A review on singular spectrum analysis for economic and financial time series

Hossein Hassani* and Dimitrios Thomakos

In recent years Singular Spectrum Analysis (SSA), a relatively novel but powerful technique in time series analysis, has been developed and applied to many practical problems across different fields. In this paper we review recent developments in the theoretical and methodological aspects of the SSA from the perspective of analyzing and forecasting economic and financial time series, and also represent some new results. In particular, we (a) show what are the implications of SSA for the, frequently invoked, unit root hypothesis of economic and financial times series; (b) introduce two new versions of SSA, based on the minimum variance estimator and based on perturbation theory; (c) discuss the concept of causality in the context of SSA; and (d) provide a variety of simulation results and real world applications, along with comparisons with other existing methodologies.

AMS 2000 subject classifications: 92C55, 94A12.
Keywords and phrases: Singular spectrum analysis, Cointegration, Economic/Finacial Time Series, Filtering, Forecasting, Smoothing, Unit root, Causality.

1. INTRODUCTION

Econometric methods have been widely used to model, smooth and forecast the evolution of all kinds of economic and financial time series, such as quarterly and annual national account data sets like real Gross Domestic Product, inflation or unemployment, stock indices, interest rates etc. Econometric methods and models for such series are accompanied by a huge literature (which we cannot possibly review here) but in many instances have failed to produce statistically accurate, and economically meaningful, forecasts.

There are several reasons that a classical model does not have a good performance for modelling and forecasting economic and financial series. The most prominent are (a) their pronounced non-stationarity, in the mean and also in the variance, and (b) the frequent structural breaks that they present. Both these features can be traced to a number of factors, such as measurement noise, exogenous shocks, technological change, policy changes, changes in consumer preferences, news and announcements etc. Many econometric models used for forecasting have problems dealing with these

*Corresponding author.

features, at least simultaneously, and a number of techniques have been developed to address them. Such techniques include applications of Kalman-filter like procedures with time-varying coefficients, semi and non-parametric models, models with allowance for unit root non-stationarity, models with cointegration, intercept corrections and many others. It is evident that the assumptions of stationarity, linearity and normality are, in many cases, poor approximations to the real world data that we are given in economics and finance.

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, [1, 2] and references therein). We should note that other distributional assumptions, beyond normality, are frequently found in the context of financial time series but they are made on empirical grounds and are many times ad hoc.

Furthermore, it is well known that noise can seriously limit accuracy of time series prediction. In general, there are two main approaches for forecasting noisy time series. According to the first one, we ignore the presence of noise and fit a forecasting model directly from noisy data hoping to extract the underlying deterministic dynamics. According to the second approach, which is often more effective than the first one, we start with filtering the noisy time series in order to reduce the noise level and then forecast the new data points (see, for example, [3, 4] and references therein). There are several linear and nonlinear noise reduction methods such as ARMA model, local projective, singular value decomposition (SVD) and simple nonlinear filtering. It is currently accepted that SVD-based methods are very effective for the noise reduction in deterministic time series and correspondingly for forecasting [4].

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, it has been shown that most of the financial time series are nonlinear (see, for example, [3–6]); in these cases, we should use nonlin-

ear methods. Having a method that works well for both linear and nonlinear, stationary and non-stationary time series is ideal for modelling and forecasting. The Singular Spectrum Analysis (SSA) technique meets all conditions stated above. 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 [7]. Note also that SSA naturally incorporates the filtering of the series and the SVD.

It is worth mentioning that methods such as SSA that have been proposed outside economics and have been shown to be successful in forecasting, do make assumptions about the presence of deterministic dynamics (possibly buried in noise) that can be extracted and forecasted with relative accuracy. Such assumptions cannot really stand in the context of economic dynamics, which are driven by a variety of random factors and rarely have deterministic components in them (possibly apart from seasonality). However, this does not preclude from asking the question as to whether these methods can be successfully applied to economic and financial time series. To understand why SSA can be a potent tool for such a series, we need to understand the features that these series exhibit. The purpose of this paper is to review some of the recent developments on, the relatively novel, but powerful methodology of SSA, and discuss results associated with SSA from the perspective of analyzing economic and financial time series.

The rest of the paper is structured as follows: in section 2 we provide a literature review on SSA, its methodology and applications; in section 3 we go over the generic SSA methodology, univariate and multivariate, and list a variety of associated references; in section 4 we go over some modifications that apply to SSA when we have time series with stochastic trends and discuss the implications for modeling and forecasting and also a cointegration problem; in section 5 we consider SSA-based causality measures; the SSA technique based on the minimum variance estimator and based on the perturbation theory are considered in sections 6 and 7; the problem of signal extraction and smoothing are considered in section 8; in section 9 we present a variety of real world examples from forecasting point of view; finally, in section 10 we offer some concluding remarks and directions for future research.

2. LITERATURE REVIEW

The appearance of SSA is usually associated with the publication of papers by Broomhead and King [8]. A thorough description of the theoretical and practical foundations of the SSA technique (with many examples) can be found in [7, 9].

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. There are different ways of modifying SSA leading to different versions such as SSA with single and double centering, Toeplitz SSA, and sequential SSA [7].

Another important feature of SSA is that it can be used for analyzing relatively short series. It has been shown that SSA works very well for short time series as well as for long time series in forecasting macro-economics data [10].

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 extensions. First, the multivariate version of the method permits the simultaneous expansion of several time series; see, for example [9]. Second, the SSA ideas lead to several forecasting procedures for time series; see [7, 9]. The same ideas are also used in [7] and [11] for change-point detection in time series. For comparison with classical methods, ARIMA, ARAR algorithm and Holt-Winter, see [12] and [13]. The SSA has also been used for signal extraction and forecasting of the UK tourism income. The results showed that SSA outperforms SARIMA and time-varying parameter State Space models in terms of several forecasting accuracy criteria [14].

Two new versions of SSA, based on the minimum variance estimator and based on the perturbation theory, have been introduced in [15, 16]. These versions of SSA often yield more accurate reconstructing and forecasting results than the basic SSA.

In the area of nonlinear time series analysis, SSA was considered as a technique that could compete with more standard methods. The SSA technique has been used as a filtering method in [17]. A family of the causality test based on the multivariate SSA technique has been introduced in [18]. An important feature of the multivariate version of SSA is that it can be used for several time series with varied series length. For example, SSA has been used (univariate and multivariate) for forecasting the Index of Industrial Production which is one of the most widely scrutinised and intensively studied of all economic time series for the U.K [19]. The index is revised several times before a final figure is published which makes a set of the series with different series length. The SSA technique has also been used for forecasting exchange rate series [20, 21].

There is substantial previous literature that dealt with filtering and smoothing of non-stationary (including unit root)

processes in economics (and of course other fields) but its focus was that of trend ("signal") extraction and smoothing based on mainly cyclical (e.g. business cycles) considerations and was related to the extraction of components of certain frequencies. Thomakos [22, 24] has a line of research where he derives several new results of SSA-based analysis under the assumptions of time series that have a unit root and cointegration. His work expands that of Phillips [25–27] on smoothing and trend extraction for unit root processes. Thomakos [22, 24] derives asymptotically optimal linear filters for smoothing and trend extraction for unit root processes and their m-period differences and also suggests a method for selecting the degree of smoothing based on data considerations. Related work on smoothing economic time series has been considered in [28–32].

The SSA technique has also been used to model the realized volatility and logarithmic standard deviations of two important futures return series in [33]. Furthermore, SSA was used in forecasting price movement in Forex and predicting market behaviour in [34–36].

3. GENERIC SSA METHODOLOGY

3.1 Univariate SSA

The SSA technique is a decomposition-based approach and its usefulness lies in extracting information from the (auto)covariance structure of a time series. In the original formulation of SSA it was assumed that the time series under analysis has a deterministic component (such as a trend and/or a seasonal) with noise superimposed and that the deterministic component can be successfully extracted from the noise. This formulation is not, of course, confined to SSA; the decomposition-based approach to time series analysis is very old. What SSA brings into the picture, that identify it as a novel method, is that it accounts for the (auto)covariance structure of the time series without imposing a parametric model for it: it is thus a non-parametric, and from a practical perspective a model-free, approach for time series analysis.

The decomposition of the time series into components is followed by a step called reconstruction via the process of diagonal averaging. Both the decomposition and reconstruction/diagonal average steps have some optimality characteristics in terms of matrix operations. It is important to note that these characteristics are independent of the true data generating process. We next go over the generic SSA methodology for a single time series.

3.1.1 Stage 1: Reconstruction

Step 1: Embedding & the trajectory matrix

Consider a univariate stochastic processes $\{Y_t\}_{t\in\mathbb{Z}}$ and suppose that a realization of size N from this process is available $Y_N = [y_1, y_2, \ldots, y_N]$. The single parameter of the embedding is the *window length* L, an integer such that

$2 \leq L \leq N$. Embedding can be regarded as a mapping operation that transfers a one-dimensional time series Y_N into the multidimensional series $X_1, \ldots, X_K$ with vectors

$$(1) \qquad X_i = [y_i, y_{i+1}, x_{i+2}, \ldots, y_{i+L-1}]^T$$

for $i = 1, 2, \ldots, K$ where $K = N-L+1$. These vectors group together L time-adjacent observations and are supposed to describe the local state of the underlying process. Vectors X_i are called *L-lagged vectors* (or, simply, *lagged vectors*). The result of this step is the trajectory matrix

$$(2) \qquad \mathbf{X} = [X_1, \ldots, X_K] = (x_{ij})_{i,j=1}^{L,K}.$$

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.

