entropy

MDPI

Article

The Decomposition and Forecasting of Mutual Investment Funds Using Singular Spectrum Analysis

Paulo Canas Rodrigues [1,2,*], Jonatha Pimentel [1] and Patrick Messala [1] and Mohammad Kazemi [3]

1 Department of Statistics, Federal University of Bahia, 40170-110 Salvador, Brazil; jsppimentel9@gmail.com (J.P.); patrickmessala@gmail.com (P.M.)
2 CAST, Faculty of Information Technology and Communication Sciences, Tampere University, FI-33014 Tampere, Finland
3 Department of Statistics, Faculty of Mathematical Sciences, Shahrood University of Technology, P.O. Box 3619995161 Shahroud, Iran; m.kazemie64@gmail.com
* Correspondence: paulocanas@gmail.com or paulo.canas@ufba.br

Received: 7 November 2019; Accepted: 7 January 2020; Published: 9 January 2020

Abstract: Singular spectrum analysis (SSA) is a non-parametric method that breaks down a time series into a set of components that can be interpreted and grouped as trend, periodicity, and noise, emphasizing the separability of the underlying components and separate periodicities that occur at different time scales. The original time series can be recovered by summing all components. However, only the components associated to the signal should be considered for the reconstruction of the noise-free time series and to conduct forecasts. When the time series data has the presence of outliers, SSA and other classic parametric and non-parametric methods might result in misleading conclusions and robust methodologies should be used. In this paper we consider the use of two robust SSA algorithms for model fit and one for model forecasting. The classic SSA model, the robust SSA alternatives, and the autoregressive integrated moving average (ARIMA) model are compared in terms of computational time and accuracy for model fit and model forecast, using a simulation example and time series data from the quotas and returns of six mutual investment funds. When outliers are present in the data, the simulation study shows that the robust SSA algorithms outperform the classical ARIMA and SSA models.

Keywords: singular spectrum analysis; robust singular spectrum analysis; time series forecasting; mutual investment funds

1. Introduction

Mutual investment funds provide management services to institutional and individual investors, besides great liquidity for financial investments made in them and low transactional costs [1,2]. These funds can be of fixed or variable income and allow to diversify the assets while reducing unsystematic risk. Fixed income mutual investment funds are of low risk, whereas variable-income mutual investment funds vary in terms of risk but also in terms of returns. In this study, we were interested in analyzing the quotas and returns of six of the largest Brazilian based mutual investment funds—three purely based on stocks: (i) Alaska Black, (ii) APEX Long Biased, and (iii) Brasil Capital; and three balanced funds (usually combining a stock component, a bond component, and sometimes a money market component in a single portfolio): (iv) ADAM Strategy, (v) Gavea Macro, and (vi) SPX Nimitz.

A natural framework for analyzing mutual investment funds, due to its underlying structure, is a time series method.

Singular spectrum analysis (SSA) is a powerful non-parametric technique for time series analysis and forecasting, which incorporates elements of classical time series analysis, multivariate statistics, and matrix algebra. Its main aim is to decompose the original time series into a set of components that can be interpreted as trend components, seasonal components, and noise components [3–6]. SSA has proven both wide usefulness and applicability across many applications [7–17], being that its scope of application ranges from parameter estimation to time series filtering, synchronization analysis, and forecasting [18].

The SSA methodology for model fit can be summarized in four steps: (i) embedding, which maps the original univariate time series into a trajectory matrix; (ii) singular value decomposition (SVD), which helps decomposing the trajectory matrix into the sum of rank-one matrices; (iii) eigentriple grouping, which helps deciding which of the components are associated to the signal and which are associated to the noise; and (iv) diagonal averaging, which maps the rank-one matrices, associated to the signal, back to time series that can be interpreted as trend, seasonal, or other meaningful components.

SSA results and interpretation, similarly to many other classical time series methods, can be sensitive to data contamination with outliers [19,20]. In those cases, even a small percentage of outliers can make a big difference on the results for model fit and model forecast. Very few attempts have been made in order to access the effect of the presence of outliers in the data while conducting a SSA. One study [21,22] presented some preliminary results on the effect of outliers in singular spectrum analysis, and [23] made a first attempt to robustify the SSA by considering an SVD based on a robust L_1 norm [24] instead of the L_2 norm used in the classical algorithm, which they used for model fit.

In this paper we go one step further than [23] and propose a new robust algorithm for SSA that considers the SVD based on the Huber function [25]. Moreover, we propose two robust SSA forecasting algorithms, one based on the the L_1 norm and another based on the Huber function. Comparisons are made between the classical SSA algorithm, the robust SSA algorithm based on the L_1 norm (RLSSA), the robust SSA algorithm based on the Huber function (RHSSA), and the classical autoregressive integrated moving average (ARIMA) model, in terms of computational time and accuracy for model fit and model forecast. These comparisons for decomposing and forecasting time series were done by considering a simulation example and the six mutual investment funds mentioned above.

The rest of this paper is organized as follows. Section 2 provides the materials and methods containing the data description, a brief introduction to the ARIMA and SSA methodologies, and the details of the proposed robust SSA algorithm that uses the SVD based on the Huber function. Section 3 presents the results and discussion, wherein the ARIMA, SSA, and robust SSA algorithms are compared in terms of model fit and model forecast, using the six mutual investment funds and the simulation example. The paper closes in Section 4, wherein some conclusions are drawn.

2. Materials and Methods

2.1. Data

In this paper we consider a dataset that includes daily observations of six mutual investment funds, three based purely on stocks and three balanced funds:

Stock funds

- Alaska Black: 3 January 2017–30 August 2019 (N = 666 observations).
- APEX Long Biased: 15 April 2013–30 August 2019 (N = 1604 observations).
- Brasil Capital: 27 August 2012–30 August 2019 (N = 1760 observations).

Balanced funds

- ADAM Strategy: 29 April 2016–30 August 2019 (N = 838 observations).
- Gavea Macro: 30 June 2008–30 August 2019 (N = 2809 observations).
- SPX Nimitz: 01 December 2010–30 August 2019 (N = 2199 observations).

The datasets were collected from https://infofundos.com.br/carteira.

2.2. ARIMA Model

The autoregressive integrated moving average (ARIMA) models are among the most widely used techniques for time series analysis and forecasting. Such a model depends on three parameters: p is the number of lagged observations in the model, i.e., the autoregressive (AR) order; d is the number of times that the original observations are differenced, i.e., the integrated (I) degree; and q is the size of the moving average window, i.e., the order of the moving average (MA) [26]. This parametric model can then be written as $ARIMA(p, d, q)$, with p, d, and q non-negative integers. Given a time series $Y_N = y_1, \ldots, y_N$, the $ARIMA(p, d, q)$ model can be written as:

$$(1 - \phi_1 B^1 - \cdots - \phi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B^1 + \cdots + \theta_q B^q)\varepsilon_t, \tag{1}$$

where $\phi_1, \ldots, \phi_p$ are the parameters or coefficients of the p autoregressive terms; B is the time lag operator, or backward shift, which is a linear operator denoted by B^k such that $L^k y_t = y_{t-k}$, $t \in \mathbb{Z}$; y_t is the observation at the time point t; $c = \mu(1 - \phi_1 - \cdots - \phi_p)$; μ is the mean of $(1 - B)^d y_t$; $\beta_1, \ldots, \beta_q$ are the parameters or coefficients of the q moving average terms; and ε_t is an error term, usually white noise with variance σ^2.

Alternatively, the model can be written as:

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B)^d (y_t - \mu t^d / d!) = (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t, \tag{2}$$

which is the parametization used in the "arima" function of the software R [27].

2.3. Singular Spectrum Analysis

Singular spectrum analysis is a non-parametric technique for model fit and model forecasting that decomposes a time series into a number of components that are summed and interpreted as trend, periodicity, and noise. Similarly to many other time series techniques, SSA can be used for solving a wide range of problems, some of the most relevant being its ability to smooth the original time series, and to separate the signal (i.e., trend and oscillatory components with different amplitudes) from the noise components. Therefore, SSA can be used to analyze and reconstruct smoother noise-free time series that can then be used for model forecasting.

SSA is divided into two interconnected stages: decomposition and reconstruction of the time series. These stages are divided into two sets each, forming a total of four steps: embedding, singular value decomposition (SVD), grouping, and diagonal averaging. The complete algorithm for model fit is described in the following sub-section. Further details can be found in, e.g., [5,6,28].

2.3.1. Decomposition

In the first stage, the (univariate) time series is converted into a high-dimensional matrix called a trajectory matrix, which is then decomposed into the sum of rank-one matrices based on the SVD.

(1) Embedding:

Consider a non-zero time series $Y_N = \{y_1, \ldots, y_n\}$ with size $N > 2$. Let $L(1 < L < N)$ be an integer value called window length and K an integer such that the trajectory matrix includes all values; i.e., $K = N - L + 1$. The embedding step is achieved by mapping the original time series into a sequence of K vectors with length L:

$$Y_i = (y_i, \cdots, y_{i+L-1})^T, 1 \leq i \leq K. \tag{3}$$

Then, the trajectory matrix $\mathbf{X}$, that includes the vectors Y_i, $i = 1, \ldots, K$, in its columns can be written as:

$$\mathbf{X} = [Y_1, \cdots, Y_K] = (y_i j)_{i,j=1}^{LK} = \begin{bmatrix} y_1 & y_2 & \cdots & y_K \\ y_2 & y_3 & \cdots & y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & \cdots & y_N \end{bmatrix}. \tag{4}$$

(2) Singular value decomposition:

Let $\mathbf{S} = \mathbf{X}\mathbf{X}^T$, $\mathbf{U}_1, \ldots, \mathbf{U}_L$ be the eigenvectors of $\mathbf{S}$, and $\lambda_1 \geq \cdots \geq \lambda_L$, its corresponding eigenvalues. If d is the number of non-null eigenvalues of $\mathbf{S}$, and considering $\mathbf{V}_i = \mathbf{X}^T U_i \sqrt{\lambda_i}$, we can decompose the trajectory matrix $\mathbf{X}$ as:

$$\mathbf{X} = \sum_{i=1}^{d} \mathbf{X}_i = \sum_{i=1}^{d} \sqrt{\lambda_i} U_i V_i^T. \tag{5}$$

The decomposition stage can be accomplished either by the eigendecomposition of $\mathbf{X}^T\mathbf{X}$ or by the SVD of $\mathbf{X}$ ($\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, $\mathbf{D} = diag(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d})$). A comparison between both decompositions can be found in [29].

