

# Smoothing Matrices in Time Series Short-Term Trend Analysis: Algebraic and Statistical Properties

## *Le Matrici di Smoothing nell'Analisi del Trend di Breve periodo di Serie Storiche: Proprietà Algebriche e Statistiche*

Estela Bee Dagum, Alessandra Luati  
Dipartimento di Scienze Statistiche  
Università di Bologna  
beedagum@stat.unibo.it; luati@stat.unibo.it

**Riassunto:** Le matrici di smoothing associate a filtri lineari per la stima delle componenti latenti nell'analisi delle serie storiche non sono in generale simmetriche. Di conseguenza, le proprietà inferenziali sviluppate per le matrici di smoothing utilizzate nell'ambito dei modelli di regressione non valgono per matrici di smoothing utilizzate nell'ambito delle serie storiche. Queste ultime, tuttavia, hanno una struttura algebrica ben definita, sono infatti centrosimmetriche, da cui è possibile derivare proprietà algebriche e statistiche analoghe a quelle valide per matrici simmetriche. E' possibile inoltre definire misure di distorsione, varianza ed errore quadratico medio per filtri simmetrici e per filtri asimmetrici per la stima delle ultime osservazioni.

**Keywords:** smoothing matrices, linear filters, bias and variance measures.

## 1. Smoothing Matrices in Time Series

Smoothing matrices associated to linear filters acting on  $N$ -dimensional time series to produce smooth estimates of the underlying trends have got the following structure

$$\mathbf{S} = \begin{bmatrix} \mathbf{W}_{(m \times 2m)}^a & & \mathbf{O}_{(m \times N-2m)} \\ & \mathbf{W}_{(N-2m \times N)}^s & \\ \mathbf{O}_{(m \times N-2m)} & & \mathbf{W}_{(m \times 2m)}^{a*} \end{bmatrix}$$

where  $\mathbf{W}^s$  is the submatrix whose row elements are the  $2m + 1$ -symmetric weights to be applied to central observations, while  $\mathbf{W}^a$  and  $\mathbf{W}^{a*}$  are the submatrices whose rows contain the asymmetric weights to be applied to the first and last observations, respectively; the number into parentheses indicate the dimension of each submatrix.

The matrix  $\mathbf{S}$  is centrosymmetric, that is,  $w_{ij} = w_{n+1-i, n+1-j}$ , where  $w_{ij}$  is the generic  $ij$ -th element of  $\mathbf{S}$ ;  $\mathbf{W}^s$  is rectangular centrosymmetric, whereas  $\mathbf{W}^a$  and  $\mathbf{W}^{a*}$  are one the  $t$ -transform of the other, where  $t$  is a linear transformation introduced by Dagum and Luati (2004) that consists in the pre- and post-multiplication of a matrix for the permutation matrix  $\mathbf{E}_k \in \mathbb{R}^{k \times k}$  having ones on the cross diagonal (bottom left to top right) and zeros elsewhere. In other words,  $t$  acts on a matrix reversing the order of its rows by pre-multiplication for  $\mathbf{E}_m$  and then reversing the order of its columns by postmultiplication for  $\mathbf{E}_n$ . A centrosymmetric matrix is invariant with respect to such a transformation.

Centrosymmetric matrices inherit desirable properties from the properties of the permutation matrices  $\mathbf{E}_k$  which are: (a) symmetric (b) orthogonal and (c) reflections, *i.e.*  $\mathbf{E}_k \stackrel{(a)}{=} \mathbf{E}_k' \stackrel{(b)}{=} \mathbf{E}_k^{-1}$  and  $\mathbf{E}_k^2 \stackrel{(c)}{=} \mathbf{I}_k$ , where the superscripts “'” and “ $-1$ ” indicate transposition and inversion, respectively.

On the other hand, centrosymmetric smoothing matrices are not necessarily symmetric. That is the reason for the need of generalizing to centrosymmetric smoothing matrices the theory of symmetric smoother matrices developed for linear regression models and generalized linear models by Hastie and Tibshirani (1990). The authors define  $\text{tr}\{\mathbf{S}\}$  as the “degrees of freedom” of a smoother, a measure of the smoothness of an estimator, which is minimum when  $\mathbf{S} = \mathbf{I}$ , that is interpolation. Applied to non symmetric smoothing matrices this measure does not make sense for two reasons: first, the eigenvalues of a non-symmetric matrix can be complex and their algebraic and geometric multiplicities can be different; secondly it does not take into account the problem of asymmetric weights for ending observations. The problem of complex eigenvalues can be overcome considering, as Hastie and Tibshirani themselves argue,  $\text{tr}\{\mathbf{S}\mathbf{S}'\}$  instead of  $\text{tr}\{\mathbf{S}\}$  but the asymmetric weights problem still cannot. A solution can be considering the decomposition

$$\text{tr}\{\mathbf{S}\mathbf{S}'\} = \text{tr}\{\mathbf{W}^s\mathbf{W}^{s'}\} + 2\text{tr}\{\mathbf{W}^a\mathbf{W}^{a'}\} \quad (1)$$

where  $\text{tr}\{\mathbf{W}^s\mathbf{W}^{s'}\}$  accounts for the degrees of freedom of the symmetric filter applied to central observations whereas  $\text{tr}\{\mathbf{W}^a\mathbf{W}^{a'}\}$  accounts for the degrees of freedom of the asymmetric filters for the first and last observations.

