Comparison of Singular Spectrum Analysis and ARIMA Models

Zokaei, Mohammad
Shahid Beheshti University, Department of Statistics
Evin
Tehran 19839-63113, Iran.
E-mail: zokaei@sbu.ac.ir

Mahmoudvand, Rahim
Shahid Beheshti University, Department of Statistics
Evin
Tehran 19839-63113, Iran.
E-mail: r.mahmodvand@gmail.com

Najari, Nader
Shahid Beheshti University, Department of Statistics
Evin
Tehran 19839-63113, Iran.
E-mail: nader_najari85@yahoo.com

Introduction and Preliminaries

Singular Spectrum Analysis (SSA) is a relatively new powerful non-parametric technique for time series analysis incorporating the elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing (Golyandina et al, 2001).

SSA is designed to look for nonlinear, non–stationary, and intermittent or transient behavior in an observed time series, and has gained successful application in the various sciences such as meteorological, bio-mechanical, hydrological, physical sciences, economics and finance, engineering and so on (see Khan and Poskitt (2010) and references therein).

In spite of the successful application of SSA for different data in comparison with other models such as ARIMA (see for example Hassani (2007)), little of the current literature analyzes the statistical properties of SSA from an abstract theoretical perspective. So, to obtain a theoretical view of the forecasting with SSA, accuracy of forecasting with SSA is compared with ARIMA by means of statistical simulations.

In what follows we give a brief explanation of the SSA method (for more details see for example Golyandina et al, 2001). Consider the real-valued non-zero time series $Y_T = (y_1, \ldots, y_T)^T$ of sufficient length T. Let $K = T - L + 1$, where L $(L \le T/2)$ is some integer called the window length. The general structure of the SSA can be described in four basic steps:

Step1, Embedding: Define the trajectory matrix $\mathbf{X}$ where:

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

Note that X is a Hankel matrix, which means that all the elements along the off diagonal are equal.

Step2, Decomposition: Define the matrix $\mathbf{XX}^T$. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L \geq 0$ denote eigenvalues of $\mathbf{XX}^T$ and U_i be the normalized eigenvector corresponding to the eigenvalue λ_i $(i = 1,\ldots,L)$. Then singular value decomposition (SVD) of the trajectory matrix X can be written as:

$$\mathbf{X} = \mathbf{E_1} + \ldots + \mathbf{E_L}$$

Where, $\mathbf{E}_i = \sqrt{\lambda_i}\, U_i V_i^T$, $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$ (i = 1, . . . , L).

Step3, Separation: Partition the set of indices $\{1,\ldots,L\}$ into m disjoint subsets $\{I_1,\ldots,I_m\}$. This separation leads to the following decomposition:

$$\mathbf{X} = \mathbf{X}_{I_1} + \ldots + \mathbf{X}_{I_m}$$

Where, $\mathbf{X}_{I_j} = \sum_{\ell \in I_j} \mathbf{E}_\ell$, $j = 1,\ldots,m$.

Step4, Reconstruction: The last step in SSA, transforms each matrix of the grouped decomposition (step3) into a new series of length *T*. This will be doing with diagonal averaging as follow. If z_{ij} stands for an element of a matrix **Z**, then the *k*-th term of the resulting series is obtained by averaging z_{ij} over all *i; j* such that $i+j = k+1$. By performing the diagonal averaging of all matrix components of $\mathbf{X}_{I_j}$ in the expansion of **X** above, we obtain another expansion: $\mathbf{X} = \tilde{\mathbf{X}}_{I_1} + \ldots + \tilde{\mathbf{X}}_{I_m}$, where $\tilde{\mathbf{X}}_{I_j}$ is the diagonalized version of the matrix $\mathbf{X}_{I_j}$. This is equivalent to the decomposition of the initial series $Y_T = (y_1,\ldots,y_T)^T$ into a sum of *m* series; $y_t = \sum_{j=1}^{m} \tilde{y}_t^{(j)}$, where $\tilde{Y}_T^{(j)} = (\tilde{y}_1^{(j)},\ldots,\tilde{y}_T^{(j)})^T$ corresponds to the matrix $\tilde{\mathbf{X}}_{I_j}$.