2.3.2. Reconstruction

In the second stage, after a separating signal from noise components, a diagonal averaging procedure is conducted in the matrices associated to the signal resulting into the sum of time series components that can then be interpreted as trend or oscillatory components:

(1) Eigentriple grouping:

This step consists of identifying the first r eigentriples associated with the signal and discarding the $d - r$ eigentriples associated with the noise. Formally, let $I = 1, \ldots, r$ and $I^c = r + 1, \ldots, d$. The goal of this step is to choose I such that the trajectory matrix can be written as:

$$\mathbf{X}_I = \sum_{i \in I} \sqrt{\lambda_i} \mathbf{U}_i \mathbf{V}_i^T + \epsilon, \tag{6}$$

where ϵ is the noise term.

The number of eigentriples to conduct the reconstruction is often decided based on w-correlations. We shall say that two series $Y^{(1)}$ and $Y^{(2)}$ are approximately separable if all correlations between the rows and the columns of the corresponding trajectory matrices obtained from series $Y^{(1)}$ and $Y^{(2)}$ are close to zero. In [5] they considered other characteristics of the quality of separability; namely, the weighted correlation or **w**-correlation, which is a natural measure of deviation of two series $Y_T^{(1)}$ and $Y_T^{(2)}$ from **w**-orthogonality:

$$\rho_{12}^{(w)} = \frac{\left(Y_T^{(1)}, Y_T^{(2)}\right)_w}{\|Y_T^{(1)}\|_w \|Y_T^{(1)}\|_w}, \tag{7}$$

where $\|Y_T^{(i)}\|_w = \sqrt{\left(Y_T^{(i)}, Y_T^{(i)}\right)_w}$, $i = 1, 2$, and $\left(Y_T^{(1)}, Y_T^{(2)}\right)_w = \sum_{t=1}^{T} w_t y_t^{(1)} y_t^{(2)}$ with $w_t = \min\{t, L, T - t + 1\}$. If the absolute value of the **w**-correlation is small, the two series are almost **w**-orthogonal. If the absolute value of the **w**-correlation is large, the series are far from being **w**-orthogonal and are, therefore, badly separable. Further explanation and intuition about this measure can be found in [5,28]. Other proposals for this choice were proposed by, e.g., [30,31].

(2) Diagonal averaging:

In this step, using anti-diagonal averaging on the matrices included in $\mathbf{X}_I$, the noise-free time series is reconstructed. First, the approximate trajectory matrix $\mathbf{X}_I$ is transformed into a Hankel matrix. Let $A_s = \{(l,k) : l+k = s, 1 \leq l \leq L, 1 \leq k \leq K\}$ and $\#(A_s)$ be the number of elements in A_s. The element $\widetilde{x}_{ij}$ of the new Hankel matrix $\widetilde{\mathbf{X}}$ is given by:

$$\widetilde{x}_{ij} = \sum_{(l,k)\in A_s} \frac{x_{lk}}{\#(A_s)}. \tag{8}$$

Next, the Hankel matrix $\widetilde{\mathbf{X}}_I$ is transformed into a new series of dimension N, and the original time series $\mathbf{Y}_N$ can be approximated by:

$$\widetilde{y}_i = \begin{cases} \widetilde{x}_{i1} & \text{for } i = 1, \ldots, L, \\ \widetilde{x}_{Lj} & \text{for } i = L+1, \ldots, N, \end{cases} \tag{9}$$

where $j = i - L + 1$.

The reconstructed noise-fee time series can then be used for out-of-sample forecasting.

2.4. Robust SSA

Despite knowing that SSA has shown to be superior to traditional model-based methods in many applications, the singular value decomposition (second step of the SSA algorithm) is highly sensitive to data contamination with outliers. Very few studies were made in order to access effects of outliers in SSA and to generalize this methodology [21,22]. A first attempt to robustify the SSA by considering an SVD based on a robust L_1 norm [24] instead of the L_2 norm used in the classical algorithm, was proposed by [23]. That robust generalization was compared with the classical SSA algorithm for model fit by these authors. In this subsection we review that robust SSA algorithm proposed by [23] and propose a new robust algorithm for SSA that considers the SVD based on the Huber function [25] and also propose an algorithm for robust SSA model forecasting. While the robust algorithms based on the L_1 norm are very popular, they have difficulties in handling heavy tail outliers. The robust algorithms based on the Huber function combine the sum of squares loss and the least absolute deviation loss, that is, a quadratic on small errors, but grows linearly for large errors. As a result, the Huber loss function is not only more robust against outliers but also more adaptive for different types of data [32]. Further details and comparisons between the L_1 and Huber loss functions, among others, can be found in [33]. The R source code is available upon request from the first author of this paper.

2.4.1. Robust SSA Based on the L_1 Norm

The robust SSA algorithm proposed by [23] replaces the classical SVD based on the least squares L_2 norm, by the robust SVD algorithm based on the L_1 norm [24]. This robust SVD is performed iteratively, starting with an initial estimate of the first left singular vector U_1 and leading to an outlier-resistant approach that also allows for missing data. The robust SVD based on the L_1 norm is implemented under the function "robustSVD()" from the R package "pcaMethods".

2.4.2. Robust SSA based on the Huber Function

Here we propose a new alternative to robustify the SSA algorithm, where the least squares SVD in the step two is replaced by the robust SVD based on the Huber function [25]. The Huber loss function [34] can be defined as:

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{if } |a| \leq \delta \\ \delta\left(|a| - \frac{1}{2}\delta\right) & \text{if } |a| > \delta \end{cases}, \tag{10}$$

where δ is a parameter that controls the robustness level, and a smaller value of δ usually leads to more robust estimation.

The robust SVD based on the Huber function is a special case of robust regularized SVD and can be obtained with the function "RobRSVD" of the "RobRSVD" R package, in the following way: RobRSVD (data, rough = TRUE, uspar = 0, vspar = 0). In this R implementation, the authors consider $\delta = 1.345$, the value commonly used in robust regression that produces 95% efficiency for normal errors [35]. However, numerical studies suggested that the RobRSVD function is not very sensitive to the choice of δ [25]. More details about this robust SVD can be found in [25].

2.5. Robust SSA Forecasting Algorithm

The standard recurrent SSA forecasting algorithm assumes that a given observation can be written as a linear combination of the $L-1$ previous observations [5,6,30]. The coefficients of those linear combinations in the classical SSA forecasting algorithm are obtained based on the left singular vectors, U, of the trajectory matrix $\mathbf{X}$. This is valid for SSA because of the orthogonality of the vectors in U and of the full rank decomposition of $\mathbf{X}$, which is not the case for the robust SVD algorithms because of their construction and specific properties. To overcome this limitation for the robust SSA algorithms and to be able to obtain out-of-sample forecasts using a robust SSA algorithm, a three stages approach can be conducted:

(i) Use the robust SSA algorithm to obtain a robust approximation for the signal in the trajectory matrix; i.e., conduct the two stages of the robust SSA algorithms, decomposition (using the robust SVD algorithm) and reconstruction, to obtain the noise free (i.e., the signal) trajectory matrix $\widetilde{\mathbf{X}}$;

(ii) Apply the standard SVD to the matrix $\widetilde{\mathbf{X}}$ obtained in (i) and obtain U_j^{∇}, the vector of the first $L-1$ components of U_j and π_j, the last component of the vector U_j, $j = 1, \cdots, r$. Then, we can write the coefficient vector $\widehat{a}$ as

$$\widehat{a} = (\widehat{a}_{L-1}, \cdots, \widehat{a}_1)' = \frac{1}{1-\gamma^2} \sum_{j=1}^{r} \pi_j U_j^{\nabla}, \tag{11}$$

where $\gamma^2 = \sum_{j=1}^{r} \pi_j^2$.

(iii) The h-steps-ahead out-of-sample recurrent robust SSA forecasts $\widehat{y}_{N+1}, \ldots, \widehat{y}_{N+h}$, can be obtained as

$$\widehat{y}_t = \begin{cases} \tilde{y}_t, & \text{for } t = 1, \cdots, N \\ \sum_{j=1}^{L-1} \widehat{a}_j \widehat{y}_{t-j}, & \text{for } t = N+1, \cdots, N+h \end{cases} \tag{12}$$

where $\tilde{y}_1, \ldots, \tilde{y}_N$, are the fitted values for the reconstructed time series, as obtained from the robust SSA algorithm in (i).

2.6. Accuracy Measures

There are several methods and measures for assessing model accuracy based on the behavior of model errors. Here, there are two types of errors:

- Sample errors, called tuning errors;
- Out-of-sample errors, called forecast errors.

Typically, the root mean squared error (RMSE) is used as a criterion for accessing the precision of a model. The RMSE to investigate the quality of the model fit can be written as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \widetilde{y}_t)^2}, \tag{13}$$

where y_t are the observed values and $\widetilde{y}_t$ the fitted values by the considered model/algorithm (i.e., ARIMA, SSA, robust SSA).

