SPRINGER BRIEFS IN STATISTICS

Nina Golyandina
Anatoly Zhigljavsky

Singular Spectrum Analysis for Time Series

Springer

SpringerBriefs in Statistics

For further volumes:
http://www.springer.com/series/8921

Nina Golyandina · Anatoly Zhigljavsky

Singular Spectrum Analysis for Time Series

Springer

Nina Golyandina
Department of Mathematics
St. Petersburg University
St. Petersburg
Russia

Anatoly Zhigljavsky
School of Mathematics
Cardiff University
Cardiff
UK

ISSN 2191-544X ISSN 2191-5458 (electronic)
ISBN 978-3-642-34912-6 ISBN 978-3-642-34913-3 (eBook)
DOI 10.1007/978-3-642-34913-3
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012953018

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Chapter 1
Introduction

1.1 Preliminaries

Singular spectrum analysis (SSA) is a technique of time series analysis and forecasting. It combines elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing. SSA aims at decomposing the original series into a sum of a small number of interpretable components such as a slowly varying trend, oscillatory components and a 'structureless' noise. It is based on the singular value decomposition (SVD) of a specific matrix constructed upon the time series. Neither a parametric model nor stationarity-type conditions have to be assumed for the time series. This makes SSA a model-free method and hence enables SSA to have a very wide range of applicability.

The present book is fully devoted to the methodology of SSA. It exhibits the huge potential of SSA and shows how to use SSA both safely and with maximum effect.

Potential readers of the book. (a) Professional statisticians and econometricians; (b) specialists in any discipline where problems of time series analysis and forecasting occur; (c) specialists in signal processing and those needed to extract signals from noisy data; (d) PhD students working on topics related to time series analysis; (e) students taking appropriate MSc courses on applied time series analysis; (f) anyone interested in the interdisciplinarity of statistics and mathematics.

Historical remarks. The first publication, which can be considered as one of the origins of SSA (and more generally of the subspace-based methods of signal processing), can be traced back to the eighteenth century [28].

The commencement of SSA is usually associated with publication in 1986 of the papers [4, 5] by Broomhead and King. Since then SSA has received a fair amount of attention in literature. Additionally to [4, 5] the list of most cited papers on SSA published in the 1980s and 1990s includes [2, 10, 32, 33].

There are three books fully devoted to SSA, [8, 9, 14]. The book [9] is well written but it only provides a very elementary introduction to SSA. The volume [8] is a collection of papers written entirely by statisticians based at that time at St.Petersburg university. All these papers are devoted to the so-called 'Caterpillar'

N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*,
SpringerBriefs in Statistics, DOI: 10.1007/978-3-642-34913-3_1,

methodology (the words 'Caterpillar' or 'Gusenitsa' is due to the association with the moving window). This methodology is a version of SSA that was developed in the former Soviet Union independently (the 'iron curtain effect') of the mainstream SSA. The work on the 'Caterpillar' methodology has started long after publication of [28] but well before 1986, the year of publication of [4] and [5].

The main difference between the main-stream SSA of [2, 4, 5, 10, 32, 33] and the 'Caterpillar' SSA is not in the algorithmic details but rather in the assumptions and in the emphasis in the study of SSA properties. To apply the mainstream SSA, one often needs to assume some kind of stationarity of the time series and think in terms of the 'signal plus noise' model (where the noise is often assumed to be 'red'). In the 'Caterpillar' SSA, the main methodological stress is on separability (of one component of the series from another one) and neither the assumption of stationarity nor the model in the form 'signal plus noise' are required.

The main methodological principles described in [8] have been further developed in the monograph [14]. The publication of [14] has helped to attract much wider attention to SSA from the statistical circles as well as many other scientific communities. During the last 10 years much new SSA-related research has been done and many new successful applications of SSA have been reported. A recent special issue of 'Statistics and Its Interface' [35] gives an indication of how much progress in theoretical and methodological developments of SSA, as well as its applications, has been achieved in recent years. The SSA community regularly organizes international workshops on SSA. The latest SSA workshop was held in Beijing in May 2012, see http://www.cefs.ac.cn/express/SSA.html.

The research on the theory and methodology of SSA performed in the last two decades has resulted in a rather pleasing state of affairs: (i) the existence of an active SSA community and (ii) the existence of a general methodology of SSA rather than simply a collection of many different SSA algorithms. This methodology unifies different versions of SSA into a very powerful tool of time series analysis and forecasting. Description of SSA methodology is the sole purpose of the present book.

