ARTICLE IN PRESS

International Journal of Forecasting ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting stochastic processes using singular spectrum analysis: Aspects of the theory and application

M. Atikur Rahman Khan[a], D.S. Poskitt[b,*]

[a] Commonwealth Scientific and Industrial Research Organisation, Research Way, Clayton, VIC 3168, Australia

[b] Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia

ARTICLE INFO

Keywords:
Linear recurrent formula
Mean squared forecast error
Signal dimension
Window length

ABSTRACT

This paper presents theoretical results on the properties of forecasts obtained by using singular spectrum analysis to forecast time series that are realizations of stochastic processes. The mean squared forecast errors are derived under broad regularity conditions, and it is shown that, in practice, the forecasts obtained will converge to their population ensemble counterparts. The theoretical results are illustrated by examining the performances of singular spectrum analysis forecasts when applied to autoregressive processes and a random walk process. Simulation experiments suggest that the asymptotic properties developed are reflected in the behaviour of observed finite samples. Empirical applications using real world data sets indicate that forecasts based on singular spectrum analysis are competitive with other methods currently in vogue.

1. Introduction

Singular spectrum analysis (SSA) is a nonparametric technique that is designed for use in signal extraction and the prediction of irregular time series that may exhibit non-stationary and nonlinear properties, as well as intermittent or transient behaviour. The development of SSA is often attributed to researchers working in the physical sciences, namely Broomhead and King (1986), Vautard and Ghil (1989) and Vautard, Yiou, and Ghil (1992), although many of the basic building blocks were outlined by Basilevsky and Hum (1979) in a socioeconomic setting, and an early formulation of some of the key ideas can be found in the work of Prony (1795). An introduction to SSA is presented by Elsner and Tsonis (1996), and a more detailed examination of the methodology, with an emphasis on the algebraic structure and algorithms, is provided by Golyandina, Nekrutkin, and Zhigljavski (2001).

The application of SSA to forecasting has gained popularity over recent years (see for example Hassani, Heravi, & Zhigljavsky, 2009, Hassani, Soofi, & Zhigljavsky, 2010, Hassani & Zhigljavsky, 2009 and Thomakos, Wang, & Wille, 2002, for applications in business and economics), and the general finding appears to be that SSA performs well. These studies have examined SSA forecasts by investigating real world applications and comparing the performance of SSA to those of other benchmarks like ARIMA models and Holt–Winters procedures. However, with real world data the true data generating mechanism is not known, and making a comparison with such benchmarks does not convey the full picture: knowing that SSA outperforms a benchmark serves only to show that the benchmark is suboptimal, and therefore does not provide an appropriate baseline.

In this paper, our purpose is to provide what we believe to be the first theoretical analysis of the forecasting performance of SSA under appropriate regularity conditions

* Corresponding author.
E-mail addresses: Atikur.Khan@csiro.au (M.A.R. Khan), Donald.Poskitt@monash.edu (D.S. Poskitt).

http://dx.doi.org/10.1016/j.ijforecast.2016.01.003

concerning the true data generating mechanism. We present a formulation of the SSA mean squared forecast error (MSFE) for a general class of processes. The usefulness of such formulae lies not only in the fact that they provide a neat mathematical characterization of the SSA forecast error, but also in the fact that they allow a comparison to be made between SSA and the optimal mean squared error solution for a known random processes. The minimal mean squared error (MMSE) predictor obviously provides a (*gold*) standard against which all other procedures can be measured.

Irrespective of the actual structure of the observed process, SSA forecasts are obtained by calculating a linear recurrence formula (LRF) that is used to construct a prediction of the future value(s) of the realized time series. Given a univariate time series of length N, the coefficients of the LRF are computed from a spectral decomposition of an $m \times n$ dimensional Hankel matrix known as the trajectory matrix. The dimension m is called the window length, and $n = N - m + 1$ is referred to as the window width. The Gramian of the trajectory matrix is constructed for a known window length, and the eigenvalue decomposition of the Gramian evaluated. This is then used to decompose the observed series into a signal component, constructed from k eigentriples of the Hankel matrix (the first k left and right hand eigenvalues and their associated singular values), and a residual. The resulting signal plus noise decomposition is then employed to produce a forecast via the LRF coefficients. Details are presented in the following section, where we outline the basic structure of the calculations underlying the construction of a SSA(m, k) model and the associated forecasts.

Section 3 presents the theoretical MSFE of a SSA(m, k) model under very broad assumptions. The formulae that we derive indicate how the use of different values of m, a tuning parameter, and k, a modeling parameter, will interact to influence the MSFE obtained from a given SSA(m, k) model. In Section 4, it is shown that, when appropriate regularity conditions are satisfied, the SSA forecasts constructed in practice, and their associated MSFE estimates, will converge to their theoretical population ensemble counterparts.

Section 5 illustrates the theoretical results obtained in Sections 3 and 4. In forecasting applications, it is common practice to assume implicitly that the fitted model is correct, and therefore that the forecasting formulae derived from the model are appropriate; however, such an assumption rarely holds true. In general, the true data generating process (DGP) is unknown, and the fitted model will only provide, at best, a close approximation to the true DGP. Hence, the expectation is that the forecasting performance of a fitted model will be sub-optimal, and therefore it is natural to ask in what ways and to what extent the forecasting performance of the fitted model will fall short. In an attempt to address this question, Section 5 examines the MSFE performances of different SSA(m, k) models and compares them with those of the optimal MSE predictors for known DGPs.