To investigate the forecasting accuracy, let us assume that the last g observations are used as a reference (i.e., as test set). Let $N_0 = N - h - g$. The RMSE to investigate the quality of the forecasting model can be written as:

$$RMSE = \sqrt{\frac{1}{g} \sum_{t=N_0+h+1}^{N} (y_t - \widetilde{y}_t)^2}, \tag{14}$$

where y_t are the last g observed values and $\widetilde{y}_t$ the respective h-steps-ahead forecast values.

3. Results and Discussion

In this section, comparisons are made between the classical ARIMA model, the classical SSA algorithm, and the robust SSA algorithms, in terms of computational time and accuracy for model fit and model forecast. These comparisons for decomposing and forecasting time series are done by considering a simulation example and the time series of six mutual investment funds.

Table 1 shows the descriptive statistics for the six mutual investment funds, including the minimum, maximum, and mean returns, being clear that Alaska Black is the fund that shows the largest variation and with the highest mean daily return. On the other end there are Gavea Macro and SPX Nimitz, which show the smallest variations among the considered funds, and low mean returns.

In addition to the descriptive measures, Figure 1 shows the behavior of the six investment funds over time. From these plots, it is possible to observe that all funds have an overall growing tendency, with similar patterns for Gavea Macro and SPX Nimitz.

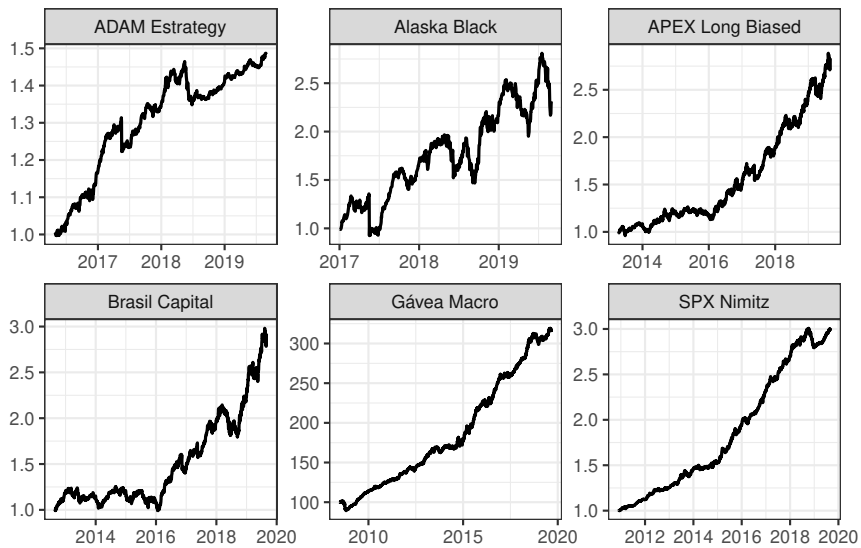

Figure 1. Time series for the returns of the six mutual investment funds, ADAM Strategy, Alaska Black, APEX Long Biased, Brasil Capital, Gávea Macro and SPX Nimitz, from left to right and from top to bottom. The vertical axes show the quota values; i.e., the total net assets of a fund divided by the total number of quotas existing.

Table 1. Descriptive measures for returns of the six mutual investment funds.

Investment Fund	Minimum	Mean	Maximum	Standard deviation
ADAM Strategy	−6.26%	0.05%	1.63%	0.0045%
Alaska Black	−29.62%	0.16%	9.80%	0.0240%
APEX Long Biased	−8.60%	0.07%	3.72%	0.0085%
Brasil Capital	−7.55%	0.07%	3.42%	0.0094%
Gavea Macro	−2.22%	0.04%	2.36%	0.0033%
SPX Nimitz	−1.92%	0.05%	1.42%	0.0030%

3.1. Model Fit

The models/algorithms under comparison for model fit are: (i) ARIMA, (ii) SSA, (iii) robust SSA based on the L_1 norm (RLSSA), and (iv) robust SSA based on the Huber function (RHSSA).

The parameters of the ARIMA model for each of the six mutual investment funds were estimated with the function "auto.arima" from the R package "forecast" [36].

For the SSA and robust SSA algorithms, there are two choices to be made by the researcher: (i) the window length L; and (ii) the number of eigentriples used for reconstruction r. Three values of L were chosen for each time series, as defined in Table 2—$L_1 = N/20$, $L_2 = N/2$, and L_p—being the L_p obtained from the periodogram, based on the largest cycle for each time series [37] (i.e., about one trimester for ADAM Strategy, one semester for Alaska Black, one year for APEX Long Biased, one quadrimeter for Brasil Capital, one quadrimeter for Gavea Macro, and one quadrimester for SPX Nimitz), and N being the time series length. The choice of the number of eigentriples used for reconstruction r, for each of the considered window lengths and each of the time series, was done by taking into consideration the the w-correlations among components [5]. Figure 2 shows the w-correlation matrices for each of the six mutual investment funds, considering an window length $L = N/20$, and Figure A1 of the appendix shows the w-correlation matrices for each of the six mutual investment funds, considering an window length $L = N/2$. The w-correlation matrices can be obtained with the function "wcor" of the R package "Rssa" [38] and the number of eigentriples r should be chosen in order to maximize the separability between signal and noise components; i.e., maximize the w-correlation among signal components, maximize the w-correlation among noise components, and minimize the w-correlation between signal and noise components. A summary of the number of eigentriples used for the reconstruction of each time series for each of the window length considered can be seen in Table 2.

Since one of the objectives in SSA is to decompose the original time series into interpretable components such as trend and seasonality, plus the noise component that is then discarded, Figure 3 shows the original time series for the Alaska Black mutual investment fund, its trend component (sum of individual trend components), its seasonal component (sum of individual seasonal components), and its residuals (sum of the remaining components associated to noise), considering an window length $L = N/20 = 33$ and $r = 12$ eigentriples for reconstruction. Similar SSA decompositions for ADAM Strategy, APEX Long Biased, Brasil Capital, ADAM Strategy, Gavea Macro, and SPX Nimitz—considering the values of window length L_1 and r_1 eigentriples used for reconstruction, as defined in Table 2—can be found in Figures A2–A6 of the appendix, respectively.

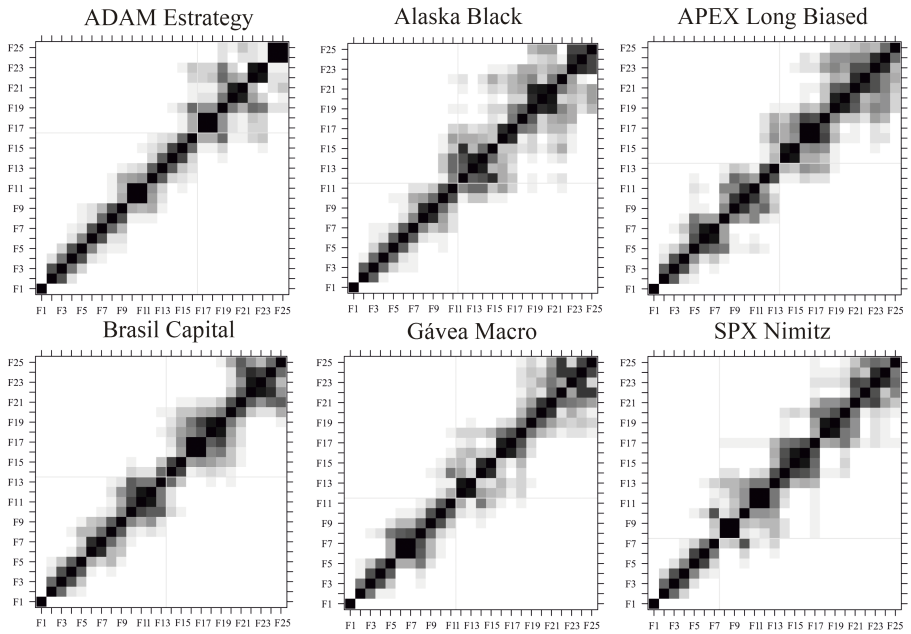

Figure 2. W-correlation matrices for each of the six mutual investment funds, ADAM Strategy, Alaska Black, APEX Long Biased, Brasil Capital, Gávea Macro and SPX Nimitz, from left to right and from top to bottom, considering an window length $L = N/20$.

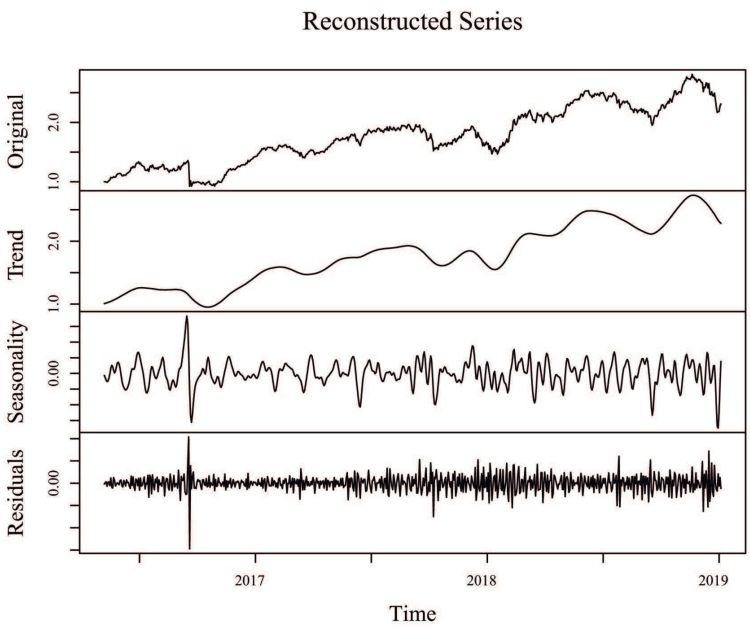

Figure 3. Decomposition of the original time series for the Alaska Black mutual investment fund (top panel), with a trend component (sum of individual trend components, second panel), a seasonal component (sum of individual seasonal components, third panel), and a residual (sum of the remaining components associated to noise, bottom panel), considering an window length $L = N/20 = 33$ and $r = 12$ eigentriples for reconstruction.

