

Singular Spectrum Analysis Based on the Minimum Variance Estimator

Hossein Hassani* and Anatoly Zhigljavsky

Abstract—In recent years Singular Spectrum Analysis (SSA), used as a powerful technique in time series analysis, has been developed and applied to many practical problems. In this paper, we introduce the SSA technique based on the minimum variance estimator. We also consider the SSA technique based on the minimum variance and structured total least squares estimators in reconstructing and forecasting time series. A well-known time series data set, namely, monthly accidental deaths in the USA time series, is used in examining the performance of the technique. The results are compared with several classical methods namely, Box-Jenkins SARIMA models, the ARAR algorithm and the Holt-Winter algorithm.

Keywords: singular spectrum analysis, structured total least squares, minimum variance, reconstruction, forecasting.

1 Introduction

It is well known that errors can seriously limit the performance of the methods and techniques. Effective methods for dealing with noisy data, especially noisy time series are currently still lacking. In general, there are two main approaches for either fitting a model to a noisy time series or forecasting new data points of a noisy time series. According to the first one, ignoring the presence of noise, we fit a model directly from noisy data (such as ARIMA type models [1]) and use the fitted model for forecasting future data points. According to the second approach, we start by filtering the noisy time series in order to reduce the noise level (such as SSA [2]) and fit a model to noise-reduced data and then use the fitted model for forecasting new data points.

In the case of the former approach, the fitted model and therefore the forecasting performance are often poor if the noise level is relatively high, especially for the economics and financial time series. Consider a noisy time series $\mathbf{y}_T = (y_1, \dots, y_T)'$, where $'$ denotes the transpose, and assume that the signal s_t is deterministic and the noise ω_t is additive:

$$y_t = s_t + \omega_t \quad t = 1, \dots, T \quad (1)$$

Assume the noise-reduced time series has been obtained by some noise reduction method, e.g., SSA and denote it by $\mathbf{z}_T = (z_1, \dots, z_T)$. The ideal result of noise reduction is $z_t = s_t$ for each t . In this case, the fitted model on the noise-reduced time series should be optimal, as we remove the noise term ω_t from noisy time series. If the noise has been significantly reduced, then the latter approach is expected to give better results than the former approach. There are several noise reduction methods. It is currently accepted that singular value decomposition (SVD) based methods and signal subspace (SS) methods are more effective than many others for noise reduction and forecasting in financial and economics time series [3].

Having a method for decomposing the vector space of the noisy time series into a subspace that is generated by the noise free series and a subspace for the noise series, we can construct the noise free time series. Approximate decomposition of the vector space of the noisy time series into noise free time series and noise series subspace can be done with, for example, the orthogonal matrix factorization technique such as SVD.

The idea to perform SS method was originally proposed in [4] where a modified SVD is used for reconstruction of noise free series. A general framework for recovering noise free series has been presented in [5]. The method forms the basis for a very general class of subspace-based noise reduction algorithms, is based on the assumption that the original time series exhibits some well-defined properties or obeys a certain model. Noise free series is therefore obtained by mapping the original time series onto the space of series that possess the same structure as the noise free series.

In this context, the SSA technique, which is SVD and SS based method, can be considered as a proper method for noise reduction and forecasting time series data sets. The SSA technique incorporates the elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing. The aim of SSA is to make a decomposition of the original series into the sum of a small number of independent and interpretable components such as a slowly varying trend, oscillatory components and a structureless noise.

The appearance of SSA is usually associated with the

*Statistics Group, Cardiff School of Mathematics, Cardiff University, CF24 4AG, UK. Email: hassanih@cf.ac.uk

publication of [6]. Possible application areas of SSA are diverse: from mathematics and physics to economics and financial mathematics, from metrology and oceanology to social science and market research (see, for example, [7, 8, 9, 10, 11] and references therein). Any seemingly complex series with a potential structure could provide an example of a successful application of SSA [2]. A thorough description of the theoretical and practical foundations of the SSA technique (with several examples) can be found in [2] and [12]. An elementary introduction to the subject can be found in [13].

