A Comprehensive Causality Test Based on the Singular Spectrum Analysis

Hossein Hassani[a]*, Anatoly A. Zhigljavsky[a], Kerry Patterson[b], Abdol S. Soofi[c]

[a] *Group of Statistics, Cardiff School of Mathematics, Cardiff University, CF24 4AG, UK.*

[b] *School of Economics, University of Reading, UK.*

[c] *Department of Economics, University of Wisconsin-Platteville, Platteville, WI 53818-3099, USA.*

Abstract

In this paper, we consider the concept of casual relationship between two time series based on the singular spectrum analysis. We introduce several criteria which characterize this causality. The criteria are based on the forecasting accuracy and the predictability of the direction of change. The performance of the proposed tests is examined using different real time series.

Keywords: Causality, Singular spectrum analysis, Time series, Forecasting.

1 Introduction

A question that frequently arises in time series analysis is whether one economic variable can help in predicting another economic variable. One way to address this question was proposed by Granger (1969). Granger (1969) formalized a causality concept as follows: process X does not cause process Y if (and only if) the capability to predict the series Y based on the histories of all observables is unaffected by the omission of X's history (see also Granger (1980)). Testing causality, in the Granger sense, involves using F-tests to test whether lagged information on one variable, say X, provides any statistically significant information about another variable, say Y, in the presence of lagged Y. If not, then "Y does not Granger-cause X."

Criteria for Granger causality typically have been realized in the framework of multivariate Gaussian statistics via vector autoregressive (VAR) models. It is worth mentioning that the linear Granger causality is not causality in a broader sense of the word. It just considers linear prediction and time-lagged dependence between two time series. The definition of Granger causality does not mention anything about possible instantaneous correlation between two series X_T and Y_T. (If the innovation to X_T and the innovation to Y_T are correlated then it is sometimes called instantaneous causality.) It is not rare when

*Corresponding author. Tel.:+44 (0)29 208 74811; Fax: +44 (0)29 208 74199 .
E-mail addresses: hassanih@cf.ac.uk (H. Hassani), Soofi@uwplatt.edu (A.S. Soofi), ZhigljavskyAA@cardiff.ac.uk (A.A. Zhigljavsky)

instantaneous correlation between two time series can be easily revealed, but since the causality can go either way, one usually does not test for instantaneous correlation. In this paper, several of our causality tests incorporate testing for the instantaneous causality. One more drawback of the Granger causality test is the dependence on the right choice of the conditioning set. In reality one can never be sure that the conditioning set selected is large enough (in short macro-economic series one is forced to choose a low dimension for the VAR model). Moreover, there are special problems with testing for Granger causality in co-integrated relations (see Toda and Phillips (1991)).

The original notion of Granger causality was formulated in terms of linear regression, but there are some nonlinear extensions in the literature (see, for example, Chu et al. (2004)). Hiemstra and Jones (1994) also propose a nonparametric test which seems to be most used test in testing nonlinear causality. However, this method also has several drawbacks: *i*) the test is not consistent, at least against a specific class of alternatives (Diks and Panchenko (2005)), *ii*) there are restrictive assumptions in this approach (Bosq (1998)) and *iii*) the test can severely over-reject the null hypothesis of non-causality (Diks and Panchenko (2006)).

It is also important to note that Granger causality attempts to capture an important aspect of causality, but it is not meant to capture all. A method based on the information theory has realized a more general Granger causality measure that accommodates in principle arbitrary statistical processes (Diks and DeGoede (2001)). Su and White (2008) propose a nonparametric test of conditional independence based on the weighted Hellinger distance between the two conditional densities. There are also a number of alternative methods, but they are rarely used.

We overcome many of these difficulties by implementing a different technique for capturing the causality; this technique uses the singular spectrum analysis (SSA) technique; a nonparametric technique that works with arbitrary statistical processes, whether linear or nonlinear, stationary or non-stationary, Gaussian or non-Gaussian.

The general aim of this study is to assess the degree of association between two arbitrary time series (these associations are often called causal relationships as they might be caused by the genuine causality) based on the observation of these time series. We develop new tests and criteria which will be based on the forecasting accuracy and predictability of the direction of change of the SSA algorithms.

The structure of the paper is as follows. Section 2 briefly describes the SSA technique. The proposed criteria and statistical tests are considered in Section 3. Empirical results are presented in Section 4. Conclusions are given in Section 5. Appendix contains some necessary technical details about SSA.

2 Singular Spectral Analysis

A thorough description of the theoretical and practical foundations of the SSA technique (with many examples) can be found in Golyandina, *et al.* (2001) and Danilov and A. Zhigljavsky (1997). There are many papers where SSA has been applied to real-life time series. In particular, the performance of the SSA technique has been compared with other techniques for forecasting economics time series (Hassani (2007) and Hassani, *et*

al. (2009a–d)), and see also Hassani (2009e) for a new SSA-based algorithm and its application for forecasting.

Consider the real-valued nonzero time series $Y_T = (y_1, \ldots, y_T)$ of sufficient length T. The main purpose of SSA is to decompose 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. This is followed by a reconstruction the original series.

The state of a process at time t is considered to capture the relevant information of the process up to time t. Moreover, it is the state of a process that is to be predicted. Assume that the process is governed by some linear recurrent formula (LRF), then having the LRF and embedding theory, forecasting the process at time t may be regarded as forecasting the state vector. According to the SSA terminology, the problem of forecasting a new vector requires (a) a window of some suitable length and (b) the number of eigenvalues.

The SSA technique consists of two complementary stages: decomposition and reconstruction, both of which include two separate steps. At the first stage we decompose the series and at the second stage we reconstruct the original series and use the reconstructed series (which is without noise) for forecasting new data points. Below we provide a brief discussion on the methodology of the SSA technique (for more description of the SSA algorithm, forecasting procedure and parameter estimation, see Appendix A).

2.1 A short description of the Basic SSA

We consider a time series $Y_T = (y_1, \ldots, y_T)$. Fix L ($L \leq T/2$), the window length, and let $K = T - L + 1$.

(*Basic SSA*)

Step 1. (*Computing the trajectory matrix*): this transfers a one-dimensional time series $Y_T = (y_1, \ldots, y_T)$ into the multi-dimensional series $X_1, \ldots, X_K$ with vectors $X_i = (y_i, \ldots, y_{i+L-1})' \in \mathbf{R}^L$, where $K = T - L + 1$. Vectors X_i are called *L-lagged vectors* (or, simply, *lagged vectors*). The single parameter of the embedding is the *window length* L, an integer such that $2 \leq L \leq T$. The result of this step is the trajectory matrix $\mathbf{X} = [X_1, \ldots, X_K] = (x_{ij})_{i,j=1}^{L,K}$.

Step 2. (*Constructing a matrix for applying SVD*): compute the matrix $\mathbf{X}\mathbf{X}^T$.

Step 3. (*SVD of the matrix* $\mathbf{X}\mathbf{X}^T$): compute the eigenvalues and eigenvectors of the matrix $\mathbf{X}\mathbf{X}^T$ and represent it in the form $\mathbf{X}\mathbf{X}^T = P\Lambda P^T$. Here $\Lambda = diag(\lambda_1, \ldots, \lambda_L)$ is the diagonal matrix of eigenvalues of $\mathbf{X}\mathbf{X}^T$ ordered so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L \geq 0$ and $P = (P_1, P_2, \ldots, P_L)$ is the corresponding orthogonal matrix of eigen–vectors of $\mathbf{X}\mathbf{X}^T$.

Step 4. (*Selection of eigen–vectors*): select a group of l ($1 \leq l \leq L$) eigen–vectors $P_{i_1}, P_{i_2}, \ldots, P_{i_l}$.

The grouping step corresponds to splitting the elementary matrices $\mathbf{X}_i$ into several groups and summing the matrices within each group. Let $I = \{i_1, \ldots, i_l\}$ be a group of indices $i_1, \ldots, i_l$. Then the matrix $\mathbf{X}_I$ corresponding to the group I is defined as $\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_l}$.

Step 5. (*Reconstruction of the one-dimensional series*): compute the matrix $\tilde{\mathbf{X}} = ||\tilde{x}_{i,j}|| = \sum_{k=1}^{l} P_{i_k} P_{i_k}^T \mathbf{X}$ as an approximation to $\mathbf{X}$. Transition to the one–dimensional series can now be achieved by averaging over the diagonals of the matrix $\tilde{\mathbf{X}}$.