Table 2. Window length L_1, L_2, and L_p, and number of eigentriples r considered for model fit and model forecast for each of the mutual investment funds.

Investment Fund	n	L_1	r_1	L_2	r_2	L_p	r_p
ADAM Strategy	838	41	17	419	18	60	13
Alaska Black	666	33	12	333	11	125	8
APEX Long Biased	1604	80	14	802	11	250	11
Brasil Capital	1760	88	12	880	12	80	13
Gavea Macro	2809	140	12	1404	12	80	12
SPX Nimitz	2199	109	8	1099	8	80	11

In order to evaluate and compare the ability for model fit using the four models, ARIMA, SSA, robust SSA based on the L_1 norm (RLSSA), and robust SSA based on the Huber function (RHSSA), the root mean square error (RMSE) was calculated for each time series. Table 3 shows the RMSE for model fit by each of the four models applied to each of the six mutual investment funds, considering a window length $L = N/2$ (Table 2). Table 4 shows the RMSE for model fit by each of the four models applied to each of the six mutual investment funds, considering a window length $L = N/20$ (Table 2). Table 5 shows the RMSE for model fit by each of the four models applied to each of the six mutual investment funds, considering a window length obtained based on the largest cycle for each time series (Table 2). From the analyzes of these tables, we can conclude that the ARIMA model shows an overall better performance when the window length in the SSA related algorithms is set to be half of the time series (Table 3). However, when the window length is set to be $L_1 = N/20$ or L_p (i.e., equal to the length of the largest cycle), the classical SSA provides the best results, while the ARIMA model and the robust SSA algorithms alternate for the second best performances. For all choices of window length, the two robust SSA algorithms behaved similarly.

Table 3. Root mean square error for each of the six mutual investment funds, considering each of the four models, ARIMA, SSA, robust SSA based on the L_1 norm (RLSSA), and robust SSA based on the Huber function (RHSSA), for the window length $L_2 = N/2$ and considering r_2 engentriples for reconstruction as defined in Table 2.

Investment Fund	ARIMA	SSA	RLSSA	RHSSA
ADAM Strategy	0.0057	0.0075	0.0088	0.0076
Alaska Black	0.0402	0.0450	0.0508	0.0476
APEX Long Biased	0.0160	0.0294	0.0318	0.0320
Brasil Capital	0.0170	0.0338	0.0429	0.0346
Gavea Macro	0.6756	1.9758	2.1486	2.0016
SPX Nimitz	0.0063	0.0197	0.0239	0.0207

Table 4. Root mean square error for each of the six mutual investment funds, considering each of the four models, ARIMA, SSA, robust SSA based on the L_1 norm (RLSSA), and robust SSA based on the Huber function (RHSSA), for the window length $L_1 = N/20$ and considering r_1 engentriples for reconstruction as defined in Table 2.

Investment Fund	ARIMA	SSA	RLSSA	RHSSA
ADAM Strategy	0.0057	0.0024	0.0034	0.0034
Alaska Black	0.0402	0.0190	0.0244	0.0234
APEX Long Biased	0.0160	0.0107	0.0124	0.0116
Brasil Capital	0.0170	0.0124	0.0143	0.0133
Gavea Macro	0.6756	0.6508	0.7716	0.7432
SPX Nimitz	0.0063	0.0066	0.0078	0.0077

Table 5. Root mean square error for each of the six mutual investment funds, considering each of the four models, ARIMA, SSA, robust SSA based on the L_1 norm (RLSSA), and robust SSA based on the Huber function (RHSSA), for the window length L_p (i.e., the length of the largest cycle) and considering r_p engentriples for reconstruction as defined in Table 2.

Investment Fund	ARIMA	SSA	RLSSA	RHSSA
ADAM Strategy	0.0057	0.0038	0.0046	0.0045
Alaska Black	0.0402	0.0415	0.0482	0.0459
APEX Long Biased	0.0160	0.0185	0.0196	0.0190
Brasil Capital	0.0170	0.0123	0.0139	0.0132
Gavea Macro	0.6756	0.5049	0.5997	0.5986
SPX Nimitz	0.0063	0.0049	0.0058	0.0057

Tables 6–8 show the computational times for each combination of model/algorithm and mutual investment fund, as presented in Tables 3–5, respectively. From the analyzes of these tables, we can conclude that the best performance was obtained by the ARIMA and SSA algorithms. The computational time, for the classic and robust SSA algorithms, increases with the increase of the length L. Moreover, for larger trajectory matrices (i.e., considering $L = N/2$) the robust SSA algorithm based on the Huber function has a lower computational time than the robust SSA algorithm based on the L_1 norm (Table 6). However, when the trajectory matrices are more rectangular (i.e., considering $L = N/20$, Table 7, or $L = L_p$, Table 8), the robust SSA algorithm based on the L_1 norm has a much lower computational time (comparable to the ARIMA and SSA computational times) than the robust SSA algorithm based on the Huber function).

Table 6. Computational time, in minutes, for each of the six mutual investment funds, considering each of the four models, ARIMA, SSA, robust SSA based on the L_1 norm (RLSSA), and robust SSA based on the Huber function (RHSSA), for the window length $L_2 = N/2$ and considering r_2 engentriples for reconstruction as defined in Table 2.

Investment Fund	ARIMA	SSA	RLSSA	RHSSA
ADAM Strategy	0.0010	0.0052	15.563	14.232
Alaska Black	0.0018	0.0042	7.5859	6.8834
APEX Long Biased	0.0175	0.0320	195.27	61.031
Brasil Capital	0.0226	0.0366	287.80	83.821
Gavea Macro	0.0057	0.1584	1605.2	632.84
SPX Nimitz	0.0022	0.0618	616.75	120.83

Table 7. Computational time, in minutes, for each of the six mutual investment funds, considering each of the four models, ARIMA, SSA, robust SSA based on the L_1 norm (RLSSA), and robust SSA based on the Huber function (RHSSA), for the window length $L_1 = N/20$ and considering r_1 engentriples for reconstruction as defined in Table 2.

Investment Fund	ARIMA	SSA	RLSSA	RHSSA
ADAM Strategy	0.0010	0.0025	0.1257	68.384
Alaska Black	0.0018	0.0031	0.0669	16.794
APEX Long Biased	0.0175	0.0039	1.2952	530.43
Brasil Capital	0.0226	0.0048	1.9145	629.79
Gavea Macro	0.0057	0.0088	10.823	1441.1
SPX Nimitz	0.0022	0.0050	3.7450	375.29

Table 8. Computational time, in minutes, for each of the six mutual investment funds, considering each of the four models, ARIMA, SSA, robust SSA based on the L_1 norm (RLSSA), and robust SSA based on the Huber function (RHSSA), for the window length L_p (i.e., the length of the longest cycle) and considering r_p engentriples for reconstruction as defined in Table 2.

Investment Fund	ARIMA	SSA	RLSSA	RHSSA
ADAM Strategy	0.0010	0.0024	0.3371	65.149
Alaska Black	0.0018	0.0026	1.6994	3.3270
APEX Long Biased	0.0175	0.0078	26.826	115.14
Brasil Capital	0.0226	0.0099	2.0020	804.16
Gavea Macro	0.0057	0.0126	3.4485	1718.4
SPX Nimitz	0.0022	0.0078	3.4937	905.16

Figure 4 shows the original time series and the model fit by the SSA model with $L = N/20$ and by the ARIMA model. We can confirm that both fits are almost overlapped and very near to the original time series, which was expected from the small RMSE showed in Table 4.

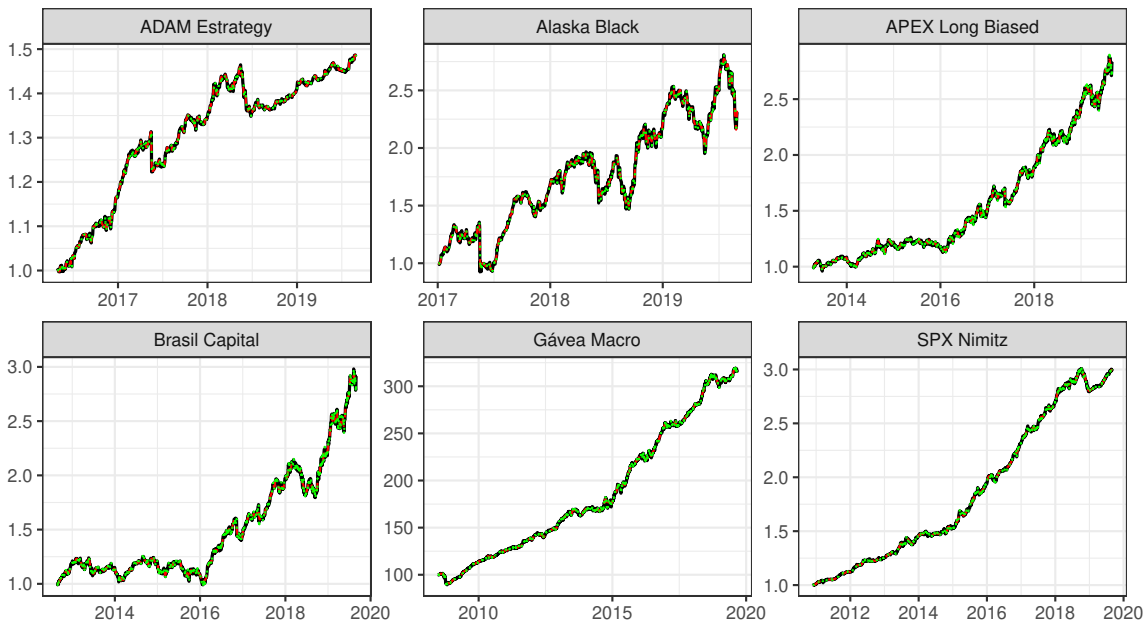

Figure 4. Original time series (black line); smoothed time series after applying the SSA considering $L = N/20$, with the number of eigentriples r as they are defined in Table 2 (red line); and model fit by the ARIMA model (green line), for each of the six mutual investment funds, ADAM Strategy, Alaska Black, APEX Long Biased, Brasil Capital, Gávea Macro and SPX Nimitz, from left to right and from top to bottom. The vertical axes show the quota values.