## 2. Local Measures of Bias and Variance for a Smoother

So decomposed,  $\text{tr}\{\mathbf{S}\mathbf{S}'\}$  allows to differently treat the asymmetric from the symmetric weights and to define matrix-based measures of bias, variance and mean square error. The local bias provides information on the concentration of the smoother respect to the observation to be estimated; the local variance is an approximation of the variance left in the estimated series after the smoothing procedure. These measures agree with the classical spectral ones and provide information on the empirical fitting and smoothing properties of the estimators. One major advantage of these measures relative to the classical spectral one is their easiness of calculation. Furthermore, it is possible to have a local measure of bias for symmetric filters which cannot be obtained by means of spectral techniques. The local bias and variance can be defined based on the decomposition (1), normalized such that it depends neither on the length of the time series nor on the length of that of the filter, and their properties can be easily analysed. For space reasons, we will expose here the theory for symmetric fitters only.

Hastie and Tibshirani (1990) state that a filter is completely unbiased or concentrated if it has got maximum number of degrees of freedom, *i.e.*  $1 - \frac{\text{tr}\{\mathbf{S}\}}{N} = 0$ . Considering  $\mathbf{W}^s$  instead of  $\mathbf{S}$  and “correcting” the measure by enlarging to the  $2k + 1$  central weights the span of the bias interval, we define a measure of **local bias for a symmetric filter** as

$$1 - \frac{\text{tr}\{\mathbf{W}^s\mathbf{D}'\}}{(N - 2m)}$$

where  $\mathbf{D} \in \mathbb{R}^{N-2m \times N}$  is the  $2k + 1$ -band matrix with ones corresponding to the  $2k + 1$  central weights of the filter and zeros elsewhere with generic element  $d_{ij} = 1$  if  $j - i = m - k, \dots, m + k$  and  $d_{ij} = 0$  otherwise.

The justification of the above definition lies in the fact that to properly estimate points of maxima and minima locally present in a given input, the filter weights distribution

cannot be flat and the more weight (relative to the total) is concentrated in the restricted local neighborhood of the observation to be estimated, the less unbiased the output will be. The problem is to define an appropriate span of the restricted neighborhood. With no prior information on both the characteristics of the input series and the shape of the symmetric filters weighting system, the local neighborhood should not be larger than half of the total filter span. The constraints that the restricted neighborhood is local and its number of weights should be odd lead to the choice of  $k = 2$  (representing 40% of the total filter span). We have found that for weight systems following an unimodal function that decreases rapidly away from the central point, using  $k = 2$  is appropriate even if the filter is much longer and obviously, in such case, the number of weights represents a much smaller percentage of the total filter.

The **local variance** of the smoothed series with a symmetric filter is defined as

$$\frac{\text{tr}\{\mathbf{W}^s \mathbf{W}^{s'}\}}{N - 2m}.$$

Given that the global variance introduced in the output after the filtering process, standardized for the input variance is  $\frac{1}{N-2m} [\text{tr}\{\mathbf{W}^s \mathbf{W}^{s'}\} + \text{tr}\{\mathbf{W}^s (\mathbf{P} - \mathbf{I}) \mathbf{W}^{s'}\}]$  where  $\mathbf{P}$  is the  $N$ -Dimensional Toeplitz matrix of the process, then the local variance gives *exactly* the amount of variance reduction in the output *only* if the input is a white noise *or* if the weights correspond to interpolation or arithmetic mean (minimum and maximum smoothing, respectively). Although for most socio-economic time series this is not the case, the approximation is appropriate. In fact, for weights adding up to unity and decreasing autocorrelation function, the term  $\text{tr}\{\mathbf{W}^s (\mathbf{P} - \mathbf{I}) \mathbf{W}^{s'}\}$  becomes negligible with respect to the term  $\text{tr}\{\mathbf{W}^s \mathbf{W}^{s'}\}$ . It is evident that the latter depends only on the filter while the former depends also on the autocovariance structure of the series.

As a synthesis of the bias and variance measures, a **local mean square error** for symmetric and asymmetric filters is obtained as the sum of square bias plus variance.