All the aforementioned research is based on the standard SVD and the structured total least squares (some times referred to simply as least squares (LS)). The LS estimate (we shall use this short abbreviation for the structured total least squares) of the noise free series can be obtained by truncating the singular values. The LS estimator projects the noisy time series onto the perturbed signal (noise + signal) subspace. The reconstructed series using LS estimator has the lowest possible (zero) signal distortion and the highest possible residual noise level. In this paper, we consider an alternative method which is based on the minimum variance (MV) estimator for reconstruction and forecasting noisy time series. The MV estimator is the optimal linear estimator, which gives the minimum total residual power [14, 15].

The structure of the paper is as follows. The next section briefly describes least squares and minimum variance estimators. The reconstruction and forecasting algorithm is presented in Section 3. Our forecast results are then presented and described in Section 4 and some conclusions are given in Section 5.

2 LS and MV Estimators

Consider a noisy signal vector $\mathbf{y}_T = (y_1, \dots, y_T)'$ of length T . We will add the additive white noise to the noise free series (signal) and assume that the noise is uncorrelated with the signal:

$$\mathbf{y}_T = \mathbf{s}_T + \mathbf{n}_T; \quad (2)$$

here \mathbf{s}_T represents the signal component and \mathbf{n}_T noise component. Let $K = T - L + 1$, where L is some integer called the window length (we can assume $L \leq T/2$). Define the so-called 'trajectory matrix' $\mathbf{X} = (x_{ij})_{i,j=1}^{L,K}$, where $x_{ij} = y_{i+j-1}$. Note that \mathbf{X} is a Hankel matrix (by the definition, these are the matrices such that their (i, j) -th entries depend only the sum $i + j$).

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & y_3 & \dots & y_K \\ y_2 & y_3 & y_4 & \dots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \dots & y_T \end{pmatrix} \quad (3)$$

We then consider \mathbf{X} as a multivariate data with L characteristics and $K = T - L + 1$ observations. The columns X_j of \mathbf{X} , considered as vectors, lie in an L -dimensional space \mathbb{R}^L . It is obvious that:

$$\mathbf{X} = \mathbf{S} + \mathbf{N}, \quad (4)$$

where \mathbf{S} and \mathbf{N} represent Hankel matrices of the signal \mathbf{s}_T and noise \mathbf{n}_T , respectively. The Singular value decomposition (SVD) of the trajectory matrix \mathbf{X} can be written as:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}', \quad (5)$$

where $\mathbf{U} \in \mathbb{R}^{L \times K}$ is the matrix consists of the normalized eigenvector U_i corresponding to the eigenvalue λ_i ($i = 1, \dots, L$), $\mathbf{V} \in \mathbb{R}^{K \times K}$, is the matrix contains the principal components defined as $V_i = \mathbf{X}'U_i/\sqrt{\lambda_i}$, and $\mathbf{\Sigma} = \text{diag}(\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L)$. The diagonal elements of $\mathbf{\Sigma}$ are called singular value of \mathbf{X} , and their set is called the singular value spectrum.

The SS methods are based on the assumption that the vector space of the noisy time series (signal) can be split in mutually orthogonal noise and signal+noise subspaces. The components in the noise subspace are suppressed or in the ideal form removed completely. Therefore, one can reconstruct the noise free series from signal+noise subspace by choosing the weight. Thus, by adapting the weights of the different singular components, an estimate of the Hankel matrix \mathbf{X} , which corresponds to noise reduced series, can be achieved:

$$\mathbf{X} = \mathbf{U}(\mathbf{W}\mathbf{\Sigma})\mathbf{V}', \quad (6)$$

where \mathbf{W} is the diagonal matrix containing the weights. Now, the problem is choosing weight matrix \mathbf{W} . In the following we consider the problem of choosing this matrix using different criteria. The SVD of the matrix \mathbf{X} can be written as:

$$\mathbf{X} = [\mathbf{U}_1 \quad \mathbf{U}_2] \begin{bmatrix} \mathbf{\Sigma}_1 & 0 \\ 0 & \mathbf{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}'_1 \\ \mathbf{V}'_2 \end{bmatrix} \quad (7)$$

where $\mathbf{U}_1 \in \mathbb{R}^{L \times r}$, $\mathbf{\Sigma}_1 \in \mathbb{R}^{r \times r}$ and $\mathbf{V}_1 \in \mathbb{R}^{K \times r}$. We can also represent SVD of the Hankel matrix of the signal \mathbf{s}_T as:

$$\mathbf{S} = [\mathbf{U}_{1s} \quad \mathbf{U}_{2s}] \begin{bmatrix} \mathbf{\Sigma}_{1s} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}'_{1s} \\ \mathbf{V}'_{2s} \end{bmatrix} \quad (8)$$

It is clear that the Hankel matrix \mathbf{S} can not be reconstructed exactly if it is perturbed by noise.

