

MODELAGEM DE SÉRIES TEMPORAIS VIA SINGULAR SPECTRUM ANALYSIS - UMA OTIMIZAÇÃO DO WINDOW LENGTH

Renata de Miranda Esquivel

Mestranda do Programa de Modelagem Computacional, SENAI CIMATEC,
Av. Orlando Gomes, 1845- Piatã – Salvador- BA
esquivel.renata@hotmail.com

Valter de Senna

Programa de Modelagem Computacional, SENAI CIMATEC,
Av. Orlando Gomes, 1845- Piatã – Salvador- BA
valter.senna@gmail.com

Gecynalda S. da S. Gomes

Departamento de Estatística da UFBA
Av. Adhemar de Barros, S/N- Ondina – Salvador- BA
gecynalda@yahoo.com

RESUMO

Os métodos estatísticos para análise de séries temporais encontram-se desenvolvidos na literatura e vários modelos estão implementados em *software*. Contudo, cada modelo possui suposições relacionadas às características dos dados e o uso adequado dos modelos exigirá verificações dessas suposições, que podem demandar esforços na etapa de pré-análise. Como alternativa, pode-se utilizar uma técnica poderosa e relativamente moderna conhecida como *Singular Spectrum Analysis* (SSA). Este método não exige o conhecimento sobre o modelo paramétrico da série temporal. A técnica SSA básica fundamenta-se em dois estágios complementares: decomposição e reconstrução. Na decomposição, a série original é representada como uma série multidimensional com dimensão L denominada *window length*, que representa a quantidade de componentes independentes e interpretáveis extraídos da série. Tal *window length* deve ser um valor inteiro suficientemente grande, mas não superior a metade do tamanho da série ($T/2$). No segundo estágio, representa-se o resultado da decomposição como uma soma de matrizes, as quais são transformadas em uma nova série de tamanho T e que pode ser considerada como uma aproximação da série original. O trabalho objetiva apresentar a potencialidade da SSA na identificação de componentes de uma série, utilizando um L ótimo. Com esta finalidade, analisou-se a série do IBOVESPA ao longo de 1314 dias e os dados de companhias aéreas no decorrer de 144 meses (*AirPassengers*). As análises foram realizadas na versão 2.13 do *software* R através do pacote *simsalabim*. As funções *decompSSA* e *reconSSA* deste pacote implementam a SSA de forma padrão, em que L deve ser especificado em *decompSSA*. O parâmetro L foi otimizado de forma que minimizasse o erro quadrático médio (EQM) cometido na reconstrução da série através da SSA, fundamentando-se na condição $1 < L < T/2$. Tal otimização foi realizada considerando o *default* da função *optimize* do R. A otimização resultou em $L = 656$ e $L = 71$ para as séries do IBOVESPA e *AirPassengers*, respectivamente. Após realizar a decomposição da série *AirPassengers*, observou-se que o maior autovalor destaca-se dos demais, representando claramente a tendência principal da série pelo primeiro componente. A partir do terceiro componente os autovalores tendem a se estabilizarem em torno do zero, e assim o sinal da série é formado pelos dois primeiros componentes. Para a série do IBOVESPA foi possível verificar que o primeiro componente indica a tendência da série e a partir do segundo componente o padrão de sazonalidade começa a ser identificado. Comparando as séries originais

com as reconstruídas nota-se os seguintes EQM: $2,417936 \times 10^{-22}$ (IBOVESPA) ; $6,387 \times 10^{-27}$ (AirPassengers). Como esperado, a modelagem das séries via SSA discriminou satisfatoriamente os principais componentes constituintes, demonstrando que com apenas dois componentes foi possível recompor a estrutura principal de ambas as séries investigadas.

PALAVRAS CHAVE. Série temporal, Singular Spectrum Analysis, Otimização.

Área principal (Estatística).

ABSTRACT

Statistical methods for time series analysis are developed in the literature and several models are implemented in software. However, each model has assumptions concerning the characteristics of data and appropriate use of the models requires verification of these assumptions, which may require efforts in the pre-analysis. Alternatively, you can use a relatively modern and powerful technique known as Singular Spectrum Analysis (SSA). This method does not require knowledge about the parametric model of the time series. The SSA basic technique is based on two complementary stages: decomposition and reconstruction. In decomposition, the original series is represented as a multidimensional series with dimension L called window length, which represents the amount of independent and interpretable components extracted from the series. This window length must be an integer large enough, but not more than half the size of the series ($T/2$). In the second stage, the result of decomposition is represented as a sum of matrices, which are transformed into a new series of size T and can be considered as an approximation of the original series. The paper aims to present the potentiality of the SSA to identify components of a series, using an optimized L . To this end, we analyzed the IBOVESPA time series over 1314 days and the data from airlines over 144 months (AirPassengers). Analyses were performed in version 2.13 software R through package *simsalabim*. The functions *decompSSA* and *reconSSA* this package implement the SSA by default, where L must be specified in *decompSSA*. The parameter L was optimized so that minimized the mean square error (MSE) in the reconstruction of the series made by SSA and are based on the condition $1 < L < T/2$. This optimization was performed considering the default the function *optimize* of R. The optimization resulted in $L = 656$ and $L = 71$ IBOVESPA and AirPassengers time series, respectively. After performing the decomposition AirPassengers time series observed that the largest eigenvalue stands out among the other, representing the main trend of the time series by the first component. From the third component of the eigenvalues tend to stabilize around zero, and thus the signal of the series is formed by the first two components. For the IBOVESPA's series was possible to verify that the first component indicates the trend of the series and from the second component the pattern of seasonality begins to be identified. After comparing the original series with the reconstructed note the following MSE: 2.417936×10^{-22} (IBOVESPA) 6.387×10^{-27} (AirPassengers). As expected, the modeling of the series via the SSA discriminated satisfactorily the main constituent components, showing that with only two components was possible to reconstruct the main structure of both investigated series.

KEYWORDS. Time series. Singular Spectrum Analysis. Optimization.

Main area (Statistics).