Section 6 demonstrates the application of SSA forecasting to different real world time series. It shows that SSA forecasts can provide considerable improvements in empirical MSFE performances over the conventional benchmark models that have been used previously to characterize these series. Section 7 presents a brief conclusion.

2. The mechanics of SSA forecasting

Singular spectrum analysis (SSA) is based on the basic idea that there is an isomorphism between an observed time series $\{x(t) : t = 1, \ldots, N\}$ and the vector space of $m \times n$ Hankel matrices, defined by the mapping

$$\{x(t) : t = 1, \ldots, N\} \mapsto \boldsymbol{X} = \begin{bmatrix} x(1) & x(2) & \cdots & x(n) \\ x(2) & x(3) & \cdots & x(n+1) \\ \vdots & \vdots & & \vdots \\ x(m) & x(m+1) & \cdots & x(N) \end{bmatrix} = [\boldsymbol{x}_1 : \ldots : \boldsymbol{x}_n], \quad (1)$$

where m is a preassigned window length, $n = N - m + 1$, $\boldsymbol{x}_t = (x(t), x(t+1), \ldots, x(t+m-1))'$, and the so called trajectory matrix $\boldsymbol{X} = [x(i + t - 1)]$ for $i = 1, \ldots, m$ and $t = 1, \ldots, n$. Let $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_m > 0$ denote the eigenvalues of $\boldsymbol{XX}'$ arranged in descending order of magnitude, and $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m$ the corresponding orthonormal system of eigenvectors. The trajectory matrix can be expressed as $\boldsymbol{X} = \sum_{i=1}^{m} \boldsymbol{X}_i$, the sum of m rank one projections $\boldsymbol{X}_i = \sqrt{\ell_i}\boldsymbol{u}_i\boldsymbol{v}_i' = \boldsymbol{u}_i\boldsymbol{u}_i'\boldsymbol{X}$, where $\boldsymbol{u}_i$ and $\boldsymbol{v}_i = \boldsymbol{X}'\boldsymbol{u}_i/\sqrt{\ell_i}$, $i = 1, \ldots, m$, are the left and right eigenvectors of $\boldsymbol{X}$. Now suppose that a large proportion of the total variation in $\boldsymbol{XX}'$ can be associated with a subset of dominant eigentriples $\{\ell_i, \boldsymbol{u}_i, \boldsymbol{v}_i\}$, $i = 1, \ldots, k$. The projection of $\boldsymbol{X}$ onto the space spanned by $\boldsymbol{u}_i$, $i = 1, \ldots, k$, $\boldsymbol{S}_k = \sum_{i=1}^{k} \boldsymbol{X}_i$, can then be viewed as the component of $\boldsymbol{X}$ that is due to the presence of a signal in the original series, where k is the designated dimension of the signal, and the remainder $\boldsymbol{E}_k = \sum_{i=k+1}^{m} \boldsymbol{X}_i$ can be taken as the component due to noise. This will be referred to henceforth as an SSA(m, k) model.

Suppose that a SSA(m, k) model has been fitted to time series data $x(1), x(2), \ldots, x(N)$. Since $\boldsymbol{S}_k$ has rank $k < m$, there exists an $m \times (m-k)$ matrix $\boldsymbol{P}$ with columns that span the null space of $\boldsymbol{S}_k$, implying that $\boldsymbol{P}'\boldsymbol{S}_k = \boldsymbol{0}$, and hence, that the last row of $\boldsymbol{S}_k$ can be expressed as a linear combination of the first $m - 1$ rows. This in turn implies that, in the terminology of SSA, the signal satisfies a linear recurrent formula (LRF), namely $s(t) = \sum_{j=1}^{m-1} a_j s(t - m + j)$, where the coefficients $a_1, \ldots, a_{m-1}$ in the LRF are calculated by forming the projection of $\boldsymbol{S}_k^l$, the last row of the signal component $\boldsymbol{S}_k$, onto $\boldsymbol{S}_k^u$, its first $m - 1$ rows.

Lemma 1. *Let* $\boldsymbol{U}_k = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k]$ *denote the matrix containing the first* k *eigenvectors of* $\boldsymbol{XX}'$, *and let* $a_1, \ldots, a_{m-1}$ *denote the coefficients formed by projecting* $\boldsymbol{S}_k^l$, *the last row of* $\boldsymbol{S}_k = \boldsymbol{U}_k\boldsymbol{U}_k'\boldsymbol{X}$, *onto* $\boldsymbol{S}_k^u$, *its first* $m - 1$ *rows. Then,* $(a_1, \ldots, a_{m-1}) = (1 - \boldsymbol{U}_k^l\boldsymbol{U}_k^{l'})^{-1}\boldsymbol{U}_k^l\boldsymbol{U}_k^{u'}$, *where* $\boldsymbol{U}_k^l$ *is the last row of* $\boldsymbol{U}_k$ *and* $\boldsymbol{U}_k^u$ *is the matrix containing the first* $m - 1$