Besides the application in SSA, the trajectory matrix can be used to unify a number of common time series procedures, such as filtering and autoregressive modeling. For example, let $\boldsymbol{\beta}$ denote any known, fixed $(L \times 1)$ vector and consider the following:

- For $L = 2$ and $\boldsymbol{\beta} = [-1, 1]$ we can obtain the first differences of the realization as $\boldsymbol{\beta}\mathbf{X}$.
- For any $L \geq 2$ and $\boldsymbol{\beta} = [1/L, 1/L, \ldots, 1/L]$ we can obtain a L-order moving average for the realization as $\boldsymbol{\beta}\mathbf{X}$.
- For autoregressive modeling let $\boldsymbol{\beta}^*$ denote the parameter vector and E denote the vector of innovations. Write $\mathbf{X}^T\boldsymbol{\beta}^* = E$ and define the $(L \times 1)$ and $(L \times L)$ restriction matrices:

$$(3) \qquad R = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \qquad \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{I}_{L-1} \\ \\ \boldsymbol{0}_{L-1}^T \end{bmatrix}$$

so that the restricted parameter vector $\boldsymbol{\beta}$ is written as $\boldsymbol{\beta}^* = R - \boldsymbol{Q}\boldsymbol{\beta}$. Then, the least-squares problem for estimating $\boldsymbol{\beta}$ is given by:

$$(4) \qquad \min_{\boldsymbol{\beta}} E^T E = (R - \boldsymbol{Q}\boldsymbol{\beta})^T \mathbf{X}\mathbf{X}^T (R - \boldsymbol{Q}\boldsymbol{\beta})$$

with solution $\widehat{\boldsymbol{\beta}} = (\boldsymbol{Q}^T\mathbf{X}\mathbf{X}^T\boldsymbol{Q})^{-1}\boldsymbol{Q}^T\mathbf{X}\mathbf{X}^T R$.

Step 2: Singular Value Decomposition (SVD)

The second step, the SVD step, makes the singular value decomposition of the trajectory matrix $\mathbf{X}$ and represents it as a sum of rank-one bi-orthogonal elementary matrices. Denote by $\lambda_1, \ldots, \lambda_L$ the eigenvalues of $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ in decreasing order of magnitude ($\lambda_1 \geq \cdots \lambda_L \geq 0$). 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:

$$(5) \qquad \mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d,$$

where $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ $(i = 1, \ldots, d)$. The matrices $\mathbf{X}_i$ have rank 1; therefore they are elementary matrices, U_i (in SSA literature they are called 'factor empirical orthogonal functions' or simply EOFs) and V_i (often called 'principal components') stand for the left and right eigenvectors of the trajectory matrix. The collection $(\sqrt{\lambda_i}, U_i, V_i)$ is called the i-th eigentriple of the matrix $\mathbf{X}$, $\sqrt{\lambda_i}$ $(i = 1, \ldots, d)$ are the singular values of the matrix $\mathbf{X}$ and the set $\{\sqrt{\lambda_i}\}_{i=1}^d$ is called the spectrum of the matrix $\mathbf{X}$. If all the eigenvalues have multiplicity one, then the expansion (5) is uniquely defined.

SVD (5) is optimal in the sense that among all the matrices $\mathbf{X}^{(r)}$ of rank $r < d$, the matrix $\sum_{i=1}^r X_i$ provides the best approximation to the trajectory matrix $\mathbf{X}$, so that $\|\mathbf{X} - \mathbf{X}^{(r)}\|$ is minimum. Note that $\|\mathbf{X}\|^2 = \sum_{i=1}^d \lambda_i$ and $\|\mathbf{X}_i\|^2 = \lambda_i$ for $i = 1, \ldots, d$. Thus, we can consider the ratio $\ell_i = \lambda_i / \sum_{i=1}^d \lambda_i$ as the characteristic of the contribution of the matrix $\mathbf{X}_i$ to expansion (5). Consequently, $\ell_{1:r} = \sum_{i=1}^r \lambda_i / \sum_{i=1}^d \lambda_i$, the sum of the first r ratios, is the characteristic of the optimal approximation of the trajectory matrix by the matrices of rank r.

3.1.2 Stage 2: Reconstruction

Step 1: Grouping

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

$$(6) \qquad \mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}.$$

The procedure of choosing the sets $I_1, \ldots, I_m$ is called the eigentriple grouping.

Step 2: Diagonal averaging

Diagonal averaging transfers each matrix $\mathbf{X}_I$ into a time series, which is an additive component of the initial series Y_N. If z_{ij} stands for an element of a matrix $\mathbf{Z}$, then the t-th term of the resulting series is obtained by averaging z_{ij} over all i, j such that $t = i + j - 1$. This procedure is called *diagonal averaging*, or Hankelization of the matrix $\mathbf{Z}$, $\mathcal{H}\mathbf{Z}$.

3.2 Forecasting

An important advantage of SSA is that it allows, upon reconstruction of the series under study, to produce forecasts for either the individual components of the series and/or the reconstructed series itself. This is useful if ones want to make predictions about, for example, the deterministic/trending component of the series without taking into account the variability due to other sources. Below we give a short description of the h-step ahead predictor based on the SSA method. The SSA can be applied in forecasting the time series that approximately satisfy linear recurrent formulae (LRF):

$$(7) \qquad y_{i+d} = \sum_{k=1}^d \alpha_k y_{i+d-k}, \qquad 1 \le i \le N - d$$

of some dimension d with the coefficients $\alpha_1, \ldots, \alpha_d$. It should be noted that $d = \text{rank }\mathbf{X} = \text{rank }\mathbf{XX}^T$ (the order of the SVD decomposition of the trajectory matrix $\mathbf{X}$). An important property of the SSA decomposition is that, if the original time series Y_N satisfies the LRF (7), then for any N and L there are at most d nonzero singular values in the SVD of the trajectory matrix $\mathbf{X}$; therefore, even if the window length L and $K = N - L + 1$ are larger than d, we only need at most d matrices $\mathbf{X}_i$ to reconstruct the series. Let us now formally describe the algorithm for the SSA forecasting method. The SSA forecasting algorithm, as proposed in [7], is as follows:

1. Consider a time series $Y_N = (y_1, \ldots, y_N)$ with length N.
2. Fix the window length L.
3. Consider the linear space $\mathfrak{L}_r \subset \mathbf{R}^L$ of dimension $r < L$. It is assumed that $e_L \notin \mathfrak{L}_r$, where $e_L = (0, 0, \ldots, 1) \in \mathbf{R}^L$.
4. Construct the trajectory matrix $\mathbf{X} = [X_1, \ldots, X_K]$ of the time series Y_N.
5. EOF step; construct vectors U_i $(i = 1, \ldots, r)$ form the SVD of $\mathbf{X}$. Note that U_i is an orthonormal basis in $\mathfrak{L}_r$.
6. Orthogonal projection step; estimate matrix $\widehat{\mathbf{X}} = [\widehat{X}_1 : \cdots : \widehat{X}_K] = \sum_{i=1}^r U_i U_i^T \mathbf{X}$. The vector $\widehat{X}_i$ is the orthogonal projection of X_i onto the space $\mathfrak{L}_r$.
7. Hankellization step; construct matrix $\widetilde{\mathbf{X}} = \mathcal{H}\widehat{\mathbf{X}} = [\widetilde{X}_1 : \cdots : \widetilde{X}_K]$ is the result of the Hankellization of the matrix $\widehat{\mathbf{X}}$.
8. Set $v^2 = \pi_1^2 + \cdots + \pi_r^2$, where π_i is the last component of the vector U_i $(i = 1, \ldots, r)$. For any vector $U \in \mathbf{R}^L$ denote by $U^\nabla \in \mathbf{R}^{L-1}$ the vector consisting of the first $L - 1$ components of the vector U. Moreover, assume that $e_L \notin \mathfrak{L}_r$. This implies that $\mathfrak{L}_r$ is not a vertical space. Therefore, $v^2 < 1$.
9. Determine vector $A = (\alpha_1, \ldots, \alpha_{L-1})$:

$$A = \frac{1}{1 - v^2} \sum_{i=1}^r \pi_i U_i^\nabla.$$

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 y_{L-1} components, i.e.

$$y_L = \alpha_1 y_{L-1} + \cdots + \alpha_{L-1} y_1,$$

and this does not depend on the choice of a basis $U_1, \ldots, U_r$ in the linear space $\mathfrak{L}_r$.

10. The h-step ahead forecasting procedure. In the above notations, define the time series $Y_{N+h} = (y_1, \ldots, y_{N+h})$ by the formula

$$(8) \quad y_i = \begin{cases} \widetilde{y}_i & \text{for } i = 1, \ldots, N \\ \sum_{j=1}^{L-1} \alpha_j y_{i-j} & \text{for } i = N+1, \ldots, N+h \end{cases}$$

where $\widetilde{y}_i$ $(i = 1, \ldots, N)$ are the reconstructed series. Thus, $y_{N+1}, \ldots, y_{N+h}$ form the h terms of the SSA recurrent forecast.

3.3 Multivariate singular spectrum analysis

The use of MSSA for multivariate time series was proposed theoretically in the context of nonlinear dynamics in [8]. There are numerous examples of successful application of MSSA (see, for example, [9, 18–20], and [39]). Multivariate (or multichannel) SSA is an extension of the standard SSA to the case of multivariate time series.