2.1 LS Estimate of S

Let us consider the assumption that the matrix $\mathbf{X}_{L \times K}$ is rank deficient, i.e., $\text{rank } \mathbf{X} = r$ and $r < L < K$. The

simplest estimate of \mathbf{S} is obtained when we approximate \mathbf{S} by a matrix of rank r in the LS sense:

$$\min \|\mathbf{X} - \hat{\mathbf{S}}_{LS}\|_F^2 \quad (9)$$

where $\|\cdot\|_F$ is *Frobenius* norms. That is, the LS estimate is obtained by setting smallest singular value to zero ($\lambda_{r+1} = 0, \dots, \lambda_L = 0$) in (7):

$$\hat{\mathbf{S}}_{LS} = [\mathbf{U}_1 \ \mathbf{U}_2] \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}'_1 \\ \mathbf{V}'_2 \end{bmatrix} = \mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}'_1 \quad (10)$$

The \mathbf{S}_{LS} estimate removes the noise subspace, but keeps the noisy signal uncorrelated in the signal+noise subspace. Among different weighting methods, the LS estimate contains the highest possible residual noise level, only the noise from the noise subspace is filtered out, but has the lowest signal distortion (it keeps signal+noise subspace). The disadvantage of LS is that the performance of the LS estimator is crucially dependent on the estimation of the signal rank r . That is, selecting singular value in LS is a binary approach. The main advantage of the LS estimate is that one does not need consider any assumptions either about the signal or noise. For example, if the noise is not white, many other methods need prewhitening and dewhitening steps [16].

2.2 MV Estimate of \mathbf{S}

The aims of the noise reduction can be considered as follows: (1) separate the (signal+noise) subspaces from the (noise only) subspace, (2) remove the (noise-only) subspace, (3) ideally, remove the noise components in the (signal + noise) subspace. The first two steps can be achieved by the least squares estimate, while the MV estimate allows us to have the third one as well. However, one should consider the following assumptions to obtain the MV estimate:

- i) The signal is orthogonal to the noise: $\mathbf{S}'\mathbf{N} = \mathbf{0}$.
- ii) $\mathbf{N}'\mathbf{N} = \sigma_{noise}^2 \mathbf{I}$, where \mathbf{I} is a identity matrix. That is, every column of \mathbf{N} has norm σ_{noise} .
- iii) The smallest singular value of $\boldsymbol{\Sigma}_1$, λ_r , is larger than largest singular value of $\boldsymbol{\Sigma}_2$, λ_{r+1} .

If the assumptions i-iii are met, one can obtain the MV estimate as follows [14, 15]. Given the matrix \mathbf{X} , with rank $\mathbf{X} = \text{rank } \mathbf{N} = L$ and also rank $\mathbf{S} = r$. Find the matrix $\mathbf{T} \in \mathbb{R}^{K \times K}$ that minimizes:

$$\min \|\mathbf{X}\mathbf{T} - \mathbf{S}\|_F^2 \quad (11)$$

The solution is obtained by

$$\mathbf{T} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{S} \quad (12)$$

Therefore, the MV estimate of \mathbf{S} is:

$$\mathbf{X}\mathbf{T} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{S} \quad (13)$$