Correspondence between the present book and [14]. Some entirely new topics are included (for example, Sect. 3.7–3.9) but a few topics thoroughly described in [14] are not considered at all (see, for example, [14, Chap. 3]). This volume is fully devoted to the methodology of SSA unlike [14], where many theoretical issues were also considered. The material is correspondingly revised in view of the new objectives. The main aim of [14] is to establish SSA as a serious subject. There is no need to do it now and the aspiration of this book is to show the power and beauty of SSA to as wide audience as possible.

Several reasons why SSA is still not very popular among statisticians. First reason is tradition: SSA is not a classical statistical method, and therefore many people are simply not aware of it. Second, SSA demands more computing power than the traditional methods.Third, many people prefer model-based statistical techniques where calculations are automatic and do not require the computer-analyst interaction. Finally, SSA is sometimes too flexible (especially when analyzing multivariate series) and therefore has too many options which are difficult to formalize.

Links between SSA and other methods of time series analysis. SSA has no links with ARIMA, GARCH and other methods of this type and also with wavelets. However, SSA has very close links with some methods of signal processing and with methods of multivariate statistics like principal component analysis and projection pursuit; see Sect. 2.5.4 and 3.8.The so-called Empirical Mode Decomposition (EMD), see [20], is intended to solve similar problems to SSA but there are significant conceptual and methodological differences between SSA and EMD.

Structure of the next three sections. In the next section, we give a short introduction into SSA methodology and simultaneously into the material of the present book. Then we mention important issues related to SSA, which did not find their way into this book. Finally, we provide a list of most common symbols and acronyms.

Acknowledgements. The authors are very much indebted to Vladimir Nekrutkin, their coauthor of the monograph [14]. His contribution to the methodology and especially theory of SSA cannot be underestimated. The authors very much acknowledge many useful comments made by Jon Gillard. The authors are also grateful to former and current Ph.D. students and collaborators of Nina Golyandina: Konstantin Usevich (a specialist in algebraic approach to linear recurrence relations), Theodore Alexandrov (automatic SSA), Andrey Pepelyshev (SSA for density estimation), Anton Korobeynikov (fast computer implementation of SSA), Eugene Osipov and Marina Zhukova (missing data imputation), and Alex Shlemov (SSA filtering). Help of Alex Shlemov in preparation of figures is very much appreciated.

1.2 SSA Methodology and the Structure of the Book

The present volume has two chapters. In Chap. 2, SSA is typically considered as a model-free method of time series analysis. The applications of SSA dealt with in Chap. 3 (including forecasting) are model based and use the assumption that the components of the original time series extracted by SSA satisfy linear recurrence relations.

The algorithm of Basic SSA (Sect. 2.1). A condensed version of Basic SSA (which is the main version of SSA) can be described as follows.

Let $\mathbb{X}_N = (x_1, \ldots, x_N)$ be a time series of length N. Given a window length L $(1 < L < N)$, we construct the L-lagged vectors $X_i = (x_i, \ldots, x_{i+L-1})^{\mathrm{T}}$, $i = 1, 2, \ldots, K$, where $K = N - L + 1$, and compose these vectors into the trajectory matrix $\mathbf{X}$.

The columns X_j of $\mathbf{X}$ can be considered as vectors in the L-dimensional space R^L. The eigendecomposition of the matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ (equivalently, the SVD of the matrix $\mathbf{X}$) yields a collection of L eigenvalues and eigenvectors. A particular combination of a certain number r of these eigenvectors determines an r-dimensional subspace $\mathcal{L}_r$ in R^L, $r < L$. The L-dimensional data $\{X_1, \ldots, X_K\}$ is then projected onto the subspace $\mathcal{L}_r$ and the subsequent averaging over the diagonals yields some Hankel matrix $\widetilde{\mathbf{X}}$. The time series $(\tilde{x}_1, \ldots, \tilde{x}_N)$, which is in the one-to-one correspondence with matrix $\widetilde{\mathbf{X}}$, provides an approximation either the whole series $\mathbb{X}_N$ or a particular component of $\mathbb{X}_N$.