Assume that we have an M-variate time series $(y_j^{(1)}, \ldots, y_j^{(M)})$, where $j = 1, \ldots, N$ 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 $\mathbf{X} = (\mathbf{X}^{(1)} \ldots \mathbf{X}^{(M)})^T$. Now, it is straightforward to expand the univariate approach of SSA to the multivariate domain. The other stages of MSSA are similar to the stages of the univariate SSA; the main difference is in the structure of the trajectory matrix $\mathbf{X}$ and its approximation; these matrices are now block-Hankel rather than simply Hankel.

4. SSA & STOCHASTIC TRENDS

4.1 Theoretical results

As noted in the introduction, many economic and financial time series are assumed to have stochastic trends, i.e. non-stationary characteristics both in the mean and in the variance. The presence of such stochastic trends has long been acknowledged and there is a vast amount of literature that deals with the problems that their presence creates for econometric modeling. Here we are going to see what are the implications for SSA modeling and forecasting when there is such a stochastic trend presence in the series. See [22] for a complete presentation on this topic.

Consider then a stochastic process $\{y_t\}$ with a stochastic trend (or a unit root), that is:

$$(9) \qquad y_t = y_{t-1} + \eta_t \rightarrow y_t = \sum_{j=1}^{t} \eta_j$$

with $y_0 = 0$ and η_t a sequence of stationary and ergodic random variables, with mean zero and variance σ_η^2. A realization $Y_N = [y_1, y_2, \ldots, y_N]$ is again available for this process. What are the implications of the unit root assumption

for the resulting decomposition and minimum norm approximation? As in the related literature, all results given here are asymptotic, i.e. as $N \rightarrow \infty$.

It can be shown, see [22], that the matrix of sample (auto)covariances $\mathbf{S}$ converges to a simple stochastic matrix whose decomposition is known, but only when appropriately scaled. To see this consider the following. Assume for simplicity that η_t forms a sequence of $i.i.d.$ random variables, although all results continue to hold for η_t obeying mixing conditions. If we denote by $\widehat{M}(L)$ the sample cross-moment of order L, $\widehat{M}(L) = N^{-1} \sum_{t=L+1}^{N} y_t y_{t-L}$, we seek the asymptotic limit of $\widehat{M}(L)$. The required limit is obtained from known results as follows [23]. First, rewrite $\widehat{M}(L)$ as:

$$\widehat{M}(L) = \frac{1}{2N} \sum_{t=L+1}^{N} y_t^2 + \frac{1}{2N} \sum_{t=L+1}^{N} y_{t-L}^2 - \frac{1}{2N} \sum_{t=L+1}^{N} (y_t - y_{t-L})^2$$

and then note that both the first and second term on the right-hand side of the above equation converge when scaled by N to the following stochastic integral:

$$\frac{1}{2N^2} \sum_{t=L+1}^{N} y_t^2 + \frac{1}{2N^2} \sum_{t=L+1}^{N} y_{t-L}^2 \Rightarrow \sigma_\eta^2 \int_0^1 W(r)^2 dr$$

where $W(r)$ denotes standard Brownian motion and $\Rightarrow$ signifies weak convergence in the appropriate space. Since the last term on the right-hand side of the above equation converges in probability to a constant we have that, for all L,

$$N^{-1} \widehat{M}(L) \Rightarrow \sigma_\eta^2 \int_0^1 W(r)^2 dr.$$

If the stochastic process has an added drift then the above results change as follows. Let $y_t = \mu + y_{t-1} + \eta_t \rightarrow \mu t + \sum_{j=1}^{t} \eta_j$ with $\mu \neq 0$. As before expand $\widehat{M}(L)$ as:

$$\widehat{M}(L) = \frac{\mu^2}{N} \sum_{t=L+1}^{N} t^2 - \frac{\mu^2 L}{N} \sum_{t=L+1}^{N} t + \frac{\mu}{N} \sum_{t=L+1}^{N} t\varphi_{t-L}$$
$$+ \frac{\mu}{N} \sum_{t=L+1}^{N} t\varphi_t - \frac{\mu L}{N} \sum_{t=L+1}^{N} \varphi_t + \frac{1}{N} \sum_{t=L+1}^{N} \varphi_t \varphi_{t-L}$$

where $\varphi_t = \sum_{j=1}^{t} \eta_j$ and note that the leading term is of order $O(N^2)$. But upon scaling by N^2 all other terms vanish and we obtain that $N^{-2} \widehat{M}(L) \Rightarrow \mu^2/3$. We can summarize the above in the following propositions:

Proposition 1. *Under the assumptions of equation (9) we have that the matrix $N^{-2}\mathbf{S}$ of scaled-by-N^2 (auto)covariances $\widehat{M}(L)$, its eigenvalues and eigenvectors obey the following:*

1. $N^{-2}\mathbf{S} \Rightarrow \sigma_\eta^2 \boldsymbol{J}_{L,L} \int_0^1 W(r)^2 dr$, *where $W(r)$ is standard Brownian motion and $\boldsymbol{J}_{L,L}$ is a $(L \times L)$ matrix of ones.*

2. $\lambda_1 \Rightarrow \sigma_\eta^2 L \int_0^1 W(r)^2 dr$, $\lambda_j \Rightarrow 0$, for $2 \leq j \leq L$, and $\ell_1 = \ell_{1:1} \Rightarrow 1$.
3. $U_1 \Rightarrow J_L/\sqrt{L}$, where J_L is a $(L \times 1)$ vector of ones.

The theoretical implications of the above proposition are straightforward. If the series under study exhibits stochastic trends, in the form of a unit root, then SSA applied to it simplifies considerably as there is only one significant eigenvalue and the corresponding eigenvector takes on a very simple form.[1]

Proposition 2. *If the stochastic process of equation (9) contains a non-zero drift term $\mu \neq 0$ the the above results are modified as follows:*

1. $N^{-2}\mathbf{S} \Rightarrow \frac{\mu^2}{3} \mathbf{J}_{L,L}$
2. $\lambda_1 \Rightarrow L\mu^2/3$, $\lambda_j \Rightarrow 0$, for $2 \leq j \leq L$, and $\ell_1 = \ell_{1:1} \Rightarrow 1$.
3. $U_1 \Rightarrow J_L/\sqrt{L}$, where J_L is a $(L \times 1)$ vector of ones.
4. *Let $\widehat{\lambda}_1$ denote the estimator of the first eigenvalue of $N^{-2}\mathbf{S}$. It follows that a consistent estimator for μ is given by: $\widehat{\mu}_L = \sqrt{3\lambda_1/L} \Rightarrow \mu$*

Note that what is important for the application of the SSA-based filter is that the eigenvector corresponding to the leading eigenvalue is the same in both propositions, namely $U_1 = J_L/\sqrt{L}$. Since the application of the SSA-based filter depends only on the limit eigenvector, and not on the eigenvalue, we immediately ascertain that the presence of the drift is irrelevant for successful smoothing.

There are practical implications as well. The form that the leading eigenvector takes implies that we can obtain explicit expressions for the reconstructed series and examine the properties of the resulting weights. This is important since, as we will now illustrate, it provides an SSA-based justification for the use of moving averages in economic and financial data. In fact, after doing some algebra we can arrive at the following form for the reconstructed series:

$$
\widetilde{y}_s = \begin{cases} \dfrac{1}{sL}\displaystyle\sum_{j=1}^{s}\sum_{t=j}^{L+(j-1)} y_t, & s \leq L-1 \\[2ex] \dfrac{1}{L^2}\displaystyle\sum_{j=1}^{L}\sum_{t=s-L+j}^{s+(j-1)} y_t, & L \leq s \leq N-L+1 \\[2ex] \dfrac{1}{(N-s+1)L}\displaystyle\sum_{j=s}^{N}\sum_{t=j-L+1}^{j} y_t, & s > N-L+1 \end{cases}
$$

The reconstructed series takes the form of a symmetric moving average with weights that decline (increase) linearly from the center value of the average, the weights summing-up to one. In addition, it automatically takes care of the end of the series so that the first smoothed value is a forward moving average and the last smoothed value is a backward

moving average. For the middle part of the series, i.e. for $L \leq s \leq n-L+1$ we can show that we have:

$$
(10) \quad \widetilde{y}_s = \frac{1}{L^2}\sum_{j=-L+1}^{L-1}(L-|j|)y_{s+j}, \quad \text{for } L \leq s \leq N-L+1
$$

and, furthermore, this average is the solution to the local optimization problem:

$$
(11) \quad \widetilde{y}_s = \operatorname*{argmin}_{\mu_s}\sum_{j=-L+1}^{L+1} f_j(y_{s+j}-\mu_s)^2
$$

where f_j is the frequency of occurrence of each y_{s+j} in the average.

4.2 Connection with signal extraction problems

Many signal extraction problems are formulated as having two components: a signal in the form of a stochastic trend and an observable series given as the signal plus added noise. Clearly this framework can be handled by the SSA methodology described in the previous section and, moreover, one does not need to estimate any coefficients in doing so. This is important for economic and financial data that are plagued by structural breaks.

