Forecasting European Industrial Production with Multivariate Singular Spectrum Analysis

Anatoly Zhigljavsky[a*], Hossein Hassani[a], and Saeed Heravi[b]

[a] *Centre for Optimisation and Its Applications, Cardiff School of Mathematics, CF24 4AG*

[b] *Cardiff Business School, Cardiff University, CF10 3EU, UK.*

Abstract

In recent years the Singular Spectrum Analysis (SSA) technique, used as a powerful technique in time series analysis, has been developed and applied to many practical problems. The aim of this research is to develop theoretical and methodological aspects of the multivariate SSA (MSSA) technique and to demonstrate that MSSA can also be considered as a powerful method of time series analysis and forecasting. We use the UK Industrial Production series to illustrate the main findings. The performance of the SSA technique is assessed by applying it to eight series measuring the monthly seasonally unadjusted industrial production for the main sectors of the UK economy. The results are compared with those obtained using the ARIMA and VAR models.

We also develop the concept of casual relationship between two time series based on the SSA techniques. 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 the same data, the UK industrial production series.

Keywords: Singular Spectrum Analysis, Forecasting, Causality, The UK industrial production series.

1 Introduction

Econometric methods have been widely used to forecast the evolution of quarterly and yearly national account data. However, many of the structural or time series forecasting models have failed to accurately predict economic time series. This is due to technological advances, change in government policies and also change in consumer preferences. These shocks cause structural changes in these time series making them nonstationary. Development of a methodology which is robust under these changes is of paramount importance in accurate prediction of macroeconomic time series.

There are several reasons that classical model does not have a good performance for modelling and forecasting economic and financial time series. First, an economic model that has been established to have validity in explaining a relationship under one set of assumptions is useless if the assumptions are not valid. Model assumptions include not

*E-mail addresses: Zhigljavskyaa@cf.ac.uk, hassanih@cf.ac.uk, Heravis@cf.ac.uk

only those that can be expressed as predicates on model parameters but others with more qualitative or asymptotic form (for more information see [1]).

Moreover, many structural econometric and time series models devised for forecasting macroeconomic time series are based on restrictive assumptions of normality and linearity of the observed data. The methods that do not depend on these assumptions could be very useful for modelling and forecasting economics data. On the other hand classical methods of forecasting such as ARIMA type models are based on the assumption such as stationarity of the series and normality of residuals (see, for example, [2], [3] and references therein) .

Additionally, some of the previous research have considered economic and financial time series as deterministic, linear dynamical systems. In this case, the linear models can be used for modelling and forecasting. However, many financial time series exhibit nonlinear behavior (see, for example, [4, 5, 6, 7]); and therefore, we should use nonlinear methods. In addition a method that works well for both linear and nonlinear, stationary and non stationary time series is ideal for modelling and forecasting the real time series data. The Singular Spectrum Analysis (SSA) in a technique that is free from all these assumptions. The SSA technique is a nonparametric technique of time series analysis incorporating the elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing[8]. Note also that SSA naturally incorporates the filtering of the series and the SVD.

The basic SSA method consists of two complementary stages: decomposition and reconstruction; both stages 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 for forecasting new data points. The main concept in studying the properties of SSA is 'separability', which characterizes how well different components can be separated from each other. The absence of approximate separability is often observed in series with complex structure. For these series and series with special structure, there are different ways of modifying SSA leading to different versions such as SSA with single and double centering, Toeplitz SSA, and sequential SSA [8].

It is worth noting that although some probabilistic and statistical concepts are employed in the SSA-based methods, we do not have to make any statistical assumptions such as stationarity of the series or normality of the residuals. SSA is a very useful tool which can be used for solving the following problems:

finding trends of different resolution;
smoothing;
extraction of seasonality components;
simultaneous extraction of cycles with small and large periods;
extraction of periodicities with varying amplitudes;
simultaneous extraction of complex trends and periodicities;
finding structure in short time series.

Solving all these problems correspond to the so-called basic capabilities of SSA. In addition, the method has several essential extensions. First, the multivariate version of the method permits the simultaneous expansion of several time series; see, for example [10]. Second, the SSA ideas lead to several forecasting procedures for time series; see [8, 10]. Also, the same ideas are used in [8] and [14] for change-point detection in time series. For comparison with classical methods, ARIMA, ARAR algorithm and Holt-

Winter, see [15] and [16], and for comparison between multivariate SSA and VAR model see [17]. For automatic methods of identification within the SSA framework see [18] and for recent work in 'Caterpillar'-SSA software as well as new developments see [19]. A family of causality tests based on the SSA technique has also been considered in [20].

In the area of nonlinear time series analysis SSA was considered as a technique that could compete with more standard methods. There are a number of research that considered SSA as a filtering method in (see, for example, [21] and references therein). In another research, the noise information extracted using the SSA technique, has been used as a biomedical diagnostic test [22]. The SSA technique also used as a filtering method for longitudinal measurements. It has been shown that noise reduction is important for curve fitting in growth curve models, and that SSA can be employed as a powerful tool for noise reduction for longitudinal measurements [23].

The monthly industrial production indices for the UK, have been previously analysed in linear and nonlinear contexts in [16], [24] and [25]. The eight series examined for the UK, are interesting and important since they ranging from traditional industrial sections (Basic metals) to Food and Electricity and Gas. These eight time series contribute for at least 50% to the aggregate industrial production in the UK economics.

Osborn et al. [24] have considered the extent and nature of seasonality in these series. Seasonality accounts for at least 80% of variation in all series (except vehicles) in the UK. In our recent research [16], we used Singular Spectrum Analysis (SSA), ARIMA and Holt-Winter methods for forecasting seasonally unadjusted monthly data on industrial production indicators in Germany, France and the UK. We have demonstrated that SSA is a very powerful tool for analyzing and predicting these series. The SSA technique decomposes the original time series into a sum of small number of independent and interpretable components such as slowly varying trend, oscillatory components and noise. Theoretical and practical foundations of the SSA technique can be found in [8].

Hassani et al. [16] showed that the quality of 1-step ahead forecasts are similar for ARIMA and SSA; Holt-Winter forecasts being slightly worse. The quality of SSA forecasts at horizons $h = 3, 6$ and 12 is much better than the quality of ARIMA and Holt-Winter forecasts. As h increases, the quality of ARIMA and Holt-Winter forecasts becomes worse. Also the standard deviation of the ARIMA and Holt-Winter forecasts increases almost linearly with h. The situation is totally different for the SSA forecasts: the quality of SSA forecasts is almost independent of the value of h (for the values of h considered in our research).

Another important aspect of the SSA (which can be very useful in economics) is that, unlike many other methods, it works well even for small sample size (see, for example, [12] and [15]). We found that SSA works well for small sample sizes, as for the UK with the sample size of 84 observations [16].

Hassani et al. [16] also showed that three methods perform similarly well in predicting the direction of change for small h. However, SSA outperforms the Holt-Winter and ARIMA models at longer horizons and hence can be considered as a reliable method for predicting recessions and expansions.

The present project aims to predict the monthly industrial production indices for the UK using multivariate singular spectrum analysis (MSSA). Here we develop the methodology of forecasting these series based on MSSA. Preliminary results indicate that the SSA can further improve the results of univariate SSA for the these series. The quality of MSSA forecasts can be higher when the analysed series are interdependent and therefore

highly correlated. We also use MSSA to investigate the causality among these series. As these series are non-stationary and non-linear, we use nonlinear correlation which is based on the Mutual information introduced in [26] and [27], and has also been used in [21].

We are motivated to use SSA because of its capability in dealing with non-stationary series. Given that the dynamics of some industrial production series has gone through structural changes during the time period under consideration, one needs to make sure that the method of prediction is not sensitive to the dynamical changes. Moreover, contrary to the traditional methods of time series forecasting (both autoregressive or structural models that assume normality and stationarity of the series), SSA method is non-parametric and makes no prior assumptions about the data. The data considered in this study has a complex structure of this kind; as a consequence, we found superiority of SSA over classical techniques. Additionally, SSA method decomposes a series into its component parts, and reconstruct the series without including the random (noise) component.

The structure of this report is as follows. A brief introduction of the SSA method is represented in Section 2. The descriptive statistics of the series and the results of various tests (such as normality, nonlinearity, stationarity) are presented in Section 3. The performance of ARIMA, SSA, MSSA and VAR is considered in Section 4. A new casuality test based on the SSA technique is introduced in Section 5. Finally, Section 6 presents a summary of the study and some concluding remarks.

2 Singular Spectrum Analysis

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 of the original series. The Basic SSA technique is performed in two stages, both of which include two separate steps as follows:

$$\begin{cases} \text{Stage 1 : Decomposition} & \begin{cases} \text{Step 1 : Embedding} \\ \text{Step 2 : Singular Value Decomposition (SVD)} \end{cases} \\ \text{Stage 2 : Reconstruction} & \begin{cases} \text{Step 1 : Grouping} \\ \text{Step 2 : Diagonal Averaging} \end{cases} \end{cases}$$

A short description of the SSA technique is given as follows (for more information see [8]).

2.1 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 multidimensional series $X_1, \ldots, X_K$ with vectors $X_i = (y_i, \ldots, y_{i+L-1})^T \in \mathbf{R}^L$, 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 window length L should be sufficiently large. The result of this step is the trajectory matrix $\mathbf{X} = [X_1, \ldots, X_K] = (x_{ij})_{i,j=1}^{L,K}$. 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.

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 $\mathbf{X}\mathbf{X}^T$ in decreasing order of magnitude ($\lambda_1 \geq \ldots \lambda_L \geq 0$) and by $U_1, \ldots, U_L$ the orthonormal system of the eigenvectors of the matrix $\mathbf{X}\mathbf{X}^T$ corresponding to these eigenvalues. Set

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

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

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d, \tag{1}$$

where $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$.

SVD (1) is optimal in the sense that among all the matrices $\mathbf{X}^{(r)}$ of rank $r < d$, the matrix $\sum_{i=1}^{r} \mathbf{X}_i$ provides the best approximation to the trajectory matrix $\mathbf{X}$, so that $\| \mathbf{X} - \mathbf{X}^{(r)} \|$ is minimum. Here the norm of a matrix $\mathbf{Y}$ is defined as $\sqrt{\langle \mathbf{Y}, \mathbf{Y} \rangle}$, where the scalar product of two matrices $\mathbf{Y} = (y_{ij})_{i,j=1}^{q,s}$ and $\mathbf{Z} = (z_{ij})_{i,j=1}^{q,s}$ is $\langle \mathbf{Y}, \mathbf{Z} \rangle = \sum_{i,j=1}^{q,s} y_{ij} z_{ij}$. Note that $\| \mathbf{X} \|^2 = \sum_{i=1}^{d} \lambda_i$ and $\| X_i \|^2 = \lambda_i$ for $i = 1, \ldots, d$. Thus, we can consider the ratio $\lambda_i / \sum_{i=1}^{d} \lambda_i$ as the characteristic of the contribution of the matrix $\mathbf{X}_i$ to expansion (1). Consequently, $\sum_{i=1}^{r} \lambda_i / \sum_{i=1}^{d} \lambda_i$, the sum of the first r ratios, is the characteristic of the optimal approximation of the trajectory matrix by the matrices of rank r.