It must be noted that the general purpose of the SSA analysis is decomposition of original series with additive components that are 'independent' and 'identifiable' time series. Sometimes, however, one can also be interested in particular tasks, such as ‘extraction of signal from noise,’ ‘extraction of oscillatory components’ and ‘smoothing’.

Forecasting by SSA

The Basic SSA recurrent forecasting algorithm discussed in Golyandina et al (2001) should be regarded as the main forecasting algorithm. Although, there exist several natural modifications to this algorithm that can give better forecasts in specific situations (see, e.g Golyandina et al (2001)), here we consider recurrent forecasting algorithm and call it by R-SSA. Let $\hat{y}_T(s)$ denotes the R-SSA forecast at time T for lead time s or s steps ahead. According to the R-SSA, the following recursive formula can be used to obtain forecasts:

$$\hat{y}_T(s) = \begin{cases} \tilde{y}_s & s = 1, \dots, T \\ \sum_{t=1}^{L-1} a_t \hat{y}_T(s-t) & s = T+1, \dots, T+M \end{cases}$$

Where,

$$\mathcal{R} = (a_{L-1}, \dots a_1)^T = \frac{1}{1-(\pi_1^2+\cdots+\pi_r^2)} \sum_{i=1}^{r} \pi_i U_i^{\nabla}$$

Such that π_i is the last component of the vector U_i and U_i^{∇} is the vector consisting of the first $L-1$ components of U_i, (i = 1, . . . , L).

Comparison Criteria

First of all recall that an ARIMA(p,d,q) model can be explained with an equation such as:

$$\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)\left(1-L^d\right)y_t = \left(1-\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$$

Where L is the lag operator, the α_i 's are the parameters of the autoregressive part of the model, the θ_i 's are the parameters of the moving average part and the ε_t 's are error terms. The error terms are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean.

There are several criteria to measure the accuracy of the forecasting model and all of them are based on the behavior of the forecast errors. As there is no one best criteria, we use root mean square errors (RMSE) which commonly used for the evaluation of alternative forecasting models. The MSE of some ARIMA models with respect to the number of step ahead, are shown in Table 1.

Table 1: MSE of several ARIMA models with respect to the number of steps ahead

Model \ Steps ahead	1	2	3	s, where s>3
AR(1)	σ^2	$\sigma^2\left(1+\alpha_1^2\right)$	$\sigma^2\left(1+\alpha_1^2+\alpha_1^4\right)$	$\sigma^2\left(1+\alpha_1^2+\ldots+\alpha_1^{2(s-1)}\right)$
AR(2)	σ^2	$\sigma^2\left(1+\alpha_1^2\right)$	$\sigma^2\left(1+\alpha_1^2+\left(\alpha_1^2+\alpha_2\right)^2\right)$	Depends on s
MA(1)	σ^2	$\sigma^2\left(1+\theta_1^2\right)$	$\sigma^2\left(1+\theta_1^2\right)$	$\sigma^2\left(1+\theta_1^2\right)$
MA(2)	σ^2	$\sigma^2\left(1+\theta_1^2\right)$	$\sigma^2\left(1+\theta_1^2+\theta_2^2\right)$	$\sigma^2\left(1+\theta_1^2+\theta_2^2\right)$

Simulation Studies

In order to evaluate the performance of the SSA procedure in comparison with ARIMA models, data were simulated using a set of stationary autoregressive and moving average models of orders 1 and 2 where the coefficients were changed stepwise (with variance of the white noise process be one). For all data generating processes, N = 10000 series of length T = 100 were simulated by means of ***arima.sim*** function in statistical software **R**. The simulated data were used to forecast, applying the R-SSA algorithm presented above. Then standard simulation procedure was used to obtain estimates of root mean square errors (RMSE) for forecasting.