Using the SVD of the \mathbf{X} , we can obtain:

$$\mathbf{X}\mathbf{T} = \mathbf{U}\mathbf{U}'\mathbf{S} \quad (14)$$

That is, the MV estimate of \mathbf{S} can be interpreted as a orthogonal projection of \mathbf{S} onto the column space of \mathbf{X} because $\mathbf{U}\mathbf{U}'$ is the associated projection matrix. Note also that rank $(\mathbf{X}\mathbf{T}) = \text{rank}(\mathbf{S}) = r$. In real application the matrix \mathbf{S} is not known, but it is possible to achieve the MV estimate, from SVD of \mathbf{X} , if assumption i-iii are satisfied. Let us now consider an alternative form of the SVD of the matrix \mathbf{X} using the SVD of \mathbf{S} (8) as follows:

$$\begin{aligned} \mathbf{X} &= \mathbf{S} + \mathbf{N} = \mathbf{U}_{1s} \boldsymbol{\Sigma}_{1s} \mathbf{V}'_{1s} + \mathbf{N}\mathbf{V}_{1s} \mathbf{V}'_{1s} + \mathbf{N}\mathbf{V}_{2s} \mathbf{V}'_{2s} \\ &= [(\mathbf{U}_{1s} \boldsymbol{\Sigma}_{1s} + \mathbf{N}\mathbf{V}_{1s})(\boldsymbol{\Sigma}_{1s}^2 + \sigma_{noise}^2 \mathbf{I})^{-1/2} \quad \sigma_{noise}^{-1} \mathbf{N}\mathbf{V}_{2s}] \\ &\times \begin{bmatrix} (\boldsymbol{\Sigma}_{1s}^2 + \sigma_{noise}^2 \mathbf{I})^{1/2} & \mathbf{0} \\ \mathbf{0} & \sigma_{noise} \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{V}'_{1s} \\ \mathbf{V}'_{2s} \end{bmatrix} \end{aligned} \quad (15)$$

As it appears from (15), the middle matrix is diagonal, and the left and right matrices have orthonormal columns. Therefore, (15) can be considered as an alternative form of the SVD of \mathbf{X} , and the singular values of \mathbf{X} are:

$$\begin{aligned} \boldsymbol{\Sigma}_1 &= (\boldsymbol{\Sigma}_{1s}^2 + \sigma_{noise}^2 \mathbf{I})^{1/2}, \\ \boldsymbol{\Sigma}_2 &= \sigma_{noise} \mathbf{I}. \end{aligned} \quad (16)$$

Hence, the singular values in $\boldsymbol{\Sigma}_2$ can be interpreted as a noise threshold, which permits estimating σ_{noise} from $\boldsymbol{\Sigma}_2$ in (16). We can also consider the following submatrices:

$$\begin{aligned} \mathbf{U}_1 &= (\mathbf{U}_{1s} \boldsymbol{\Sigma}_{1s} + \mathbf{N}\mathbf{V}_{1s})(\boldsymbol{\Sigma}_{1s}^2 + \sigma_{noise}^2 \mathbf{I})^{-1/2} \\ &= (\mathbf{U}_{1s} \boldsymbol{\Sigma}_{1s} + \mathbf{N}\mathbf{V}_{1s}) \boldsymbol{\Sigma}_1^{-1}, \\ \mathbf{U}_2 &= \sigma_{noise}^{-1} \mathbf{N}\mathbf{V}_{2s}, \\ \mathbf{V}_1 &= \mathbf{V}_{1s}, \\ \mathbf{V}_2 &= \mathbf{V}_{2s}. \end{aligned} \quad (17)$$

Now, using (16-17) and also $\mathbf{S}'\mathbf{N} = \mathbf{0}$, $\mathbf{U}'_1 \mathbf{N} = \mathbf{0}$, we then obtain the MV estimate of \mathbf{S} :