2.2 Reconstruction

1st Step: Grouping

The grouping step corresponds to splitting the elementary matrices 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 $J = \{1, \ldots, d\}$ into disjoint subsets $I_1, \ldots, I_m$ corresponds to the representation

$$\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}. \tag{2}$$

The procedure of choosing the sets $I_1, \ldots, I_m$ is called the eigentriple grouping. For a given group I the contribution of the component $\mathbf{X}_I$ in the expansion (2) is measured by the share of the corresponding eigenvalues: $\sum_{i \in I} \lambda_i / \sum_{i=1}^{d} \lambda_i$.

2nd Step: Diagonal averaging

The purpose of diagonal averaging is to transform a matrix to the form of a Hankel matrix which can be subsequently converted to a time series. 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 $i + j = k + 1$. 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 $\mathcal{H}\mathbf{Z}$. 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 (see [8], Sect. 6.2). 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 (2), we obtain another expansion $\mathbf{X} = \widetilde{\mathbf{X}}_{I_1} + \ldots + \widetilde{\mathbf{X}}_{I_m}$, 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; $y_t = \sum_{k=1}^{m} \widetilde{y}_t^{(k)}$, where $\widetilde{Y}_T^{(k)} = (\widetilde{y}_1^{(k)}, \ldots, \widetilde{y}_T^{(k)})$ corresponds to the matrix $\mathbf{X}_{I_k}$. A sensible grouping leads to the decomposition (1) where the resultant matrices $\mathbf{X}_{I_k}$ are almost Hankel ones.

2.3 Forecasting Algorithm

Let us now describe the SSA forecasting algorithm (for more information see [8]):

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.

Notations and comments:
(a) $\mathbf{X} = [X_1, \ldots, X_K]$ is the trajectory matrix of the time series Y_T.
(b) $P_1, \ldots, P_r$ is an orthonormal basis in $\mathfrak{L}_r$.
(c) $\widehat{\mathbf{X}} = [\widehat{X}_1 : \ldots : \widehat{X}_K] = \sum_{i=1}^{r} P_i P_i^T \mathbf{X}$. 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 P_i ($i = 1, \ldots, r$).
(g) Suppose that $e_L \notin \mathfrak{L}_r$(This implies 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 components $(y_1, \ldots, y_{L-1})$:

$$y_L = a_1 y_{L-1} + \ldots + a_{L-1} y_1.$$

Vector $A = (a_1, \ldots, a_{L-1})$ can be expressed as

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

and dose not depend on the choice of a basis $P_1, \ldots, P_r$ 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

$$y_i = \begin{cases} \widetilde{y}_i & \text{for } i = 1, \ldots, T \\ \sum_{j=1}^{L-1} a_j y_{i-j} & \text{for } i = T+1, \ldots, T+M \end{cases} \tag{3}$$

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

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

If setting

$$Z_i = \begin{cases} \widetilde{X}_i & \text{for } i = 1, \ldots, K \\ \mathcal{P}^{(r)} Z_{i-1} & \text{for } i = K+1, \ldots, K+M \end{cases} \tag{4}$$

the matrix $\mathbf{Z} = [Z_1, \ldots, Z_{K+M}]$ is the trajectory matrix of the series Y_{T+M}. Therefore, (4) can be regard as the vector form of (3).

2.4 Multivariate singular spectrum analysis (MSSA)

The use of multivariate singular spectrum analysis (MSSA) for multivariate time series was proposed theoretically in the context of nonlinear dynamics in [9]. There are numerous examples of successful application of the multivariate SSA (see, for example, [1] and [10]). Multivariate (or multichannel) SSA is an extension of the standard SSA to the case of multivariate time series. We give a short description of MSSA method as follows.

Assume that we have an M-variate time series $y_j = \left(y_j^{(1)}, \ldots, y_j^{(M)}\right)$, where $j = 1, \ldots, T$ and let L be window length. Similar to univariate version, we can define the trajectory matrices $\mathbf{X}^{(i)}$ $(i=1, \ldots, M)$ of the one-dimensional time series $\{y_j^{(i)}\}$ $(i = 1, \ldots, M)$. The trajectory matrix $\mathbf{X}$ can then be defined as

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \vdots \\ \mathbf{X}^{(M)} \end{pmatrix}. \tag{5}$$

3 Descriptive Analysis

In this section, we consider descriptive statistics of the 8 series measuring the monthly seasonally unadjusted industrial production for important sectors of the UK economies.

3.1 The data

The data in this study are taken from Eurostat, the office for National statistics and represents eight major components of industrial production in the UK. The two digit category industrials examined in this research are seasonally unadjusted monthly indices for real output in Food Products, Chemicals, Basic Metals, Fabricated Metals, Machinery, Electrical Machinery, Vehicles and Electricity and Gas industries. The two-digit categories examined in this research are as foloows:

The same 8 series, ending in 1995, have been previously analysed in [24], [25] and ending in 2007 in [16]. As explained in these papers, these industries have been chosen primarily because of their importance and that are the basis of the contribution to total industrial production. Here we have updated the data and in all cases the sample period starts from January 1978 and ends in July 2009, giving a long series of 380 observations. In line with the usual convention for economic time series and for comparability, all time

Short name	Detail
Food product	Manufacture of food products and beverages
Chemicals	Manufacture of chemicals and chemical product
Basic metals	Manufacture of basic metals
Fabricated metal	Manufacture of fabricated metal products
Machinery	Manufacture of machinery and equipment N.E.C.
Electrical machinery	Manufacture of electrical machinery and apparatus N.E.C.
Vehicles	Manufacture of motor vehicles, trailers and semi-trailers
Electricity and gas	Electricity, gas and water supply

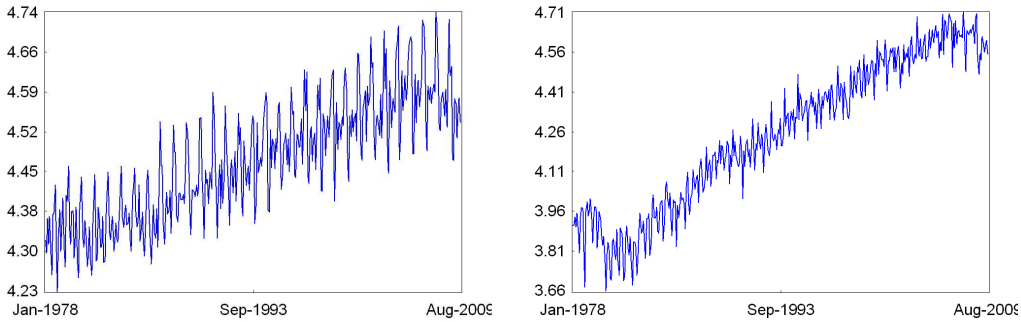
Table 1: Industrial production series.

series are analysed in the logarithmic form. Plots of these time series are presented below, their movements are dominated by seasonality. They broadly represent periods of growths in the 1980s and during 2000-2007s and periods of stagnation or recessions during the early 1990s and 2008-2009.

To investigate short term movements and the degree of seasonality, Table 2 shows the descriptive statistics for the first difference of each series. Sample means and sample standard deviations are scaled by 100 and since the data are logarithmically transformed, they effectively refer to monthly percentage changes in the original series. Table 2 shows that, with the exception of Food, Chemicals and Electricity and Gas, the production series have declined over the period. In particular, Chemicals shows substantial growth over the period with an average increase of about 0.2% per month (2.24% per year) and Basic Metals and Vehicles show an average decline of 0.17% per month or 2% per year. The sample standard deviations indicate that the series have broadly similar variability, with Food less volatile and Vehicles more volatile than the others. In the table, R^2 is a measure of seasonality and is calculated from a regression of the first differences against twelve monthly dummy variables. The highest seasonality for these series is for production of Electricity and Gas. These values which are obtained on longer time series are lower than the R^2 reported in [24].

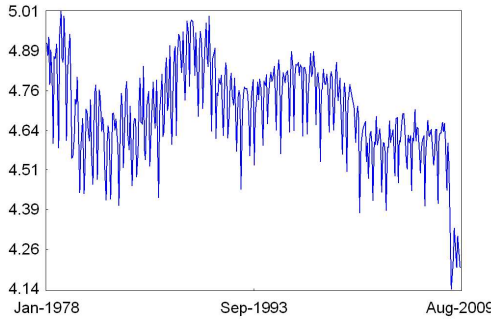
Series	Weight	Mean	S.D.	seas R^2
Food products	7.6	0.07	6.99	0.52
Chemicals	8.6	0.19	8.10	0.55
Basic metals	4.5	-0.17	11.98	0.74
Fabricated metal	7.2	-0.08	10.06	0.63
Machinery	13.6	-0.09	10.05	0.72
Electrical machinery	5.6	-0.06	10.46	0.68
Vehicles	10.4	-0.17	15.69	0.68
Electricity and Gas	6.5	0.01	10.36	0.79

Table 2: Descriptive statistics of the data.

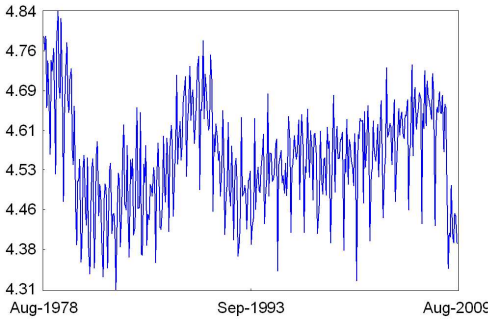


Food

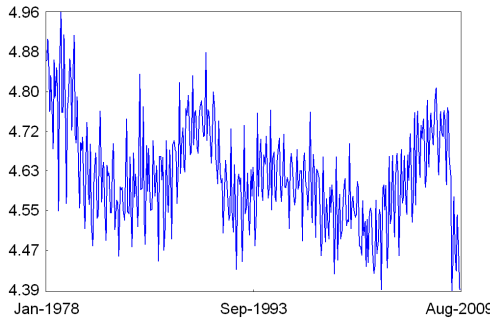
Chemical



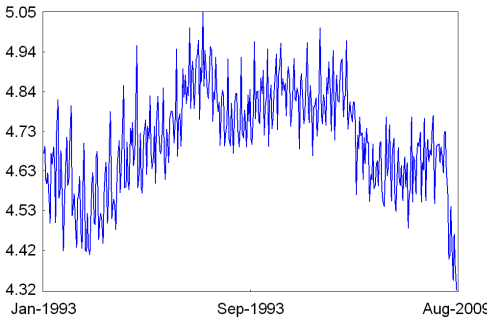
Basic Metal



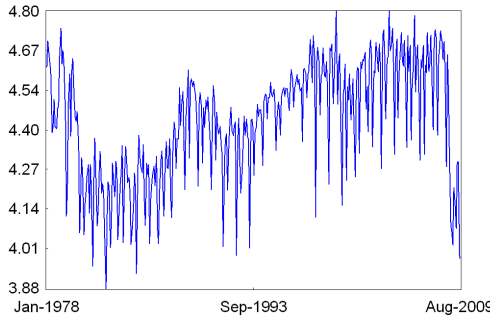
Fabricated Metal



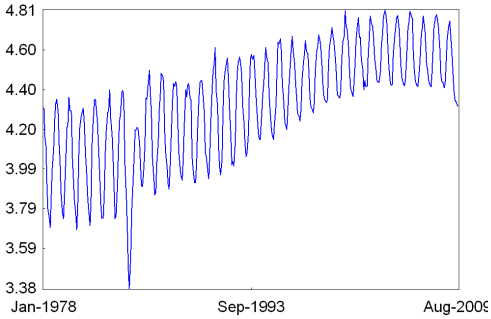
Machinery



Electrical Machinery



Vehicle



Electricity and Gas

Figure 1: The UK industrial production series.

