, we look for the moving average with coefficient that is closest to the results of the kernel approach and which keeps parabolas.

Let x_t be the coefficients of this kernel, with $\sum_{-p}^{f} x_t = 1$.

It is necessary to solve the following optimization problem:

$$\underset{\theta}{\mathrm{Min}}\left\{\sum_{-p}^{f}(\theta_t - x_t)^2 / \sum_{-p}^{f}\theta_t = 1,\ \sum_{-p}^{f} t\theta_t = 0,\ \sum_{-p}^{f} t^2\theta_t = 0\right\}. \quad (5)$$

According to the Kuhn and Tucker conditions for ordinary convex programming (Rockafellar, 1970, section 28), we obtain:

$$\begin{pmatrix}\lambda_3\\ \lambda_2\\ \lambda_1\end{pmatrix} = A^{-1}\begin{pmatrix}\sum_{-p}^{f} t^2 x_t\\ \sum_{-p}^{f} t\ x_t\\ 0\end{pmatrix}, \text{ with } \quad A = \begin{pmatrix} S_4 & S_3 & S_2\\ S_3 & S_2 & S_1\\ S_2 & S_1 & S_0\end{pmatrix} \text{ where:}$$

$$S_0 = p + f + 1,$$

$$S_1 = \sum_{-p}^{f} t = \sum_{1}^{f} t - \sum_{1}^{p} t = \frac{f(f+1) - p(p+1)}{2},$$

$$S_2 = \sum_{-p}^{f} t^2 = \sum_{1}^{p} t^2 + \sum_{1}^{f} t^2 = \frac{p(p+1)(2p+1) + f(f+1)(2f+1)}{6}, \quad (6)$$

$$S_3 = \sum_{-p}^{f} t^3 = \sum_{1}^{f} t^3 - \sum_{1}^{p} t^3 = \frac{f^2(f+1)^2 - p^2(p+1)^2}{4},$$

$$S_4 = \sum_{-p}^{f} t^4 = \sum_{1}^{p} t^4 + \sum_{1}^{f} t^4 = \frac{p(p+1)(6p^3 + 9p^2 + p - 1) + f(f+1)(6f^3 + 9f^2 + f - 1)}{30},$$

and

$$\lambda_3 = \frac{(S_0S_2 - S_1^2)(\sum_{-p}^{f} t^2 x_t) + (S_1S_2 - S_0S_3)(\sum_{-p}^{f} tx_t)}{\det(A)},$$

$$\lambda_2 = \frac{(S_1S_2 - S_0S_3)(\sum_{-p}^{f} t^2 x_t) + (S_0S_4 - S_2^2)(\sum_{-p}^{f} tx_t)}{\det(A)}, \quad (7)$$

$$\lambda_1 = \frac{(S_1S_3 - S_2^2)(\sum_{-p}^{f} t^2 x_t) + (S_2S_3 - S_1S_4)(\sum_{-p}^{f} tx_t)}{\det(A)}.$$

Then $\theta_t = x_t - (\lambda_3 t^2 + \lambda_2 t + \lambda_1)$.

With this moving average, the noise reduction is around 85% for each order, except for MM-12-0 (only 50%) and MM-11-1 (73%) (Table 3).

Table 3

Coefficients of moving averages according to the Epanechnikov kernel fitting a parabola

t	MM-12-0	MM-11-1	MM-10-2	MM-9-3	MM-8-4	MM-7-5	MM-6-6
–12	0.120879						
–11	0.032967	0.032967					
–10	–0.032967	0.000000	–0.032967				
–9	–0.076923	–0.021978	–0.021978	–0.076923			
–8	–0.098901	–0.032967	–0.008991	–0.032967	–0.098901		
–7	–0.098901	–0.032967	0.005994	0.005994	–0.032967	–0.098901	
–6	–0.076923	–0.021978	0.022977	0.039960	0.022977	–0.021978	–0.07692
–5	–0.032967	0.000000	0.041958	0.068931	0.068931	0.041958	0.00000
–4	0.032967	0.032967	0.062937	0.092907	0.104895	0.092907	0.06294
–3	0.120879	0.076923	0.085914	0.111888	0.130869	0.130869	0.11189
–2	0.230769	0.131868	0.110889	0.125874	0.146853	0.155844	0.14685
–1	0.362637	0.197802	0.137862	0.134865	**0.152847**	0.166832	0.16783
0	**0.516484***	0.274725	0.166833	**0.138861**	0.148851	**0.167833**	**0.17483**
1		**0.362637**	0.197802	0.137862	0.134865	0.152847	0.16783
2			**0.230769**	0.131868	0.110889	0.125874	0.14685
3				0.120879	0.076923	0.085914	0.11189
4					0.032967	0.032967	0.06294
5						–0.032967	0.00000
6							–0.07692

* In this table, bold font indicates the highest value in each column (coefficients of a moving average)

This approach preserves parabolas but has the same disadvantage as the Henderson moving average (the greatest value of the coefficients does not correspond to the current time). The constraints are too strong and as a result the coefficients do not depend on the reference moving average with respect to which we minimize the distance.

When the points at both ends of a time series have to be estimated with asymmetric moving averages, this filter should be rather short and have a gain $G(\omega)$ close to one for small frequency ω (for example, between 0 and $\pi/6$) and near to zero for higher frequencies. The gain function $G(\omega)$ describes how much the amplitude of the time series components is changed by the filtering. In the annex we present the gain function for some asymmetric moving averages (Table 2 and 3); it has the same form for the other asymmetric filters. We have two groups: from 6-6 to 9-3 and from 10-2 to 12-0. In each group, we obtain roughly the same curve shape. It is apparent that the asymmetric filter does not amplify the signal and converges faster to the final one. There exists a trade-off between the amplitude and phase shift effects induced by an asymmetric filter.