Basic SSA and models of time series (Sect. 2.3). As a non-parametric and model-free method, Basic SSA can be applied to any series. However, for interpreting results of analysis and making decisions about the choice of parameters some models may be useful. The main assumption behind Basic SSA is the assumption that the time series can be represented as a sum of different components such as trend (which we define as any slowly varying series), modulated periodicities, and noise. All interpretable components can be often approximated by time series of small rank, and hence can be described via certain LRRs (linear recurrence relations). Separating the whole series into these components and analysis of the LRRs for interpretable components helps in getting reliable and useful SSA results.

Potential of Basic SSA (Sect. 2.2 and also Sect. 3.7, 3.8 and 3.9). The list of major tasks, which Basic SSA can be used for, includes smoothing, noise reduction, extraction of trends of different resolution, extraction of periodicities in the form of modulated harmonics, estimation of volatility, etc. These tasks are considered in Sect. 2.2. The following more advanced abilities (but model-based) of SSA are considered in the final three sections of Chap. 3: the use of SSA for filling in missing values is considered in Sect. 3.7; add-ons to Basic SSA permitting estimation of signal parameters are considered in Sect. 3.8; finally, Basic SSA as a filtration tool is studied in Sect. 3.9. Methodologically, these last three topics are closely linked with the problem of SSA forecasting. Note also that all major capabilities of Basic SSA are illustrated on real-life time series.

Choice of parameters in Basic SSA (Sect. 2.4). There are two parameters to choose in Basic SSA: the window length L and the group of r indices which determine the subspace $\mathcal{L}_r$. A rational or even optimal choice of these parameters should depend on the task we are using SSA for. The majority of procedures require interactive (including visual) identification of components. An automatic choice of parameters of Basic SSA could be made, see Sect. 2.4.5. However, the statistical procedures for making this choice are modelbased. Success in using the corresponding versions of SSA depends on the adequacy of the assumed models and especially on achieving good separability of the time series components.

Toeplitz SSA (Sect. 2.5.3). Basic SSA can be modified and extended in many different ways, see Sect. 2.5. As a frequently used modification of Basic SSA, consider a common application of SSA for the analysis of stationary series, see Sect. 2.5.3. Under the assumption that the series $\mathbb{X}_N$ is stationary, the matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ of Basic SSA can be replaced with the so-called lag-covariance matrix $\mathbf{C}$ whose elements are $c_{ij} = \frac{1}{N-k}\sum_{t=1}^{N-k} x_t x_{t+k}$ with $i, j = 1, \ldots, L$ and $k = |i - j|$. In the book, this version of SSA is called 'Toeplitz SSA'.[1] Unsurprisingly, if the original series is stationary then Toeplitz SSA slightly outperforms Basic SSA. However, if the series

[1] In the literature on SSA, Basic SSA is sometimes called BK SSA and what we call 'Toeplitz SSA' is called VG SSA; here BK and VG stand for Broomhead & King [4, 5] and Vautard & Ghil [32], respectively.

is not stationary then the use of Toeplitz SSA may yield results which are simply wrong.

SSA–ICA (Sect. 2.5.4). Another important modification of Basic SSA can be viewed as a combination of SSA and Independent Component Analysis (ICA), see Sect. 2.5.4. This algorithm, called SSA–ICA, helps to separate time series components that cannot be separated with the help of the SVD alone, due to the lack of strong separability. Despite we deal with deterministic time series components but ICA is developed for dealing with random series and processes, the algorithm of ICA can be formally applied for achieving a kind of independence of components. Note that it is not a good idea to use the ICA as a full replacement of the SVD since the ICA is a much less stable procedure than the SVD. Therefore, a two-stage procedure is proposed, where the SVD is performed for the basic decomposition and then some version of the ICA (or the projection pursuit) is applied to those components which are produced by the SVD but remain mixed up.

Computational aspects of SSA (Sects. 2.5.6 and 2.5.7). If L is very large then the conventional software performing SVD decomposition may be computationally costly. In this case, the Partial SVD and other techniques can be used for performing fast computations. This can be achieved either by using clever implementations (Sect. 2.5.6) or by replacing the SVD with simpler procedures (Sect. 2.5.7).

SSA forecasting (Sects. 3.1 – 3.6). Time series forecasting is an area of huge practical importance and Basic SSA can be very effective for forecasting. The main idea of SSA forecasting is as follows.