3.2. Model Forecasting

In this section we compare the forecasting abilities of ARIMA, SSA with $L = N/2$, SSA with $L = N/20$, SSA with $L = L_p$ based on the largest cycle for each time series, and robust SSA based on the L_1 norm with $L = N/20$ and L_p. The decision for not considering the robust SSA algorithm based on the Huber function was because of its similarity in terms of RMSE with the robust SSA based on the L_1 norm (Tables 3–5) and the much higher computational time (Tables 6–8). A similar argument was considered for not presenting the results for the robust SSA algorithm based on the L_1 norm with $L = N/2$.

Table 9 shows the RRMSE for model forecasting for each of the six mutual investment funds, considering each of the four models, ARIMA, SSA with $L = N/2$, SSA with $L = N/20$, SSA with $L = L_p$, and robust SSA based on the L_1 norm (RLSSA) with $L = N/20$ and L_p, considering the

window length and engentriples used for reconstruction as defined in Table 2. These values were obtained based on the forecasting of the $g = 12$ observations from each time series, obtained for one, five, and ten steps ahead out-of-sample forecast; i.e., one day ahead, one week ahead, and two weeks ahead.

The overall best performance was obtained with the classic SSA algorithm that considers a lower value for the window length, either $L = N/20$ or $L = L_p$, followed closely by ARIMA and the robust SSA algorithm based on the L_1 norm. The ARIMA model obtained the best performance in three cases for one-step-ahead forecasting, and the robust SSA algorithm based on the L_1 norm with $L = N/20$ yielded the best performance in a couple of time series for five-steps-ahead forecasting. As expected, the RMSE shows an overall increase when increasing the number of steps ahead to be forecast. A possible justification for the similarity between the SSA and robust SSA algorithm can be explained by the possible lack of outliers in the data. Table 10 shows the computational time for model forecasting for each of the six mutual investment funds, considering each of the five models shown in Table 9. As expected, after analyzing the computational times for model fit (Tables 6–8), the best performance in terms of computational time for model forecasting was obtained by the the ARIMA and SSA (with lower values for the window length) models and the worse by the robust SSA algorithm based on the L_1 norm.

Table 9. Root mean square error for model forecasting for each of the six mutual investment funds, considering the models ARIMA, SSA with $L = N/2$, SSA with $L = N/20$, SSA with L_p, robust SSA based on the L_1 norm (RLSSA) with $L = N/20$, and RSSA with L_p, and their respective engentriples, as defined in Table 2.

Investment Fund	**ARIMA**	**SSA $\frac{N}{2}$**	**SSA $\frac{N}{20}$**	**SSA L_p**	**RLSSA $\frac{N}{20}$**	**RLSSA L_p**
			one-step-ahead			
ADAM Strategy	0.0027	0.0036	0.0029	0.0047	0.0048	0.0048
Alaska Black	0.0712	0.2118	0.0638	0.1357	0.1138	0.178
APEX Long Biased	0.0426	0.1778	0.0544	0.0646	0.0663	0.0576
Brasil Capital	0.0436	0.0496	0.0590	0.0573	0.0545	0.0512
Gavea Macro	1.1670	2.3104	1.5536	1.2571	1.1532	1.6582
SPX Nimitz	0.0081	0.0278	0.0061	0.0061	0.0061	0.0074
			five-step-ahead			
ADAM Strategy	0.0056	0.0047	0.0058	0.0038	0.0089	0.0057
Alaska Black	0.2031	0.2990	0.1800	0.1848	0.2120	0.2365
APEX Long Biased	0.1184	0.1965	0.0578	0.0724	0.0830	0.0577
Brasil Capital	0.1277	0.0481	0.0704	0.0669	0.0693	0.0615
Gavea Macro	2.4007	2.8585	2.0509	1.8165	1.2367	2.3534
SPX Nimitz	0.0275	0.0292	0.0075	0.0077	0.0076	0.0108
			ten-step-ahead			
ADAM Strategy	0.0057	0.0087	0.0055	0.0086	0.0111	0.0091
Alaska Black	0.2958	0.3795	0.2201	0.0263	0.3311	0.3329
APEX Long Biased	0.2012	0.2162	0.0929	0.0706	0.1020	0.0555
Brasil Capital	0.1998	0.0460	0.1100	0.1101	0.0844	0.0700
Gavea Macro	3.2948	3.6784	2.6578	2.7515	2.8015	2.5541
SPX Nimitz	0.0467	0.0314	0.0166	0.0120	0.0103	0.0170

Table 10. Computational time, in minutes, for the model for each of the six mutual investment funds, considering the models ARIMA, SSA with $L = N/2$, SSA with $L = N/20$, SSA with L_p, robust SSA based on the L_1 norm (RLSSA) with $L = N/20$, and RSSA with L_p, and their respective engentriples, as defined in Table 2.

Investment Fund	**ARIMA**	**SSA $\frac{N}{2}$**	**SSA $\frac{N}{20}$**	**SSA L_p**	**RLSSA $\frac{N}{20}$**	**RLSSA L_p**
			one-step-ahead			
ADAM Strategy	0.0123	0.1231	0.0277	0.0253	39.768	58.804
Alaska Black	0.0222	0.0549	0.0183	0.0267	30.516	45.948
APEX Long Biased	0.2106	0.4888	0.0613	0.1752	176.18	692.18
Brasil Capital	0.2712	0.8409	0.0644	0.0648	212.60	295.20
Gavea Macro	0.0681	2.7338	0.1687	0.0976	698.34	857.58
SPX Nimitz	0.0265	1.2750	0.0774	0.0740	420.23	584.59
			five-step-ahead			
ADAM Strategy	0.0129	0.0879	0.0222	0.0256	44.019	56.524
Alaska Black	0.0181	0.0531	0.0150	0.0246	32.351	58.674
APEX Long Biased	0.2203	0.4909	0.0682	0.1840	250.85	675.41
Brasil Capital	0.2620	0.6400	0.0764	0.0675	314.59	290.72
Gavea Macro	0.0702	2.7839	0.1460	0.1034	988.02	858.96
SPX Nimitz	0.0348	1.3029	0.0805	0.0755	537.94	572.93
			ten-step-ahead			
ADAM Strategy	0.0089	0.0924	0.0344	0.0261	45.729	46.518
Alaska Black	0.0156	0.0469	0.0184	0.0263	28.140	54.289
APEX Long Biased	0.1775	0.5057	0.0678	0.1906	198.27	638.13
Brasil Capital	0.2103	0.6628	0.0726	0.0679	244.06	307.30
Gavea Macro	0.0532	2.6942	0.1724	0.1060	761.49	520.66
SPX Nimitz	0.0243	1.2388	0.0634	0.0786	407.61	316.60

3.3. Simulation Example

To verify the hypothesis raised in the previous subsection that the similarity between the results from SSA and the robust SSA algorithm can be due to the lack of outliers in the time series, in this subsection we present a simulation example where the methods are compared while analyzing a time series contaminated with outlying observations. The synthetic data were obtained by generating random values from the following function, and then we transformed them into a time series (right-hand plot in Figure 5):

$$f(t) = \exp\{0.02t + 0.5\sin(2\pi t/5)\} + \epsilon, \quad t = 1, ..., 100,$$

where ϵ is the noise generated from the $N(0, 0.1)$. A total of 100 simulated time series were considered.

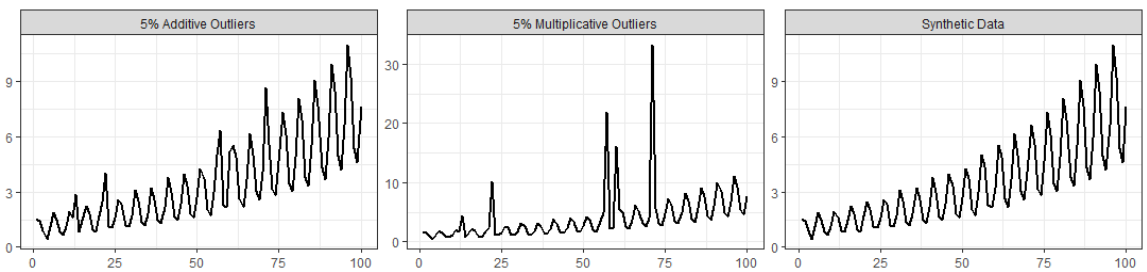

Figure 5. Synthetic data without contamination (**right**), data with 5% additive outliers (**left**), and data with 5% multiplicative outliers (**center**). The vertical axes show the simulated value and the horizontal axes show the index of the simulated observation.

The data contamination, for illustration purposes, was made by considering additive outliers and magnitude increase outliers in the following way:

- *Additive outliers:* 2%, 5%, and 10% of the time points y_i are randomly chosen to be replaced by $2 + y_i$; i.e., the values of y_i are increased by a constant value of 2, resulting in a mild contamination scenario (e.g., (left-hand plot in Figure 5));
- *Magnitude increase:* 2%, 5%, and 10% of the time points y_i are randomly chosen to be replaced by $5 \times y_i$; i.e., the time point magnitude of y_i is increased by a factor of 5, resulting in an a quite extreme contamination scenario (e.g., central plot in Figure 5).