2.2 Multivariate Singular Spectrum Analysis: MSSA

Multivariate (or multichannel) SSA is an extension of the standard SSA to the case of multivariate time series (see e.g. Broomhead and King (1986)). It can be described as follows. Assume we have two time series $X_T = x_1, \ldots, x_T$ and $Y_T = y_1, \ldots, y_T$ simultaneously (a bivariate approach), and let L be window length. Using embedding terminology, we can define the trajectory matrices $\mathbf{M}_X$ and $\mathbf{M}_Y$ of the one-dimensional time series X_T and Y_T, respectively. The trajectory matrix $\mathbf{M}$ can then be defined as

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_X \\ \mathbf{M}_Y \end{pmatrix} . \tag{1}$$

The other stages of the Basic Multivariate SSA (or MSSA) procedure are identical to the Basic SSA. The generalization to the case of several series is straightforward.

There are numerous examples of successful application of the multivariate SSA (see, for example, Plaut and Vautard, 1994; Danilov and Zhigljavsky, 1997).

3 Causality Criteria

3.1 Forecasting accuracy based criterion

The first criterion we use here is based on the out-of-sample forecasting, which is very common in the framework of Granger causality. The question behind Granger causality is whether forecasts of one variable can be improved using the history of another variable. Here, we compare the forecasted value obtained using the univariate procedure, SSA, and also the multivariate one, MSSA. We then compare the forecasted values with the actual values to evaluate the forecasting error. If the forecasting error using MSSA is significantly smaller than the forecasting error of the univariate SSA, we then conclude that there is a casual relationship between these series.

Let us consider in more detail the procedure of constructing a vector of forecasting error for an out-of-sample test. In the first step we divide the series $X_T = (x_1, \ldots, x_T)$ into two separate subseries X_R and X_F: $X_T = (X_R, X_F)$ where $X_R = (x_1, \ldots, x_R)$, and $X_F = (x_{R+1}, \ldots, x_T)$. The subseries X_R is used in reconstruction step to provide the noise free series $\tilde{X}_R$. The noise free series $\tilde{X}_R$ is then used for forecasting the subseries X_F using either the recurrent or vector forecasting algorithm, see Appendix A. The subseries X_F will be forecasted using the recursive h-step ahead forecast with SSA and MSSA. The forecasted points $\hat{X}_F = (\hat{x}_{R+1}, \ldots, \hat{x}_T)$ are then used for computing the forecasting error, and the vector $(x_{R+2}, \ldots, x_T)$ is forecasted using the new subseries $(x_1, \ldots, x_{R+1})$. This procedure is continued recursively up to the end of series, yielding the series of h-step-ahead forecasts for univariate and multivariate algorithms. Therefore, the vector of h-step-ahead forecasts obtained can be used in examining the association (or order h) between the two series. Let us now consider a formal procedure of constructing a criterion of SSA causality of order h between two arbitrary time series.

Criterion

Let $X_T = (x_1, \ldots, x_T)$ and $Y_T = (y_1, \ldots, y_T)$ denote two different time series of length T. Set window lengths L_x and L_y for the series X_T and Y_T, respectively. Here, for simplicity assume $L_x = L_y$. Using the embedding terminology, we construct trajectory matrices $\mathbf{X} = [X_1, \ldots, X_K]$ and $\mathbf{Y} = [Y_1, \ldots, Y_K]$ for the series X_T and Y_T.

Consider an arbitrary loss function $\mathcal{L}$. In econometrics, the loss function $\mathcal{L}$ is usually selected so that it minimizes the mean square error of the forecast. Let us first assume that the aim is to forecast the series X_T. Thus, the aim is to minimize $\mathcal{L}(X_{K+H_x} - \hat{X}_{K+H_x})$, where the vector $\hat{X}_{K+H_x}$ is an estimate, obtained using a forecasting algorithm, of the vector X_{K+H_x} of the trajectory matrices $\mathbf{X}$. Note that, for example, when $H_x = 1$, $\hat{X}_{K+1}$ is an estimate of the vector $X_{K+1} = (x_{T+1}, \ldots, x_{T+h})$ where h varies between 1 and L. In a vector form, this means that an estimate of X_{K+1} can be obtained using the trajectory matrix $\mathbf{X}$ consisting of vectors $[X_1, \ldots, X_K]$. The vector X_{K+H_x} can be forecasted using either univariate SSA or MSSA. Let us first consider the univariate approach. Define

$$\Delta_{X_{K+H_x}} \equiv \mathcal{L}(X_{K+H_x} - \hat{X}_{K+H_x}), \tag{2}$$

where $\hat{X}_{K+H_x}$ is obtained using univariate SSA; that is, the estimate $\hat{X}_{K+H_x}$ is obtained only from the vectors $[X_1, \ldots, X_K]$.

Let $X_T = (x_1, \ldots, x_T)$ and $Y_{T+d} = (y_1, \ldots, y_{T+d})$ denote two different time series to be considered simultaneously and consider the same window length L for both series. Now, we forecast $x_{T+1}, \ldots, x_{T+h}$ using the information provided by the series Y_{T+d} and X_T. Next, compute the following statistic:

$$\Delta_{X_{K+H_x}|Y_{K+H_y}} \equiv \mathcal{L}(X_{K+H_x} - \tilde{X}_{K+H_x}). \tag{3}$$

where $\tilde{X}_{K+H_x}$ is an estimate of X_{K+H_x} obtained using multivariate SSA. This means that we simultaneously use vectors $[X_1, \ldots, X_K]$ and $\left[Y_1, \ldots, Y_{K+H_y}\right]$ in forecasting vector X_{K+H_x}. Now, define the criterion:

$$F^{(h,d)}_{X|Y} = \frac{\Delta_{X_{K+H_x}|Y_{K+H_y}}}{\Delta_{X_{K+H_x}}} \tag{4}$$

corresponding to the h step ahead forecast of the series X_T in the presence of the series Y_{T+d}; here d shows the lagged difference between series X_T and Y_{T+d}, respectively. Note that d is any given integer (even negative). For example, $F^{(h,0)}_{X|Y}$ indicates that we use the same series length in h step ahead forecasting series X; we use the series X_T and Y_T simultaneously. $F^{(h,0)}_{X|Y}$ can be considered as a common multivariate forecasting system for time series with the same series length. The criterion $F^{(h,0)}_{X|Y}$ can then be used in evaluating two instantaneous causality. Similarly, $F^{(h,1)}_{X|Y}$ indicates that there is an additional information for series Y and that this information is one step ahead of the information for the series X; we use the series X_T and Y_{T+1} simultaneously.

If $F^{(h,d)}_{X|Y}$ is small, then having information obtained from the series Y helps us to have a better forecast of the series X. This means there is a relationship between series X and Y of order h according to this criterion. In fact, this measure of association shows how

much more information about the future values of series X contained in the bivariate time series (X, Y) than in the series X alone. If $F^{(h,d)}_{X|Y}$ is very small, then the predictions using the multivariate version are much more accurate than the predictions by the univariate SSA. If $F^{(h,d)}_{X|Y} < 1$, then we conclude that the information provided by the series Y can be regarded as useful or *supportive* for forecasting the series X. Alternatively, if the values of $F^{(h,d)}_{X|Y} \geq 1$, then either there is no detectable association between X and Y or the performance of the univariate version is better than the multivariate version (this may happen, for example, when the series Y has structural breaks which may misdirect the forecasts of X).

To asses which series is more *supportive* in forecasting, we need to consider another criteria. We obtain $F^{(h,d)}_{Y|X}$ in a similar manner. Now, these measures tell us whether using extra information about time series Y_{T+d} (or X_{T+d}) supports X_T (or Y_T) in h-step forecasting. If $F^{(h,d)}_{Y|X} < F^{(h,d)}_{X|Y}$, we then conclude that X is more *supportive* than Y, and if $F^{(h,d)}_{X|Y} < F^{(h,d)}_{Y|X}$, we then conclude that Y is more *supportive* than X.

Let us now consider a definition for a feedback system according to the above criteria. If $F^{(h,d)}_{Y|X} < 1$ and $F^{(h,d)}_{X|Y} < 1$, we then conclude that there is a feedback system between series X and Y. We shall call it F-feedback (forecasting feedback) which means that using a multivariate system improves the forecasting for both series. For a F-feedback system, X and Y are mutually supportive.