As an illustration, consider a simple example: the random walk plus noise model (also known as local level model). Let y_t denote the observable stochastic process and let x_t denote the unobservable signal, which has a unit root. They are assumed to be related by the following state space model:

$$
(12) \quad \begin{aligned} y_t &= x_t + \epsilon_t \\ x_t &= x_{t-1} + \eta_t \end{aligned}
$$

where $\epsilon_t \sim i.i.d.(0,\sigma_\epsilon^2)$ is the observational noise and where $\eta_t \sim i.i.d.(0,\sigma_\eta^2)$ is the signal noise, with ϵ_t independent of η_s for all (t,s). The properties of this model depend on the "signal-to-noise" ratio $q = \sigma_\eta^2/\sigma_\epsilon^2$: as $q \to 0$ the signal is buried in noise and is difficult to recover; as $q \to \infty$ the model collapses to the standard unit root model of equation (9). Accurate MSE extraction of the signal component requires estimation (via the Kalman filter) of the two variance parameters and then fixed point smoothing. Is the method proposed in this paper capable of separating the signal from the noise in this set-up? To examine this, let us construct the relevant trajectory matrices and the corresponding autocovariance matrices. Denoting the trajectory matrices in standard fashion as $\boldsymbol{X}_y$, $\boldsymbol{X}_x$ and $\boldsymbol{X}_\epsilon$ and the corresponding matrices of sample (auto)covariances as $\mathbf{S}_j$ for $j = y, x, \epsilon$ we immediately obtain:

$$
(13) \quad \begin{aligned} \boldsymbol{X}_y &= \boldsymbol{X}_x + \boldsymbol{X}_\epsilon \\ \mathbf{S}_y &= \mathbf{S}_x + \mathbf{S}_\epsilon + \mathbf{S}_{x,\epsilon} + \mathbf{S}_{\epsilon,x} \end{aligned}
$$

---

[1]Note that this leading eigenvector is the same one appearing in many types of orthogonal transforms, such as the discrete cosine transform, see [40].

where $\mathbf{S}_{x,\epsilon}, \mathbf{S}_{\epsilon,x}$ are the cross-covariances between the signal and the noise trajectories. Under the rest of the assumptions of equation (12), and using standard results (see [22]), we have that asymptotically $N^{-1}\mathbf{S}_\epsilon \Rightarrow \sigma_\epsilon^2 \mathbf{I}_L$ and $N^{-1}\mathbf{S}_{\epsilon,x}(L) \Rightarrow \chi$, where χ is a stochastic matrix. However, since $\mathbf{S}_x$ does not converge unless scaled by N^{-2} we end up having $N^{-2}\mathbf{S}_\epsilon \Rightarrow \mathbf{0}_{L,L}$, $N^{-2}\mathbf{S}_{x,\epsilon} \Rightarrow \mathbf{0}_{L,L}$ and therefore:

$$(14) \qquad N^{-2}\mathbf{S}_y \approx N^{-2}\mathbf{S}_x \Rightarrow \sigma_\eta^2 \mathbf{J}_{L,L} \int_0^1 W(r)^2 dr$$

exactly as in Proposition 1. This result is of practical significance for the case of a unit root process which is contaminated with noise. Using SSA, we can extract the underlying non-stationary signal directly, at least asymptotically. Combining previous results on stationary SSA with the results from the previous sections we can also select L, the degree of smoothing appropriately: as $q \to 0$ then $L \to N/2$ with $N \to \infty$; as $q \to \infty$ then $L = o(N)$, e.g. $L = \sqrt{N}$.

For comparison with the SSA approach, we reproduce below McElroy's [41] matrix-based formulas for Kalman fixed point smoothing for the local level model. Letting $\boldsymbol{\Delta}$ denote the $(N \times N-1)$ matrix with -1 on its principal diagonal and 1 in its first lower diagonal and $Y_N = [y_1, y_2, \ldots, y_N]$ denote the $(1 \times N)$ vector of observations we have that the $(N \times N)$ matrix of optimal smoothing coefficients is given by:

$$(15) \qquad \boldsymbol{F}_N(\sigma_\epsilon^2, \sigma_\eta^2) \equiv \boldsymbol{F}_N = (\boldsymbol{\Delta}\boldsymbol{\Delta}^T q + \boldsymbol{I}_N)^{-1}$$

so that the optimal MSE signal estimate is given by $\widehat{Y}_N = \mathsf{E}[X|Y] = Y_N \boldsymbol{F}_N$.

Figure 1 illustrates the above using a sample realization from equation (12) with both ϵ_t and η_t being normally distributed and $q = 1\%$, the noise variance being 100 times greater than the signal variance, and with $L = \sqrt{N}$ as before. The lower panel of the figure shows the true signal and the two smoothed series, the one based on SSA and the other on the application of fixed point smoothing (with the parameters estimated). It is clear from the figure that the non-parametric SSA smoother performs on par with the parametric fixed point smoother. We further explore the performance of the proposed methodology in the context of signal extraction in the simulation section.

In the signal extraction framework one can accommodate a comparison between the SSA method and the Hodrick-Prescott [28] HP-filter that is used frequently in trend extraction and smoothing in economics. It can be shown, see for example [42] and [43], that the HP filter can be derived from a signal extraction model similar to equation (12), but the signal process y_t has two instead of one unit roots, i.e. $(1 - B)^2 y_t = \eta_t$, with B being the backshift operator. The filtered values can be computed using exactly the same formula as in equation (15) before, with $\boldsymbol{\Delta}$ being defined as $(N \times N - 2)$ with -2 on its principal diagonal and 1's on the first upper and lower diagonals. Several methods have been proposed in [42] and [43] to consistently estimate σ_η^2, σ_ϵ^2 and q from the data.
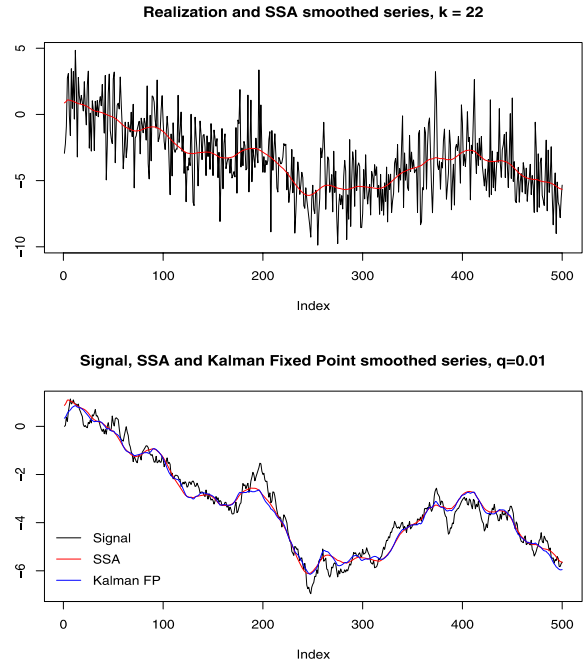

Figure 1. Sample realization from a local level model, signal and smooth series.

4.3 Cointegration & other developments

A natural extension of the material in the previous sections is to consider multiple series with stochastic trends and perform MSSA on them. This has not been extensively considered, although [22] has provided a thorough treatment of the bivariate case that includes SSA-based estimation of the cointegration coefficient and SSA-based smoothing of the common stochastic trends. There is, however, room for additional work on this topic using MSSA techniques.

Furthermore, one can define and use a number of related predictors for the reconstructed component of the leading eigenvalue, for both the univariate case and the multivariate case. For the univariate case in particular, it can be shown that the predictor of the smoothed reconstructed component coincides with a simple backward moving average — this is useful as it fully justifies (on SSA grounds) the use of moving averages as trend predictors in financial time series.

Finally, in related work by Thomakos [24] the problem of SSA-based smoothing of the m-period differences from a process with a stochastic trend is studied extensively. That is, one considers the differences $y_t - y_{t-m}$ of the model in equation (9). This is a non-trivial exercise with theoretical and practical significance. In the theoretical front it is shown that SSA applied to these differences coincides with a version of the discrete cosine transform. In the practical front, we have that such differences appear all the time in empirical analysis of economic and financial time series: for example, annual inflation is defined as the 12-month difference of the (log of) monthly prices; or, consider overlapping

multiperiod returns of a financial asset that are obtained by such differencing as monthly (20-day) returns obtained by differencing of daily prices.

5. SSA-BASED CAUSALITY TESTS

A question that frequently arises in time series analysis is whether one economic variable can help in predicting another economic variable. One way to address this question was proposed by Granger in [44] where he formalized the 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 [45]). 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 a causality in the broad meaning of the word. It just considers linear prediction and time-lagged dependence between two time series. The definition of Granger causality does not include a possibility of an instantaneous dependence between the two series. It is not rare when instantaneous dependence between two time series can be easily revealed, but since the causality can go either way, one usually does not test for instantaneous dependence. In this paper, we introduced several criteria based on the SSA technique to test for the instantaneous causality.

The general aim of the SSA-based causality tests 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 technique. For similarity and dissimilarity between SSA based tests and Granger Causality tests, and empirical results see [18] and [46].

5.1 Causality criteria based on forecasting accuracy

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 by the univariate procedure, SSA, and by the multivariate one, MSSA. If the forecasting errors using MSSA are significantly smaller than the forecasting errors 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 $X_N = (x_1, \ldots, x_N)$ into two separate subseries X_R and X_F: $X_T = (X_R, X_F)$, where $X_R = (x_1, \ldots, x_R)$ and $X_F = (x_{R+1}, \ldots, x_N)$. The subseries X_R is used in the 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 with the help of the recursive h-step ahead forecast with SSA and MSSA. The forecasted points $\hat{X}_F = (\hat{x}_{R+1}, \ldots, \hat{x}_N)$ are then used for computing the forecasting error, and the vector $(x_{R+2}, \ldots, x_N)$ 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 SSA. Therefore, the two vectors 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.