Table 11 shows the mean of the root mean square errors for model fit, computed for each of the four models, ARIMA, SSA, robust SSA based on the L_1 norm, and robust SSA based on the Huber function, for the simulated data, based on 100 runs, using $L = 24$ and $r = 5$, and considering both contamination scenarios with 2, 5, and 10% outliers. As expected, when there is no data contamination, the classic SSA model is the most appropriated. For the mild contamination scenario with additive outliers, the robust SSA algorithms outperform both ARIMA and SSA models, the better performance being more evident when the percentage of the outliers increases. For the more extreme contamination scenario with multiplicative outliers, a similar patters was obtained, the RLSSA being the best robust algorithm, in this simulation example.

Appendix B includes a second simulation scenario where robust SSA algorithm based on the Huber function (RHSSA) outperforms the classic ARIMA and SSA models and the robust SSA algorithm based on the L_1 norm (RLSSA).

Table 12 shows mean of the root mean square errors for model forecasting ($M = 1, 5, 10$ steps-ahead), computed for each of ARIMA, SSA, and robust SSA based on the L_1 norm, for the simulated data, based on 100 runs, using $L = 24$ and $r = 5$. The results for the robust SSA based on the Huber function were not included because of their computational cost and out-performance when compared with the robust SSA based on the L_1 norm. Again, as expected, the SSA model yielded the best performance for no data contamination. For scenarios with data contamination, the best performance was obtained by the robust SSA forecasting algorithm, with a very large decrease in RMSE in many scenarios.

Table 11. Mean of the root mean square errors for model fit, computed for each of the four models, ARIMA, SSA, robust SSA based on the L_1 norm, and robust SSA based on the Huber function, for the simulated data, based on 100 runs, using $L = 24$ and $r = 5$.

% of Data Contamination	Shift	ARIMA	SSA	RLSSA	RHSSA
0%	-	0.715	0.083	0.109	0.127
2%	$y_i + 2$	0.612	0.149	0.119	0.133
5%	$y_i + 2$	0.640	0.236	0.134	0.148
10%	$y_i + 2$	0.675	0.364	0.179	0.232
2%	$y_i \times 5$	1.206	1.235	0.126	0.389
5%	$y_i \times 5$	1.828	2.289	0.167	0.929
10%	$y_i \times 5$	2.384	3.404	0.425	1.463

Table 12. Mean of the root mean square errors for model forecasting ($M = 1, 5$, and 10 steps-ahead), computed for each of the four models, ARIMA, SSA, robust SSA based on the L_1 norm, and robust SSA based on the Huber function, for the simulated data, based on 100 runs, using $L = 24$ and $r = 5$.

M	% of Cont.	Shift	Method		
			ARIMA	SSA	RLSSA
M = 1	0%	-	1.685	0.125	0.245
	5%	$y_i + 2$	0.843	0.475	0.330
	10%	$y_i + 2$	0.793	0.596	0.426
	5%	$y_i \times 5$	3.960	8.461	0.358
	10%	$y_i \times 5$	4.359	9.692	0.652
M = 5	0%	-	1.631	0.122	0.222
	5%	$y_i + 2$	0.984	0.475	0.307
	10%	$y_i + 2$	0.768	0.586	0.413
	5%	$y_i \times 5$	3.789	538.447	0.323
	10%	$y_i \times 5$	3.853	17.670	0.720
M = 10	0%	-	1.381	0.127	0.244
	5%	$y_i + 2$	1.320	0.601	0.358
	10%	$y_i + 2$	1.148	0.698	0.474
	5%	$y_i \times 5$	3.486	22.695 *	4.015
	10%	$y_i \times 5$	3.694	622.783	2.320

* 10% trimed mean. The mean value is 1.566×10^6.

4. Conclusions

In this paper we considered the problem of model fit and model forecasting in time series. In particular, we analyzed six mutual investment funds. Following up on [23], who proposed a robust SSA algorithm by replacing the standard least squares SVD by a robust SVD algorithm based on the L_1 norm [24] for model fit, we proposed another robust SSA algorithm where the robust SVD based on the Huber function is considered [25]. Moreover, we propose a forecasting strategy for the robust SSA algorithms, based on the linear recurrent SSA forecasting algorithm.

Comparisons were made between the classical SSA algorithm, the robust SSA algorithms, and the classical ARIMA model, both in terms of computational time and accuracy for model fit and model forecast. Those comparisons were made by using daily observations of six mutual investment funds, and a synthetic data set where the time series were contaminated with outlying observations.

For model fit of the six mutual investment funds, the best results were obtained for the SSA model when the window length L was set to be equal to the length of the time series divided by 20, or when the window length is defined as the length of the largest cycle in the time series. The ARIMA model and the robust SSA algorithms alternated for the second best performance. For model forecasting of the six mutual investment funds, the best overall performance was obtained for the classic SSA model considering a lower value for the window length, $L = N/20$ or L_p, followed closely by the ARIMA model and the robust SSA algorithm based on the L_1 norm.

Based on the similarity between the results from the classic SSA model and the robust SSA algorithms, both for model fit and model forecasting, one may assume that the time series data from the six mutual investment funds had no or little data contamination. To access that hypothesis and to better illustrate the usefulness of the robust SSA algorithms, using a scenario with known and controlled outliers, a simulation study and its results were presented in this article. For both mild and and more extreme contamination scenarios, the robust SSA algorithms clearly outperformed the classical AMMI and SSA models, both for model fit and for model forecasting. Another important advantage of the robust SSA algorithms, because of their use of the robust SVD, is that they allow for missing values.

In terms of computational time, the SSA model gives the best performance, the robust algorithms being the most time consuming. A possible future development to reduce the computational time in the robust SSA algorithms is to consider a similar strategy as in [39], where a randomized SVD algorithm was used to speed up the SSA algorithm.

The usefulness of the proposed approach, regarding the forecasting case, can be assessed based on forecasting competitions (e.g., [40]) or large scale forecasting studies (see, e.g., [41]).

The methodology and results presented in this paper are of great generality and can be applied to other time series applications.

Author Contributions: Conceptualization, P.C.R.; Formal analysis, P.C.R., J.P. and P.M.; Methodology, P.C.R. and M.K.; Software, P.C.R., J.P., P.M. and M.K.; Supervision, P.C.R.; Visualization, J.P. and P.M.; Writing—original draft, P.C.R., J.P., P.M. and M.K.; Writing—review and editing, P.C.R., J.P. and M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors thank the associate editor and three anonymous reviewers for providing helpful suggestions which contributed to the improvement of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ARIMA	autoregressive integrated moving average
SSA	singular spectrum analysis
SVD	singular value decomposition
RHSSA	robust SSA algorithm based on the Huber function
RLSSA	robust SSA algorithm based on the L_1 norm
RMSE	root mean squared error

Appendix A

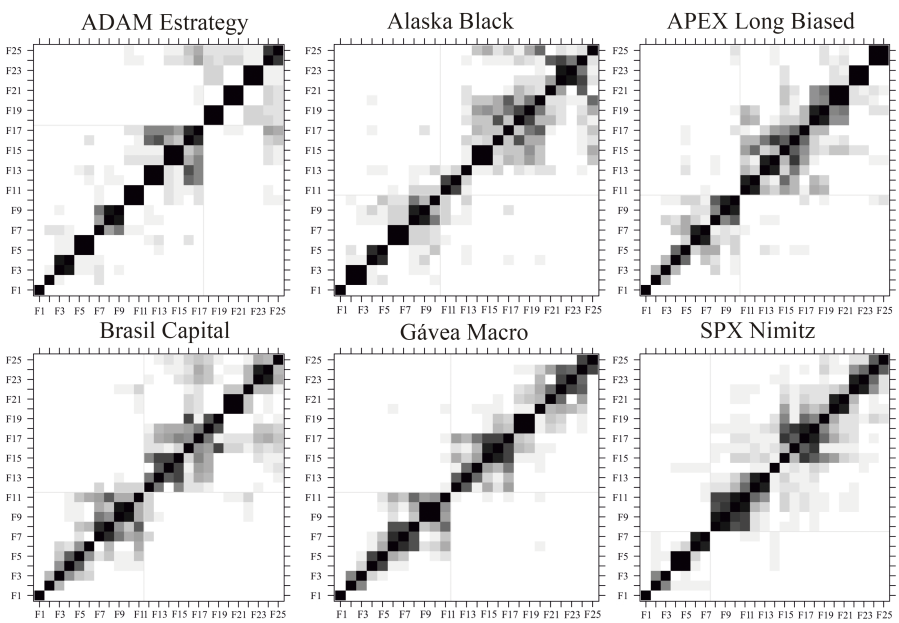

Figure A1. W-correlation matrices for each of the six mutual investment funds, ADAM Strategy, Alaska Black, APEX Long Biased, Brasil Capital, Gávea Macro and SPX Nimitz, from left to right and from top to bottom, considering an window length $L = N/2$.

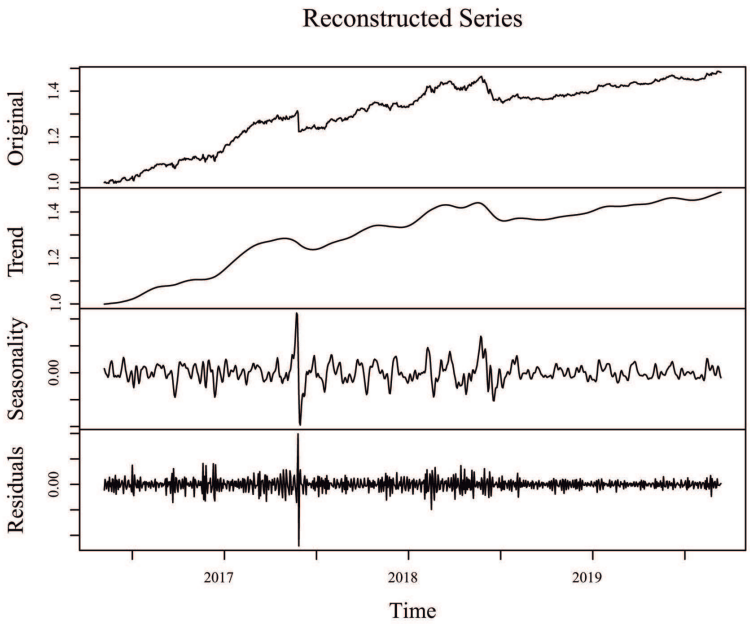

Figure A2. Decomposition of the original time series for the ADAM Strategy mutual investment fund (top panel), with a trend component (sum of individual trend components, second panel), a seasonal component (sum of individual seasonal components, third panel), and a residual (sum of the remaining components associated to noise, bottom panel), considering an window length $L = N/20 = 41$ and $r = 17$ eigentriples used for reconstruction.