Statistical test

To check if the discrepancy between the two forecasting procedures are statistically significant we may apply the Diebold and Mariano (1995) test statistic, with the corrections suggested by Harvey et al. (1997). The quality of a forecast is to be judged on some specified function $\mathcal{L}$ as a loss function of the forecast error. Then, the null hypothesis of equality of expected forecast performance is $E(D_t) = 0$, where $D_t = (D_{X_{K+H_x}|Y_{K+H_y}} - D_{X_{K+H_x}})$ and $D_{X_{K+H_x}|Y_{K+H_y}}$ and $D_{X_{K+H_x}}$ are the vectors of the forecast errors obtained with the univariate and multivariate approaches, respectively. In our case, $\mathcal{L}$ is the quadratic loss function. The modified Diebold and Mariano statistic for a h step ahead forecast and the number of n forecasted points is

$$S = \bar{D} \sqrt{\frac{n + 1 - 2h + h(h-1)/n}{n\, \widehat{var}(\bar{D})}}$$

where $\bar{D}$ is the sample mean of the vector D_t and $\widehat{var}(\bar{D})$ is, asymptotically $n^{-1}\left(\widehat{\gamma}_0 + 2\sum_{k=1}^{h-1}\widehat{\gamma}_k\right)$, where $\widehat{\gamma}_k$ is the k-th autocovariance of D_t and can be estimated by $n^{-1}\sum_{t=k+1}^{n}(D_t - \bar{D})(D_{t-k} - \bar{D})$. The S statistic has an asymptotic standard normal distribution under the null hypothesis and its correction for a finite samples follows the Student's t distribution with $n - 1$ degrees of freedom.

3.2 Direction of change based criterion

Ash et al. (1997) argue that for some purposes, it may be more harmful to make a smaller prediction error yet fail in predicting the direction of change, than to make a

larger directionally correct error. Clements and Smith (1999) discuss that the value of a model's forecasts may be better measured by the direction of change. Heravi et al. (2004) argue that the direction of change forecasts are particularly important in economics for capturing the business cycle movement relating to expansion versus contraction phases of the cycle. Thus as another measure of forecasting performance, we also compute the percentage of forecasts that correctly predict the direction of change.

Criterion

The direction of change criterion shows the proportion of forecasts that correctly predict the direction of the series movement. For the forecasts obtained using only X_T (univariate case), let Z_{X_i} take the value 1 if the forecast observations correctly predicts the direction of change and 0 otherwise. Then $\bar{Z_X} = \sum_{i=1}^{n} Z_{X\,i}/n$ shows the proportion of forecasts that correctly predict the direction of the series movement (in forecasting n data points). The Moivre-Laplace central limit theorem implies that, for large samples, the test statistic $2(\bar{Z_X} - 0.5)N^{1/2}$ is approximately distributed as standard normal. When $\bar{Z_X}$ is significantly larger than 0.5, then the forecast is said to have the ability to predict the direction of change. Alternatively, if $\bar{Z_X}$ is significantly smaller than 0.5, the forecast tends to give the wrong direction of change.

For the multivariate case, let $Z_{X|Y,i}$ takes a value 1 if the forecast series correctly predicts the direction of change of the series X having information about the series Y and 0 otherwise. Then, we define the following criterion:

$$D_{X|Y}^{(h,d)} = \frac{\bar{Z_X}}{\bar{Z}_{X|Y}} \tag{5}$$

where h and d have the same interpretation as for $F_{X|Y}^{(h,d)}$. The criterion $D_{X|Y}^{(h,d)}$ characterizes the improvement we are getting from the information contained in Y_{T+h} (or X_{T+h}) for forecasting the direction of change in the h step ahead forecast.

If $D_{X|Y}^{(h,d)} < 1$, then having information about the series Y helps us to have a better prediction of the direction of change for the series X. This means that there is an association between the series X and Y with respect to this criterion. This criterion informs us how much more information we have in the bivariate time series relative to the information contained in the univariate time series alone with respect to the prediction of the direction of change. Alternatively, if $D_{X|Y}^{(h,d)} > 1$, then the univariate SSA is better than the multivariate version.

To find out which series is more supportive in predicting the direction of change, we consider the following criterion. We compute $D_{Y|X}^{(h,d)}$ in a similar manner. Now, if $D_{Y|X}^{(h,d)} < D_{X|Y}^{(h,d)}$, then we conclude that that X is more supportive (with respect to predicting the direction) to Y than Y to X.

Similar to the consideration of the forecasting accuracy criteria, we can define a feedback system based on the criteria characterizing the predictability of the direction of change. Let us introduce a definition for a feedback system according to $D_{X|Y}^{(h,d)}$ and $D_{Y|X}^{(h,d)}$. If $D_{Y|X}^{(h,d)} < 1$ and $D_{X|Y}^{(h,d)} < 1$, we conclude that there is a feedback system between the series X and Y for prediction of the direction of change. We shall call this type of feedback

D-feedback. The existence of a D-feedback in a system yields that the series in the system help each other to capture the direction of the series movement with higher accuracy.

Statistical test

Let us describe a statistical test for the criterion $D_{X|Y}^{(h,d)}$. As in the comparison of two proportions, when we test the hypothesis about the difference between two proportions, first we need to know whether the two proportions are dependent. The test is different depending on whether the proportions are independent or dependent. In our case, obviously, Z_X and $Z_{X|Y}$ are dependent. We therefore consider this dependence in the following procedure. Let us consider the test statistic for the difference between Z_X and $Z_{X|Y}$. Assume that Z_X and $Z_{X|Y}$, in forecasting n future points of the series X, are arranged as Table 1.

$Z_{X\|Y}$	Z_X	number
1	1	a
1	0	b
0	1	c
0	0	d
Total		$n = a + b + c + d$

Table 1: An arrangement of Z_X and $Z_{X|Y}$ in forecasting n future points of the series X.

Then the estimated proportion using the multivariate system is $P_{X|Y} = (a+b)/n$, and the estimated proportion using the univariate version is $P_X = (a + c)/n$. The difference between the two estimated proportions is

$$\pi = P_{X|Y} - P_X = \frac{a+b}{n} - \frac{a+c}{n} = \frac{b-c}{n} \tag{6}$$

Since the two population probabilities are dependent, we cannot use the same approach for estimating the standard error of the difference that is used for independent case. The formula for the estimated standard error for the dependent case was given in Fleiss (1981):

$$\hat{SE}(\pi) = \frac{1}{n}\sqrt{(b+c) - \frac{(b-c)^2}{n}}. \tag{7}$$

Let us consider the related test for the difference between two dependent proportions, then the null and alternative hypotheses are:

$$\begin{aligned} H_0 &: \pi_d = \Delta_0 \\ H_a &: \pi_d \neq \Delta_0 \end{aligned} \tag{8}$$

The test statistic, assuming that the sample size is large enough for the normal approximation to the binomial to be appropriate, is:

$$T_{\pi_d} = \frac{\pi - \Delta_0 - 1/n}{\hat{SE}(\pi)} \tag{9}$$

where $1/n$ is the continuity correction. In our case $\Delta_0 = 0$. The test statistic then becomes:

$$T_{\pi_d} = \frac{(b-c)/n - 1/n}{1/n\sqrt{(b+c) - (b-c)^2/n}} = \frac{b-c-1}{\sqrt{(b+c) - (b-c)^2/n}} \tag{10}$$

The test is valid when the average of the discordant cell frequencies, $(b+c)/2$, is equal or more than 5. However, then it is less than 5, a binomial test can be used. Note that under the null hypothesis of no difference between samples Z_X and $Z_{X|Y}$, T_{π_d} is asymptotically distributed as standard normal.

3.3 Comparison with Granger causality test

Linear Granger causality test

Let X_T and Y_T be two stationary time series. To test for Granger causality we compare the full and the restricted model. The full model is given by

$$x_t = \phi_0 + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + \psi_1 y_{t-1} + \ldots + \psi_p y_{t-p} + \varepsilon_{t_{x|y}} \tag{11}$$

where $\{\varepsilon_{t_{x|y}}\}$ is an *iid* sequence with zero mean and variance $\sigma_{x|y}$, ϕ_i and ψ_i are model parameters. The null hypothesis stating that Y_T does not Granger cause X_T is:

$$H_0 = \psi_{L+1} = \psi_2 = \ldots = \psi_p = 0 \tag{12}$$

If the null hypothesis holds, the full model (11) is reduced to the restricted model as follows:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-L+1} + \varepsilon_{t_x} \tag{13}$$

where ε_{t_x} is *iid* sequence with zero mean and variance σ_x. The forecasting results obtained by the restricted model (13) are compared to those obtained using the full model (11) to test for Granger causality. We then apply the F-test (or some other similar test) to obtain a p-value for whether the full model results are better than the restricted model results. If the full model provides a better forecast, according to the standard loss functions, we then conclude that Y_T Granger causes X_T. Thus, Y_T would Granger cause X_T if Y_T occurs before and contains information useful in forecasting X_T. As the formula of Granger causality shows, the test, in fact, is a mathematical formulation based on the linear regression modeling of two time series. Therefore, the above formulation of Granger causality can only give information about linear features of the series.