$$\begin{aligned} \hat{\mathbf{S}}_{MV} &= \mathbf{U}\mathbf{U}'\mathbf{S} = \mathbf{U}_1 \mathbf{U}'_1 \mathbf{U}_{1s} \boldsymbol{\Sigma}_{1s} \mathbf{V}'_{1s} + \mathbf{U}_2 \mathbf{U}'_2 \mathbf{U}_{1s} \boldsymbol{\Sigma}_{1s} \mathbf{V}'_{1s} \\ &= \mathbf{U}_1 \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{\Sigma}_{1s} \mathbf{U}'_{1s} + \mathbf{V}'_{1s} \mathbf{N}') \mathbf{U}_{1s} \boldsymbol{\Sigma}_{1s} \mathbf{V}'_{1s} \\ &\quad + \sigma_{noise}^{-1} \mathbf{U}_2 \mathbf{V}'_{2s} \mathbf{N}' \mathbf{U}_{1s} \boldsymbol{\Sigma}_{1s} \mathbf{V}'_{1s} \\ &= \mathbf{U}_1 \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_{1s}^2 \mathbf{V}'_{1s} \\ &= \mathbf{U}_1 \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{\Sigma}_1^2 - \sigma_{noise}^2 \mathbf{I}) \mathbf{V}'_1. \end{aligned} \quad (18)$$

It should be noted that the success of SS methods essentially depends on assumptions i-iii, which in practice, except probably for condition iii, are never satisfied exactly. Let us consider the first assumption. If, for example, $\mathbf{S}'\mathbf{N} \neq \mathbf{0}$ but $\|\mathbf{S}'\mathbf{N}\|$ is small, we can still use the SVD of \mathbf{X} . The smaller $\mathbf{S}'\mathbf{N}$ gets, the better will be the approximations. For the second assumption we can assume that $E(\mathbf{N}'\mathbf{N}) = \sigma_{noise}^2 \mathbf{I}$. However, it has been shown that, under some weak conditions, the assumptions i-iii can be considered true asymptotically [14]. We can therefore still use the robustness feature of SVD with respect to weak violations of these conditions.

2.3 Weight matrix \mathbf{W}

Let us consider again the weight matrix \mathbf{W} based on the LS and MV estimates. As it appears from (10) and (18), the left and right singular vector, \mathbf{U}_1 and \mathbf{V}_1 , of LS and MV estimates are the same, but the singular values are different. The LS and MV estimates can be defined based on the weight matrix $\mathbf{W}_{r \times r}$ as follows:

$$\begin{aligned} \hat{\mathbf{S}}_{LS} &= \mathbf{U}_1(\mathbf{W}_{LS}\boldsymbol{\Sigma}_1)\mathbf{V}'_1 \\ \hat{\mathbf{S}}_{MV} &= \mathbf{U}_1(\mathbf{W}_{MV}\boldsymbol{\Sigma}_1)\mathbf{V}'_1 \end{aligned} \quad (19)$$

where

$$\begin{aligned} \mathbf{W}_{LS} &= \mathbf{I}_{r \times r} \\ \mathbf{W}_{MV} &= \text{diag} \left(\left(1 - \frac{\sigma_{noise}^2}{\lambda_1^2}\right), \dots, \left(1 - \frac{\sigma_{noise}^2}{\lambda_r^2}\right) \right) \end{aligned} \quad (20)$$

3 Reconstruction and Forecasting

3.1 Reconstruction

The matrices $\hat{\mathbf{S}}_{LS}$ and $\hat{\mathbf{S}}_{MV}$ are not Hankel matrices, and thus one can not obtain noise free time series from these estimates. We thus need a formal procedure of transformation an arbitrary matrix into a Hankel matrix and therefore into a series. This procedure is the so-called Hankelization and is obtained by diagonal averaging, where every element from antidiagonal of either $\hat{\mathbf{S}}_{LS}$ or $\hat{\mathbf{S}}_{MV}$ is replaced by the average value along the antidiagonal. Let us consider the Hankelization procedure in detail.

If z_{ij} stands for an element of a matrix \mathbf{Z} , then the k -th term of the resulting time series is obtained by averaging z_{ij} over all i, j such that $i + j = k + 2$. The result of the Hankelization of a matrix \mathbf{Z} is the Hankel matrix $\mathcal{H}\mathbf{Z}$, which is the trajectory matrix corresponding to the time series obtained as a result of the diagonal averaging.