Results

The results are summarized in Figures 1, 2 and Tables 2 and 3. Figures 1 and 2 show RMSE of the out of sample forecasting by R-SSA and AR(1) (MA(1)). The parameters α_1 and θ_1 are varied from 0.1 to 0.9 in order to generate different AR and MA processes for comparison. Moreover, 1, 2, 3 and 4 step ahead out-of-sample forecasts are depicted in the figures (from left to right) to assess the forecasting power of the considered techniques in the short and long lead. As it indicated in both of figures 1 and 2, ARIMA models gives accurate forecasting just for the parameters less than around 0.6 in 1 step ahead, but generally, SSA has superiority to ARIMA models for 2, 3 and 4 steps ahead. It means that ARIMA acts well for short period forecasting, whereas SSA is good for both short and long forecasting. Similar results presented in Tables 2 and 3 for AR(2) and MA(2). Of Course it is worth mention that, according to the results of Table 2 and 3, R-SSA is better than ARIMA for all considered values of the parameters.

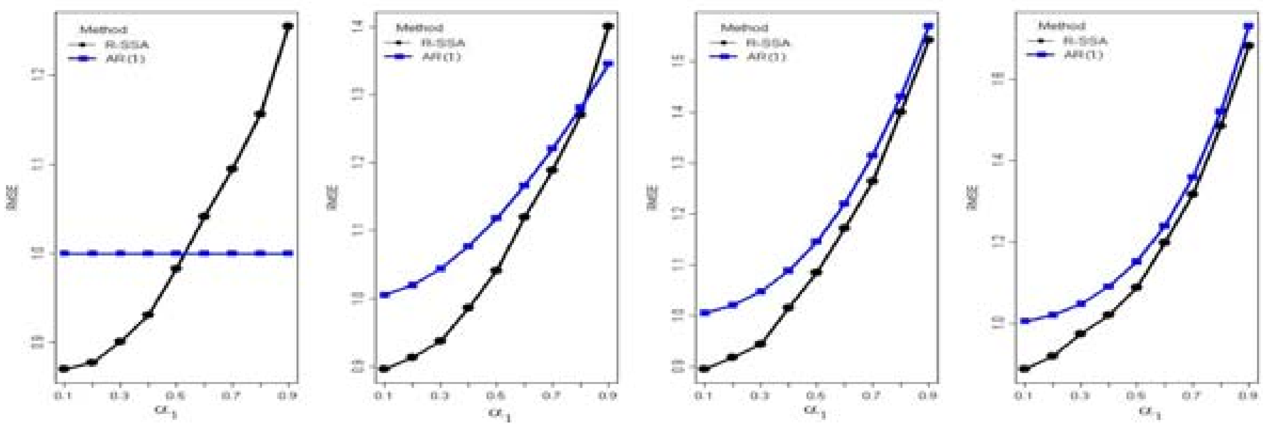

Fig 1: RMSE of the out of sample forecasting by R-SSA and AR(1)

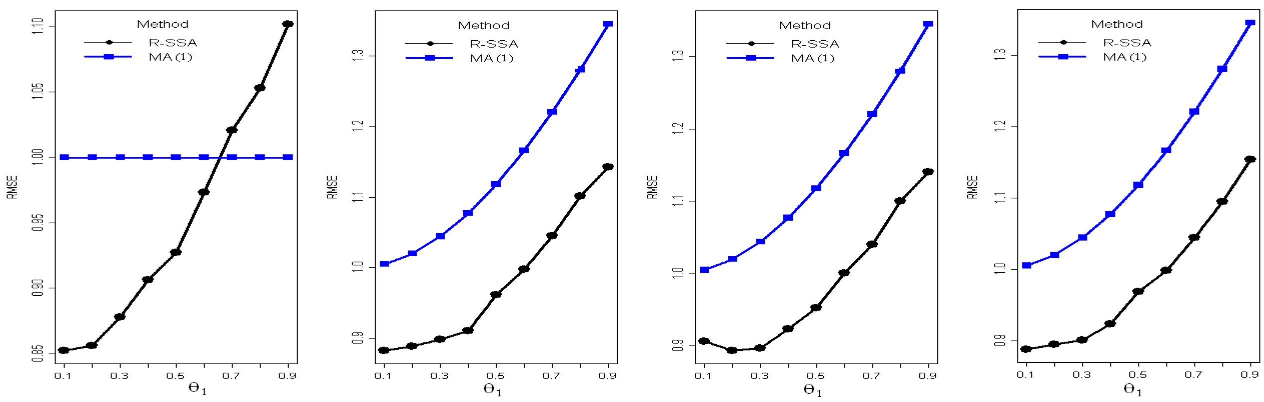

Fig 2: RMSE of the out of sample forecasting by R-SSA and MA(1)