Next, we calculate Hastie and Tibshirani (HT) bias and Dagum and Luati (DL) measures on several widely applied symmetric smoothers such as locally weighed regression smoother of degree 1 and 2 (L1 and L2, respectively), Gaussian kernel (GK), cubic smoothing spline (CSS), Henderson filter (H13), all of fixed length equal to 13 terms.

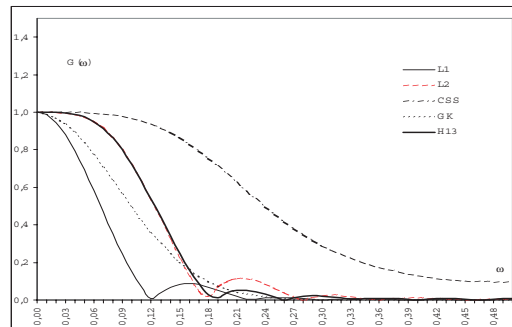
### 3. Analysis of Results and Concluding Remarks

Table 1 shows HT bias and DL measures calculated on the smoothing matrices associated to 13-term linear filters of second-order (L1, GK) and higher-order (L2, H13, CSS). In general, second-order kernels tend to oversmooth the input series whereas higher-order filters often produce estimates highly affected by variance.

| Symmetric Filters                                        | L1          | L2          | CSS           | GK          | H13      |
|----------------------------------------------------------|-------------|-------------|---------------|-------------|----------|
| $HT \text{ bias } 1 - \frac{\text{tr}\{\mathbf{S}\}}{N}$ | 0.86 (1.15) | 0.76 (1.01) | 0.46 (0.61)   | 0.78 (1.04) | 0.75 (1) |
| <i>Local bias</i>                                        | 0.40 (10)   | 0.04 (1)    | -0.07 (-1.75) | 0.18 (4.50) | 0.04 (1) |
| <i>Local variance</i>                                    | 0.10 (0.50) | 0.20 (1)    | 0.40 (2)      | 0.15 (0.75) | 0.20 (1) |
| <i>LMSE</i>                                              | 0.26 (1.30) | 0.20 (1)    | 0.40 (2.03)   | 0.18 (0.91) | 0.20 (1) |

**Table 1** HT bias and DL measures for 13-term filters (*values standardized by H13*).

Both HT and DL bias values for symmetric filters show the same ranking, in the sense that L1 is the most biased filter whereas CSS is practically unbiased; GK is less biased than L1 but more than H13 and L2 which give similar results. However, looking at the values normalized by H13, considered as a benchmark, it is clear that DL measures better evidence the differences among the smoothers. In fact, HT bias has almost the same value for H13, L2 and GK and such value is very close to that of L1. On the other hand, DL bias clearly discriminates among second order kernels (L1 and GK) and higher order kernels (L2 and H13), and also within the groups, since L1 is highly biased respect to GK, whereas L2 and H13 show the same values. What can be inferred by DL bias is coherent with what can be deduced by the gain functions of the symmetric estimators, shown in Figure 1. In fact, corresponding to the range  $0 \leq \omega \leq 0.06$ , associated with cycles of 16 months or longer attributed to the signal of the series, higher-order kernels have a gain  $G(\omega)$  close to one, *i.e.* they perfectly reproduce the signal. On the other hand, GK and L1 have smaller gain, *i.e.* they suppress part of the signal; L1 in particular oversmooths the series.



**Figure 1** Gain functions of 13-term symmetric filters.

Concerning the variance measure, the ranking is opposite to that of bias: second order kernels suppress a great amount of noise; on the contrary, higher order kernels are less biased but leave some disturbance in the estimates. In particular, CSS severely undersmooths the input, L2 and H13 leave the same amount of noise, whereas L1 is the smoother which mostly suppresses the variance, followed by GK. These informations agree with those provided by Figure 1, where now the range to be considered is  $0.10 \leq \omega \leq 0.50$ , corresponding to short frequencies associated with the noise. In that interval, the closer to zero the gain of the estimator is, the best is the variance suppression.

Considering the local mean square error, it helps in selecting the subset of smoothers where to choose the ‘optimal’ one, but in the final choice the bias and variance contribution must be taken into account. In fact, the *LMSE* values in Table 1 lead to choose either L2, H13 or GK. The latter presents the smallest *LMSE* but it is more biased than L2 and H13, where the bias and variance contributions are better distributed.

## References

- Dagum E. B., Luati A. (2004), *A Linear Transformation and its Properties with Special Applications in Time Series Filtering*, Linear Algebra and its Applications, 338, 107-117.  
Hastie T. J., Tibshirani R.J. (1990), *Generalized Additive Models*, Chapman and Hall.