Let $X_T = (x_1, \ldots, x_N)$ and $Y_T = (y_1, \ldots, y_N)$ denote two different time series of length N. Set the window lengths L common for both series. Using the embedding terminology, we construct the trajectory matrices $\mathbf{X} = [X_1, \ldots, X_K]$ and $\mathbf{Y} = [Y_1, \ldots, Y_K]$ for the series X_N and Y_N.

Consider an arbitrary loss function $\mathcal{L}$. In econometrics, the loss function $\mathcal{L}$ is usually selected as the mean square error. Let us first assume that the aim is to forecast the series X_T h-step ahead. Thus, the aim is to minimize $\mathcal{L}(X_{K+h} - \hat{X}_{K+h})$, where the vector $\hat{X}_{K+h}$ is an estimate, obtained using a forecasting algorithm, of the vector $X_{K+h} = (x_{K+h}, \ldots, x_{T+h})$. The vector X_{K+h} can be forecasted using either univariate SSA or MSSA. Let us first consider the univariate approach. Define $\Delta_{X_{K+h}} \equiv \mathcal{L}(X_{K+h} - \hat{X}_{K+h})$, where $\hat{X}_{K+h}$ is obtained using univariate SSA; that is, the estimate $\hat{X}_{K+h}$ is obtained only from the vectors $[X_1, \ldots, X_K]$.

Let $X_N = (x_1, \ldots, x_N)$ and $Y_{N+d} = (y_1, \ldots, y_{N+d})$ denote two different time series to be considered simultaneously and consider the same window length L for both series. Here d is an integer, not necessary non-negative. Now, we forecast $x_{N+1}, \ldots, x_{N+h}$ using the information provided by the series Y_{N+d} and X_N. Next, compute the statistic $\Delta_{X_{K+h}|Y_{K+d}} \equiv \mathcal{L}(X_{K+h} - \tilde{X}_{K+h})$, where $\tilde{X}_{K+h}$ is an estimate of X_{K+h} obtained using MSSA. This means that we simultaneously use vectors $[X_1, \ldots, X_K]$ and $[Y_1, \ldots, Y_{K+d}]$ in forecasting the vector X_{K+h}. Now, define the criterion $F_{X|Y}^{(h,d)} = \Delta_{X_{K+h}|Y_{K+d}}/\Delta_{X_{K+h}}$ corresponding to the h step ahead forecast of the series X_T in the presence of the series Y_{T+d}. Note that d is any given integer (even negative). For example, $F_{X|Y}^{(h,0)}$ means that $d = 0$ and that we use the series X_N and Y_N simultaneously (with zero lag). The criterion $F_{X|Y}^{(h,0)}$ can be used in evaluating the instantaneous causality.

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 that there is 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 SSA is better than of the MSSA (this may happen, for example, when the series Y has structural breaks misdirecting the forecasts of X).

To assess 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_{N+d} (or X_{N+d}) supports X_N (or Y_N) in h-step forecasting. If $F_{Y|X}^{(h,d)} < F_{X|Y}^{(h,d)}$, we then conclude that X is more supportive to Y than Y to X. Otherwise, if $F_{X|Y}^{(h,d)} < F_{Y|X}^{(h,d)}$, we conclude that Y is more supportive to X than X to Y.

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 between series X and Y. We shall call it F-feedback (forecasting feedback) which means that the use of a multivariate system improves the forecasting for both series. For a F-feedback system, X and Y are mutually supportive. To check if the discrepancy between the two forecasting procedures are statistically significant, one can employ the test performed in [18] and [46].

5.2 Direction of change based criteria

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_N (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: $D_{X|Y}^{(h,d)} =$

$\bar{Z}_X / \bar{Z}_{X|Y}$, 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. 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 criterion, we can define a feedback system based on the criterion characterizing the predictability of the direction of change. To test the significance of the values of $D_{X|Y}^{(h,d)}$, we use the test developed in [18] and [46].

6. SSA BASED ON THE MINIMUM VARIANCE ESTIMATOR

The SSA algorithms that have been considered in literature are based on the standard SVD and the least squares (LS) estimate (see, for example, [7] and references therein). The LS estimator projects the noisy time series onto the perturbed signal (noise + signal) subspace. Therefore, the reconstructed series using LS estimator has the lowest possible (zero) signal distortion and the highest possible residual noise level. Moreover, the disadvantage of LS is that the performance of the LS estimator is crucially dependent on the estimation of the signal rank r. That is, selecting singular values in LS is a binary approach. To overcome this problem, Hassani [15] considered an alternative method which is based on the minimum variance (MV) estimator. The MV estimator is the optimal linear estimator, which gives the minimum total residual power [47, 48].

Consider a noisy vector Y_N of length N. Consider again the signal-plus-noise series without any assumptions about the nature of the signal:

$$(16) \qquad Y_N = X_N + E_N;$$

here X_N represents the signal component and E_N the noise component and, as before, we can write the corresponding trajectory matrices as:

$$(17) \qquad \mathbf{S}_y = \mathbf{S}_x + \mathbf{S}_e,$$

The SVD of the matrix $\mathbf{S}_y$ can be written as:

$$(18) \qquad \mathbf{S}_y = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = [\mathbf{U}_x \ \ \mathbf{U}_e] \begin{bmatrix} \boldsymbol{\Sigma}_x & 0 \\ 0 & \boldsymbol{\Sigma}_e \end{bmatrix} \begin{bmatrix} \mathbf{V}_x^T \\ \mathbf{V}_e^T \end{bmatrix}$$

where $\mathbf{U}_x \in \mathbb{R}^{L \times r}$, $\boldsymbol{\Sigma}_x \in \mathbb{R}^{r \times r}$ and $\mathbf{V}_x \in \mathbb{R}^{K \times r}$. We can also represent SVD of the Hankel matrix of the signal $\mathbf{S}_x$ as:

$$(19) \qquad \mathbf{S}_x = [\mathbf{U}_x \;\; \mathbf{U}_x] \begin{bmatrix} \boldsymbol{\Sigma}_x & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}_x^T \\ \mathbf{V}_x^T \end{bmatrix}$$

It is clear that the Hankel matrix $\mathbf{S}_x$ can not be reconstructed exactly if it is perturbed by noise. To remove the effect of the noise term, we assume that the vector space of the noisy time series (signal) can be split in mutually orthogonal noise and signal + noise subspaces. The components in the noise subspace are suppressed or even removed completely. Therefore, one can reconstruct the noise free series from signal+noise subspace by choosing the weight. Thus, by adapting the weights of the different singular components, an estimate of the Hankel matrix $\mathbf{S}_x$, which corresponds to a noise-reduced series, can be achieved:

$$(20) \qquad \mathbf{S}_x = \mathbf{U}(\mathbf{W}\boldsymbol{\Sigma})\mathbf{V}^T,$$

where $\mathbf{W}$ is the diagonal matrix containing the weights. Now, the problem is choosing the weight matrix $\mathbf{W}$.

Let us now consider the weight matrix $\mathbf{W}$ based on the LS and MV estimates. The LS and MV estimates can be defined based on the weight matrix $\mathbf{W}_{r \times r}$ as follows:

$$(21) \qquad \begin{aligned} \widehat{\mathbf{S}}_{xLS} &= \mathbf{U}_x(\mathbf{W}_{LS}\boldsymbol{\Sigma}_x)\mathbf{V}_x^T \\ \widehat{\mathbf{S}}_{xMV} &= \mathbf{U}_x(\mathbf{W}_{MV}\boldsymbol{\Sigma}_x)\mathbf{V}_x^T \end{aligned}$$

where

$$(22) \qquad \begin{aligned} \mathbf{W}_{LS} &= \mathbf{I}_{r \times r} \\ \mathbf{W}_{MV} &= \mathrm{diag}\left(\left(1 - \frac{\sigma_{noise}^2}{\lambda_1^2}\right), \ldots, \left(1 - \frac{\sigma_{noise}^2}{\lambda_r^2}\right) \right) \end{aligned}$$

7. SSA BASED ON THE PERTURBATION THEORY

As appears from LS and MV estimate, the singular values are different, but the left and right singular vector of the LS and MV estimates are the same and both contain noise term. It is clear that we can obtain a better reconstructed series if we remove the noise term from these matrices. Let us define the following matrices:

$$(23) \qquad \begin{aligned} \tilde{\mathbf{U}}_x &= \mathbf{U}_x + \delta\mathbf{U}_x \\ \tilde{\mathbf{V}}_x &= \mathbf{V}_x + \delta\mathbf{V}_x \\ \tilde{\boldsymbol{\Sigma}}_x &= \boldsymbol{\Sigma}_x + \delta\boldsymbol{\Sigma}_x \end{aligned}$$

Therefore, the ideal situation is that we remove the noise terms $\delta\mathbf{U}_x$, $\delta\boldsymbol{\Sigma}_x$, and $\delta\mathbf{V}_x$. Note that in minimum variance estimator, we tried to remove $\delta\boldsymbol{\Sigma}_x$, and in basic SSA we kept all these noisy terms in SVD expansion. An approximation of $\mathbf{U}_x$, $\boldsymbol{\Sigma}_x$ and $\mathbf{V}_x$, up to second order perturbation

Table 1. Matrices $\mathbf{G}_U$, $\mathbf{G}_\Sigma$ and $\mathbf{G}_V$ for different estimator

Estimation Method	$\mathbf{G}_U$	$\mathbf{G}_\Sigma$	$\mathbf{G}_V$
Least Square	$\mathbf{I}$	$\mathbf{I}$	$\mathbf{I}$
Minimum Variance	$\mathbf{I}$	$\mathbf{W}_{MV}$	$\mathbf{I}$
Perturbation Theory	$\mathbf{P}_U$	$\mathbf{P}_\Sigma$	$\mathbf{P}_V$

theorem, has been introduced in [16]. We do not provide the mathematical formula here and refer the interested reader to [16]. In general, an estimation of signal matrix can be represented as follows:

$$(24) \qquad \widehat{\mathbf{S}}_x = (\mathbf{G}_U\tilde{\mathbf{U}}_x)(\mathbf{G}_\Sigma\tilde{\boldsymbol{\Sigma}}_x)(\mathbf{G}_V\tilde{\mathbf{V}}_x^T)$$

where matrices $\mathbf{G}_U$, $\mathbf{G}_\Sigma$ and $\mathbf{G}_V$, based on the least square, minimum variance and perturbation theory, are represented in Table 1 where $\mathbf{I}$ is an identity matrix, $\mathbf{W}_{MV}$ was introduced in (22), and matrices $\mathbf{P}_U$, $\mathbf{P}_\Sigma$ and $\mathbf{P}_V$ are obtained up to second order perturbation theory.