The operator \mathcal{H} acts on an arbitrary $L \times K$ -matrix $\mathbf{Z} = (z_{ij})$ with $L \leq K$ in the following way: for $i + j = s$ and $N = L + K - 1$ the element \tilde{z}_{ij} of the matrix $\mathcal{H}\mathbf{Z}$ is

$$\begin{cases} \frac{1}{s-1} \sum_{l=1}^{s-1} z_{l,s-l} & 2 \leq s \leq L-1, \\ \frac{1}{L} \sum_{l=1}^L z_{l,s-l} & L \leq s \leq K+1, \\ \frac{1}{K+L-s+1} \sum_{l=s-K}^L z_{l,s-l} & K+2 \leq s \leq K+L. \end{cases}$$

Note that the Hankelization is an optimal procedure in the sense that the matrix $\mathcal{H}\mathbf{Z}$ is the nearest to \mathbf{Z} (with respect to the Frobenius norm) among all Hankel matrices of the corresponding size [2]. The Hankel matrix $\mathcal{H}\mathbf{Z}$

uniquely defines the time series by relating the values in the diagonals to the values in the series.

By applying the Hankelization procedure to matrices $\hat{\mathbf{S}}_{LS}$ or $\hat{\mathbf{S}}_{MV}$, we obtain:

$$\tilde{\mathbf{S}}_{LS} = \mathcal{H}\hat{\mathbf{S}}_{LS} \quad , \quad \tilde{\mathbf{S}}_{MV} = \mathcal{H}\hat{\mathbf{S}}_{MV} \quad (21)$$

Therefore, the reconstructed series is:

$$y_t = \sum_{k=1}^m \tilde{y}_t^{(k)} \quad (22)$$

where $\tilde{Y}_T^{(k)} = (\tilde{y}_1^{(k)}, \dots, \tilde{y}_T^{(k)})$ corresponds to the matrix $\tilde{\mathbf{S}}_{LS}$ and $\tilde{\mathbf{S}}_{MV}$.

3.2 Forecasting

Forecasting by SSA can be applied to the time series that approximately satisfies linear recurrent formulae (LRF):

$$y_{i+d} = \sum_{k=1}^d a_k y_{i+d-k}, \quad 1 \leq i \leq T-d \quad (23)$$

of some dimension d with the coefficients a_1, \dots, a_d . An important property of the SSA decomposition is that, if the original time series Y_T satisfies a LRF (23), then for any T and L there are at most d nonzero singular values in the SVD of the trajectory matrix \mathbf{X} ; therefore, even if the window length L and $K = T - L + 1$ are larger than d , we only need at most d matrices \mathbf{X}_i to reconstruct the series.

Let us briefly describe the so-called SSA recurrent forecasting algorithm (for more information see [2]). Define the original series $Y_T = (y_1, \dots, y_T)$ and the reconstructed series $\tilde{Y}_T = (\tilde{y}_1, \dots, \tilde{y}_T)$ (which can be obtained by either LS or MV estimates). For an eigenvector $U \in \mathbb{R}^L$ we denote the vector of the first $L-1$ components of the vector U as $U^\nabla \in \mathbb{R}^{L-1}$. Set $v^2 = \pi_1^2 + \dots + \pi_r^2 < 1$, where π_i is the last component of the eigenvector U_i ($i = 1, \dots, r$). It can be proved that the last component y_L of any vector $Y = (y_1, \dots, y_L)^T$ is a linear combination of the first components (y_1, \dots, y_{L-1}) ; that is, $y_L = a_1 y_{L-1} + \dots + a_{L-1} y_1$ where the vector of coefficients $A = (a_1, \dots, a_{L-1})$ can be expressed as $A = \sum_{i=1}^r \pi_i U_i^\nabla / (1 - v^2)$. The forecasts y_{T+1}, \dots, y_{T+h} are then obtained as

$$y_i = \begin{cases} \tilde{y}_i & \text{for } i = 1, \dots, T \\ \sum_{j=1}^{L-1} a_j y_{i-j} & \text{for } i = T+1, \dots, T+h. \end{cases}$$

4 Comparison

Let us now consider the performance of the SSA technique based on the MV and LS estimates by applying it to a well-known time series data set, namely, monthly accidental deaths in the USA. The results are compared

with those obtained using Box-Jenkins SARIMA models (Model I and II), the ARAR algorithm and the Holt-Winter algorithm (as described in [17]):

Model I:

$$\nabla_{12}y_t = 28.831 + (1 - 0.478B)(1 - 0.588B^{12})Z_t \quad (24)$$

$$Z_t \sim WN(0, 94390)$$

Model II:

$$\nabla_{12}y_t = 28.831 + Z_t - 0.596Z_{t-1} - 0.407Z_{t-6} - 0.685Z_{t-12} + 0.460Z_{t-13} \quad (25)$$

$$Z_t \sim WN(0, 94390)$$

where backward shift operator B is: $B^j Z_t = Z_{t-j}$.