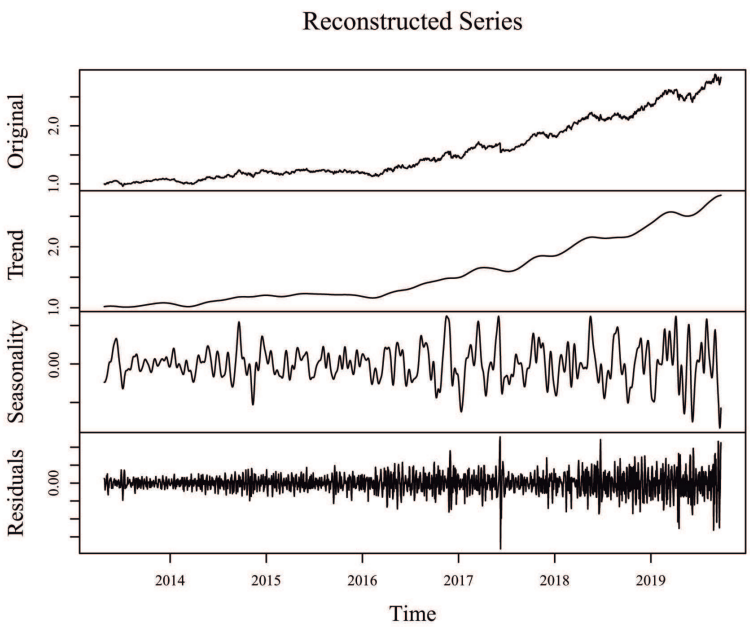

Figure A3. Decomposition of the original time series for the APEX Long Biased mutual investment fund (top panel), with a trend component (sum of individual trend components, second panel), a seasonal component (sum of individual seasonal components, third panel), and a residual (sum of the remaining components associated to noise, bottom panel), considering an window length $L = N/20 = 80$ and $r = 14$ eigentriples used for reconstruction.

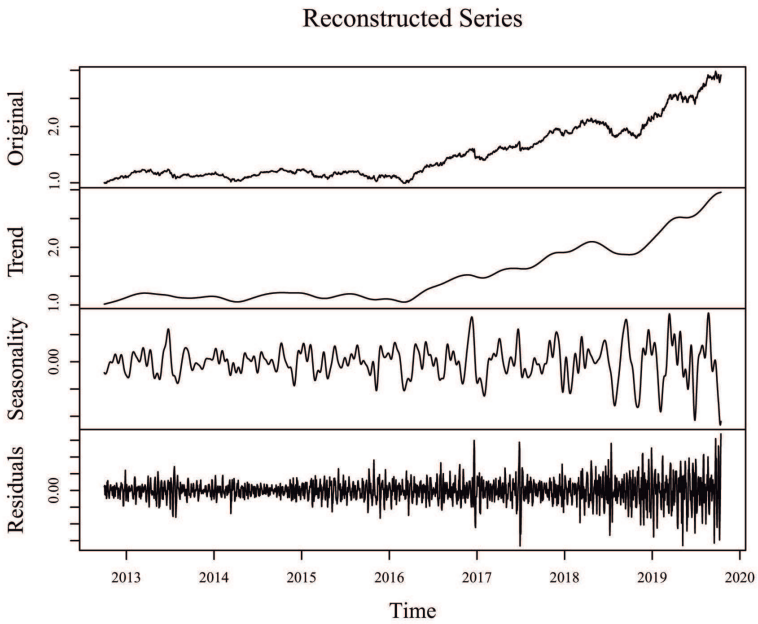

Figure A4. Decomposition of the original time series for the Brasil Capital mutual investment fund (top panel), with a trend component (sum of individual trend components, second panel), a seasonal component (sum of individual seasonal components, third panel), and a residual (sum of the remaining components associated to noise, bottom panel), considering an window length $L = N/20 = 88$ and $r = 12$ eigentriples used for reconstruction.

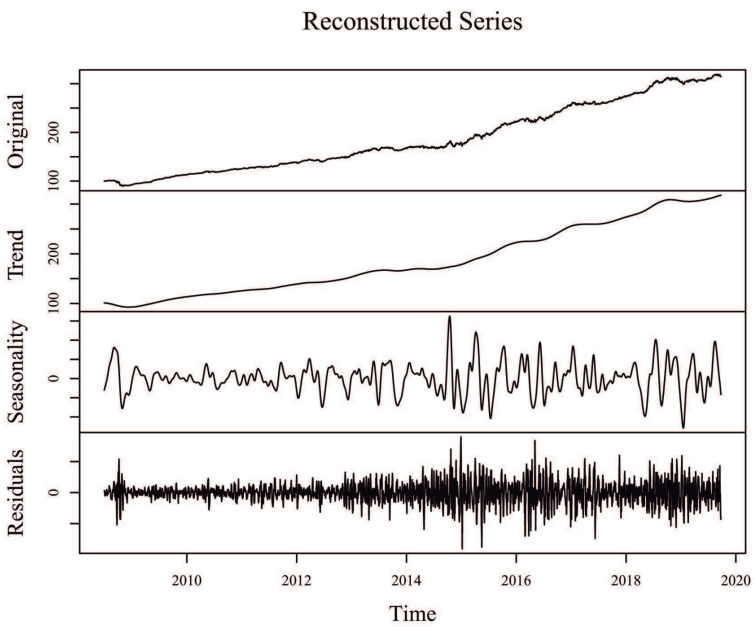

Figure A5. Decomposition of the original time series for the Gavea Macro mutual investment fund (top panel), with a trend component (sum of individual trend components, second panel), a seasonal component (sum of individual seasonal components, third panel), and a residual (sum of the remaining components associated to noise, bottom panel), considering an window length $L = N/20 = 140$ and $r = 12$ eigentriples used for reconstruction.

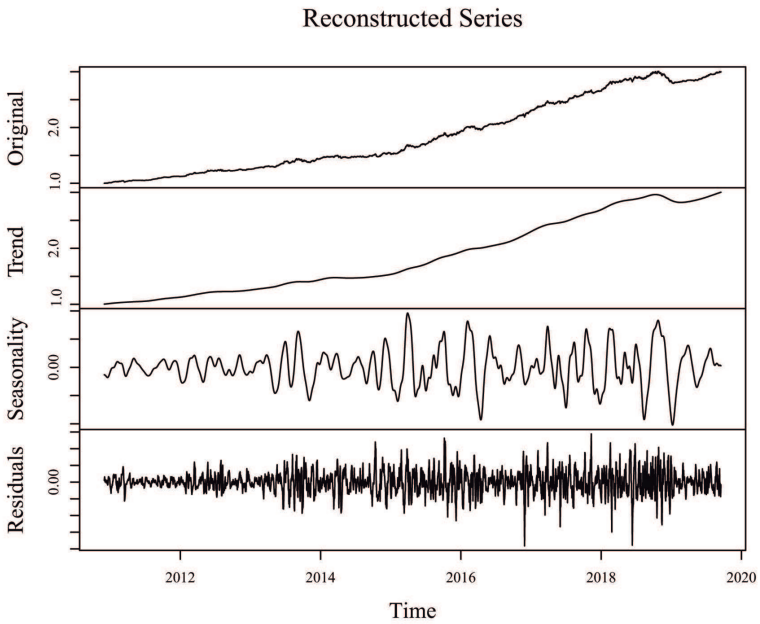

Figure A6. Decomposition of the original time series for the SPX Nimitz mutual investment fund (top panel), with a trend component (sum of individual trend components, second panel), a seasonal component (sum of individual seasonal components, third panel), and a residual (sum of the remaining components associated to noise, bottom panel), considering an window length $L = N/20 = 109$ and $r = 8$ eigentriples used for reconstruction.

Appendix B

A second synthetic dataset was obtained by generating random values from the following function and then transforming them into a time series:

$$f(t) = cos\,(2\pi wt + \phi) + \epsilon, \quad t = 1, ..., 100,$$

with $w = 3/8$, $\phi = \pi/8$ and ϵ the noise generated from the $N(0, 0.1)$ (right-hand side of Figure A7). A total of 100 simulated time series were considered.

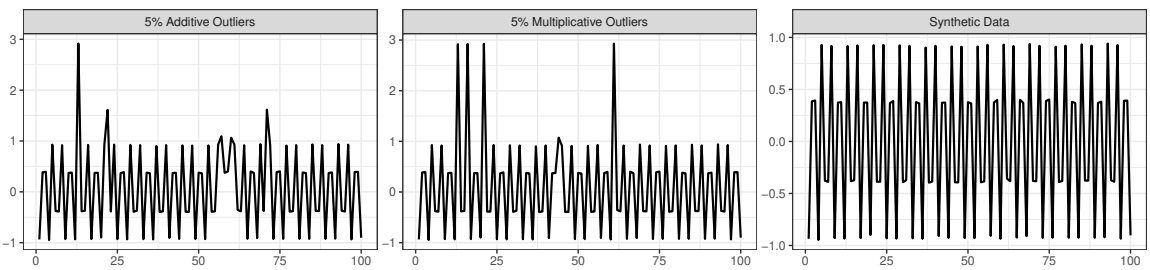

Figure A7. Synthetic data without contamination (**right**), data with 5% additive outliers (**left**), and data with 5% multiplicative outliers (**center**). The vertical axes show the simulated value and the horizontal axes show the index of the simulated observation.

