Statistics and Its Interface Volume 10 (2017) 1–3

Introduction to this special issue

Shouyang Wang, Heping Zhang, and Anatoly Zhigljavsky

This special issue can be roughly divided into three parts. The first part includes papers [1–6] and is devoted to the methodology and theory of the time series analysis technique which is called Singular Spectrum Analysis (SSA). The second part contains papers [7–10] and deals with some other methodological aspects of time series analysis. The third part is composed of papers [11–14], where the main attention is paid to applications of time series analysis techniques.

The key methodology, which the majority of papers in this issue are related to, is SSA. The most common version of SSA is called 'Basic SSA'. A short description of it is given below.

Let $x_1, \ldots, x_N$ be a time series of length N. Given a window length L $(1 < L < N)$, we construct the L-lagged vectors $X_i = (x_i, \ldots, x_{i+L-1})^T$, $i = 1, 2, \ldots, K = N-L+1$, and compose these vectors into the matrix $\mathbf{X} = (x_{i+j-1})_{i,j=1}^{L,K} = [X_1 : \ldots : X_K]$. This matrix has size $L \times K$ and is often called 'trajectory matrix'. It is a Hankel matrix, which means that all the elements along the diagonal $i+j = \text{const}$ are equal. The columns X_j of $\mathbf{X}$ can be considered as vectors in the L-dimensional space $\mathbb{R}^L$. The singular value decomposition (SVD) of the matrix $\mathbf{X}\mathbf{X}^T$ yields a collection of L eigenvalues and eigenvectors. A particular combination of a certain number l of these eigenvectors determines an l-dimensional subspace $\mathbb{L}$ in $\mathbb{R}^L$, $l < L$. The L-dimensional data $\{X_1, \ldots, X_K\}$ is then projected onto the subspace $\mathbb{L}$ and the subsequent averaging over the diagonals yields some Hankel matrix $\tilde{\mathbf{X}}$ which is considered as an approximation to $\mathbf{X}$. The time series $\tilde{x}_1, \ldots, \tilde{x}_N$, which is in the one-to-one correspondence with the matrix $\tilde{\mathbf{X}}$, provides an approximation to the original series $x_1, \ldots, x_N$.

The reconstructed time series $\tilde{x}_1, \ldots, \tilde{x}_N$ can be used for extraction of trends and periodics, for constructing a 'sum of dumped sinusoids' model, see [6] in this issue, for forecasting, for monitoring structure of time series, for testing causality in multivariate series and for many other tasks. Methodological aspects of SSA have been discussed in many publications including the majority of papers in the special issue of the Statistics and Its Interface (2010, vol. 3, No. 3), which was fully devoted to SSA and its applications. For a comprehensive discussion of SSA methodology we refer to the monograph [15]; for a short 4-page introduction to SSA, see [16].

In this issue, SSA is being developed in the following directions.

In [1], the basic version of SSA is modified by considering a different matrix norm, in place of the standard Frobenius norm, at the SVD step of SSA. It is demonstrated on a few examples that the quality of SSA approximation improves if the matrix weights are chosen to obtain approximately equal weights for the individual observations in the equivalent problem of time-series least-squares approximation. Moreover, the question of convergence of the iterated SSA (the method known in signal processing as 'Cadzow iterations') is investigated. This question was a source of some controversy in the literature in signal processing, see [1] for references.

In [2], the SVD step of the basic version of SSA is suggested to be replaced by a solution of an low-rank approximation problem in the nuclear norm rather than in the standard Frobenius norm. The main suggestion in [2] tries to emulate the success of so-called matrix completion problem (the problem of imputing missing values of a matrix), where the use of the nuclear norm is now a common thing. Application of the modified in this way SSA to to imputing missing data and forecasting are discussed. A method similar to the Cadzow iterations but based on the nuclear norm is discussed and illustrated on a number of real-world time series.

Multivariate SSA is a direct extension of the standard SSA for simultaneous analysis of several time series. Assume that we have two series, $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_N\}$. The (joint) trajectory matrix of the two-variate series (X, Y) can be defined as either $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ or $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})^T$, where $\mathbf{X}$ and $\mathbf{Y}$ are the trajectory matrices of the individual series X and Y. Matrix $\mathbf{Z}$ is block-Hankel rather than simply Hankel. Other stages of MSSA are identical to the ones of the univariate SSA except that we build a block-Hankel (rather than ordinary Hankel) approximation $\tilde{\mathbf{Z}}$ to the trajectory matrix $\mathbf{Z}$. Multivariate SSA may be very useful for analyzing several series with common structure and for establishing causality between two series. Multivariate SSA is analysed in [3]. This paper is a follow-up of [17] where a general approach to asymptotic proximity of unperturbed and perturbed signal subspaces in Multivariate SSA has been developed. It is assumed in [3], that each coordinate of a multidimensional signal produces the same signal subspace. For such signals, the authors of [3] suggest a solution for an asymptotic extraction of this subspace from the perturbed multidimensional signal series and illustrate this technique on several examples.