The window $L = 24$ and the first 12 singular values have been used in reconstructing and forecasting the series and singular values 13–24 have been considered as noise components (for more information about parameters selection, for this series, see [9]). Here, we used the same parameters and recurrent forecasting algorithm as for the vector forecasting algorithm that was used in [9]. To calculate the precision we use Root Mean Squared Error (RMSE):

$$RMSE = \left(\frac{\sum_{i=1}^n (y_{T+i} - \tilde{y}_{T+i})^2}{\sum_{i=1}^n (y_{T+i} - \hat{y}_{T+i})^2} \right)^{1/2} \quad (26)$$

Here n represents the number of forecasted points, \tilde{y}_{T+i} are the forecasted values of y_{T+i} obtained by SSA and \hat{y}_{T+i} are the forecasted values of y_{T+i} obtained by other methods. If $RMSE < 1$, then SSA procedure outperforms alternative prediction method. Alternatively, $RMSE > 1$ would indicate that the performance of the corresponding SSA procedure is worse than the predictions of the competing method.

The methods are arranged based on the performance of forecasting. The results are presented in Table 1. The values of RMSE show performance of forecasting. The last two columns, labeled Ratio, show the ratios of RMSEs SSA/other methods. As it appears in Table 1, the forecasting performance using the SSA technique based on the LS estimate (SSA_{LS}) and based on the MV estimate (SSA_{MV}) are much better than other forecasting methods and also the SSA_{MV} is the best among the methods considered, for example, the value of RMSE for the SSA_{MV} is 9 times less than the first one (model I) and almost 3 times less than the ARAR algorithm. From the table, one can see that the SSA_{MV} performance is better than the SSA_{LS} . Let us consider the performance of the SSA forecasting results with respect to different values of r . We choose the same window length L but different eigenvalues r . The results are presented in Table 2, for the first 13 eigenvalues ($r = 13$), and Table 3, for the first 14 eigenvalues ($r = 14$). As the tables show,

again, the SSA technique outperforms the other classical methods. It can be seen that the quality of the forecast is changed when one changes the number of eigenvalues in the reconstruction step. Of course, forecasting accuracy and reconstruction quality are related. By selecting a group of eigenvalues, and considering other eigenvalues as noise, some frequencies may be filtered out completely. This destroys the signal structure and then gives a poorer reconstruction. In general, a high signal to noise ratio will result in good forecasting and vice-versa.

Table 1: RMSE of the Post-sample forecasts, first 12 eigenvalues ($r = 12$).

Method	RMSE	Ratio	
		LS	MV
Model I	582.63	0.21	0.12
Model II	500.50	0.24	0.14
H-W	401.26	0.30	0.18
ARAR	253.20	0.47	0.28
SSA_{LS}	119.61	1.00	0.59
SSA_{MV}	70.90	1.69	1.00

Table 2: RMSE of the Post-sample forecasts, first 13 eigenvalues ($r = 13$).

Method	RMSE	Ratio	
		LS	MV
Model I	582.63	0.18	0.11
Model II	500.50	0.21	0.13
H-W	401.26	0.26	0.17
ARAR	253.20	0.41	0.26
SSA_{LS}	104.48	1.00	0.64
SSA_{MV}	66.96	1.56	1.00

Table 3: RMSE of the Post-sample forecasts, first 14 eigenvalues ($r = 14$).

Method	RMSE	Ratio	
		LS	MV
Model I	582.63	0.27	0.13
Model II	500.50	0.31	0.16
H-W	401.26	0.39	0.20
ARAR	253.20	0.62	0.31
SSA_{LS}	156.157	1.00	0.50
SSA_{MV}	78.49	1.99	1.00

5 Conclusion

Classical time series methods such ARIMA type models fit a model directly from noisy data and use the fitted model for forecasting future data points. Forecasting results are typically better if one fits a model to noise reduced time series and then use the fitted model for forecasting new data points. The signal subspace and SVD based methods such as SSA can be applied as powerful

tools for finding the noise free series and using it for forecasting future data points.