Let us now examine the capability of the SSA technique based on the perturbation theory (SSA$_{PT}$), in reconstructing and forecasting simple simulated sinusoidal series:

$$(25) \qquad \begin{aligned} S012 &= \beta_0 + \beta_1 Y_1 + \beta_2 Y_2 + \beta_3 Y_3 + \epsilon_t \\ S01 &= \beta_0 + \beta_1 Y_1 + \beta_2 Y_2 + \epsilon_t \\ S1 &= \beta_1 Y_1 + \beta_2 Y_2 + \epsilon_t \end{aligned}$$

where $Y_1 = Sin(2t\pi/12)$, $Y_2 = Sin(2t\pi/7)$, $Y_3 = Sin(2t\pi/5)$, and ϵ_t is a white noise series. In total 300 data were generated and we added different normally distributed noise to each point of the original series. The simulation was repeated 1,000 times. Note that usually every harmonic component with a different frequency produces two eigentriples with close singular values (except for frequency 0.5 which provides one eigentriples with saw-tooth singular vector). For example, one needs to select the first five eigenvalues for reconstruction of the series $S012$, and the first three for the series $S01$. It should be noted that we need to consider one eigentriple for the intercept, which is the first one in this particular example. To calculate the precision we use the ratio of Root Mean Square Error (RRMSE): Note that, if RRMSE < 1, then SSA$_{PT}$ procedure outperforms alternative prediction method. Alternatively, RRMSE > 1 would indicate that the performance of the corresponding SSA$_{PT}$ procedure is worse than the predictions of the competing method.

Let us now consider the effect of noise reduction with respect to different window length L which is the single parameter in decomposition stage. Certainly, the improper choice of L would imply an inferior decomposition. It should be noted that variations in L may influence the separability features when applying SSA, specifically the orthogonality and
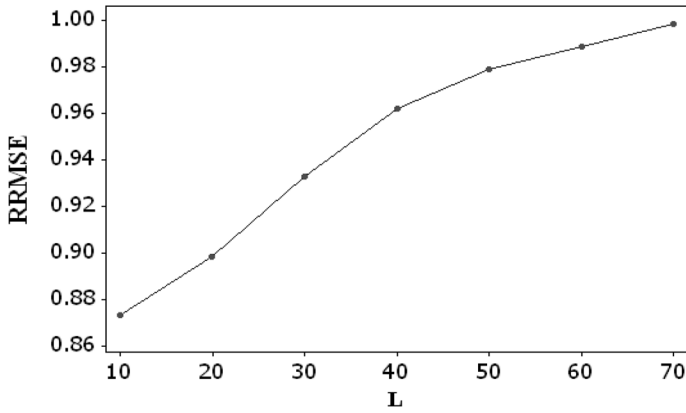
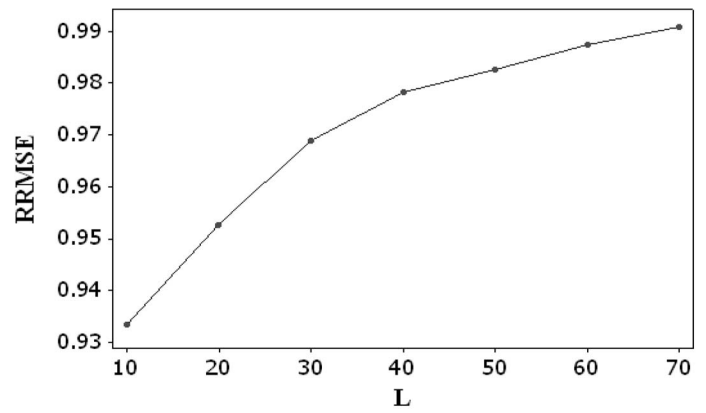Figure 2. The value of RRMSE in reconstructing of noisy series $S012$ for different window length.

Figure 4. The value of RRMSE in reconstructing of noisy series $S1$ for different window length.
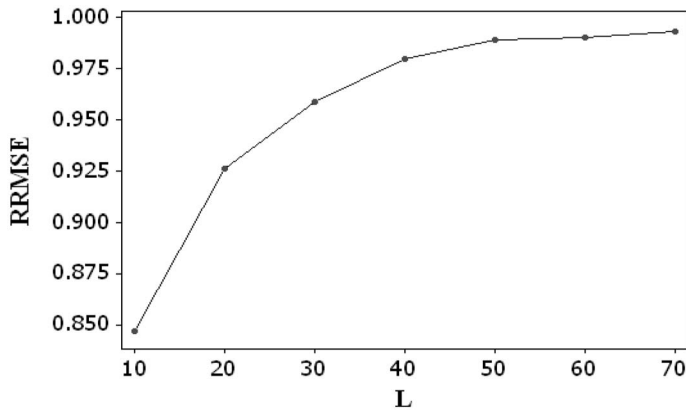

Figure 3. The value of RRMSE in reconstructing of noisy series $S01$ for different window length.
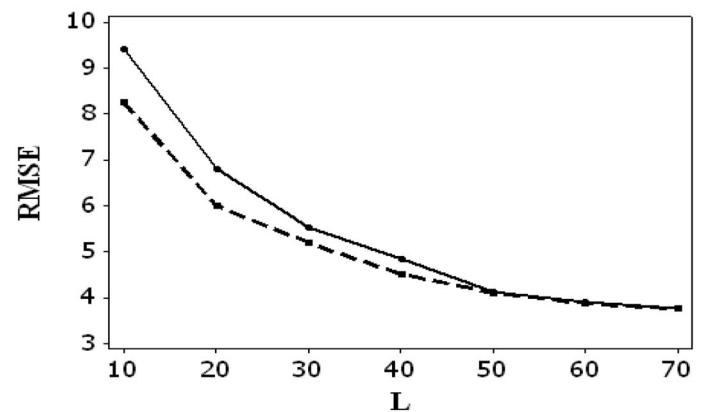

Figure 5. The value of RMSE in reconstructing of noisy sin for different window length using SSA_{PT} (dashed line) and SSA_{LS} (thick line).

closeness of the singular values. Here we consider L between 10 and 70 which is approximately $N/3$ (here $N = 200$).

Figures 2–4 show the RRMSE of reconstructed series for different simulated series. As it appears from these figures, SSA_{PT} has a better performance in reconstruction noisy series, particulary for small window length. The performance of both methods are similar for a large window length.

As the figures show, RRMSE tends to 1 as the window length increases confirming that both methods have similar performance for a large window length. The graphs also show that there is a gradual increase in RRMSE with window length. For example for window length 10, the performance of SSA_{PT} is up to 15% better than SSA_{LS} in reconstruction noisy series $S01$. However, there is not a significant discrepancy between the performance of SSA_{PT} and SSA_{LS} for window length greater than 50.

Note that the minimum value of RMSE for both SSA_{PT} and SSA_{LS} occurs for a large window length. Let us, for example, consider the RMSE of SSA_{PT} and SSA_{LS} in reconstructing $S012$ in more detail. Figure 5 shows the RMSE

of SSA_{PT} and SSA_{LS}. As it can be seen from the figure, there is a gradual decrease in RMSE with window length. In fact, the maximum accuracy in reconstruction, using both methods, occurs for a large window length. The figure also shows that the RMSE of SSA_{PT} is smaller than those obtained using SSA_{LS}. Moreover, the figure indicates that the discrepancy between SSA_{PT} and SSA_{LS} reduces as window length increases.