3.2 Normality Tests

Univariate Normality test

The Anderson-Darling (A-D), Ryan-Joiner (R-J), and Kolmogorov-Smirnov (K-S) tests are used to test if a sample of data came from a population with a specific distribution. The A-D and K-S tests are based on the empirical distribution function and the R-J (similar to Shapiro-Wilk) is based on regression and correlation [28].

All three tests tend to work well in identifying a distribution as not normal when the distribution under consideration is skewed. All three tests are less discriminating when the underlying distribution is a t-distribution and non-normality is due to kurtosis. In general, among the tests based on the empirical distribution function, the A-D tends to be more effective in detecting departures in the tails of the distribution. In practice, if departure from normality at the tails is the major concern, many statisticians would use the A-D test as the first choice. Here we use all the above mentioned tests in order to have a comprehensive view of non-normality test results for the series.

The tests reject the hypothesis of normality when the p-value is less than or equal to 0.05. Failing the normality test allows us to state with 95% confidence the data does not fit the normal distribution. Table 3 represents the results of the normality test. The symbol $*$ indicates the 5% levels of significance. As can be seen from the results, except the Fabricated series, the other series are not distributed normally. Note also that, the results of non normality test for the series Machinery and Electrical Machinery are different.

Series	A-D	R-J	K-S
Food products	1.597*	0.993*	0.056*
Chemicals	5.182*	0.978*	0.089*
Basic metals	2.053*	0.985*	0.058*
Fabricated metal	0.326	0.999	0.031
Machinery	0.725	0.996*	0.039
Electrical machinery	0.876*	0.995*	0.041
Vehicles	3.091*	0.986*	0.069*
Electricity and gas	3.034*	0.985*	0.078*

Table 3: Normality tests.

Multivariate Normality test

Almost all of the industrial production series have complex structure with nonlinear trends and complex seasonality. It is worth mentioning that the SSA technique does not assume linearity or normality of the data either in finite samples or asymptotically. To assess the normality aspect of our data set, we used the Doornik-Hansen Omnibus, DHO, multivariate normality test [29]. This is a multivariate version of Shenton and Bowman's [30] univariate omnibus test for normality, based on transformed skewness and kurtosis coefficients.

The DHO(p) test statistic is approximately distributed as let p be the number of time series under consideration, then the DHO statistic is:

$$\text{DHO(p)} = Z_1^{'} Z_1^{'} + Z_2^{'} Z_2^{'} \tag{6}$$

where $Z_1' = (z_{1_1}, \ldots, z_{1_p})$ and $Z_2' = (z_{2_1}, \ldots, z_{2_p})$; z_{1_i} is a transformation of the standard univariate skewness coefficient $\sqrt{b_1}$, applied to the i-th series, due to D'Agostino [28], and z_{2_i} is a transformation of the standard kurtosis coefficient $\sqrt{b_2}$, from a gamma distribution to and then to standard normal, applied to the i-th series; see [29] for details.

Two cases of interest reported here. In the first case, the test is applied pairwise between each two series in each country. The results are represented in Table 4. As the results show, all table entries exceed the 99% quantile, indicating that there is strong evidence of non-normality.

The second case is when all 8 variables of each country are considered together, so that the null hypothesis corresponds to multivariate normality with p = 8. Again, the results with strong evidence, DHO statistics is 72.96*, confirm that the multivariate data are distributed from a multivariate non-normal distribution.

Series	FP	CH	BM	FM	MA	EM	VE	EG
Food products (FP)	26.28*	43.06*	303.5*	195.4*	70.27*	51.36*	97.83*	25.09*
Chemicals (CH)		15.33*	440.5*	474.3*	62.89*	19.01*	110.3*	44.99*
Basic metals (BM)			275.7*	231.8*	335.9*	397.8*	179.1*	234.8*
Fabricated metal (FM)				163.8*	125.5*	195.4*	115.9*	136.6*
Machinery (MA)					33.83*	37.63*	39.36*	45.74*
Electrical machinery (EM)						11.66*	69.05*	32.27*
Vehicles (VE)							67.28*	76.63*
Electricity and gas (EG)								14.28*

Table 4: Multivariate normality test.

3.3 Assessing nonlinear dependence

If a time series is the output of a non-deterministic, linear dynamic system, then measures of linear association such as the standard correlation coefficient can be used for measuring dependencies, for example, between two time series. However, if the data are outputs

from a nonlinear process, a measure should be used that has the ability to capture the nonlinearities of series. Granger and Lin [26] defined the following measure

$$\lambda = \Big(1 - exp[-2I(X,Y)]\Big)^{\frac{1}{2}}. \tag{7}$$

where I is the mutual information between the series (for more information see [27], [31]).

Table 5 shows the nonlinear correlation (the value of λ) and and the linear correlation (the second value in each cell) between the series for each country. All table entries are significant at the 5% level as indicated by *. These results strongly suggest that there is a significant degree of nonlinear dependence among the series.

Series	CH	BM	FM	MA	EM	VE	EG
Food	0.87*	0.51*	0.55*	0.42*	0.54*	0.53*	0.65*
products (FP)	0.86*	-0.16*	0.26*	0.005	0.24	0.53*	0.63*
Chemicals (CH)		0.62*	0.65*	0.56*	0.69*	0.68*	0.72*
		-0.14*	0.28*	-0.14*	0.25*	0.63*	0.72*
Basic			0.66*	0.56*	0.71*	0.62*	0.49*
metals (BM)			0.58*	0.53*	0.69*	0.42*	-0.12*
Fabricated				0.66*	0.76*	0.64*	0.44*
metal (FM)				0.73*	0.44*	0.69*	0.10*
Machinery (MA)					0.51*	0.47*	0.45*
					0.35*	0.32*	-0.16*
Electrical						0.56*	0.54*
machinery (EM)						0.55*	0.52*
Vehicles (VE)							0.55*
							0.53*

Table 5: Linear and nonlinear correlation.

Therefore, taken together, the measures of nonlinear dependence and tests for non normality show that it would be unwise to adopt a linear forecasting framework, whether univariate or multivariate, for these series.

3.4 Non-linearity tests

The performance of various non-linearity tests has been investigated in [32]. It is known that no single test dominates all the others for a variety of data generating process [32]. In the light this finding, various tests for non linearity, have been performed. The test applied are those of McLeod and Li [33], Tsay [35], Ramsey (RESET) test [34] and White [36]. To remove seasonality all tests were computed using models based on seasonal differences (D_{12}). Before testing for non linearity data were filtered by an AR model and tests are calculated from the estimated residual of a linear fit. The choice of antoregressive lag length can be made on various criteria such as maximising R^2 or minimising the Akaike

Information Criterion (AIC) or Bayesian Information Criterion (BIC) (see, for example, [37]). Alternatively the minimum number of lags that ensures the Ljung Box, Q, statistic is insignificant can be used. Table 6 presents the results for the UK. Here we have selected the lag length to a maximum of 6 by BIC; lag 12 was always included to allow for stochastic seasonality.

Series	McLeod	Tsay	RESET	NN	Lags
Food	0.163	0.007*	0.034	0.215	1-4,12
Chemicals	0.000*	0.0005*	0.490	0.002*	1-3,12
Basic metals	0.000*	0.0002*	0.156	0.0001*	1-3,12
Fabricated metal	0.000*	0.0027*	0.025	0.104	1-3,12
Machinery	0.008*	0.0028*	0.693	0.177	1-3,12
Electrical machinery	0.04	0.0030*	0.281	0.002*	1-3,12
Vehicles	0.196	0.0000*	0.001*	0.000*	1-4,12
Electricity and Gas	0.000*	0.000*	0.0002*	0.000*	1-2,12

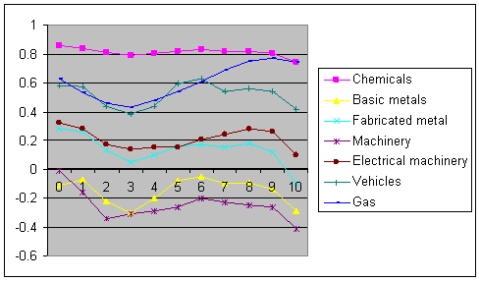
Table 6: Nonlinearity tests.

Applying multiple tests for the same null hypothesis (linearity against possible non-linearity) implies that our overall level of significance may be substantially higher than the nominal one. Therefore for a sign of non-linearity we look at either for strong rejection of linearity (at 1%) or rejections in at least two tests. There is strong evidence of non-linearity for all production series, except food products, in that two or more tests indicate non-linearity at the 1% level of significance. The results for Food products are marginal with two rejections, one at 1% and one at 5% level of significance.

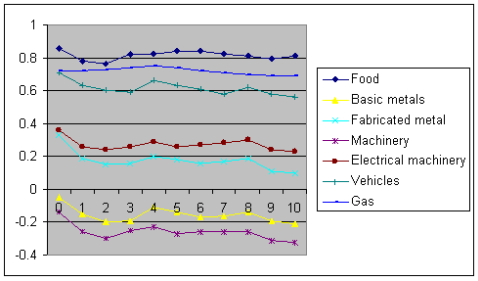
3.5 Cross Correlations for industries

We also assessed the potential of other production series for forecasting a specific industry by computing the cross correlations of each production series with lags of other productions. Figure 2 presents these cross correlations. For these figures, first the productions series were smoothed by combining the seasonal differences with differences over one quarter. This is the same as summing three values of the first and seasonal differenced series.