In this paper we introduced the SSA technique based on the minimum variance estimator. The results has illustrated that the SSA technique performs well in forecasting time series. The comparison of the forecasting results showed that SSA, based on the minimum variance (MV) and structured total least squares (LS) estimates, are much more accurate than several well-known classical methods, in forecasting of a well know time series. We also find that the SSA forecasting results based on MV are better than based on LS for forecasting this series. However, comparison between these two estimates depends on the choice of the SSA parameters, the window length L and the number of eigenvalues r , the data we have and also the analysis we have to perform. We believe that SSA based on both estimates, LS and MV, gives accurate results.

References

- [1] Box, G.E.P. and Jenkins, G.M. (1970). *Time series analysis: Forecasting and control*, Holden-Day.
- [2] Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and related techniques*. Chapman & Hall/CRC.
- [3] Soofi, A. and L. Cao. (2002). Nonlinear Forecasting of Noisy Financial Data, in Soofi and Cao (eds.) *Modeling and Forecasting Financial Data: Techniques of Nonlinear Dynamics*, Kluwer Academic Publishers, Boston.
- [4] Tufts, D. W., Kumaresan, R., and Kirsteins, I. (1982). Data adaptive signal estimation by singular value decomposition of a data matrix, *Proceedings of the IEEE*, **70**, no. 6, pp. 684-685.
- [5] Cadzow, J. A. (1988). Signal enhancement a composite property mapping algorithm, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **36**, no. 1, pp. 49-62.
- [6] Broomhead, D.S. and King, G.P. (1986). Extracting Qualitative Dynamics from Experimental Data. *Physica D* **20**, 217-236.
- [7] Vautard, R., Yiou, P. and Ghil, M. (1992). Singular-spectrum analysis: A toolkit for short, noisy chaotic signal. *Physica D* **58**, 95-126.
- [8] Ghil, M. and Taricco, C. (1997). Advanced spectral analysis methods. In G.C. Castagnoli and Provenzale, A. (Eds.), *Past and present Variability of the Solar-Terrestrial system: Measurement, Data Analysis and Theoretical Model*, 137-159. IOS Press.
- [9] Hassani, H. (2007). Singular Spectrum Analysis: Methodology and Comparison, *Journal of Data Science*, **5(2)**, 239-257.
- [10] Moskvina, V.G. and Zhigljavsky, A. (2003). An algorithm based on singular spectrum analysis for change-point detection. *Communication in Statistics - Simulation and Computation* **32**, No. 2, 319-352.
- [11] Golyandina, N. and Osipov, E. (2006). Caterpillar-SSA method for analysis of time series with missing values, *Journal of Statistical Planning and Inference*, **137**, pp. 2642-2653.
- [12] Danilov, D. and Zhigljavsky, A. (Eds.). (1997). *Principal Components of Time Series: the 'Caterpillar' method*, University of St. Petersburg Press. (In Russian.)
- [13] Elsner, J.B. and Tsonis, A.A. (1996). *Singular Spectral Analysis. A New Tool in Time Series Analysis*. Plenum Press.
- [14] De Moor, B. (1993). The singular value decomposition and long and short spaces on noisy matrices, *IEEE Transaction on Signal Processing*, Vol, 41, no 9, 2826-2838.
- [15] Van Huffel, S. (1993). Enhanced resolution based on minimum variance estimation and exponential data modeling, *Signal Processing*, **33**, no. 3, pp. 333-355, 1993.
- [16] Jensen, S.H., Hansen, P.C., Hansen, S.D. and Sørensen, J.A. (1995). Reduction of Broad-Band Noise in Speech by Truncated QSVD, *IEEE Transactions on Speech and Audio Processing*, November, **6**, pp. 439-448.
- [17] Brockwell, P.J. and Davis R.A. (2002). *Introduction to Time Series and Forecasting*, 2nd edition. Springer.