Let us now compare the similarity and dissimilarity of the proposed algorithm which is based on the SSA forecasting algorithm with the Granger causality procedure. As mentioned in the description of the SSA forecasting algorithm, the last component X_L of any vector $X = (x_1, \ldots, x_L)^T \in \mathfrak{L}_r$ is a linear combination of the first $L-1$ components $(x_1, \ldots, x_{L-1})$ such that:

$$x_L = \alpha_1 x_{L-1} + \ldots + \alpha_{L-1} x_1.$$

where $A = (\alpha_1, \ldots, \alpha_{L-1})$ can be estimated using equation (23) of the Appendix A. Thus, the univariate version of SSA is given by

$$x_t = \alpha_1 x_{t-1} + \ldots + \alpha_{L-1} x_{t-L+1} \tag{14}$$

As can be seen from (14), a univariate SSA forecasting formula is similar to the restricted model. However, the procedure of parameter estimation in the SSA technique and the Granger model are quite different. Both are linear combinations of previous observations, and from this point of view both are similar. The multivariate version of SSA is a system in which X_T and Y_T are considered simultaneously to estimate vectors A and B as follows. The multivariate forecasting system is:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \alpha_1 x_{t-1} + \ldots + \alpha_{L-1} x_{t-L+1} \\ \beta_1 y_{t-1} + \ldots + \beta_{L-1} y_{t-L+1} \end{pmatrix} \tag{15}$$

where the vectors $A = (\alpha_1, \ldots, \alpha_{L-1})$ and $B = (\beta_1, \ldots, \beta_{L-1})$ are estimated using the multivariate system. As equation (15) shows, the multivariate SSA is not similar to the Granger full model. An obvious discrepancy is that we use the value of the series Y in parameter estimation and also in forecasting series X in the Granger based test, while we use the information provided in the subspaces generated by Y in multivariate SSA and not the observed values. More specifically, the Granger causality test uses a linear combination of the values of both series X and Y in the full model, whereas multivariate SSA uses the information provided by X and Y in construction of the subspace and not the observations themselves.

Nonlinear Granger causality test

It is worth mentioning that the simultaneous reconstruction of the trajectory matrices $\mathbf{X}$ and $\mathbf{Y}$ in the MSSA technique is also used in testing for Granger causality between two nonlinear time series. Let us consider the concept of nonlinear Granger causality in more detail. Let $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}]$ be the joint trajectory matrix with lagged difference zero (same value of K in the trajectory matrix $\mathbf{X}$ and $\mathbf{Y}$). In the joint phase space consider a small neighborhood of any vector. The dynamics of this neighborhood can be described via a linear approximation and a linear autoregressive model can be used to predict the dynamics within the neighborhood. Assume that the vectors of prediction errors are given by $\mathbf{e}_{X|Y}$ and $\mathbf{e}_{Y|X}$. The reconstruction and the fitting procedure are now employed for the individual time series X_T and Y_T in the same neighborhood and the vector of prediction errors $\mathbf{e}_X$ and $\mathbf{e}_Y$ are then computed. Now, we compute the following criteria

$$\frac{Var(\mathbf{e}_{X|Y})}{Var(\mathbf{e}_X)}, \quad \frac{Var(\mathbf{e}_{Y|X})}{Var(\mathbf{e}_Y)} \tag{16}$$

The above procedure is then repeated for various regions on the attractor, each column of trajectory matrices $\mathbf{X}$ and $\mathbf{Y}$, and the average of the above criteria are used. The above criteria, clearly, can be considered as a function of neighborhood size. If the ratios are smaller than 1, we then conclude that there is a nonlinear Granger causal relation between two series. The similarity of nonlinear Granger causality test with SSA causality test is only in the construction of the trajectory matrices $\mathbf{X}$ and $\mathbf{Y}$ using embedding terminology, which is only the first step of SSA. Otherwise, the Granger nonlinear test is different from the test considered here. Moreover, the major drawback of the standard nonlinear analysis is that it requires a long time series, while the SSA technique works well for short and long time series (see, for example, Hassani and Zhigljavsky (2009)).

Further discussion of the difference between Granger causality and the SSA-based techniques

One of the main drawbacks of the Granger causality is that we need to assume that the model is fixed (we then just test for significance of some parameters in the model); model can be (and usually is) wrong. The test statistics used for testing the Granger causality are not comprehensive. In the certain case of the linear model, testing for Granger causality consists in the repeated use of the standard F-test which is sensitive to various deviations from the model, and the Granger causality is only associated with the lag difference between the two series.

In our approach, the model of dependence (or causality) is not fixed a priori; instead, this is built into the process of analysis. The models we build are non-parametric and are very broad (in particular, causality is not necessarily associated with a lag) and flexible.

The tests for Granger causality consider the past information of other series in forecasting the series. For example, in the linear Granger causality test, we use the series X up to time t and the series Y up to time $t-d$; and the series Y_{T-d} is used in forecasting series X_T. Whereas in the proposed test here, the series Y_{T+d} is employed in forecasting series X_T.

Furthermore, the tests for Granger causality are based on the forecasting accuracy. In this paper, we have also introduced another criterion for capturing causality which is based on the predictability of the direction of change. As we mentioned above, it may be more harmful to make a smaller prediction error yet fail in predicting the direction of change, than to make a larger directionally correct error (Ash et al. (1997)).

The definition of Granger causality does not mention anything about possible instantaneous correlation between two series X_T and Y_T, where the criteria introduced enable an interpretation of an instantaneous causality. In fact, the proposed test is not restricted to the lagged difference between two series. It works equally well when there is no lagged difference between series.

Furthermore, real world time series are typically noisy (e.g., financial time series), non-stationary, and can have small length. It is well known that the existence of a significant noise level reduces the efficiency of the tests (linear and nonlinear) for capturing the amount of dependence between two financial series (see, for example, Hassani et al. (2009c)).

There are mainly two different approaches to examine causality between two time series. According to the first one, that is utilized in current methods, the criteria of capturing causality is computed directly from the noisy time series. Therefore, we ignore the existence of the noise, which can lead to misleading interpretations of causal effects. In our approach, the noisy time series is filtered in order to reduce the noise level and then we calculate the criteria. It is commonly accepted that the second approach is more effective than the first one if we are dealing with the series with high noise level (Soofi and Cao (2002)).

4 Empirical results

4.1 Exchange rate

Given the high correlation between the UK (pound/dollar) and EU (euro/dollar) exchange rates, Hassani et al. (2009d) used a 2-variable vector autoregressive (VAR) model and SSA (univariate and multivariate model) in exchange rate predicting. This approach to prediction is called a-theoretical, since there is no theoretical justifications in asserting that one exchange rate is a predictor of another one. They showed that VAR model is not a good choice in predicting exchange rate series, while SSA (specifically multivariate version) decisively outperforms the VAR model. They also found that the exchange rate series has a unit root, which implies the series is non-stationary.

Moreover, using Johansen maximum-likelihood method, they also found that the exchange rates are cointegrated, and the Granger causality test showed that the UK/dollar rate does Granger cause the EU/dollar exchange rate series and vice versa.

Next we consider testing for causality between the two exchange rate series using the criteria we have introduced in previous section. First, we consider univariate SSA to forecast one step ahead of the UK and EU exchange rate series, and then compare the MSSA and SSA forecasting results to find $F^{(1,0)}_{UK|EU}$ and $F^{(1,1)}_{UK|EU}$. In this particular example, examining $F^{(h,0)}_{UK|EU}$ also shows whether exchange rate series is martingale or not.

To find the vector of forecasting errors, we forecast all observations of the UK and EU series from 1-May-2009 to 26-June-2009. Fig. 1 shows these series over the period 3-Jan-2000 to 26-Jun-2009, in these prediction exercises. Each of these series contain 2452 points. It is very clear that the UK and EU series are highly correlated (indeed, the nonlinear correlation coefficient between UK and EU series is about 0.75). It must be mentioned that this correlation only shows the relationship between the main trends of the series. One source of the relation between the UK and EU exchange rate series is obvious as the two series are each a ratio of US series.

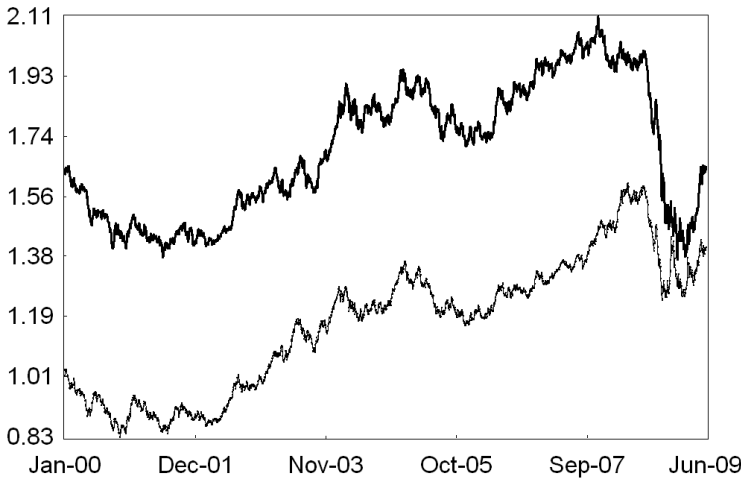

Figure 1: The exchange rate series UK (thick line) and EU (thin line) exchange rate series over the period 3-Jan-2000 to 26-Jun-2009.