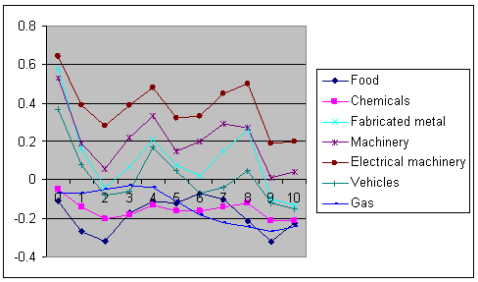
The figures show that the strongest correlations are generally at lag zero (contemporaneous information) or at very short lags. In particular, the results indicate that Food products, Chemicals and Gas have high cross correlations and Food may lead Gas by three months. Generally the strongest correlations for Basic metals are exhibited with Machinery and Fabricated metals and strongest correlations for Fabricated metal are associated with Machinery and Vehicles. The results also show that productions of machinery have strong correlations with Basic and Fabricated metals. Cross correlations for Vehicles are also relatively high with Chemicals and Food, with also a very strong contemporaneous correlation with Fabricated metal.



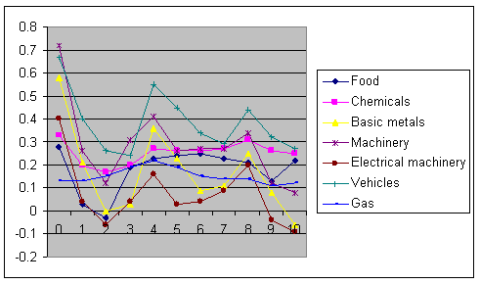
Food



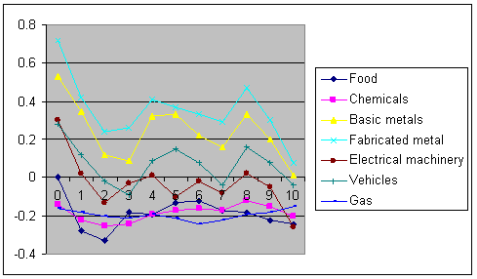
Chemical



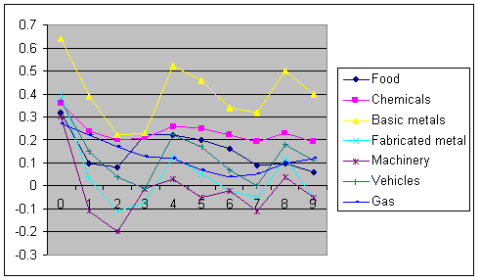
Basic Metal



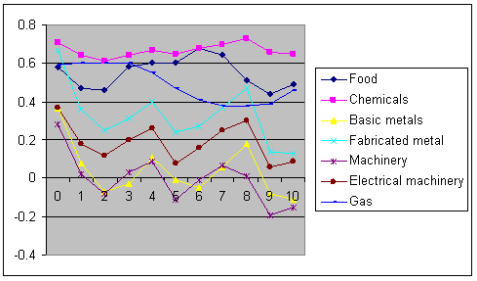
Fabricated Metal



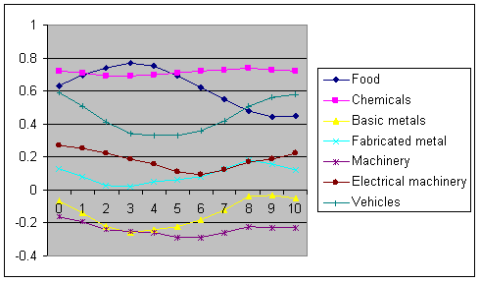
Machinery



Electrical Machinery



Vehicle



Electricity

Figure 2: The value of cross industry correlations for the UK industrial production series.

4 Forecasting Results

In this section, the performance of the SSA technique is assessed by applying it to 8 series measuring the monthly seasonally unadjusted industrial production for important sectors of the UK economies.

The SSA technique is especially useful for analyzing and forecasting series with complex seasonal components and non-stationarity. Thus, unlike ARIMA models, choosing an appropriate degree of differencing is not an issue in SSA. The data considered in this study has a complex structure; as a consequence, we found superiority of SSA over classical techniques. The results are compared with those obtained using ARIMA and VAR models.

4.1 Comparison of the accuracy of the forecasts

We consider forecasting performance of the SSA (univariate and multivariate), with ARIMA and VAR techniques at different horizons h, up to one year. The results are provided for $h = 1, 3, 6$ and 12 (months). We use the data up to the end of 2007 as in-sample (to perform SSA decomposition and to estimate parameters of ARIMA and VAR models). Thus, with approximately two year and 8 months of the out-of-sample data, we have $N = 32, 30, 27$ and 21 out-of-sample forecast errors at the horizons $h = 1, 3, 6$ and 12, respectively.

Here, we use the ratio of the root mean squared error (RRMSE) and the percentage of forecasts that correctly predict the direction of change to measure the forecast accuracy[1] (see appendix A for the definition of these measures). Note that if RRMSE < 1, then the SSA outperforms the other methods (either ARIMA or VAR).

In computing Box-Jenkins ARIMA forecasts, we need to choose the lags, the degree of differencing and the degree of seasonality (p, d, q), $(P, D, Q)_s$, where $s = 12$. To do that we use the maximum order of lags, set by the software, and apply the Bayesian Information Criterion (BIC). The SSA parameters, the window length L and the number of eigentriples r, are chosen based on the eigenvalue spectra and separability. The parameters (L, r) of the SSA and the orders $(p, d, q), (P, D, Q)_s$ of the ARIMA models are given when the models are estimated using data up to the end of 2007. Here we used $L = 24$ and $r = 14$ for all series.

Table 7 shows the out-of-sample RMSE ratios for the UK. Some summary statistics (average RRMSE of SSA models to the VAR and ARIMA models for each horizon) are also given at the bottom of each table. The summary statistics are the RRMSE averages and the scores. The score is the number of times when SSA model yields lower RMSE. For the VAR model, we use the series as a supportive series which provides the minimum RMSE in forecasting one step ahead. The series were shown in bracket. For example, for forecasting the Food products series in the VAR model, the Chemical series was used. This series was chosen as the results of one step ahead forecast of the Food series gives minimum RMSE if we use the Chemical series as the second series. The same series is also used for MSSA. Fig. 3 shows all chosen pairs of the series. The series in the bold line indicates the main series and the series in thin line indicates the supportive series.

[1] We have also computed other measures based on the magnitude of forecast errors, such as relative root mean absolute errors. These measures yield qualitatively similar results to RMSE; we thus do not report them.

Note also that the Granger causality test confirms that there exists a Granger causality between two series considered here. We have indicated this with $\leftrightarrows$ in Table 7.

The averages and the scores for h-step ahead show that SSA forecasts are much better than the forecasts obtained by ARIMA model. The averages and the scores for 1 and 3-steps ahead forecast show that VAR forecasts are also much better than the forecasts obtained by ARIMA model. However, the advantage of the VAR, relative to ARIMA, reduces for forecasting at the horizons greater than 3. The scores also confirm that the SSA forecasts outperform the forecasts produced by ARIMA model, at all horizons. For all the series (8 cases), SSA outperforms ARIMA $7, 7, 7$ and 6 times at $h = 1, 3, 6$ and 12 horizons respectively. Similarly, for all the series (8 cases), VAR outperforms the ARIMA $8, 5, 3$ and 2 times at $h = 1, 3, 6$ and 12 horizons, respectively. MSSA also outperforms the ARIMA model $7, 7, 8$ and 8 times at $h = 1, 3, 6$ and 12 horizons.

As can be seen from Table 7, MSSA outperforms ARIMA much more than the VAR model. On average, the results obtained by MSSA are up to $43\%, 38\%, 24\%$ and 18% better than for ARIMA. The results also indicate that MSSA outperforms the VAR model. On average, the results obtained by MSSA are up to $15\%, 37\%, 34\%$ and 26% better than for the VAR model. In fact, MSSA outperforms the VAR model $5, 8, 8$ and 7 times at $h = 1, 3, 6$ and 12 horizons.

The results confirm that the quality of SSA forecasts (both univariate and multivariate) at horizons $h = 1, 3, 6$ and 12 is much better than the quality of ARIMA and VAR forecasts. This observation serves as a confirmation of the following facts:

(i) most of the series considered here have a complex structure of trend and seasonality;

(ii) this structure is well recovered by SSA;

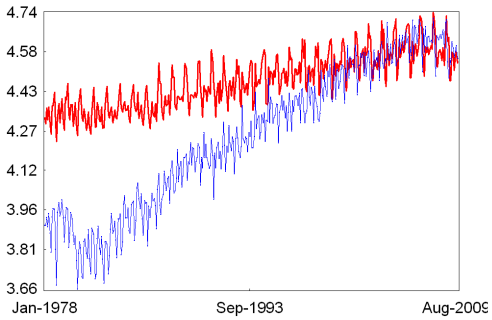
(iii) in most cases, existence of a structural break in the series reduces the performance of the ARIMA and VAR models. Note that the current recession happened at the period which we have considered as the out-of-sample forecast period. It is well known that ARIMA and VAR models are very sensitive to the structural break of this type. The quality of the forecast can be extremely high or low for classical methods depending on whether there is a structural change in the series in the out-of-sample period;

(iii) the SSA forecasts are more robust than ARIMA and VAR forecasts with respect to the presence of this shock in the series.

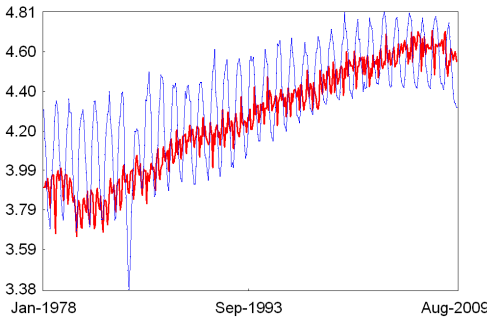
Using the modified Diebold-Marino statistics, given in [49], we test for the statistical significance of the results of the forecasts. The symbol $*$ indicates the results at the 5% level of significance or less. Comparing the SSA forecasts with the ARIMA, SSA outperforms the ARIMA significantly $6, 5, 3$ and 5 times at $h = 1, 3, 6$ and 12 horizons respectively at 5% significance level or less. MSSA also outperforms VAR significantly $3, 7, 8$ and 5 times at $h = 1, 3, 6$ and 12 horizons respectively at 5% significance level or less. In fact, the scores for all the horizons in Table 7 show that both the SSA and MSSA have outperformed the ARIMA and VAR models.