Table 2: RMSE of the out of sample forecasting by R-SSA and AR(2) for several selected AR coefficients

Lead	Method	AR parameters: (α_1, α_2)								
		(0.1,0.1)	(0.1,0.2)	(0.1,0.4)	(0.2,0.1)	(0.2,0.2)	(0.2,0.4)	(0.4,0.1)	(0.4,0.2)	(0.4,0.4)
1	R-SSA	0.799	0.889	0.911	0.864	0.883	0.921	0.897	0.938	0.980
	AR(2)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	R-SSA	0.876	0.902	0.936	0.837	0.894	0.985	0.929	0.927	1.113
	AR(2)	1.050	1.050	1.050	1.020	1.020	1.020	1.077	1.077	1.077
3	R-SSA	0.860	0.831	0.964	0.851	0.904	0.945	0.941	0.957	1.093
	AR(2)	1.011	1.011	1.011	1.048	1.048	1.048	1.214	1.214	1.214
4	R-SSA	0.863	0.882	0.926	0.863	0.915	0.977	0.956	1.036	1.147
	AR(2)	1.011	1.011	1.011	1.051	1.051	1.051	1.273	1.273	1.273

Table 3: RMSE of the out of sample forecasting by R-SSA and MA(2) for several selected MA coefficients

Lead	Method	MA parameters: (θ_1, θ_2)								
		(0.1,0.1)	(0.1,0.2)	(0.1,0.4)	(0.2,0.1)	(0.2,0.2)	(0.2,0.4)	(0.4,0.1)	(0.4,0.2)	(0.4,0.4)
1	R-SSA	0.848	0.850	0.895	0.854	0.856	0.900	0.902	0.898	0.942
	MA(2)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	R-SSA	0.849	0.871	0.909	0.866	0.880	0.907	0.902	0.906	0.962
	MA(2)	1.001	1.001	1.001	1.020	1.020	1.020	1.077	1.077	1.077
3	R-SSA	0.872	0.870	0.915	0.871	0.878	0.916	0.901	0.918	0.966
	MA(2)	1.020	1.020	1.020	1.039	1.039	1.039	1.149	1.149	1.149
4	R-SSA	0.863	0.868	0.921	0.862	0.896	0.937	0.915	0.922	0.972
	MA(2)	1.020	1.020	1.020	1.039	1.039	1.039	1.149	1.149	1.149

Conclusion

Despite the successful applications of SSA compared to simpler models, little of the current literature analyzes the statistical properties of SSA. Therefore, to obtain a theoretical view of the forecasting with SSA, basic SSA forecasting algorithm is compared with ARIMA by means of a Monte Carlo study. The evaluation criteria used is the accuracy of out of sample point forecasts via RMSE. It has been shown in the simulations that forecasting by SSA can be advantageous compared to ARIMA models.

REFERENCES

1- Golyandina N, Nekrutkin V, Zhigljavsky A. Analysis of Time Series Structure: SSA and related techniques. Chapman & Hall/CRC, 2001.

2- Rahman Khan, M, A.; and Poskitt, D.S. (2010). Description Length Based Signal Detection in Singular Spectrum Analysis. Working Paper. Department of Econometrics and Business Statistics, Monash University. Australia.

3- Hassani, H. (2007). Singular Spectrum Analysis: Methodology and Comparison. Journal of Data Science, 5(2), pp. 239-257.

ABSTRACT

Singular Spectrum Analysis (SSA) is a non-parametric method that can be applied to analyze time series of complex structure. The main purpose of SSA is a decomposition of the original series into a sum of series, so that each component in this sum can be identified as either a trend, periodic or quasi-periodic component (perhaps, amplitude modulated), or noise. In this paper, by means of statistical simulation, we compare the performance of the SSA and ARIMA models. We show that, in general, the performance of forecasting using these methods are different and depends on parameters of the models.

Keywords: SSA, ARIMA, Forecasting.