We perform one-step ahead forecasting based on the most up-to-date information available at the time of the forecast. Note that we first use SSA in prediction of a single series, e.g. in prediction of the UK series without using euro series. Next, we use both series simultaneously, e.g. we use the EU time series in forecasting the UK series and vice versa. We shall refer to the results of this step $F^{(1,0)}_{UK|EU}$ and $F^{(1,0)}_{EU|UK}$. We also use one-step ahead information of EU time series as additional information in forecasting UK series and vice versa. We shall call this version of results $F^{(1,1)}_{UK|EU}$ and $F^{(1,1)}_{EU|UK}$. Note that we select window length 3 for both single and multivariate SSA in forecasting exchange rate series. The symbol $*$ indicates the significant results on the 1% level.

$F^{(1,0)}_{UK\|EU}$	$D^{(1,0)}_{UK\|EU}$	$F^{(1,1)}_{UK\|EU}$	$D^{(1,1)}_{UK\|EU}$	$F^{(1,0)}_{EU\|UK}$	$D^{(1,0)}_{EU\|UK}$	$F^{(1,1)}_{EU\|UK}$	$D^{(1,1)}_{EU\|UK}$
0.94	0.92	0.63*	0.88	0.81*	0.92	0.45*	0.84

Table 2: The value of $F_{UK|EU}$, $D_{UK|EU}$, $F_{EU|UK}$ and $D_{EU|UK}$ in forecasting 1 step ahead of the UK and EU exchange rate series for d =0 and 1.

It can be observed from Table 2 that the difference between the MSSA predictions and SSA is significant with respect to $F^{(h,d)}_{UK|EU}$ and $D^{(h,d)}_{UK|EU}$. The results confirm with that we have improved both accuracy and direction of change of the forecasting results. For example, in forecasting one step ahead for the EU series and $d = 0$, compared to the univariate case, we have improved the accuracy and the direction of change of the forecasting results up to 19% and 8% (column 3 of Table 2), respectively. Similarly for the UK exchange rate series with zero lagged difference, MSSA enable an improvement in forecasting accuracy and prediction of the direction of change up to 6% and 8% (with respect to $F^{(1,0)}_{UK|EU}$ and $D^{(1,0)}_{UK|EU}$), respectively. Thus, using the information of the UK and EU exchange rate (with zero lagged difference) enable an improvement the results.

The results obtained so far can be considered as zero-lag correlation between two exchange rate series or multivariate version of the SSA with zero lagged difference. These results can be considered as an evidence that there is the SSA causal relationship between the UK and EU exchange rate of order zero. It should be noted that the SSA causality of zero order confirms that there exists instantaneous causality. The SSA causality of zero order, instantaneous causality, suggests that there might be SSA causal relationship of higher order.

To examine this, next we consider MSSA with one more additional observation for one series. For example, we use the UK exchange rate series up to time t, and the EU exchange rate series up to time $t + 1$ in forecasting one step ahead of the UK exchange rate series to obtain $F^{(1,1)}_{UK|EU}$. In fact, there is one lagged difference between two series in one step ahead forecasting. We use a similar procedure in forecasting the EU series. We expect this additional information gives better results in both forecasting accuracy and the direction of change prediction.

Tt can be observed from columns $F^{(1,1)}_{UK|EU}$, $D^{(1,1)}_{UK|EU}$, $F^{(1,1)}_{EU|UK}$ and $D^{(1,1)}_{EU|UK}$, thus the errors for the MSSA forecast and direction of change, with only one additional observation, are much smaller than those obtained using univariate version. These results are also better than the results obtained using the multivariate approach with zero lag difference. This is not surprising though as the additional data used for forecast is highly

correlated with the values we are forecasting. As the results show, the accuracy performance of MSSA has been significantly increased. This means using only one additional observation enable an improvement in forecasting accuracy up to 37% and 55% relative to the univariate version for the UK and EU series (according to $F^{(1,1)}_{UK|EU}$ and $F^{(1,1)}_{EU|UK}$), respectively. Similarly for the direction of change, using only one additional observation enable an improvement in predicting the direction of change up to 12% and 16% (with respect to $D^{(1,1)}_{UK|EU}$ and $D^{(1,1)}_{EU|UK}$).

These results imply that the exchange rate time series are not martingales with respect to all available information available at the currency exchange markets. In fact, the results confirm that the series are SSA causal of order 1. Moreover, $F^{(1,1)}_{EU|UK} > F^{(1,1)}_{UK|EU}$ indicates that, in forecasting this period of the series, the UK exchange rate series is more supportive than the EU series. Furthermore, $F^{(1,0)}_{EU|UK} > F^{(1,0)}_{UK|EU}$ is other evidence for this. This means for this particular example, the SSA casual of order zero then consequences SSA casual of order one as well. However, $D^{(0,1)}_{EU|UK} = D^{(0,1)}_{UK|EU}$ and the discrepancy between $D^{(1,1)}_{EU|UK}$ and $D^{(1,1)}_{UK|EU}$ is not substantially indicating that neither is more directive.

Finally, the results of Table 2 strongly confirm that there exists F-feedback and D-feedback between the UK and EU exchange rate series. This means, considering both the UK and EU exchange rate series simultaneously, with and without one additional observation, will improve both the accuracy of forecasting and predictability of the direction of change.

4.2 Index of Industrial Production Series

As the second example, we consider the index of industrial production (IIP) series. The IIP series is a key indicator of the state of the UK's industrial base and regarded as a leading indicator of the general state of the economy. The IIP series is published on a monthly basis by the Office for National Statistics (ONS). The index is first released as a provisional estimate and then revised each month to incorporate the information that was not available at the time of the preliminary release. A number of studies have been concerned with the size and nature of revisions to important economic time series. Patterson and Heravi (1991a, b, 1992) have extensively analyzed the key national income and expenditure time series. There are many other studies for modelling and forecasting of data revision. For example, Patterson (1995a, b) have used state space approach in forecasting the final vintage of the IIP series and real personal disposable income. For more information about the data revision see Patterson (1992, 1994, 1995c).

The overall data period for the study includes 423 monthly observations for 1972:1 to 2007:3 on 12 vintages of data seasonally adjusted IIP. The first vintage, which is published one month after the latest month of published data, refers to the first publication in the monthly Digest of statistics. The second vintage refers to the next published figure and so on. For this study we take the 12^{th} vintage as the final vintage (m), then having 12 vintages of data on the same variables.

Let y^v_t be the vth vintage ($v = 1, \cdots, m$) of the data on variable y for the period t, where $v = 1$ indicates the initially published data and $v = m$ the finally published data. (In practice, m may be taken to indicate the conditionally final vintage.) Here $m = 12$.

The structure of the data which is published by *Monthly Digest of Statistics* (MDS) is as follow:

$$\begin{pmatrix} y_1^1 & y_1^2 & y_1^3 & \cdots & y_1^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{t-m}^1 & y_{t-m}^2 & y_{t-m}^3 & \cdots & y_{t-m}^m \\ \vdots & \vdots & \vdots & \ddots & \\ y_{t-2}^1 & y_{t-2}^2 & y_{t-2}^3 & & \\ y_{t-1}^1 & y_{t-1}^2 & & & \\ y_t^1 & & & & \end{pmatrix}. \tag{17}$$

Thus, publication from a particular issue of MDS traces back a diagonal of this data matrix which is a composite of data of different vintages. We expect that there is a SSA causal relationship between preliminary vintage (v^{th} vintage) and final vintage (m^{th} vintage). To answer this, we need to forecast h step ahead ($h = 1, \ldots, 11$) of the final vintage, $v = m$, giving the information at time t. The forecast could be obtained using classical univariate time series methods. However, the forecasts are not optimal since other information (vintages) available at time t are not used. For example, in forecasting y_{t-m+1}^m we also have available information of y_{t-m+1}^v for $v = 1, \cdots, m-1$, each of which could itself be regarded as a forecast of y_{t-m+1}^m. This matter motivates us to use a multivariate method for forecasting h step ahead of y_t^m. For example, to obtain the final vintage value at time t, y_t^m, we can use the information for the first vintage data $y_1^1, \ldots, y_t^1$ and the final vintage data $y_1^m, \ldots, y_{t-m}^m$. If the results of h step ahead forecast MSSA are better than SSA, e.g. $F_{v^m|v^i}^{(h,m-i)} < 1$ and $D_{v^m|v^i}^{(h,m-i)} < 1$, we then conclude that there is a SSA causal relationship of order h between i^{th} vintage and final vintage. To asses this, SSA and MSSA models were estimated using data to the end of 2000 and post-sample forecasts are then computed for 64 observations of 2001:1-2006:3. Thus, we have 64 one step ahead post sample forecast errors, at horizon $h = 1$. The number of forecast errors available decreases as the forecast horizon increases, so that at horizons of $h = 2, 3, \ldots, 12$ the number of forecast errors are $63, 62, \ldots, 52$ respectively. The value of $F_{v^m|v^i}^{(h,m-i)}$ and $D_{v^m|v^i}^{(h,m-i)}$ $(i = 1, \ldots, 11)$ for each vintage and relative to single SSA are given in Table 3. The two parameters L (window length) and r (number of eigenvalues) chosen in the decomposition and reconstruction are also presented in the table.