Assume that $\mathbb{X}_N = \mathbb{X}_N^{(1)} + \mathbb{X}_N^{(2)}$ and we are interested in forecasting of $\mathbb{X}_N^{(1)}$. If $\mathbb{X}_N^{(1)}$ is a time series of finite rank, then it generates some subspace $\mathcal{L}_r \subset \mathsf{R}^L$. This subspace reflects the structure of $\mathbb{X}_N^{(1)}$ and can be taken as a base for forecasting. Under the conditions of separability between $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$ (these conditions are discussed throughout the volume; see, for example, Sects. 2.3.3, 2.4, 3.3.1 and 3.5), Basic SSA is able to accurately approximate $\mathcal{L}_r$ and hence it yields an LRR which approximates the true LRR and can be directly used as a forecasting formula. This method of forecasting is called recurrent forecasting and considered in Sect. 3.3, as well as few other sections. Alternatively, we may use the so-called 'vector forecasting' which main idea is in the consecutive construction of the vectors $X_i = (x_i, \ldots, x_{i+L-1})^{\mathrm{T}}$, for $i = K+1, K+2, \ldots$ so that they lie as close as possible to the subspace $\mathcal{L}_r$.

Short-term forecasting makes very little use of the model while responsible forecasting for long horizons is only possible when an LRR is built by SSA and the adequacy of this LRR is testified. As demonstrated in Sect. 3.3, in addition to the LRRs, SSA forecasting methods use the characteristic polynomials associated with these LRRs. The precision of SSA forecasting formulas depends on the location of the roots of these polynomials. In Sect. 3.2 we provide an overview of the relations between LRRs, the characteristic polynomials and their roots and discuss properties of the so-called min-norm LRRs which are used for estimating parameters of the signal (see Sect. 3.8), in addition to forecasting.

In forecasting methodology, the construction of confidence intervals for the forecasts is often an essential part of the procedure. Construction of these intervals for

SSA forecasts is discussed in Sect. 3.4. Despite SSA itself being a model-free technique, for building confidence intervals we need to make certain assumptions such as that the residual series is a realization of a stochastic white noise process.

In Sect. 3.5 we give recommendations on the choice of forecasting parameters and in Sect. 3.6 we discuss results of a case study. We argue that stability of forecasts is the major aim we have to try to achieve in the process of building forecasts. Forecast stability is highly related to the forecast precision and forecast reliability.

SSA for missing value imputation (Sect. 3.7). Forecasting can be considered as a special case of missing value imputation if we assume that the missing values are located at the end of the series. We show how to extend some SSA forecasting procedures (as well as methods of their analysis) to this more general case.

Parameter estimation in SSA and signal processing (Sect. 3.8). Although there are many similarities between SSA and the subspace-based methods of signal processing, there is also a fundamental difference between these techniques. This difference lies in the fact that the model of the form 'signal plus noise' is obligatory in signal processing; consequently, the main aim of the signal processing methods is the estimation of the parameters of the model (which is usually the sum of damped sinusoids). The aims of SSA analysis are different (for instance, splitting the series into components or simply forecasting) and the parameters of the approximating time series are of secondary importance. This fundamental difference between the two approaches leads, for example, to different recommendations for the choice of the window length L: a typical recommendation in Basic SSA is to choose L reasonably large while in the signal processing methods L is typically relatively small.

Causal SSA (Sect. 3.9.5). Causal SSA (alternatively, Last Point SSA) can be considered as an alternative to forecasting. In Causal SSA, we assume that the points in the time series $\mathbb{X}_\infty = (x_1, x_2, \ldots)$ arrive sequentially, one at a time. Starting at some $M_0 > 0$, we apply Basic SSA with fixed window length and the grouping rule to the series $\mathbb{X}_M = (x_1, \ldots, x_M)$ for all $M \geq M_0$. We then monitor how SSA reconstruction of previously obtained points of the series change as we increase M (this is called redrawing). The series consisting of the last points of the reconstructions is the result of Causal SSA. The delay of the Causal SSA series reflects the quality of forecasts based on the last points of the reconstructions. Additionally to the redrawings of the recent points of the reconstructions, this delay can serve as an important indicator of the proper choice of the window length, proper grouping and in general, predictability of the time series. This could be of paramount importance for the stock market traders when they try to decide whether a particular stock is consistently decreasing/increasing its value or they only observe market fluctuations.

1.3 SSA Topics Outside the Scope of This Book

Theory of SSA. For the basic theory of SSA we refer to the monograph [14]. Since the publication of that book, several influential papers on theoretical aspects of SSA have been published. The main theoretical paper on perturbations in SSA and