The topic of stability of Multivariate SSA under random perturbations of the input time series is also studied in [4] with the main emphasis on the stability of forecasts. In [4], the reconstruction kernel of SSA is considered as a convolution filter and concise formulae for the variance under perturbations of SSA forecasts are derived. On the base of the behaviour of these variances under scaling of the support series, the authors of [4] formulate and study a natural criterion of supportiveness of one series by the other; this can also be considered as a causality criterion.

In [5], a modification of the Basic SSA, called SSA with projection, is proposed. This version of SSA is able to take into consideration a structure given in advance. SSA with projection includes preliminary projection of rows and columns of the trajectory matrix $\mathbf{X}$ to given subspaces. A natural application of SSA with projection is the extraction of polynomial trends. It is demonstrated in [5] that SSA with projection can extract polynomial trends much more accurately than Basic SSA, especially in the case of linear trends. Numerical examples, including comparison with the least-squares polynomial regression, are presented which confirm that the proposed version of SSA could be extremely useful for analyzing time series where an additional information about trend is available.

As mentioned above, the main model of the signal in SSA is the 'sum of dumped sinusoids' model. The authors of [6] postulate this model and apply Lipschitz optimization methods for fitting this model. They show that the arising optimization problem is very difficult in view of presence of many local minima in the objective function. It is shown in [6] how Lipschitz-based deterministic methods can be adapted for studying these challenging global optimization problems, when a limited computational budget is given and some guarantee of the found solution is required.

In [7] a model of time series, very much related to the SSA model, is considered. It is assumed that the time series contains a trend, a seasonal component and a periodically correlated time series. For analyzing such series, a semiparametric three-step method is proposed. The seasonal component and trend are estimated in [7] by means of B-splines, and the Yule-Walker estimates of the time series model coefficient are calculated via the residuals after removing the estimated seasonality and trend. The oracle efficiency of the proposed Yule-Walker type estimators is established. Simulation studies confirm the theoretical findings. The proposed method is used for the analysis of the monthly global temperature data provided by the National Space Science and Technology Center.

The authors of [8] notice that econometric and financial data often take the form of a collection of curves observed consecutively over time. Such curves can be viewed as functional time series. A fundamental issue that must be addressed, before an attempt is made to statistically model or predict such series, is whether they can be assumed to be stationary with a possible deterministic trend. This paper extends a well-known KPSS test to the setting of functional time series. Two testing procedures are proposed in [8]: Monte Carlo and asymptotic. The limit distributions of the test statistics are specified, the procedures are algorithmically described and illustrated by application to yield curves and daily price curves.

In [9] the problem of efficient financial surveillance is considered aimed at quickest detection of structural breaks in live-monitored financial time series. A semi-parametric multi-cyclic change-point detection procedure is proposed to promptly spot anomalies as they occur in the time series under surveillance. The proposed procedure is a modification of the well-known likelihood ratio-based Shiryaev-Roberts procedure. It is compared with the celebrated CUSUM test on a set of real-world financial data. While both procedures perform well, the proposed method shows slightly better results.

In [10], the problem of estimation of parameters in the harmonic regression with cyclically dependent errors is addressed. Asymptotic properties of the least-squares estimates are analyzed by simulation experiments. In particular, the authors establish that consistency and asymptotic normality of the least-squares estimator of unknown parameters hold under different scenarios.

In [11], A hybrid transfer learning model is suggested for crude oil price forecasting. While most of the existing models for oil price forecasting only use the data in the forecasted time series, the authors of [11] propose a hybrid transfer learning model. It first selectively transfers some related time series in the source domain to assist in modeling the target time series by transfer learning technique, and then constructs the forecasting model by the so-called analog complexing method. The optimal match between two important parameters in this model is found numerically using some global optimization techniques. Two main crude oil price time series, West Texas Intermediate crude oil spot price and Brent crude oil spot price are used for empirical analysis, and the results show the effectiveness of the proposed model.

In [12], a time-heterogeneous generalised Pareto distribution is fit to the flood heights in the lower Limpopo River basin of Mozambique. The maximum likelihood method is used for parameter estimation. An in-depth review of the merits of peaks-over-threshold and block maxima is provided. A relationship between generalised extreme value distribution and the generalised Pareto distribution is studied. Nonstationary time-dependent models with a trend in the scale parameter are also considered. The results show overwhelming evidence in support of the existence of a linear trend in the scale parameter of the generalised Pareto distribution models at all the three sites in the lower Limpopo River basin. The models developed in [12] seem to be more reliable than their stationary counterparts for planning and decision making processes in Mozambique.

In [13], the forecasts of Earth temperature records made in [18] are compared with the data actually observed during 2010–2014. It is demonstrated that the forecasts made in [18] are quite accurate. In the second part of [13], the SSA-based change-point detection algorithm proposed in [19] is applied to the same temperature records data. The results show that the data does not have essential structural breaks except perhaps a small rise of the general level of temperatures at around 1998.