From Table 3, observe that there are gains to using MSSA throughout the revision process, these being between 87% and 67% for vintage up to $v = 5$, reducing to 50% or slightly less for latter vintages (according to the column labeled $F_{v^m|v^i}^{(h,m-i)}$). This is because, as the structure of the data matrix (17) shows, even one observation is very important in forecasting a new vector of the data matrix (17). All results are statistically significant at the 1% significant level.

For the direction of change results, for each preliminary vintage v, we compare the true direction of $y_t^m - y_{t+v-12}^m$ with the direction of vintage v estimate $y_t^v - y_{t+v-12}^m$ and the SSA estimate $\hat{y}_t - y_{t+v-12}^m$. Table 3 provides the percentage of forecasts that correctly predict the direction of change for each vintage. As the results show the percentage of correct signs produced by MSSA are significantly higher than those given by SSA, these

being between 55% and 45% for vintage up to $v = 5$, reducing to 18% for latter vintages (according to the column labeled $D^{(h,m-i)}_{v^m|v^i}$).

Thus, these results, without exception, confirm that there exists a SSA causal relationship between each vintage and the final vintage. In fact the results strongly indicate that there is SSA causality between i^{th} vintage and final vintage is of order $m-i$. It should be noted that here i is equal to h step ahead forecast which is the time lag difference between i^{th} vintage and final vintage. Here, as the results show, SSA causality holds for lower lag order such as in the case of the exchange rate series. This confirms that SSA causality of order $m-i$ has consequences for other orders of causality. Note that here the problem of interest is one-side causality as the final vintage is forecasted.

The results of Granger causality tests, also showed that there is a Granger causal relationship between these series. This is not surprising as each column of the data matrix is a revised version of the previous column and therefore they are high correlated. Also, it should be noted that the results of VAR model in forecasting these series are worse than the MSSA results.

i^{th} Vintage	L	r	$F^{(h,m-i)}_{v^m\|v^i}$	$D^{(h,m-i)}_{v^m\|v^i}$
1	13	5	0.22*	0.45*
2	12	5	0.24*	0.47*
3	11	5	0.27*	0.48*
4	10	5	0.31*	0.50*
5	9	5	0.33*	0.55*
6	8	4	0.36*	0.61*
7	7	4	0.39*	0.65*
8	6	3	0.41*	0.70*
9	5	3	0.45*	0.73*
10	4	3	0.49*	0.77*
11	3	2	0.55*	0.82

Table 3: The value of $F^{(h,m-i)}_{v^m|v^i}$ and $D^{(h,m-i)}_{v^m|v^i}$ in forecasting of i^{th} vintage of the index of industrial production series.

5 Conclusion

In this paper, we developed a new approach in testing for causality between two arbitrary univariate time series. We introduced a family of causality tests which are based on the singular spectrum analysis (SSA) analysis. The SSA technique accommodates, in principle, arbitrary processes, including linear, nonlinear, stationary, non-stationary, Gaussian, and non-Gaussian. Accordingly, we believe our approach to be superior to the traditional criteria used in Granger causality tests, criteria that are based on autoregressive integrated moving average (p, d, q) or multivariate vector autoregressive (VAR) representation of the data; the models that impose restrictive assumptions on the time series under investigation.

Several test statistics and criteria are introduced in testing for casuality. The criteria are based on the idea of minimizing a loss function, forecasting accuracy and predictability

of the direction of change. We use the univariate SSA and multivariate SSA in forecasting the value of the series and also prediction of the direction.

The performance of the proposed test was examined using the euro/dollar and the pound/dollar daily exchange rates as well as the index of industrial production (IIP) series for the United Kingdom. It has been shown here that the euro/dollar rate causes the pound/dollar rate and vice versa. Moreover, it has been documented that, without exception, there exists a SSA causal relationship between each vintage and final vintage of the IIP data.

References

[1] Ash, J. C. K, Smyth, D. J., and Heravi, S. (1997). The accuracy of OECD forecasts for Japan. *Pacific Economic Review*, **2** (1), pp. 25–44.

[2] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*, (2nd ed), Springer.

[3] Broomhead, D. S. and King, G. P. (1986). Extracting Qualitative Dynamics from Experimental Data, *Physica D*, **20**, pp. 217–236.

[4] Chu, T., Danks, D. and Glymour, C. (2004). *Data Driven Methods for Nonlinear Granger Causality: Climate Teleconnection Mechanisms*, 2004.

[5] Clements, M. P., and Smith, J. (1999). A Monte Carlo investigation of forecasting performance of emprical SETAR model. *Journal of Applied Econometrics*, **14**, 123–141.

[6] Danilov, D. and Zhigljavsky, A (Eds.). (1997). *Principal Components of Time Series: the 'Caterpillar' method*, University of St. Petersburg, St. Petersburg. (In Russian.).

[7] Diebold F. X., and Mariano R. S. (1995). Comparing predictive accuracy, *Journal of Business and Economic Statistics*, **13**(3), pp. 253-63.

[8] Diks, C. and DeGoede, J. (2001). A general nonparametric bootstrap test for Granger causality. In: HW broer, B Krauskopf, G Vegter (Eds.), Global Analysis of Dynamical Systems (Institute of Physics Publishing, Bristol, UK), pp. 391-403.

[9] Diks, C. and Panchenko, V. (2005). A note on the Hiemstra-Jones test for Granger non-causality. *Studies in Nonlinear Dynamics and Econometrics*, **9** , art. 4.

[10] Diks, C. and Panchenko, V. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing, *Journal of Economic Dynamics and Control*, **30**, pp. 1647-1669.

[11] Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions* (2nd ed.). New York: Wiley.

[12] Golyandina, N., Nekrutkin, V., and Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and related techniques*, Chapman & Hall/CRC, New York - London.

[13] Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, **37** (3), pp. 424-438.

[14] Granger, C. W. J. (1980). Testing for causality: A personal viewpoint, *Journal of Economic Dynamics and Control*, **2**, pp. 329–352.

[15] Harvey, D. I., Leybourne, S. J. and Newbold, P. (1997). Testing the Equality of Prediction Mean Squared Errors, *International Journal of Forecasting*, **13**, 281-291.

[16] Hassani H, and Zhigljavsky A. (2009a). Singular Spectrum Analysis: Methodology and Application to Economics Data, *journal of system science and complexity*, 22(3), pp. 372–394.

[17] Hassani H, Heravi S, and Zhigljavsky A. (2009b). Forecasting European Industrial Production with Singular Spectrum Analysis, *International Journal of Forecasting*, 25(1), pp. 103–118.

[18] Hassani H, Dionisio A, and Ghodsi M. (2009c). The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets, *Nonlinear Analysis: Real World Applications*, doi:10.1016/j.nonrwa.2009.01.004

[19] Hassani, H; Soofi, A; and Zhigljavsky, A. (2009d). Predicting Daily Exchange Rate with Singular Spectrum Analysis. *Nonlinear Analysis: Real World Applications*, doi:10.1016/j.nonrwa.2009.05.008

[20] Hassani, H. (2009e). Singular Spectrum Analysis based on the Minimum Variance Estimator, *Nonlinear Analysis: Real World Applications*, doi:10.1016/j.nonrwa.2009.05.009

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

[22] Heravi, S., Osborn, D. R., and Birchenhall, C. R. (2004). Linear Versus Neural Network Forecasts for European Industrial Production Series. *International Journal of Forecasting*, *20*, 435–446.

[23] Hiemstra, C. and Jones, J. D. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *Journal of Finance*, **49**(5), pp. 1639-1664.