8. SMOOTHING & SIGNAL EXTRACTION

8.1 Simulations results

In this section we present some simulation results on the performance of the SSA-based smoother in the context of the signal extraction problem. The data generating process is the local level model of equation (12) with different values for the signal-to-noise ratio q and different assumptions on the distribution of η_t. For a sample size of $N = 250$ observa-

Table 2. Dispersion of residuals from true signal, $N = 250$, $\epsilon_t \sim \mathcal{K}(0, \sigma_\epsilon^2)$ and $\eta_t \sim \mathcal{K}(0, \sigma_\eta^2)$

		Average Mean Squared Error				
q	KFP	SSA-$\sqrt{N}$	PP-$\sqrt{N}$	ES	SSA-O	PP-O
0.01	0.22	0.25	0.31	0.50	0.80	0.32
0.10	0.06	0.08	0.09	0.15	0.09	0.18
1.00	0.02	0.07	0.07	0.06	0.03	0.17
		Average Mean Absolute Error				
q	KFP	SSA-$\sqrt{N}$	PP-$\sqrt{N}$	ES	SSA-O	PP-O
0.01	0.38	0.40	0.44	0.55	0.68	0.45
0.10	0.20	0.22	0.24	0.30	0.24	0.33
1.00	0.11	0.20	0.21	0.20	0.14	0.31

Notes to Table 2. The column nomenclature is as follows: KFP denotes Kalman fixed point smoothing; SSA-$\sqrt{N}$ and PP-$\sqrt{N}$ denote SSA and Phillips' smoothing with L fixed to $\sqrt{N}$; SSA-O and PP-O denote SSA and Phillips' smoothing with L^* selected via equation proposed in [22].

Table 3. Selected degree of smoothing L^*

	N=250		N=500	
q	SSA-O	PP-O	SSA-O	PP-O
0.01	6	7	7	9
0.10	6	7	8	10
1.00	6	8	8	11

Note to Table 3. For column nomenclature see note of Table 2.

tions, we perform a signal extraction using: (a) the Kalman fixed point smoother; (b) the SSA approach of this paper with fixed L and L^{*2}; (c) the approach of Phillips (2005) [27] with fixed L and L^*; and (d) exponential smoothing. For (a) and (d) the relevant parameters are estimated from the data. There are 500 replications and reported are average mean-squared and mean-absolute errors for the residuals from the true signal. Specifically, if $\widetilde{y}_{t,ij}$ denotes the signal estimate for the ith replication for the jth method the tables below report:

$$
(26) \quad
\begin{aligned}
AMSE_j &= \frac{1}{500} \sum_{i=1}^{500} \left[\frac{1}{N} \sum_{t=1}^{N} (y_{t,i} - \widehat{y}_{t,ij})^2 \right], \\
AMAE_j &= \frac{1}{500} \sum_{i=1}^{500} \left[\frac{1}{N} \sum_{t=1}^{N} |y_{t,i} - \widehat{y}_{t,ij}| \right]
\end{aligned}
$$

For the same simulations the average value of the selected L^* using the methodology proposed in [22]. These results are summarized in Tables 2–4.

The results in Table 2 are extremely encouraging. The fixed $L = \sqrt{N}$ filter is very competitive to the Kalman fixed point filter, which is optimal for the data generating process. For the AMSE on the top panel of the table we can observe that as q increases the performance of the optimal

---

[2] L^* represents an 'optimally' selected value of the threshold based on the methodology in [38] where the degree of smoothing is selected so as to minimize the absolute value of the expected sign of the residuals from smoothing.

Table 4. Dispersion of residuals from true signal, $N = 250$, $\epsilon_t \sim \mathcal{K}(0, \sigma_\epsilon^2)$ and $\eta_t \sim t_{(2)}\sigma_\eta$

		Mean Squared Error				
q	KFP	SSA-$\sqrt{N}$	PP-$\sqrt{N}$	ES	SSA-O	PP-O
0.01	0.58	0.93	1.13	1.50	1.03	2.18
0.10	0.16	0.63	0.68	0.61	0.31	1.40
1.00	0.03	0.68	0.75	0.47	0.24	1.59
		Mean Absolute Error				
q	KFP	SSA-$\sqrt{N}$	PP-$\sqrt{N}$	ES	SSA-O	PP-O
0.01	0.57	0.61	0.68	0.83	0.74	0.95
0.10	0.30	0.49	0.53	0.49	0.35	0.78
1.00	0.14	0.49	0.53	0.33	0.26	0.81

Note to Table 4. Column nomenclature and table entries as in Table 2.

L^* filter also becomes very competitive to the Kalman filter, eventually matching it. For the AMAE on the bottom panel of the table we see a similar situation as with the AMSE results. Overall we can say that (a) the fixed L SSA-based filter of this paper can be used reliably for this type of data generating process; and (b) the optimal L^* SSA-based filter is to be preferred when the signal-to-noise ratio is relatively large. Phillip's smoother is trailing after the SSA smoother, its performance being slightly worse. The exponential smoother performs well only for $q = 1$, as expected. Finally, the results in Table 3 summarize the average L^* that was selected across replications.

In Table 4 we have a variation of the results in Table 2. Here the signal noise is generated by a Student $t_{(2)}$ distribution scaled by a factor σ_η. This distribution does not have finite second moments so the meaning of q is not the same as before. Nevertheless it is instructive to look at the performance of the various filters. We can see that the performance of almost all filters deviates more substantially than before from the performance of the Kalman fixed point filter and their differences do not significantly diminish as q increases (this was expected since for any value of q the signal has infinite variance). It is only the optimal L^* SSA-based filter that can come close to the Kalman fixed point filter and its performance improves as q increases.

8.2 Real data set: U.S. GDP

As an application, adapted from Thomakos [22], we compare the SSA-based smoother with the Hodrick-Prescott (HP) filter for the series of quarterly observations of the U.S. real GDP (series GDPC96 from the Federal Reserve Bank of St. Louis Database, in billions of chained 2,000 dollars). The series has $N = 244$ quarterly observations from 1947:Q1 to 2007:Q4. The HP filter has been used widely in smoothing trending economic time series and the original paper [28] was followed by much literature on optimal filtering. Here we show that the SSA-based filter of this paper performs on par with the HP filter, either when a naive value is used for the smoothing parameter q or when the value of the smoothing parameter is optimized. Consider first the
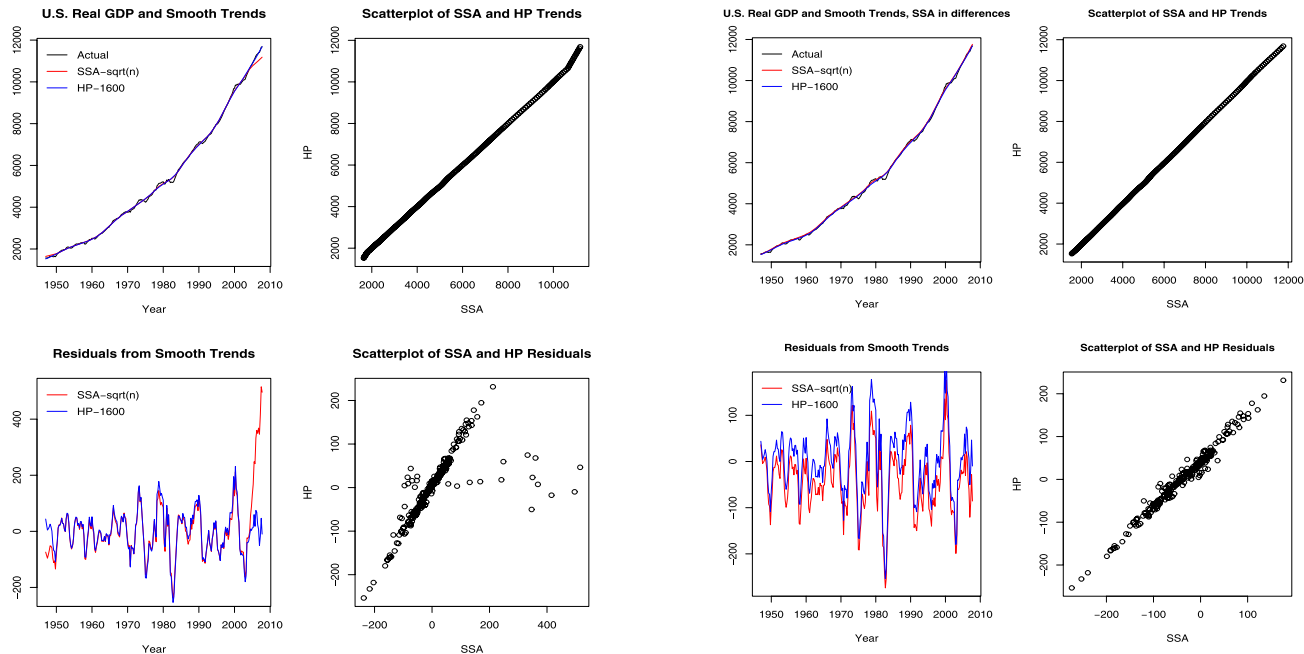
Figure 6. SSA and HP Smoothing of the U.S. Real GDP - smoothing parameters are fixed to $L = \left[\sqrt{244}\right] = 15$ and $q = 1,600$ respectively.

Figure 7. SSA and HP Smoothing of the U.S. Real GDP, SSA filter applied in differences - smoothing parameters are fixed to $L = \left[\sqrt{244}\right] = 15$ and $q = 1,600$ respectively.

case where $q = 1,600$, the value suggested by Hodrick and Prescott in their paper [28] for quarterly data, and compare it with the SSA-based filter with L fixed to $L = \sqrt{N}$.

Figure 6 contains the results. There are several things to notice: first, see that both filters achieve practically the same degree of smoothing over the entire length of the series and produce practically identical residuals, with the exception of the end of the series. This is due to the construction of the two filters: in the SSA case the last value of the smoothed trend is an unweighted backward moving average. This can be seen as a shortcoming of the SSA-based filter which brings us to a question: are the two filters, as applied, comparable? The answer is clearly no!