The authors of [14] have continued the research started in their previous paper [18]. They have applied SSA to forecast the Earth temperature records taken from the website of the National Space Science and Technology Center, USA, NASA. They have demonstrated that the forecasts of [18] were quite accurate. The Earth temperatures are also forecasted for the next several years. These forecasts show that the temperatures are not going to be too different from the ones we observe at present so that in a near future the Earth temperatures are not likely to visibly increase or decrease but will continue to be volatile. This SSA analysis of temperatures is complemented in [14] with analysis of the Oceanic Nino Index and Arctic and Antarctic sea ice extents.

We believe this special issue will be a valuable addition to the literature on time series analysis and many readers will find interesting papers related either to the methodology of time series analysis or particular applications.

REFERENCES

[1] Zvonarev, N. and Golyandina, N. (2017). *Iterative algorithms for weighted and unweighted finite-rank time-series approximations*. Statistics and Its Interface, **10**, No. 1, 5–18.

[2] Butcher, H. and Gillard, J. (2017). *Simple nuclear norm based algorithms for imputing missing data and forecasting in time series*. Statistics and Its Interface, **10**, No. 1, 19–25.

[3] Nekrutkin, V. and Vasilinetc, I. (2017). *Asymptotic extraction of common signal subspaces from perturbed signals*. Statistics and Its Interface, **10**, No. 1, 27–32.

[4] Vronskaya-Robl, M. and Schmidt, K. M. (2017). *On perturbation stability of SSA and MSSA forecasts and the supportiveness of time series*. Statistics and Its Interface, **10**, No. 1, 33–46.

[5] Golyandina, N. and Shlemov, A. (2017). *Semi-nonparametric singular spectrum analysis with projection*. Statistics and Its Interface, **10**, No. 1, 47–57.

[6] Gillard, J. W. and Kvasov, D. E. (2017). *Lipschitz optimization methods for fitting a sum of damped sinusoids to a series of observations*. Statistics and Its Interface, **10**, No. 1, 59–70.

[7] Tang, L. and Shao, Q. (2017). *Semiparametric analysis for environmental time series*. Statistics and Its Interface, **10**, No. 1, 71–79.

[8] Kokoszka, P. and Young, G. (2017). *Testing trend stationarity of functional time series with application to yield and daily price curves*. Statistics and Its Interface, **10**, No. 1, 81–92.

[9] Pepelyshev, A. and Polunchenko, A. S. (2017). *Real-time financial surveillance via quickest change-point detection methods*, **10**, No. 1, 93–106.

[10] Frías, M. P., Ivanov, A. V., Leonenko, N., Martínez, F., and Ruiz-Medina, M. D. (2017). *Detecting hidden periodicities for models with cyclical errors*. Statistics and Its Interface, **10**, No. 1, 107–118.

[11] Xiao, J., Hu, Y., Xiao, Y., Xu, L., and Wang, S. (2017). *A hybrid transfer learning model for crude oil price forecasting*. Statistics and Its Interface, **10**, No. 1, 119–130.

[12] Maposa, D., Cochran, J. J., and Lesaoana, M. (2017). *Modelling extreme flood heights in the lower Limpopo River basin of Mozambique using a time-heterogeneous generalised Pareto distribution*. Statistics and Its Interface, **10**, No. 1, 131–144.

[13] Kornikov, V. (2017). *Assessment of SSA predictions of Earth temperature records*. Statistics and Its Interface, **10**, No. 1, 145–150.

[14] Pepelyshev, A. and Zhigljavsky, A. (2017). *SSA analysis and forecasting of records for Earth temperature and ice extents*. Statistics and Its Interface, **10**, No. 1, 151–163.

[15] Golyandina, N. and Zhigljavsky, A. (2013). *Singular Spectrum Analysis for Time Series*. Springer. MR3024734

[16] Zhigljavsky, A. (2010). *Singular spectrum analysis for time series: Introduction to this special issue*. Statistics and Its Interface, **3**, No. 3, 255–258. MR2720131

[17] Nekrutkin, V. (2010). *Perturbation expansions of signal subspaces for long signals*. Statistics and Its Interface, **3**, No. 3, 297–319. MR2720134

[18] Pepelyshev, A. and Zhigljavsky, A. (2010). *Assessing stability of long-horizon SSA forecasting*. Statistics and Its Interface, **3**, No. 3, 321–327. MR2720135

[19] Moskvina, V. and Zhigljavsky, A. (2003). *An algorithm based on singular spectrum analysis for change-point detection*. Communications in Statistics: Simulation and Computation, **32**, No. 2, 319–352. MR1983338

Shouyang Wang
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
Beijing 100190
China
E-mail address: sywang@amss.ac.cn

Heping Zhang
Yale School of Public Health
60 College Street
New Haven, CT 06520-8034
USA
E-mail address: heping.zhang@yale.edu

Anatoly Zhigljavsky
School of Mathematics
Cardiff University
Cardiff CF24 4AG
UK
E-mail address: ZhigljavskyAA@cf.ac.uk