4.2 Descriptive Statistics of the Errors

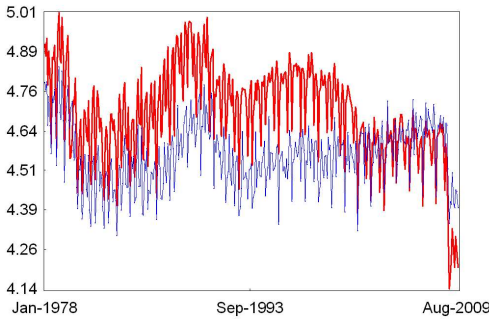
We also use descriptive statistics to describe the main features of the forecast errors, obtained by ARIMA, VAR, SSA and MSSA, in quantitative terms. We have considered



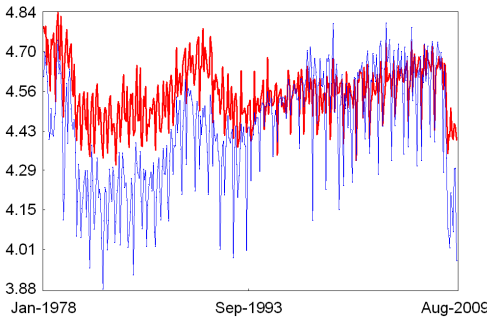
Food (bold) - Chemical (thin)



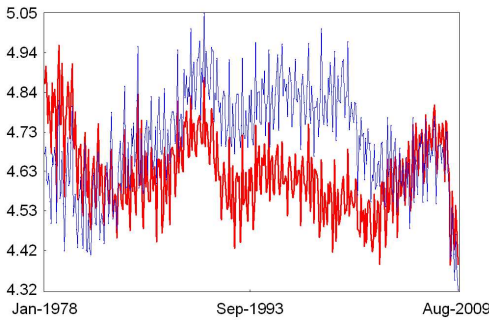
Chemical (bold) - Electricity and Gas (thin)



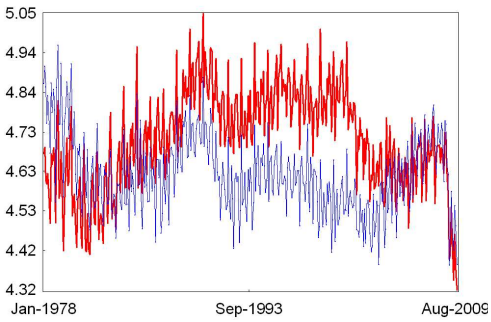
Basic (bold) - Fabricated (thin)



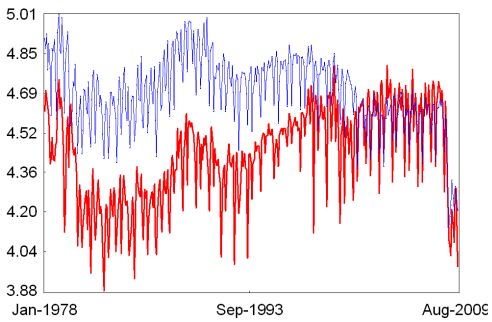
Fabricated (bold) - Vehicle (thin)



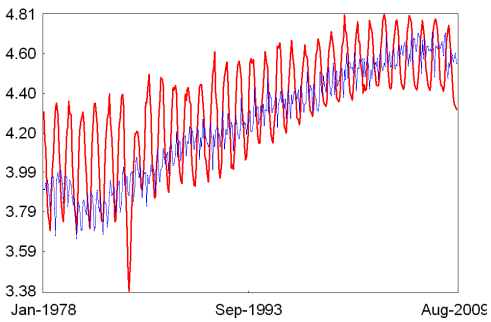
Machinery (bold) - Electrical Machinery (thin)



Electrical Machinery (bold) - Machinery (thin)



Vehicle (bold) - Basic (thin)



Electricity and Gas (bold) - Chemical (thin)

Figure 3: The series used in the multivariate forecasting approach.

two types of descriptive measures; measures of central tendency (mean and median) and measures of dispersion (standard deviation, minimum and maximum). Table 8 represents a summary statistics of the out-of-sample absolute forecast errors for ARIMA, VAR, SSA and MSSA at horizons up to one year. The first column shows that the results obtained by MSSA have smaller bias than those obtained using the VAR model. For example, for one step ahead forecast, the average of the out-of-sample forecast errors for MSSA is -0.0015 whilst this is -0.0021 for the VAR model. The performance of MSSA for $h = 3, 6$ and 12 is much better than the VAR model. The results also show that the SSA bias forecast is smaller than for ARIMA model at all considered horizons. Therefore, we can conclude that SSA and MSSA outperform the ARIMA and the VAR models in terms of bias criterion. It should be noted that the forecast bias has been reduced from ARIMA to VAR and also from SSA to MSSA. For example, for one step ahead forecast, the bias of the ARIMA forecast is -0.0623, whilst this value has been reduced to -0.0021 for the VAR model. Similarly, the bias of the SSA forecast error is -0.0105 while this has been reduced to -0.0015 for MSSA. This pattern can be seen for the other horizons too. This can be considered as a confirmation that the multivariate forecasting has a better performance than the univariate, at least for these series.

The second column represents standard deviation (SD) of the forecast errors. As the results show, MSSA results provide the minimum SD, at all considered horizons, among others. In fact, MSSA has the best performance compared to VAR, ARIMA and SSA in terms of standard deviation criterion. For example, for one step ahead forecast, the SD of the out-of-sample forecast errors for MSSA is 0.0704 whilst this is 0.0838 for the VAR model. Note also that the performance of the MSSA forecast is much better than for the VAR model at longer horizons. Similar results are obtained for SSA relative to ARIMA model. We also provide some extra descriptive statistics such as median, minimum and maximum value. Again, similar conclusion can be obtained from these statistics. The overall conclusion is that, without exception, SSA outperforms ARIMA in the univariate forecasting, and MSSA outperforms VAR model in multivariate forecasting.

4.3 Empirical Cumulative Distribution Function of the Forecasting Errors

Let us now examine the empirical cumulative distribution function (c.d.f.) for the absolute value of the forecasting errors for all the methods considered. If the c.d.f. graph produced by one method is strictly above the graph of another c.d.f., then we can conclude that the errors obtained by the first method are stochastically smaller than the errors of the second method. Note also that the Kolmogorov-Smirnov test is based on cumulative distribution functions and can be used to test to see whether two empirical distributions are different. The c.d.f. of the absolute values of the out-of-sample errors (for all eight series) obtained by ARIMA, SSA, VAR and MSSA forecasts are presented in Fig. 4. We can see from Fig. 4 that for $h = 1, 3, 6$ and 12, the SSA forecasting errors are stochastically smaller than the errors of the ARIMA model. In addition, it can be seen that the MSSA forecast errors are much smaller than the VAR forecast errors. We therefore conclude that the errors obtained by SSA are stochastically smaller than those obtained by ARIMA in univariate case. Similarly, MSSA forecast errors are stochastically smaller than those obtained by the VAR model. The results obtained from the empirical cumulative distribution function are in line with those concluded from descriptive statistics. That is, the SSA outperforms

the ARIMA model in the univariate approach, and that MSSA outperforms VAR model in multivariate approach.

4.4 Direction of change predictions

As another measure of forecast accuracy, in addition to the above criteria, we also compute the percentage of forecasts that correctly predict the direction of change (for more details see appendix A).

Table 9 provides the percentage of forecasts that correctly predict the direction of change, at $h = 1, 3, 6$ and 12 horizons. It also shows whether they are significantly greater than the pure chance ($p = 0.50$). The symbol $*$ in the table indicates the 5% levels of significance. A set of summary results is also given at the bottom of the table. The summary statistics are the average of correct signs for all eight series at $h = 1, 3, 6$ and 12 horizons. The percentage of correct signs are generally smaller than those reported in [16]. This is due to the fact that the results for directional change are particulary sensitive to structural change in the out-of-sample period. Similar to those stated for quality of the forecast, the percentage of correct signs can be very high or low for ARIMA and VAR models depending on whether there is a structural change in the series in the out-of-sample period. The overall percentage of correct signs for MSSA are $82\%, 80\%, 80\%$ and 73% at $h = 1, 3, 6$ and 12, respectively. These values for SSA are $74\%, 76\%, 76\%$ and 72% at $h = 1, 3, 6$ and 12, respectively. For ARIMA, these results are $68\%, 74\%, 77\%$ and 63%, respectively, which are slightly lower than the SSA and significantly lower than for MSSA. The VAR model has produced slightly better results (76% and 72%) at horizons $h = 1$ and $h = 12$, whilst it has produced lower (80%) at $h = 3$ and 6 horizons. For all 32 cases ($h = 1, 3, 6$ and 12 horizons) MSSA has produced 28 significant cases at the 5% level, while this value for VAR is 16 significant cases. The results indicate that MSSA enables us to improve the series movement prediction relative to SSA at all horizons. However, this is not the case for the VAR model. The VAR model only produce better results for $h = 1$ and 12 relative to ARIMA, whilst it fails for $h = 3$ and 6. Therefore, the overall conclusion from the direction of change results is: taking into account a suitable additional series for the SSA technique enables us to improve the series movement prediction in short and long horizon forecasting, whilst having extra information in the VAR model helps us to improve the results only in short term forecasting. Another remark is that, the relation between the series movement in a multivariate system is captured much better by SSA than by the VAR model.

5 the Causality Tests Using the Singular Spectrum Analysis

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 in [38]. Granger [38] 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 [39]). 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 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 [40].

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, [41]). Hiemstra and Jones [42] 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 [43], *ii*) there are restrictive assumptions in this approach [44] and *iii*) the test can severely over-reject the null hypothesis of non-causality [45].

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 [46]. Su and White [47] 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.

5.1 Causality Criteria

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. If the forecasting errors 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. 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

$$\Delta_{X_{K+H_x}} \equiv \mathcal{L}(X_{K+H_x} - \hat{X}_{K+H_x}), \tag{8}$$

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:

$$\Delta_{X_{K+H_x}|Y_{K+H_y}} \equiv \mathcal{L}(X_{K+H_x} - \tilde{X}_{K+H_x}). \tag{9}$$

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:

$$F_{X|Y}^{(h,d)} = \frac{\Delta_{X_{K+H_x}|Y_{K+H_y}}}{\Delta_{X_{K+H_x}}} \tag{10}$$

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, $F_{X|Y}^{(h,0)}$ 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. $F_{X|Y}^{(h,0)}$ can be considered as a common multivariate forecasting system for time series with the same series length. The criterion $F_{X|Y}^{(h,0)}$ can then be used in evaluating two instantaneous causality. Similarly, $F_{X|Y}^{(h,1)}$ 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 $F_{X|Y}^{(h,d)}$ 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 $F_{X|Y}^{(h,d)}$ is very small, then the predictions using the multivariate version are much more accurate than the predictions by the univariate SSA. If $F_{X|Y}^{(h,d)} < 1$, 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 $F_{X|Y}^{(h,d)} \geq 1$, 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 $F_{Y|X}^{(h,d)}$ 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 $F_{Y|X}^{(h,d)} < F_{X|Y}^{(h,d)}$, we then conclude that X is more *supportive* than Y, and if $F_{X|Y}^{(h,d)} < F_{Y|X}^{(h,d)}$, 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 $F_{Y|X}^{(h,d)} < 1$ and $F_{X|Y}^{(h,d)} < 1$, 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 [48] test statistic, with the corrections suggested in [49]. 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.

5.2 Direction of change based criterion

Ash [50] 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 [51] discuss that the value of a model's forecasts may be better measured by the direction of change. Heravi [25] 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 $\bar{Z_X} = \sum_{i=1}^{n} Z_{Xi}/n$ 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:

$$D_{X|Y}^{(h,d)} = \frac{\bar{Z_X}}{\bar{Z}_{X|Y}} \tag{11}$$

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 10.