The data contamination was done in the same manner as described before. An example of 5% additive outliers scenario can be found on the left-hand plot of Figure A7, and an example of 5% multiplicative outliers scenario can be found on the central plot of Figure A7. The results for the root mean square errors for model fit, computed for each of the four models, ARIMA, SSA, robust SSA based on the L_1 norm, and robust SSA based on the Huber function, can be found in Table A1.

Table A1. Mean of the root mean square errors for model fit, computed for each of the four models, ARIMA, SSA, robust SSA based on the L_1 norm, and robust SSA based on the Huber function, for the simulated data, based on 100 runs, using $L = 24$ and $r = 2$.

% of Data Contamination	Shift	ARIMA	SSA	RLSSA	RHSSA
0%	-	0.1045	0.0097	0.0099	0.0104
2%	$y_i + 2$	0.277	0.071	0.058	0.019
5%	$y_i + 2$	0.351	0.113	0.096	0.032
10%	$y_i + 2$	0.465	0.161	0.197	0.055
2%	$y_i \times 5$	0.279	0.108	0.026	0.018
5%	$y_i \times 5$	0.386	0.193	0.052	0.040
10%	$y_i \times 5$	0.484	0.338	0.075	0.098

References

1. Varga, G.; Wengert, M. A industria de fundos de investimentos no Brasil. *Rev. Econ. Adm.* **2011**, *10*, 66–109. [CrossRef]
2. Maestri, C.O.N.M.; Malaquias, R.F. Exposition to factors of the investment funds market in Brazil. *Rev. Contab. Financ.* **2017**, *28*, 61–76. [CrossRef]
3. Broomhead, D.S.; King, G.P. Extracting qualitative dynamics from experimental data. *Phys. D Nonlinear Phenom.* **1986**, *20*, 217–236. [CrossRef]
4. Fraedrich, K. Estimating the Dimensions of Weather and Climate Attractors. *J. Atmos. Sci.* **1986**, *43*, 419–432. [CrossRef]
5. Golyandina, N.; Nekrutkin, V.; Zhigljavsky, A. *Analysis of Time Series Structure: SSA and Related Techniques*; Chapman & Hall/CRC: New York, NY, USA, 2001.
6. Golyandina, N.; Zhigljavsky, A. *Singular Spectrum Analysis for Time Series*; Springer Science and Business Media: Berlin/Heidelberger, Germany, 2013.
7. Hassani, H. Singular spectrum analysis: Methodology and comparison. *J. Data Sci.* **2007**, *5*, 239–257.
8. Hassani, H.; Zhigljavsky, A. Singular spectrum analysis: methodology and application to economics data. *J. Syst. Sci. Complex.* **2009**, *22*, 372–394. [CrossRef]
9. Mahmoudvand, R.; Alehosseini, F.; Rodrigues, P.C. Forecasting mortality rate by singular spectrum analysis. *RevStat-Stat. J.* **2015**, *13*, 193–206.
10. Mahmoudvand, R.; Rodrigues, P.C. Missing value imputation in time series using singular spectrum analysis. *Int. J. Energy Stat.* **2016**, *4*, 1650005. [CrossRef]
11. Groth, A.; Ghil, M. Synchronization of world economic activity. Chaos: An Interdisciplinary. *J. Nonlinear Sci.* **2017**, *27*, 127002.
12. Mahmoudvand, R.; Konstantinides, D.; Rodrigues, P.C. Forecasting mortality rate by multivariate singular spectrum analysis. *Appl. Stoch. Models Bus. Ind.* **2017**, *33*, 717–732. [CrossRef]
13. Zabalza, J.; Qing, C.; Yuen, P.; Sun, G.; Zhao, H.; Ren, J. Fast implementation of two-dimensional singular spectrum analysis for effective data classification in hyperspectral imaging. *J. Frankl. Inst.* **2018**, *355*, 1733–1751. [CrossRef]
14. Mahmoudvand, R.; Rodrigues, P.C.; Yarmohammadi, M. Forecasting daily exchange rates: A comparison between SSA and MSSA. *RevStat-Stat. J.* **2019**, *17*, 599–616.
15. Mahmoudvand, R.; Rodrigues, P.C. Predicting the Brexit outcome using singular spectrum analysis. *J. Comput. Stat. Model.* **2019**, *1*, 9–15.
16. Ge, M.; Lv, Y.; Zhang, Y.; Yi, C.; Ma, Y. An effective bearing fault diagnosis technique via local robust principal component analysis and multi-scale permutation entropy. *Entropy* **2019**, *21*, 959. [CrossRef]
17. Sulandari, W.; Subanar; Lee, M.H.; Rodrigues, P.C. Indonesian electricity load forecasting using singular spectrum analysis. *Energy* **2020**, *190*, 116408. [CrossRef]
18. Mahmoudvand, R.; Rodrigues, P.C. Prediction intervals for the vector SSA forecasting algorithm in a median based singular spectrum analysis. *Comput. Math. Methods* **2020**. [CrossRef]
19. Reisen, V.A.; Molinares, F.F. Robust estimation in time series with long and short memory properties. *Ann. Math. Inform.* **2012**, *39*, 207–224.

20. Rodrigues, P.C.; Monteiro, A.; Lourenço, V.M. A Robust additive main effects and multiplicative interaction model for the analysis of genotype-by-environment data. *Bioinformatics* **2016**, *32*, 58–66.
21. Hassani, H.; Mahmoudvand, R.; Omer, H.N.; Silva, E.S. A preliminary investigation into the effect of outlier(s) on singular spectrum analysis. *Fluct. Noise Lett.* **2014**, *13*, 1450029. [CrossRef]
22. Rodrigues, P.C.; Mahmoudvand, R. Correlation analysis in contaminated data by singular spectrum analysis. *Qual. Reliab. Eng. Int.* **2016**, *32*, 2127–2137. [CrossRef]
23. Rodrigues, P.C.; Lourenço, V.M.; Mahmoudvand, R. A robust approach to singular spectrum analysis. *Qual. Reliab. Eng. Int.* **2018**, *34*, 1437–1447. [CrossRef]
24. Hawkins, D.M.; Liu, L.; Young, S. Robust singular value decomposition. *Natl. Inst. Stat. Sci.* **2001**, *122*, 1–12.
25. Zhang, L.; Shen, H.; Huang, J.Z. Robust regularized singular value decomposition with application to mortality data. *Ann. Appl. Stat.* **2013**, *7*, 1540–1561. [CrossRef]
26. Brockwell, P.J.; Davis, R.A. *Introduction to Time Series and Forecasting*; Springer: New York, NY, USA, 1996.
27. Ripley, B.D. Time Series in R 1.5.0. R News, 2/2, 2–7. Available online: https://www.r-project.org/doc/Rnews/Rnews_2002-2.pdf (accessed on 6 January 2020).
28. Rodrigues, P.C.; Mahmoudvand, R. The benefits of multivariate singular spectrum analysis over the univariate version. *J. Frankl. Inst.* **2018**, *355*, 544–564. [CrossRef]
29. Ghil, M.; Allen, M.R.; Dettinger, M.D.; Ide, K.; Kondrashov, D.; Mann, M.E.; Robertson, A.W.; Saunders, A.; Tian, Y.; Varadi, F.; et al. Advanced spectral methods for climate time series. *Rev. Geophys.* **2002**, *40*, 3.1–3.41. [CrossRef]
30. Mahmoudvand, R.; Rodrigues, P.C. A new parsimonious recurrent forecasting model in singular spectrum analysis. *J. Forecast.* **2018**, *37*, 191–200. [CrossRef]
31. Rodrigues, P.C.; Mahmoudvand, R. A new approach for the vector forecast algorithm in singular spectrum analysis. *Commun. Stat. Simul. Comput.* **2020**. [CrossRef]
32. Wen, Q.; Gao, J.; Song, X.; Sun, L.; Tan, J. RobustTrend: A Huber loss with a combined first and second order difference regularization for time series trend filtering. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 3856–3862.
33. Bouwmans, T.; Aybat, N.S.; Zahzah, E. *Handbook of Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*; CRC Press: New York, NY, USA, 2016.
34. Huber, P.J. Robust estimation of a location parameter. *Ann. Math. Stat.* **1964**, *35*, 73–101. [CrossRef]
35. Huber, P.J.; Ronchetti, E.M. *Robust Statistics*; Wiley: Hoboken, NJ, USA, 2009.
36. Hyndman, R.J.; Khandakar, Y. Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* **2008**, *26*, 1–22.
37. de Carvalho, M.; Rua, A. Real-Time Nowcasting the US Output Gap: Singular Spectrum Analysis at Work. *Int. J. Forecast.* **2017**, *33*, 185–198. [CrossRef]
38. Golyandina, N.; Korobeynikov, A.; Shlemov, A.; Usevich, K. Multivariate and 2D Extensions of Singular Spectrum Analysis with the Rssa Package. *J. Stat. Softw.* **2015**, *67*. Available online: https://www.jstatsoft.org/article/view/v067i02 (accessed on 6 January 2020). [CrossRef]
39. Rodrigues, P.C.; Tuy, P.G.S.E.; Mahmoudvand, R. Randomized singular spectrum analysis for long time series. *J. Stat. Comput. Simul.* **2018**, *88*, 1921–1935. [CrossRef]
40. Hyndman, R.J. A brief history of forecasting competitions. *Int. J. Forecast.* **2020**, *36*, 7–14. [CrossRef]
41. Papacharalampous, G.; Tyralis, H.; Koutsoyiannis, D. Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 481–514. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).