[24] Su, L., and White, H. (2008). A Nonparametric Hellinger Metric Test For Conditional Independence,*Econometric Theory*, 24(4), pp. 829-864

[25] Patterson, K. D. (1992). Revisions to the components of the trade balance for the United Kingdom, Oxford Bulletin of Economics and Statistics, **54**, pp. 103–120.

[26] Patterson, K. D. (1994). Consumers' expenditure on nondurables and services and housing equity withdrawal in the United Kingdom, Manchester School of Economic and Social Studies, **62**. pp. 251–274.

[27] Patterson, K. D. (1995a). An integrated model of the data measurement and data generation processes with an application to consumers' expenditure, *Economic Journal*, **105**, pp. 54–76.

[28] Patterson, K. D. (1995b). A State Space Approach to Forecasting the Final Vinatege of Revised Data with an Application to the Index of Industrial Production, *Journal of Forecasting*, **14**, pp. 337–350.

[29] Patterson, K. D. (1995c). Forecasting the final vintage of real personal disposable income: A state space approach, *International Journal of Forecating*, **11**, pp. 395–405.

[30] Patterson, K. D. and Heravi, , S. M. (1991a). Data revisions and the expenditure components of GDP, *Economic Journal*, **101**, pp. 887–901.

[31] Patterson, K. D. and Heravi, S. M. (1991b). Are different vintages of data on the components of GDP cointegrated?, *Economics Letters*, **35**, pp. 409–413.

[32] Patterson, K. D. and Heravi, S. M. (1992). Efficient forecasts or measurement errors? Some evidence for revisions to the United Kingdom GDP growth rates, Manchester School of Economic and Social Studies, **60**, pp. 249–263.

[33] Plaut, G. R. and Vautard, R.(1994). Spells of oscillations and weather regimes in the low-frequency dynamics of the Northern Hemisphere. *J. Atmos. Sci.*, 51, 210-236. Su, L. and White, H (2008). Nonparametric Hellinger metric test for conditional independence, *Econometric Theory*, **24**, pp. 829-864.

[34] Soofi, A., and Cao, L. Y. (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.

[35] Toda, H. Y. and Phillips, P. C. B. (1991). Vector Autoregressions and Causality: A Theoretical Overview and Simulation Study, *Econometric Reviews*, **13**, pp. 259–285.

Appendix A: Formal Description of SSA

Stage 1: Decomposition

1st step: Embedding

Embedding can be regarded as a mapping that transfers a one-dimensional time series $Y_T = (y_1, \ldots, y_T)$ into the multi-dimensional series $X_1, \ldots, X_K$ with vectors $X_i = (y_i, \ldots, y_{i+L-1})' \in \mathbf{R}^L$, where $K = T - L + 1$. Vectors X_i are called *L-lagged vectors* (or, simply, *lagged vectors*). The single parameter of the embedding is the *window length* L, an integer such that $2 \leq L \leq T$. The result of this step is the trajectory matrix $\mathbf{X} = [X_1, \ldots, X_K]$:

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_T \end{pmatrix}. \tag{18}$$

Note that the trajectory matrix $\mathbf{X}$ is a Hankel matrix, which means that all the elements along the diagonal $i + j = const$ are equal. Embedding is a standard procedure in time series analysis. With the embedding performed, future analysis depends on the aim of the investigation.

2nd step: Singular Value Decomposition (SVD)

The second step, the SVD step, makes the singular value decomposition of the trajectory matrix and represents it as a sum of rank-one bi-orthogonal elementary matrices. Denote by $\lambda_1, \ldots, \lambda_L$ the eigenvalues of $\mathbf{X}\mathbf{X}^{'}$ in decreasing order of magnitude ($\lambda_1 \geq \ldots \lambda_L \geq 0$) and by $U_1, \ldots, U_L$ the orthonormal system (that is, (U_i, U_j)=0 for $i \neq j$ (the orthogonality property) and $\|U_i\|$=1 (the unit norm property)) of the eigenvectors of the matrix $\mathbf{X}\mathbf{X}^{'}$ corresponding to these eigenvalues. (U_i, U_j) is the inner product of the vectors U_i and U_j and $\|U_i\|$ is the norm of the vector U_i. Set

$$R = \max(i, \text{ such that } \lambda_i > 0) = rank\, \mathbf{X}.$$

If we denote $V_i = \mathbf{X}^{'} U_i / \sqrt{\lambda_i}$, then the SVD of the trajectory matrix can be written as:

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_R, \tag{19}$$

where $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^{'}$ $(i = 1, \ldots, R)$. The matrices $\mathbf{X}_i$ have rank 1; therefore they are elementary matrices, U_i (in SSA literature they are called 'factor empirical orthogonal functions' or simply EOFs) and V_i (often called 'principal components') stand for the left and right eigenvectors of the trajectory matrix. The collection $(\sqrt{\lambda_i}, U_i, V_i)$ is called the i-th eigentriple of the matrix $\mathbf{X}$, $\sqrt{\lambda_i}$ $(i = 1, \ldots, R)$ are the singular values of the matrix $\mathbf{X}$ and the set $\{\sqrt{\lambda_i}\}$ is called the spectrum of the matrix $\mathbf{X}$. If all the eigenvalues have multiplicity one, then the expansion (19) is uniquely defined.

SVD (19) is optimal in the sense that among all the matrices $\mathbf{X}^{(r)}$ of rank $r < R$, the matrix $\sum_{i=1}^{r} X_i$ provides the best approximation to the trajectory matrix $\mathbf{X}$, so that $\| \mathbf{X} - \mathbf{X}^{(r)} \|$ is minimum. Note that $\| \mathbf{X} \|^2 = \sum_{i=1}^{R} \lambda_i$ and $\| \mathbf{X}_i \|^2 = \lambda_i$ for $i = 1, \ldots, d$. Thus, we can consider the ratio $\lambda_i / \sum_{i=1}^{R} \lambda_i$ as the characteristic of the contribution of the matrix $\mathbf{X}_i$ to expansion (19). Consequently, $\sum_{i=1}^{r} \lambda_i / \sum_{i=1}^{R} \lambda_i$, the sum of the first r ratios, is the characteristic of the optimal approximation of the trajectory matrix by the matrices of rank r.

Stage 2: Reconstruction

1st step: Grouping

The grouping step corresponds to splitting the elementary matrices $\mathbf{X}_i$ into several groups and summing the matrices within each group. Let $I = \{i_1, \ldots, i_p\}$ be a group of indices $i_1, \ldots, i_p$. Then the matrix $\mathbf{X}_I$ corresponding to the group I is defined as $\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_p}$. The spilt of the set of indices $J = 1, \ldots, R$ into the disjoint subsets $I_1, \ldots, I_m$ corresponds to the representation

$$\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}. \tag{20}$$

The procedure of choosing the sets $I_1, \ldots, I_m$ is called the eigentriple grouping. For given group I the contribution of the component $\mathbf{X}_I$ into the expansion (1) is measured by the share of the corresponding eigenvalues: $\sum_{i \in I} \lambda_i / \sum_{i=1}^{R} \lambda_i$.

2nd step: Diagonal Averaging

Diagonal averaging transfers each matrix I into a time series, which is an additive component of the initial series Y_T. If z_{ij} stands for an element of a matrix $\mathbf{Z}$, then the k-th term of the resulting series is obtained by averaging z_{ij} over all i, j such that $i+j = k+2$. This procedure is called *diagonal averaging*, or Hankelization of the matrix $\mathbf{Z}$. 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 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 $\widetilde{z}_{ij}$ of the matrix $\mathcal{H}\mathbf{Z}$ is

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

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 matrix norm) among all Hankel matrices of the corresponding size (for more information see Golyandina et al. (2001, chap. 6, sec. 2)). In its turn, the Hankel matrix $\mathcal{H}\mathbf{Z}$ uniquely defines the series by relating the value in the diagonals to the values in the series. By applying the Hankelization procedure to all matrix components of (20), we obtain another expansion:

$$\mathbf{X} = \widetilde{\mathbf{X}}_{I_1} + \ldots + \widetilde{\mathbf{X}}_{I_m} \tag{21}$$

where $\widetilde{\mathbf{X}}_{I_1} = \mathcal{H}\mathbf{X}$. This is equivalent to the decomposition of the initial series $Y_T = (y_1, \ldots, y_T)$ into a sum of m series:

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

where $\widetilde{Y}_T^{(k)} = (\widetilde{y}_1^{(k)}, \ldots, \widetilde{y}_T^{(k)})$ corresponds to the matrix $\mathbf{X}_{I_k}$.