Conclusions

For the kernel approaches, two elements remain constant:

– In the method which keeps the parabolas, a shift takes place between asymmetric moving averages of order 9-3 and those of order 10-2, when we move from a concave curve to a convex curve. A concave curve attributes a dominant weight coefficient to the current value that seems to be reasonable. A convex curve attributes the largest weight to the last observed value, which is less reasonable.

– With the kernel method which keeps only constants, the largest weight is always the weight of the current value and the weights decrease as we move away from the current value. This suggests that for the last values, it may be better to take moving averages which keep only constants (even the straight lines); otherwise the last observed values are over-weighted and strongly influence the trend. This is in contrast to the definition of a trend, which describes the long-term evolution of the series and must be relatively robust.

There are many methods for time series decomposition. The statistician-economist must choose the one that seems to be the best smoothing technique according to his/her experience. The present paper contributes to the choice issue by examining the properties of different moving average methods.

References

1. Bianconcini, S., Quenneville, B. (2010). Real Time Analysis Based on Reproducing Kernel Henderson Filters. Estudios de Economia Aplicada, 28 (3), 1–22.

2. Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplot. Journal of American Statistical Association, 74(368), 829–836.

3. Dagum, E. B. (1982). The Effects of Asymmetric Filters of Seasonal Factor Revisions. Journal of the American Statistical Association, 77, 732–738.

4. Dagum, E. B. (1988). The X11ARIMA/88 Seasonal Adjustment Method, Foundation and User's Manual. Technical Report 12-564E. Time Series Research and Analysis Division, Statistics Canada.

5. Dagum, E. B., Bianconcini, S. (2006). Local Polynomial Trend-cycle Predictors in Reproducing Kernel Hilbert Spaces for Current Economic Analysis. Anales de Economia Aplicada, 1–22. Retrived from http://www2.stat.unibo.it/beedagum/Papers/DagumBianconciniJBES.pdf

6. Dagum, E. B., Bianconcini, S. (2008). The Henderson Smoother in Reproducing Kernel Hilbert Space. Journal of Business & Economic Statistics, 26(4), 536–545.

7. Dagum, E. B., Bianconcini, S. (2012). Reducing Revisions in Real Time Trend-Cycle Estimation. Joint Statistical Meeting Proceedings, Section on Government Statistics. Alexandria, VA: American Statistical Association. 1830–1841.

8. Doherty, M. (1992). Surrogate Henderson filters in X-11. Technical Report, New Zealand Department of Statistics, Working Paper.

9. Doherty, M. (2001). The Surrogate Henderson Filters in X11, Australian and New Zealand Journal of Statistics, 43, 901–908.

10. Findley, D., Monsell, B., Bell, W., Otto, M., Chen, B. (1998). New Capabilities and Methods of the X12ARIMA Seasonal Adjustment Program. Journal of Business & Economic Statistics, 16, 127–152.

11. Gray, A., Thomson, P. (1996). Design of Moving-Average Trend Filters Using Fidelity and Smoothness Criteria. Time Series Analysis, Vol. 2 (in memory of E. J. Hannan). Springer Lectures Notes in Statistics, Vol. 15, 205–219.

12. Guggemos, F., Ladiray, D., Grun-Rehomme, M. (2012). New Results on Linear Filters Minimizing Phase-Shift for Seasonal Adjustment. Joint Statistical Meetings, San Diego, USA, July 28 – August 3, 2012.

13. Grun-Rehomme, M., Ladiray, D. (1994). Moyennes mobiles centrées et non centrées: construction et comparaison. Revue de Statistiques Appliquées, 42(3), 33–61.

14. Han, J., Kamber, M. (2006). Data Mining: Concepts and Techniques. Second Ed. San Diego: Academic Press, 743. (Morgan-Kaufman Series of Data Management Systems).

15. Hendreson, R. (1916). Note on Graduation by Adjusted Average. Transaction of Actuarial Society of America, 17, 43–48.

16. Ladiray, D., Quenneville, B. (2001). Seasonal Adjustment with X11-Method. New York: Springer, 227. (Lecture notes in statistics, 158).

17. Laniel, N. (1985). Design Criteria for 13 Term Henderson End-Weights. Technical Report Working paper TSRA-86-011, Statistics Canada, Ottawa K1A 0T6.

18. Musgrave, J. (1964a). A Set of End Weights to End all End Weights. US Bureau of Census. Retrieved from http://www.census.gov/ts/papers/Musgrave1964a.pdf

19. Musgrave, J. (1964b). Alternative Sets of Weights for Proposed X-11 Seasonal Factor Curve Moving Averages. Working Paper. Washington DC : US Bureau of Census.

20. Proietti, F., Luati, A. (2008). Real Time Estimation in Local Polynomial Regression, with Application to Trend-Cycle Analysis. The Annals of Applied Statistics, 2(4), 1523–1553.

21. Quenneville, B., Ladiray, D., Lefrançois, B. (2003). A Note on Musgrave Asymmetrical Trend-Cycles Filters. International Journal of Forecasting, 19, 727–734.

22. Rockafellar, R. T. (1970). Convex Analysis. Princeton: University Press, 472.

23. Weyman, D. J., Wright, I. W. (1983). Splines in Statistics. Journal of the American Statistical Association, 78 (382), 351–365.