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

$$\pi = P_{X|Y} - P_X = \frac{a+b}{n} - \frac{a+c}{n} = \frac{b-c}{n} \tag{12}$$

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 [52]:

$$\hat{SE}(\pi) = \frac{1}{n}\sqrt{(b+c) - \frac{(b-c)^2}{n}}. \tag{13}$$

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

$$\begin{aligned} H_0 &: \pi_d = \Delta_0 \\ H_a &: \pi_d \neq \Delta_0 \end{aligned} \tag{14}$$

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

$$T_{\pi_d} = \frac{\pi - \Delta_0 - 1/n}{\hat{SE}(\pi)} \tag{15}$$

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

$$T_{\pi_d} = \frac{(b-c)/n - 1/n}{1/n\sqrt{(b+c) - (b-c)^2/n}} = \frac{b-c-1}{\sqrt{(b+c) - (b-c)^2/n}} \tag{16}$$

The test is valid when the average of the discordant cell frequencies, $(b+c)/2$, is equal or more than 5. However, if this 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.

5.3 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

$$x_t = \phi_0 + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + \psi_1 y_{t-1} + \ldots + \psi_p y_{t-p} + \varepsilon_{t_{x|y}} \tag{17}$$

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:

$$H_0 = \psi_{L+1} = \psi_2 = \ldots = \psi_p = 0 \tag{18}$$

If the null hypothesis holds, the full model (17) is reduced to the restricted model as follows:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-L+1} + \varepsilon_{t_x} \tag{19}$$

where ε_{t_x} is *iid* sequence with zero mean and variance σ_x. The forecasting results obtained by the restricted model (19) are compared to those obtained using the full model (17) 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 the eigenvectors of the trajectory matrix $\mathbf{X}$. Thus, the univariate version of SSA is given by

$$x_t = \alpha_1 x_{t-1} + \ldots + \alpha_{L-1} x_{t-L+1} \tag{20}$$

As can be seen from (20), 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:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \alpha_1 x_{t-1} + \ldots + \alpha_{L-1} x_{t-L+1} \\ \beta_1 y_{t-1} + \ldots + \beta_{L-1} y_{t-L+1} \end{pmatrix} \tag{21}$$

where the vectors $A = (\alpha_1, \ldots, \alpha_{L-1})$ and $B = (\beta_1, \ldots, \beta_{L-1})$ are estimated using the eigenvectors of the trajectory matirx $\mathbf{M} = [\mathbf{X}\ \mathbf{Y}]^T$. As equation (21) 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 matrices $\mathbf{X}$ 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

$$\frac{Var(\mathbf{e}_{X|Y})}{Var(\mathbf{e}_X)}, \quad \frac{Var(\mathbf{e}_{Y|X})}{Var(\mathbf{e}_Y)} \tag{22}$$

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 [12].

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.

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 [53].

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 [5].

5.4 Empirical results

The results presented in previous section showed that the VAR model is not a suitable choice in predicting the UK industrial production series, while SSA (specifically, multivariate SSA) decisively outperforms the VAR model. We also found that the UK industrial production series series are nonlinear, non-stationary and are not normally distributed. Moreover, the Granger causality test confirms that there is a Granger causality between series considered in multivariate forecasting approach.

Next we consider the proposed test for finding causality between, the Food and the

Chemical series using the criteria we have introduced in previous section. For more application of the proposed tests on exchange rate series and the final vintage of the Index of Industrial Production series see [20]. It is very clear that the Food and Chemical series are highly correlated (indeed, the linear and nonlinear correlation coefficient between each two series are about 0.75 and 0.85, respectively). Fig. 5 shows both the Food and Chemical series over the period Feb-1978 to Aug-2009.

We perform $h = 1, 3, 6$ and 12 step ahead forecasting based on the most up-to-date information available at the time of the forecast. We use both series simultaneously, e.g. we use the Food series in forecasting the Chemical series and vice versa. We use d-step ahead information of the Food series as additional information in forecasting h step ahead of the Chemical series and vice versa. We denote these statistics $F_{F|C}^{(h,d)}$ and $F_{C|F}^{(h,d)}$, respectively. Note that we select window length 24 for both single and multivariate SSA in forecasting these series. In model selection for VAR model we use Akaike Information criterion, AIC, and Schwarz Information Criterion, SC, values to identify the VAR model order. The symbol $*$ indicates the significant results on the 1% level.

As we mentioned earlier the Granger causality test results confirm that there is a significant relationship between the Food and the Chemical series. To examine this using the SSA technique, next we consider MSSA with d additional observations for one series. For example, we use the Food series up to time t, and the Chemical series up to time $t + d$ in forecasting h step ahead of the Food series to compute $F_{F|C}^{(h,d)}$. We use similar procedure in forecasting the Chemical series. We expect this additional information to give better results in both forecasting accuracy and the direction of change prediction.

Let us first consider the results obtained for the Food series using extra information of Chemical series. As can be seen from Table 11, the accuracy of the results obtained using MSSA are better than those obtained using SSA for $h \geq 6$ with respect to $F_{F|C}^{(h,d)}$ column $\frac{MSSA}{SSA}$. The MSSA also improved predictability of the direction of change of the Food series for all considered horizon (according to $D_{C|F}^{(h,d)}$ column $\frac{MSSA}{SSA}$ in Table 12). The results show that, for the Food series, we have improved both accuracy and direction of change of the forecasting results for $h \geq 6$. Furthermore, the prediction of correct change has been improved for $h \leq 6$. This indicates that we can at least improve the direction of change predictability for the Food series using extra information of the Chemical series.

Note also that, $F_{F|C}^{(h,d)}$ for $h = 1, \ldots, 6$ is equal to 1 or slightly grater than 1 which indicates that the multivariate version could not help in short horizon forecasting. To examine this more precisely, we need to increase our forecasting period. In average, we were able to improve both forecasting accuracy and direction of change predictability by 4%. As the results show, the VAR model can not help us in forecasting the Food series using extra information of the Chemical series according to the results provided in the second column of Tables 11 and 12. To have more comparison, we have compared VAR results and those obtained by MSSA. As it can be seen from the tables, the results obtained by MSSA are much better than those provided using the VAR model. On average, the MSSA results are up to 50% and 13% better than the VAR model in forecasting accuracy and predictability of the direction of change.

Let us now consider the results of forecasting the Chemical series having extra observation of the Food series. As can be observed from columns $F_{C|F}^{(h,d)}$, $D_{C|F}^{(h,d)}$, the errors for the MSSA forecast and direction of change, with d additional observations, are much smaller than those obtained for the univariate SSA. These results are also better than the

results obtained using multivariate approach with zero lag difference. This is not surprising 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. However, it seems that this correlation, either linear and nonlinear, does not capture by the VAR model properly.

For example, in forecasting one step ahead of the Chemical series having extra information of the Food series, comparing to univariate case, we have improved the accuracy and the direction of change of the forecasting results up to 45% and 35% (column 4 of Tables 11 and 12), respectively. On average, the forecasting accuracy results have been improved up to 39% and 22% using MSSA and VAR model, respectively. However, the direction of change results, presented in table 12 for the Chemical series, confirm that even though VAR model could help us to improve the forecasting accuracy, it fails to improve the direction of change predictability. Comparing MSSA prediction of change results and forecasting accuracy results relative to VAR model indicates that the MSSA technique are up to 24% and 9% better than the VAR model, respectively.

Moreover, $F_{C|F}^{(h,d)} > F_{F|C}^{(h,d)}$ indicates that, in forecasting this period of the series, the Food series is more supportive than the Chemical series in terms of forecasting accuracy.

Furthermore, the results of Table 12 confirm that there exists D-feedback between the Food and Chemical series for $h = 1, \ldots, 12$. This means that considering both the Food and Chemical series simultaneously and using SSA, we are able to improve the predictability of the direction of change. On average, the MSSA results has improved the percentage of correct prediction of the series movement up to 4% and 22% for the Food and the Chemical series, respectively. Moreover, there exists F-feedback between the Food and the Chemical series for $h = 6, \ldots, 12$.

Finally, comparing VAR results and ARIMA results in terms of forecasting accuracy and direction of change predictability confirms that for the Food series VAR models can not help us to improve our results whilst there exists Granger causality between series. However, as the results show, VAR model can improve the results obtained for the Chemical series except for $h = 7$ and 9 (this probably happened as our out-of-sample size is short). On average, for the Chemical series, MSSA and VAR model enable us to improve the forecasting accuracy up to 34% and 12%, respectively. However, there is no improvement for the direction of change prediction using VAR model for both the Food and the Chemical series.

6 Summary and conclusion

Given that the dynamics of the economy of many countries has gone through many political and structural changes over different periods of time, one needs to make certain that the method of prediction is not sensitive to the dynamical variations.

In this regard the Singular Spectrum Analysis (SSA) technique can be considered as a technique which is not too sensitive to the structural breaks. The data considered in this study has a complex structure and contains structural changes; as a consequence, we found superiority of SSA over classical techniques. We are motivated to use SSA because of its capability in dealing with stationary as well as non-stationary series. Moreover, contrary to the traditional methods of time series forecasting (both autoregressive or structural models that assume normality and stationarity of the series), the SSA technique is non-

parametric and makes no prior assumptions about the data.

In this report, we have described the methodology of SSA and demonstrated that SSA can be successfully applied to the analysis and forecasting of the industrial production series for the United Kingdom. This research has illustrated that the SSA technique performs well in the simultaneous extraction of harmonics and trend components even in circumstances where there is a structural break in series. The comparison of forecasting results showed that SSA is more accurate than ARIMA model confirming the results obtained in [16]. The multivariate SSA also outperformed the VAR model in predicting the values and the direction of the production series according to the RMSE criterion and the direction of change results.

We also considered empirical cumulative distribution function and descriptive statistics of the errors obtained by SSA, MSSA, ARIMA and the VAR model. The results confirmed that the errors obtained by SSA are stochastically smaller than those obtained by ARIMA in the univariate case. Similarly, MSSA forecast errors are stochastically smaller than those obtained by the VAR model. The results obtained from the empirical cumulative distribution function were in line with those concluded from descriptive statistics. The overal conclusion according to these criteria is, the SSA outperforms the ARIMA model in the univariate approach, and MSSA outperforms VAR model in multivariate approach.

We therefore conclude that the SSA technique can be considered as a reliable method for predicting recessions and expansions. The series considered in this research are some examples of different seemingly complex series with potential structure which can be easily analysed by SSA and could provide a typical example of a successful application of SSA.

We also 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.

