Extracting a Smooth Trend from a Time Series: A Modification of Singular Spectrum Analysis

ANDREW R. SOLOW
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

ANAND PATWARDHAN
Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, Pennsylvania

(Manuscript received 20 June 1995, in final form 6 March 1996)

ABSTRACT

Singular spectrum analysis is commonly used in climatology to extract a trend from a noisy time series. Implicit in this method is the association of trends with high variance. In many cases, it may be more natural to associate trends with smoothness. This paper describes how singular spectrum analysis can be modified to incorporate this idea. The modified approach is illustrated using the annual central England temperature series.

1. Introduction

The problem of decomposing the temporal variability in a noisy time series arises in many fields. A method called singular spectrum analysis (Vautard et al. 1992) has been used in a number of recent climatological applications (e.g., Schlesinger and Ramanan 1994; Plaut et al. 1995). This method is closely related to principal component analysis (Jolliffe 1986). In particular, as described in more detail below, singular spectrum analysis uses a variance criterion to construct index series that reflect different components of the variability in the original time series. When interest centers on trend extraction, it may be more natural to use a smoothness criterion instead of the variance criterion. The purpose of this paper is to describe and illustrate how a multivariate method proposed by Shapiro and Switzer (1989) can be modified to extract a smooth trend from a univariate time series.

2. Singular spectrum analysis

Consider a univariate time series $Y_t, t = 1, \cdots, n$, standardized to have mean 0. For a fixed bandwidth $p$, let

$$Y_t = (Y_t \cdots Y_{t+p-1}), \quad t = 1, 2, \cdots, n - p + 1$$

be the $p$ vector, consisting of the $p$ consecutive values of the time series starting at time $t$. Interest centers on constructing an index series,

$$Z_t = w'Y_t, \quad t = 1, 2, \cdots, n - p + 1, \quad (1)$$

that best represents the trend in the original time series. Here, $w$ is a $p$ vector of weights. The index series is a linear smoothing of the original time series. Under singular spectrum analysis, the weight vector $w$ is chosen to maximize

$$\text{var}Z_t = w'Cw, \quad (2)$$

subject to the constraint that the sum of the squared weights is 1:

$$w'w = 1. \quad (3)$$

Here, $C$ is the $p$-by-$p$ covariance matrix between the elements of $Y_t$, given by

$$C = \text{cov}(Y_t, Y_{t+|i-j|})
= [C(|i-j|)]_{i,j = 1, 2, \cdots, p}. \quad (4)$$

In practice, the autocovariances in $C$ are unknown and are replaced by estimates such as

$$C(d) = \sum_{t=1}^{n-d} Y_tY_{t+d}/n.$$

Under this formulation, $w$ is given by the unit eigenvector associated with the dominant eigenvalue of $C$.

Singular spectrum analysis is equivalent to the application of principal component analysis to the multivariate time series $Y_t$. Like principal component analysis, singular spectrum analysis can be used to construct $p$ index series, such as (1), with variance decreasing from the maximum possible and with each pair of index series uncorrelated at lag 0. These series
are given by the columns of the \((n - p + 1)\)-by-\(p\) matrix:

\[
Z = YW,
\]

where \(Y\) is the \((n - p + 1)\)-by-\(p\) matrix whose rows are \(Y'_t, t = 1, 2, \ldots, n - p + 1\), and the columns of the \(p\)-by-\(p\) weight matrix \(W\) are the unit eigenvectors of \(C\) arranged in decreasing order of the corresponding eigenvalues.

In a typical application, the first few index series are treated as representing components of the trend in the original time series and retained, and the remaining index series are treated as noise and discarded. Commonly, the matrix \(Y\) is projected onto the retained index series. This projection is given by

\[
Y^* = Z^*W^*,
\]

where the columns of \(Z^*\) are equal to the columns of \(Z\) when the corresponding index series are retained and have elements that are all equal to 0 otherwise. A reconstruction of the original time series can be made from \(Y^*\) in the following way (Vautard et al. 1992). Note that \(Y\) has the following structure:

\[
Y_1Y_2Y_3 \ldots Y_p
\]
\[
Y_2Y_3Y_4 \ldots Y_{p+1}
\]
\[
Y_3Y_4Y_5 \ldots Y_{p+2}
\]
\[
\ldots
\]

so that the original time series can be recovered by averaging the reverse diagonals. The original time series can be reconstructed by applying the same procedure \(Y^*\). In fact, averaging is not necessary in the case of \(Y\) since the elements of each reverse diagonal are equal. However, this property is not shared by \(Y^*\), and so averaging is necessary.

3. An alternative method

The use of singular spectrum analysis for trend extraction rests on an implicit association of trends with variance. In many situations, it may be more natural to associate trends with smoothness. In the case of multivariate time series, a number of methods have been proposed to construct index series using a smoothness criterion (e.g., Box and Tiao 1977; Peña and Box 1987; Shapiro and Switzer 1989). For example, the method proposed by Shapiro and Switzer (1989) (see also Sowell 1994) uses lag 1 autocorrelation instead of variance as the criterion. The same idea can be used to extract a trend from a univariate time series.