Selection of parameters

Here we consider a version of SSA where we split the set if indicies $\{1, 2, \ldots, L\}$ into two groups only: $I = \{1, \ldots, r\}$ and $\bar{I} = \{r+1, \ldots, L\}$. We associate the group I with signal and the group $\bar{I}$ with noise. The SSA method requires then the selection of two parameters, the window length L and the number of elementary matrices r. There are specific rules for selecting these parameters; their choice depends on structure of the

data and the analysis we want to perform. Detailed description of parameter selection procedures is given in Golyandina et al. (2001). Here we summarize a few general rules.

The window length L is the single parameter that should be selected at the decomposition stage. Selection of the proper window length depends on the problem in hand, and on preliminary information about the time series. For the series with a complex structure, too large window length L can produce an undesirable decomposition of the series components of interest, which may lead, in particular, to their mixing with other series component. Let us, for example, consider the problem of trend extraction in GCM. Since trend is a relatively smooth curve, its separability from noise requires small values of L. It should be noted that the values of L should not be smaller than the true eigenvalues r. The chosen L also should results good separability between the reconstructed series using $I = \{1, \ldots, r\}$ and $\bar{I} = \{r+1, \ldots, L\}$. In growth curve model that we are dealing with only trend extraction, usually the first or second eigenvalue is considered for reconstruction step.

The first elementary matrix $\mathbf{X}_1$ with the norm $\sqrt{\lambda_1}$ has the highest contribution to the norm of $\mathbf{X}$ in $\mathbf{X} = \mathbf{X}_1 + \ldots, \mathbf{X}_L$ and the last elementary matrix $\mathbf{X}_L$ with the norm $\sqrt{\lambda_L}$ has the lowest contribution to the norm of $\mathbf{X}$. The plot of the eigenvalues $\lambda_1, \cdots, \lambda_L$ gives an overall view concerning the values of the eigenvalues and is essential in deciding where to truncate the summation of $\mathbf{X} = \mathbf{X}_1 + \ldots, \mathbf{X}_L$ in order to build a good approximation of the original matrix. A slowly decreasing sequence of eigenvalues typically indicate the presence of noise in the series.

A group of r (with $1 \leq r < L$) eigenvectors determine an r-dimensional hyperplane in the L-dimensional space $\mathbb{R}^L$ of vectors X_j. The distance between vectors X_j ($j = 1, \ldots, K$) and this r-dimensional hyperplane can be rather small (it is controlled by the choice of the eigenvalues) meaning that the projection of $\mathbf{X}$ into this hyperplane is a good approximation of the original matrix $\mathbf{X}$. If we choose the first r eigenvectors $U_1, \ldots, U_r$, then the squared L_2-distance between this projection and $\mathbf{X}$ is equal to $\sum_{j=r+1}^{L} \lambda_j$. According to the Basic SSA algorithm, the L-dimensional data is projected onto this r-dimensional subspace and the subsequent averaging over the diagonals allows us to obtain an approximation to the original series.

Forecasting Algorithm

Let us formally describe the forecasting algorithm under consideration (for more information see Golyandina et al.(2001)):

Algorithm input:

(a) Time series $Y_T = (y_1, \ldots, y_T)$.

(b) Window length L, $1 < L < T$.

(c) Linear space $\mathfrak{L}_r \subset \mathbf{R}^L$ of dimension $r < L$. It is assumed that $e_L \notin \mathfrak{L}_r$, where $e_L = (0, 0, \ldots, 1) \in \mathbf{R}^L$.

(d) Number M of points to forecast for.

Procedure:

(a) $\mathbf{X} = [X_1, \ldots, X_K]$ is the trajectory matrix of the time series Y_T.

(b) $U_1, \ldots, U_r$ is an orthonormal basis in $\mathfrak{L}_r$.

(c) $\widehat{\mathbf{X}} = [\widehat{X}_1 : \ldots : \widehat{X}_K] = \sum_{i=1}^{r} U_i U_i^T \mathbf{X}$. The vector $\widehat{X}_i$ is the orthogonal projection of X_i onto the space $\mathfrak{L}_r$.

(d) $\widetilde{\mathbf{X}} = \mathcal{H}\mathbf{X} = [\widetilde{X}_1 : \ldots : \widetilde{X}_K]$ is the result of the Hankellization of the matrix $\widehat{\mathbf{X}}$.

(e) For any vector $Y \in \mathbf{R}^L$ we denote by $Y_\triangle \in \mathbf{R}^{L-1}$ the vector consisting of the last $L-1$ components of the vector Y, while $Y^\triangledown \in \mathbf{R}^{L-1}$ is the vector of the first $L-1$ components of the vector Y.

(f) We set $v^2 = \pi_1^2 + \ldots + \pi_r^2$, where π_i is the last component of the vector U_i ($i = 1, \ldots, r$).

(g) Suppose that $e_L \notin \mathfrak{L}_r$. (In the other words, we assume that $\mathfrak{L}_r$ is not a vertical space). Then $v^2 < 1$. It can be proved that the last component y_L of any vector $Y = (y_1, \ldots, y_L)^T \in \mathfrak{L}_r$ is a linear combination of the first $L-1$ components $(y_1, \ldots, y_{L-1})$:

$$y_L = \alpha_1 y_{L-1} + \ldots + \alpha_{L-1} y_1.$$

Vector $A = (\alpha_1, \ldots, \alpha_{L-1})$ can be expressed as

$$A = \frac{1}{1 - v^2} \sum_{i=1}^{r} \pi_i U_i^\triangledown \tag{23}$$

and does not depend on the choice of a basis $U_1, \ldots, U_r$ in the linear space $\mathfrak{L}_r$. In the above notations, define the time series $Y_{T+M} = (y_1, \ldots, y_{T+M})$ by the formula

$$y_i = \begin{cases} \widetilde{y}_i & \text{for } i = 1, \ldots, T \\ \sum_{j=1}^{L-1} \alpha_j y_{i-j} & \text{for } i = T+1, \ldots, T+M \end{cases} \tag{24}$$

The numbers $y_{T+1}, \ldots, y_{T+M}$ from the M terms of the SSA recurrent forecast. Let us define the linear operator $\mathcal{R}^{(r)} : \mathfrak{L}_r \mapsto \mathbf{R}^L$ by the formula

$$\mathcal{R}^{(r)} Y = \begin{pmatrix} Y_\triangle \\ A^T Y_\triangle \end{pmatrix}, \quad Y \in \mathfrak{L}_r$$

If setting

$$Z_i = \begin{cases} \widetilde{X}_i & \text{for } i = 1, \ldots, K \\ \mathcal{R}^{(r)} Z_{i-1} & \text{for } i = K+1, \ldots, K+M \end{cases} \tag{25}$$

the matrix $\mathbf{Z} = [Z_1, \ldots, Z_{K+M}]$ is the trajectory matrix of the series Y_{T+M}. Therefore, (25) can be regard as the vector form of (24).

The SSA recurrent forecasting algorithm can be modified in several ways. For example, we can base our forecast on the Toeplitz SSA or SSA with centering rather than on

the basic SSA. Perhaps the most important modification is the so-called SSA vector forecasting algorithm developed in Golyandina et al. (2001).

So far we considered SSA recurrent forecasting algorithm. In the following we consider SSA vector forecasting algorithm. The SSA vector forecasting algorithm has the same inputs and conditions as the SSA recurrent forecasting algorithm. The notation in (a)-(g) is kept. Let us introduce some more notations. Consider the matrix

$$\Pi = V^{\nabla}(V^{\nabla})^T + (1 - v^2)AA^T$$

where $V^{\nabla} = [U_1^{\nabla}, \ldots, U_r^{\nabla}]$. The matrix Π is the matrix of the linear operator that performs the orthogonal projection $\mathbf{R}^{L-1} \mapsto \mathfrak{L}_r^{\nabla}$, where $\mathfrak{L}_r^{\nabla} = \mathrm{span}(U_1^{\nabla}, \ldots, U_r^{\nabla})$. We define the linear operator $\mathcal{V}^{(v)} : \mathfrak{L}_r \mapsto \mathbf{R}^L$ by the formula

$$\mathcal{V}^{(v)}Y = \begin{pmatrix} \Pi Y_{\triangle} \\ A^T Y_{\triangle} \end{pmatrix}, \quad Y \in \mathfrak{L}_r$$

In the notation above we define the vectors Z_i as follow:

$$Z_i = \begin{cases} \widetilde{X}_i & \text{for } i = 1, \ldots, K \\ \mathcal{V}^{(v)} Z_{i-1} & \text{for } i = K+1, \ldots, K+M+L-1 \end{cases}$$

By constructing the matrix $\mathbf{Z} = [Z_1, \ldots, Z_{K+M+L-1}]$ and making its diagonal averaging we obtain a series $y_1, \ldots, y_{T+M+L-1}$. The numbers $y_{T+1}, \ldots, y_{T+M}$ form the M terms of the SSA vector forecast.