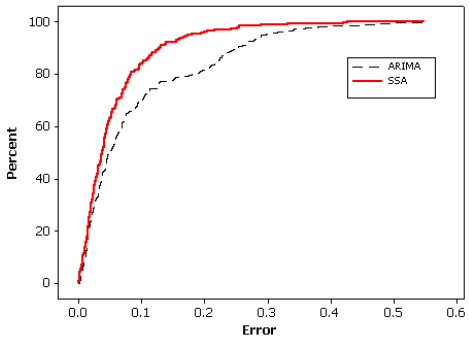
The performance of the proposed test was examined using the Food production series and the Chemical series for the United Kingdom. It has been shown that, on average, the Chemical series causes the Food series and the Food series causes the Chemical series in terms of direction of change predictability. Therefore, we conclude that there exists a SSA causal relationship between the Food and the Chemical series in forecasting direction of change. We also found that there exist one way causality between the Food and the Chemical series in terms of forecasting accuracy criterion for horizon longer than 6 months. On the other hand the VAR model fails to use the dependence between the Food and the Chemical series.

Series	h	RRMSE $\frac{SSA}{ARIMA}$	$\frac{VAR}{ARIMA}$	$\frac{MSSA}{ARIMA}$	$\frac{MSSA}{VAR}$
Food products	1	1.01	0.98	1.07	1.09
(Chemical)	3	1.00	1.36	1.03	0.75
⇆	6	0.96	1.37	0.95	0.70*
	12	0.95	1.32	0.94	0.71*
Chemicals	1	0.81*	0.54*	0.56*	1.03
(Electricity and Gas)	3	0.89	0.90	0.54*	0.60*
⇆	6	0.96	1.19	0.68*	0.57*
	12	1.03	1.38	0.52*	0.38*
Basic metals	1	0.68*	0.54*	0.51*	0.94
(Fabricated metal)	3	0.61*	1.28	0.53*	0.41*
⇆	6	0.87	1.49	0.60*	0.40*
	12	1.00	1.42	0.65*	0.45*
Fabricated metal	1	0.46*	0.82*	0.44*	0.54*
(Vehicles)	3	0.48*	0.76*	0.46*	0.61*
⇆	6	0.59*	0.76*	0.58*	0.77*
	12	0.74*	0.79*	0.65*	1.17
Machinery	1	0.62*	0.65*	0.44*	0.78*
(Electrical machinery)	3	0.68*	0.90	0.52*	0.58*
⇆	6	0.77*	0.92	0.67*	0.73*
	12	0.77*	0.80*	0.72*	0.90
Electrical machinery	1	0.56*	0.46*	0.48*	1.04
(Machinery)	3	0.65*	0.79*	0.61*	0.78*
⇆	6	1.04	1.26	0.97	0.77*
	12	0.94	1.03	0.94	0.91
Vehicles	1	0.69*	0.78*	0.66*	0.84*
(Basic metal)	3	0.86	1.14	0.82*	0.72*
⇆	6	0.99	1.22	0.92*	0.75*
	12	0.86*	1.06	0.90	0.85*
Electricity and Gas	1	0.35*	0.34*	0.33*	0.97
(Chemicals)	3	0.46*	0.74*	0.49*	0.70*
⇆	6	0.60*	0.79*	0.56*	0.76*
	12	0.66*	1.08	0.60*	0.61*
Average	1	0.61	0.63	0.53	0.85
	3	0.69	0.99	0.62	0.63
	6	0.86	1.14	0.76	0.66
	12	0.89	1.11	0.82	0.74
Score	1	7	8	7	5
	3	7	5	7	8
	6	7	3	8	8
	12	6	2	8	7

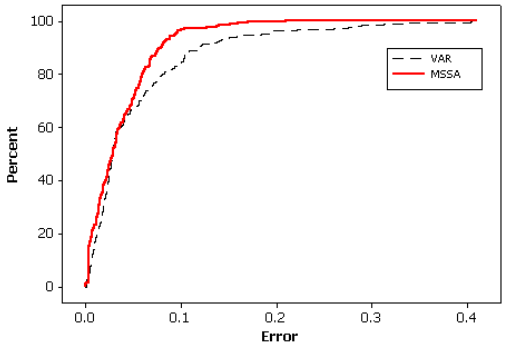
Table 7: Forecasting results for the UK industrial production series using ARIMA, SSA, VAR and MSSA methods.

Method	N	Mean	SD	Min	Median	Max
1-step ahead						
ARIMA	256	-0.0623	0.1246	-0.549	-0.0223	0.252
SSA	256	-0.0105	0.0854	-0.419	-0.0071	0.216
VAR	256	-0.0021	0.0838	-0.410	-0.0180	0.193
MSSA	256	-0.0015	0.0704	-0.293	-0.0002	0.159
3-step ahead						
ARIMA	240	-0.0760	0.1290	-0.582	-0.0259	0.246
SSA	240	-0.0217	0.1053	-0.468	-0.0112	0.225
VAR	240	-0.0492	0.1513	-0.584	-0.0221	0.261
MSSA	240	-0.0172	0.1001	-0.463	-0.0048	0.121
6-step ahead						
ARIMA	216	-0.0779	0.1367	-0.6082	-0.0289	0.177
SSA	216	-0.0424	0.1274	-0.5938	-0.0251	0.154
VAR	216	-0.0670	0.1795	-0.6816	-0.0271	0.279
MSSA	216	-0.0382	0.1182	-0.3880	-0.0125	0.124
12-step ahead						
ARIMA	168	-0.1280	0.1559	-0.6922	-0.0670	0.134
SSA	168	-0.0952	0.1496	-0.6310	-0.0579	0.129
VAR	168	-0.0869	0.2045	-0.7128	-0.0474	0.283
MSSA	168	-0.0759	0.1323	-0.5740	-0.0412	0.117

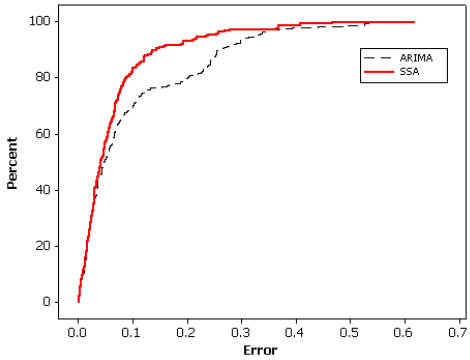
Table 8: Descriptive statistics for out-of-sample absolute forecasting errors.



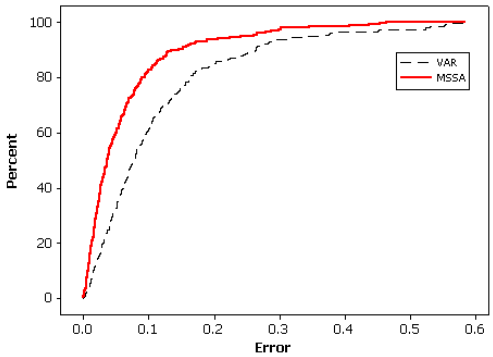
ARIMA-SSA (1 step ahead)



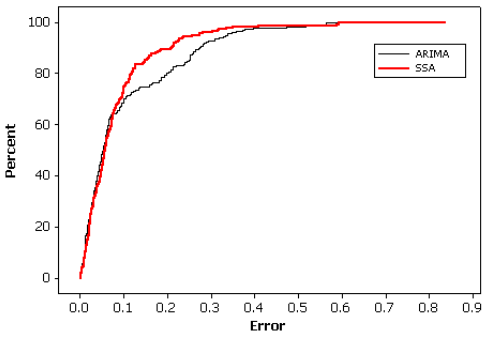
VAR-MSSA (1 step ahead)



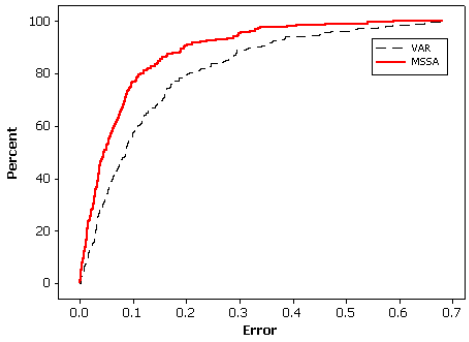
ARIMA-SSA (3 step ahead)



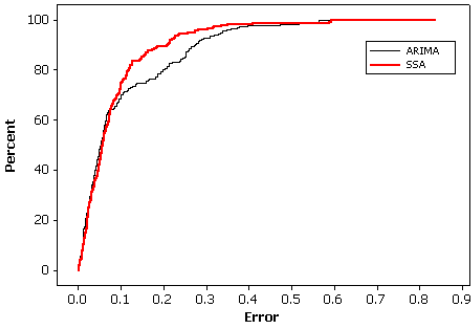
VAR-MSSA (3 step ahead)



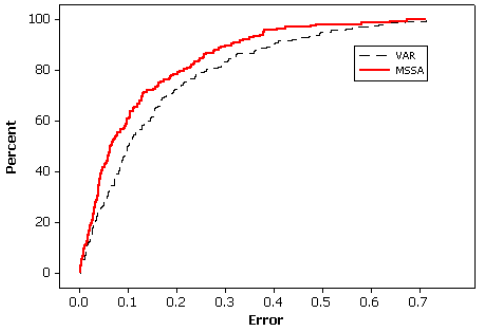
ARIMA-SSA (6 step ahead)



VAR-MSSA (6 step ahead)



ARIMA-SSA (12 step ahead)



VAR-MSSA (12 step ahead)

Figure 4: The empirical cumulative distribution functions for the absolute forecasting errors of ARIMA, SSA, VAR and MSSA.

Series	h	ARIMA	SSA	VAR	MSSA
Food products	1	0.87*	0.77*	0.77*	0.84*
	3	0.77*	0.83*	0.77*	0.77*
	6	0.89*	0.89*	0.93*	0.93*
	12	0.67	0.71*	0.76*	0.76*
Chemicals	1	0.71*	0.55	0.77*	0.71*
	3	0.70*	0.67	0.53	0.70*
	6	0.74*	0.74*	0.52	0.81*
	12	0.67	0.62	0.62	0.81*
Basic metals	1	0.68*	0.81*	0.74*	0.81*
	3	0.80*	0.80*	0.50	0.83*
	6	0.70*	0.67	0.44	0.70*
	12	0.71*	0.76*	0.29	0.76*
Fabricated metal	1	0.61	0.71*	0.81*	0.81*
	3	0.70*	0.83*	0.60	0.83*
	6	0.70*	0.67	0.59	0.70*
	12	0.48	0.71*	0.90*	0.81*
Machinery	1	0.74*	0.77*	0.74*	0.87*
	3	0.80*	0.73*	0.57	0.90*
	6	0.74*	0.74*	0.70*	0.89*
	12	0.52	0.67	0.76*	0.67
Electrical machinery	1	0.68*	0.74*	0.77*	0.77*
	3	0.77*	0.70*	0.70*	0.80*
	6	0.67	0.70*	0.56	0.59
	12	0.71*	0.76*	0.48	0.67
Vehicles	1	0.65	0.71*	0.68*	0.81*
	3	0.77*	0.67	0.70*	0.67
	6	0.74*	0.70*	0.67	0.74*
	12	0.67	0.81*	0.67	0.71*
Electricity and Gas	1	0.48	0.87*	0.81*	0.94*
	3	0.63	0.87*	0.53	0.93*
	6	0.96*	1.00*	0.52	1.00*
	12	0.57	0.71*	0.62	0.67*
Average	1	0.68	0.74	0.76	0.82
	3	0.74	0.76	0.61	0.80
	6	0.77	0.76	0.62	0.80
	12	0.63	0.72	0.64	0.73

Table 9: Direction of change results for the UK industrial production series using ARIMA, SSA, VAR and MSSA methods.