Recall that singular spectrum analysis is equivalent to applying principal component analysis to the multivariate time series \(Y\). The columns of the \((n - p + 1)\)-by-\(p\) matrix \(Y\) are the components of this multivariate time series, and the \(p\)-by-\(p\) matrix \(C\) is an estimate of their covariance matrix. Let \(D\) be the \((n - p)\)-by-\(p\) matrix whose columns are the first differences of the columns of \(Y\), and let

\[
C_D = [2C(|i-j| - C(|i-j-1|)
\]
\[
- C(|i-j+1|)] \quad i, j = 1, 2, \ldots, p
\]

be the estimated covariance matrix of the columns of \(D\). It is straightforward to show that

\[
w'C_Dw/w'Cw = 2(1 - r),
\]

where \(r\) is the sample lag one autocorrelation of the index series \(Z_t = w'Y_t\). It follows from basic matrix theory that the choice of \(w\) that maximizes \(r\) is proportional to the eigenvector associated with the smallest eigenvalue of the matrix \(C^{-1}C_D\). As in singular spectrum analysis, this method can be used to produce \(p\) index series with lag one autocorrelation decreasing from the maximum possible and with each pair of index

![Fig. 2. Weights for the first index series produced by singular spectrum analysis (solid) and the alternative method (dotted).](image-url)
series uncorrelated at lag 0. These index series are

given by the columns of

\[ Z = YW, \]

where the columns of \( W \) are proportional to the eigenvectors of \( C^{-1}C_D \) arranged in increasing (not decreasing) order of the corresponding eigenvalues. The index series constructed in this way can be treated in exactly the same way as those produced by singular spectrum analysis.

Because the matrix \( C^{-1}C_D \) is asymmetric, specialized methods are needed to extract its eigenvectors. Perhaps the simplest method is to use the following two-step application of singular spectrum analysis (Shapiro and Switzer 1989). First, apply singular spectrum analysis to \( Y \) and scale each of the resulting index series to have unit variance. Let

\[ X = YVE \]

be the \( (n - p + 1) \times p \) matrix whose columns are these scaled index series. Here, the columns of the \( p \times p \) matrix \( V \) are the unit eigenvectors of \( C \) arranged in decreasing order of the corresponding eigenvalues, and

\[ E = \text{diag}(\lambda_i^{-1/2}), \]

where \( \lambda_i, i = 1, 2, \ldots, p, \) are the eigenvalues of \( C \) arranged in decreasing order. Postmultiplication by \( E \) ensures that each of the columns of \( X \) has unit variance. For convenience, let \( U = VE. \) Next, apply singular spectrum analysis to the \( (n - p) \times p \) matrix \( \Delta \) whose columns are the first differences of the columns of \( X. \) The estimated covariance matrix of the columns of \( \Delta \) is \( C_\Delta = U'C_DU. \) Let \( T \) be the \( p \times p \) matrix whose columns are the eigenvectors of \( C_\Delta \) arranged in decreasing order of the corresponding eigenvalues. The \( (n - p + 1) \times p \) matrix \( Z \) whose columns are the final index series is given by
with the last column of \( Z \) corresponding to the first (i.e., smoothest) index series.

4. Illustration

In this section, the method outlined in section 3 is applied to the annual central England temperature series constructed by Manley (1974) and extended by Parker et al. (1992). The centered series, which covers the period of 1659–1991, is shown in Fig. 1. Plaut et al. (1995) applied singular spectrum analysis to this series, and their results are reproduced here for comparison. Following Plaut et al. (1995), a bandwidth of \( p = 40 \) years was used throughout the analysis, and only the first two components were associated with trends.

The weights for the first index series for both singular spectrum analysis and the alternative method are shown in Fig. 2. In both cases, the weights are scaled to have a unit sum of squares. The corresponding reconstructions are shown in Fig. 3. These reconstructions are very similar. The weights for the second index series for both methods are shown in Fig. 4. Each set has again been scaled to have a unit sum of squares. The projections of the original series onto the second index series are shown in Fig. 5. In this case, the projection onto the series produced by the alternative method is noticeably smoother. The reconstructions using both of the first two index series are shown in Fig. 6. Again, the reconstruction produced by the alternative method is noticeably smoother. This difference is accentuated in Fig. 7, where the same reconstructions are shown for the period of 1751–1850.

5. Discussion

Trend extraction involves a decomposition of a time series into low-frequency trends and high-frequency variability. In the absence of a physical or statistical model, the distinction between trends and variability is not always clear. Using the methods discussed in this paper, this distinction is implicit in the choice of bandwidth and the choice of the number of components that are retained. Although these are important issues, they are beyond the scope of this paper.

When interest centers on the trend, limited experience suggests that the differences between singular spectrum analysis and the method described in this paper are likely to be primarily cosmetic. In recent climatological work, interest has increasingly centered on characterizing variability. In that case, the differences between the two methods can be more significant. As illustrated in the previous section, trend estimates based on singular spectrum analysis tend to be rougher—that is, to incorporate more variability—than those based on the alternative method. This has the effect of reducing the variability in the residuals from the estimated trend, leading, in broad terms, to an underestimation of its magnitude.

Acknowledgments. The helpful comments of two anonymous reviewers are gratefully acknowledged.

REFERENCES