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

Table 10: An arrangement of Z_X and $Z_{X|Y}$ in forecasting n future points of the series X.

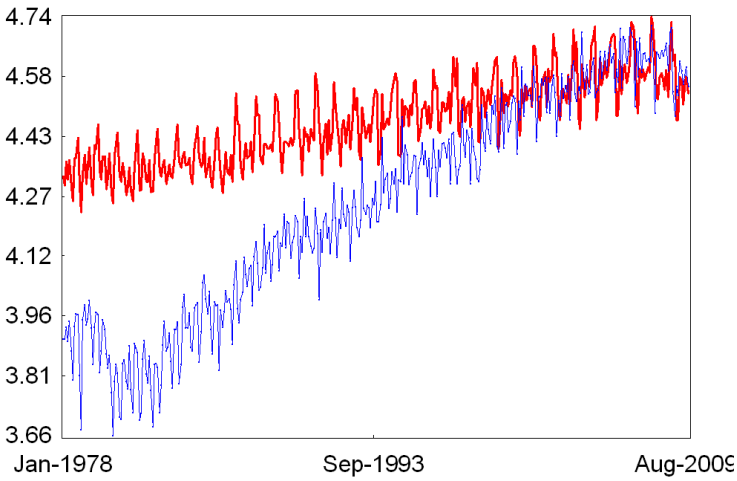


Figure 5: The UK Food production (bold) and Chemical series (thin) over the period Feb-1978 to Aug-2009.

h	$F^{(h,d)}_{F\|C}$			$F^{(h,d)}_{C\|F}$		
	$\frac{MSSA}{SSA}$	$\frac{VAR}{ARIMA}$	$\frac{MSSA}{VAR}$	$\frac{MSSA}{SSA}$	$\frac{VAR}{ARIMA}$	$\frac{MSSA}{VAR}$
1	1.05	1.08	0.62*	0.55*	0.72*	0.83*
2	1.03	1.10	0.55*	0.74*	0.61*	0.86*
3	1.02	1.11	0.50*	0.69*	0.72*	0.85*
4	1.03	1.14	0.52*	0.73*	0.80*	0.86*
5	1.00	1.12	0.53*	0.67*	0.98	0.83*
6	0.90	1.15	0.48*	0.85*	1.00	0.73*
7	0.85*	1.17	0.44*	0.60*	1.06	0.74*
8	0.87*	1.08	0.45*	0.60*	0.95	0.72*
9	0.85*	1.06	0.43*	0.53*	1.08	0.60*
10	0.83*	1.07	0.39*	0.54*	0.97	0.59*
11	0.92	1.09	0.49*	0.42*	0.92	0.68*
12	0.97	1.07	0.66*	0.45*	0.80*	0.78*
Average	0.96	1.10	0.50	0.61	0.88	0.76

Table 11: The value of RRMSE in forecasting h step aheads for Food and Chemical series for $d = 12$.

h	$D^{(h,d)}_{F\|C}$			$D^{(h,d)}_{C\|F}$		
	$\frac{MSSA}{SSA}$	$\frac{VAR}{ARIMA}$	$\frac{MSSA}{VAR}$	$\frac{MSSA}{SSA}$	$\frac{VAR}{ARIMA}$	$\frac{MSSA}{VAR}$
1	0.89	1.08	0.93	0.65*	0.96	1.00
2	0.96	1.00	0.88	0.90	0.91	1.05
3	0.93	1.29	0.78*	0.95	1.29	0.89
4	0.92	1.14	0.85*	0.90	1.11	0.95
5	0.96	1.25	0.80*	0.81*	1.25	0.76*
6	0.96	1.19	0.84*	1.00	1.67	0.63*
7	0.98	1.21	0.86*	0.57*	1.06	0.74*
8	1.00	1.17	0.78*	0.84*	1.00	0.89
9	0.95	1.05	0.86*	0.88	1.13	0.94
10	1.00	1.00	0.90	0.86*	1.07	1.07
11	1.00	0.93	1.07	0.53*	1.08	0.87*
12	0.92	0.73	0.85*	0.50*	0.67*	1.07
Average	0.96	1.09	0.87	0.78	1.10	0.91

Table 12: Post sample relative percentage of the corrected forecast sign for Food and Chemical series.

References

[1] Plaut, G., and Vautard, R. (1994). Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere, *J. Atmos. Sci.*, **51**, pp. 210-236.

[2] Box, G. E. P., and Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*, Holden-Day.

[3] Brockwell, P. J., and Davis R. A. (2002). *Introduction to Time Series and Forecasting*, 2nd edition. Springer.

[4] Cao, L. Y., and Soofi, A. (1999). Nonlinear deterministic forecasting of daily dollar exchange rates, *International Journal of Forecasting*, **15**(4), pp. 421–430.

[5] 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.

[6] Hsieh, D. A. (1991). Chaos and nonlinear Dynamics: Application to Financial Markets, *Journal of Finance*, **46**, pp. 1839–1877.

[7] Scheinkman, J., and LeBaron, B. (1989). Nonlinear Dynamics and Stock Returns, *Journal of Business*, **62**, pp. 311–337.

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

[9] Broomhead, D. S., and King, G. (1986). Extracting qualitative dynamics from experimental data, *Physica* D, **20**, pp. 217–236.

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

[11] Elsner, J. B., and Tsonis, A. A. (1996). *Singular Spectrum Analysis, A New Tool in Time Series Analysis*, Plenum Press, New York and London.

[12] Hassani, H., and Zhigljavsky, A. (2009). Singular Spectrum Analysis: Methodology and Application to Economics Data, *Journal of System Science and Complexity*, **22**(3), pp. 372–394.

[13] Vautard, R., Yiou, P., and Ghil, M. (1992). Singular spectrum analysis: a toolkit for short, noisy and chaotic series, *Physica D*, **58**, pp. 95–126.

[14] Moskvina, V. G., and Zhigljavsky, A. (2003). An algorithm based on singular spectrum analysis for change-point detection, *Communication in Statistics - Simulation and Computation*, **32**(4), pp. 319–352.

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

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

[17] Patterson, K., Hassani, H., Heravi, S., and Zhigljavsky, A. (2010). Forecasting the final vintage of the industerial production series, *Jpournal on Applied Statistics*, Forthcoming.

[18] Alexandrov, Th., and Golyandina, N. (2004). The automatic extraction of time series trend and periodical components with the help of the Caterpillar-SSA approach. *Exponenta Pro 3-4* (In Russian.), pp. 54–61.

[19] www.gistatgroup.com

[20] Hassani, H., Zhigljavsky, A., Patterson, K., and Soofi, A. (2010). A Comprehensive Causality Test Based on the Singular Spectrum Analysis, *Causality in Science*, Oxford University press, Forthcoming.

[21] Hassani, H., Dionisio, A., and Ghodsi, M. (2010). The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets, *Nonlinear Analysis: Real World Applications*, **11**(1), pp. 492–502.

[22] Ghodsi, M., Hassani, H., Sanei, S., and Hick, Y. (2009). The use of noise information for detecting temporomandibular disorder, *Biomedical Signal Processing and Control*, **4**, pp. 79–85.

[23] Hassani, H., Zokaei, M., von Rosen, D., Amiri, S., and Ghodsi, M. (2009). Does Noise Reduction Matter for Curve Fitting in Growth Curve Models?, *Computer Methods and Program in Biomedicine*, **96**(3), pp. 173–181.

[24] Osborn, D. R., Heravi, S., and Birchenhall, C. R. (1999). Seasonal unit roots and forecasts of two-digit European industrial production. *International Journal of Forecasting*, **15**, pp. 27–47.

[25] 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**, pp. 435–446.

[26] Granger, C. W. J. and Lin, J. (1994). Using the Mutual Information Coefficient to Identify Lags in Nonlinear Models, *Journal of Time Series Analysis*, **15**, pp. 371-384.

[27] Granger, C. W. J., Maasoumi, E. and Racine, J. (2004). Journal of Time Series Analysis, **25**(5), pp. 649–669.

[28] DAugostino, R. B., and Stephens, M. A. Eds. (1986). *Goodness-of-Fit Techniques*, Marcel Dekker.

[29] Doornik, J. A., and Hansen, H. (1994). An Omnibus Test for Univariate and Multivariate Normality. Discussion Paper W4&91, Nuffield College, Oxford, UK.

[30] Shenton, L. R., and Bowman, K. O. (1997). A Bivariate Model for the Distribution of sqrt(b1) and b2, *Journal of the American Statistical Association*, **72**, pp. 206-211.

[31] Darbellay, G. (1998). An Adaptive Histogram Estimator for Mutual Information, UTIA Research Report 1889. Academy of Sciences, Prague.

[32] Lee, T. H., White, H., and Granger, C. W. J. (1993). Testing for Neglected Nonlinearity in Time Series Models, *Journal of Econometrics*, **56**, pp. 269–290.

[33] McLeod, A. I. and Li, W. K. (1983). Diagnostic checking ARMA time series models using squared residual autocorrelations, *Journal of Time Series Analysis*, **4**, pp. 169–176.

[34] Ramsey, J. B. (1969). Tests for specifications errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society*, Series B, **31**(2), pp. 350–371.

[35] Tsay, R. S. (1986). Non-linearity tests for time series. *Biometrika*, **73**(2), pp. 461–466.

[36] White, H. (1989). Some asymptotic results for learning in single hidden layer feedforward network models. *Journal of the American Statistical Association*, **84**, pp. 1003–1013.

[37] Lhtkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

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

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

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

[41] Chu, T., Danks, D. and Glymour, C. (2005). Data Driven Methods for Nonlinear Granger Causality: Climate Teleconnection Mechanisms, Technical report (CMU-PHIL-171).

[42] 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.

[43] 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.

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

[45] 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.

[46] 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.

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

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

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

[50] 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.

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

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

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

[54] Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, **14**, pp. 35-62.

[55] Granger, C. W. J., and Lin, J. (1994). Using the Mutual Information Coefficient to Identify Lags in Nonlinear Models, *Journal of Time Series Analysis*, **15**, pp. 371–384.

[56] Darbellay, G., and Wuertz, D. (2000). The Entropy as a Tool for Analysing Statistical Dependence's in Financial Time Series, *Physica A* , **287**, pp. 429–439.

[57] Dionisio, A., Menezes, R., and Mendes, D. A. (2004). Mutual Information: a Measure of Dependency for Nonlinear Time Series, *Physica A*, **344**, pp. 326–329.

[58] Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, **75**, pp. 800–802.