As noted before, the underlying stochastic model on which the HP filter is based, is for a time series with two not one unit root. To make a meaningful comparison, we need to apply the SSA filter in the first differences of the GDP series and then cumulate the extracted signal. The results from this approach are presented in Figure 7.

One can now make a meaningful comparison between the two filters and we can see that they now match everywhere, both in the smooth trends and the corresponding residuals; see the differences in the scatterplots between Figure 6 and Figure 7. We therefore see that the proposed SSA-based filter achieves the same degree of smoothness as the standard HP filter, after taking into account the stochastic model under which the HP filter operates. All the above analysis was repeated with both L and q (in HP filter) selected in an optimal fashion but the results were qualitatively similar to

the ones already presented. For a more in-depth analysis on this series using methods presented in this paper see [22].

We next turn to the modeling of the annual growth rate of this series. The annual growth rate of real GDP is an important measure of economic strength and there is much interest in analyzing its properties. The actual and smoothed series are given in Figure 8. We can see that the smoothed series tracks annual growth all along except for the sharpest peaks and troughs, explaining over 90% of the variability of annual growth. Also given in the figure are the residuals from the smooth trend and the spectrum of the smooth trend. The cyclical features of growth can be seen both in the series itself and in the residuals from the smooth trend; in the latter the reduction in the volatility of growth during the 1990's is also evident.

In the bottom panel of Figure 9 we can see the spectral density of the smooth trend of the series. There are two clear power peaks at frequencies that correspond to periods of 20 and 9.2 quarters, i.e. 5 and about 2.3 years respectively. It also interesting to examine the shape and power properties of the two components of the smooth trend. These are given in Figure 9. In the top panel we can see the components of the smooth trend and in the bottom panel we can see the corresponding spectral densities. As expected, the first component is the slow moving one and has much of its power concentrated at the lower frequency of 5 years; the second component is less smooth has its power on the higher frequency of 2.3 years.
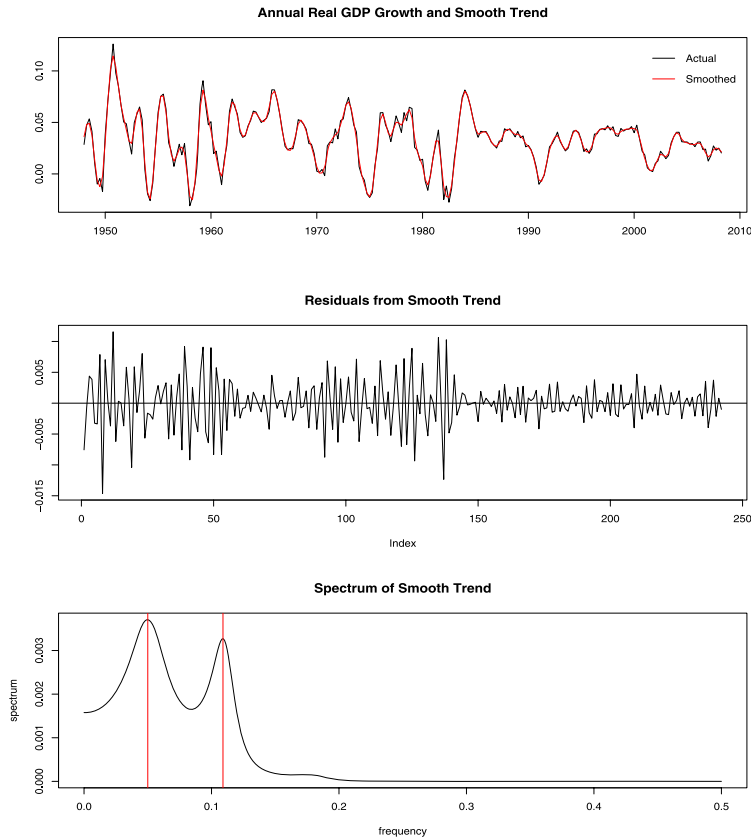
SSA, economic and financial time series: A review 389

Annual Real GDP Growth and Smooth Trend

Residuals from Smooth Trend

Spectrum of Smooth Trend

Figure 8. *Actual, smoothed and residual series, annual U.S. real GDP growth.*

Annual Real GDP Growth, 1st Component of Smooth Trend

Annual Real GDP Growth, 2nd Component of Smooth Trend

Annual Real GDP Growth Spectrum, 1st Component of Smooth Trend

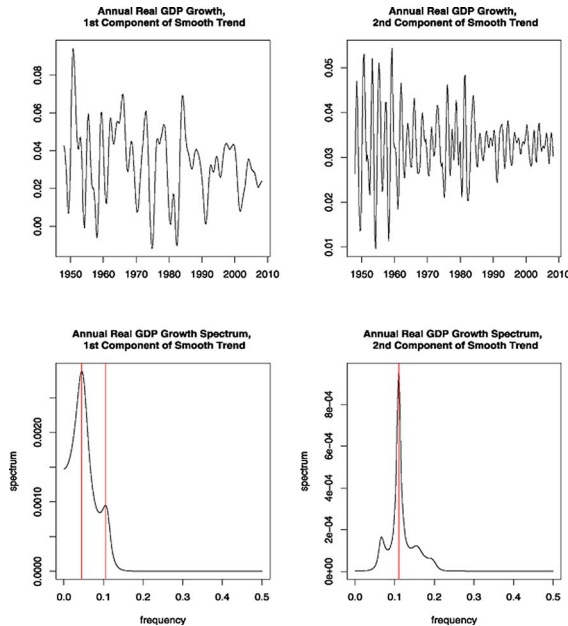**Annual Real GDP Growth Spectrum, 2nd Component of Smooth Trend**

Figure 9. *Smooth components and their corresponding spectra, annual U.S. real GDP growth.*

9. FORECASTING

9.1 Complex trend and strong seasonality

The SSA 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 important issue in SSA. Hassani et al. [12] considered the eight series of monthly industrial production indices for Germany, France and the UK. Fig. 12 (top) presents such a series, indicating a complex trend and strong seasonality. Almost all of the industrial production series have a complex structure with nonlinear trends and complex seasonality. Therefore, they found superiority of SSA over classical techniques.

Selection of the window length L

Knowing that the time series may have a periodic component with an integer period, to achieve a better separability of this periodic component it is advisable to take the window length proportional to that period. For example, the assumption that there is an annual periodicity in the series suggests that we must pay attention to the frequencies $f/12$ ($f = 1, ..., 12$). As it is advisable to choose L reasonably large (but smaller than $N/2$ which is 162 in this case), we choose $L = 120$.
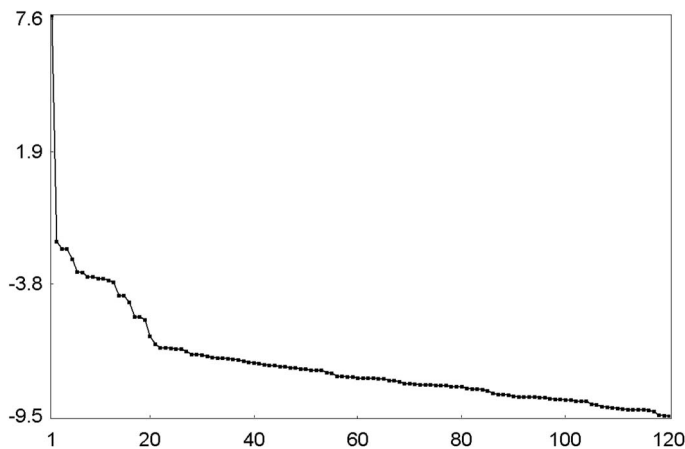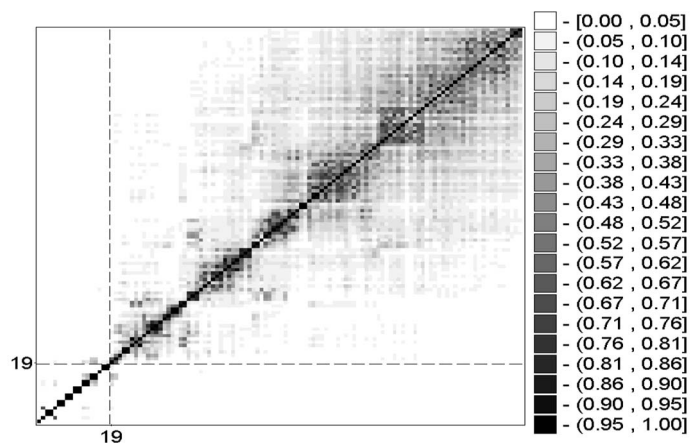
Figure 10. Logarithms of the 120 eigenvalues.

Figure 11. Matrix of w-correlations for the 120 reconstructed components.

Selection of r

Auxiliary information can be used to choose the parameters L and r. Below we briefly explain some methods that can be useful in the separation of the signal from noise. Usually a harmonic component produces two eigentriples with close singular values (except for the frequency 0.5 which provides one eigentriple with the saw-tooth singular vector). Another useful insight is provided by checking breaks in the eigenvalue spectra. Additionally, a pure noise series typically produces a slowly decreasing sequence of singular values. Choosing $L = 120$ and performing SVD of the trajectory matrix $\boldsymbol{X}\boldsymbol{X}^T$, we obtain 120 eigentriples, ordered by their contribution (share) in the decomposition. Fig. 10 depicts the plot of the logarithms of the 120 singular values. Here a significant drop in values occurs around component 19 which could be interpreted as the start of the noise floor. Six evident pairs, with almost equal leading singular values, correspond to six (almost) harmonic components of
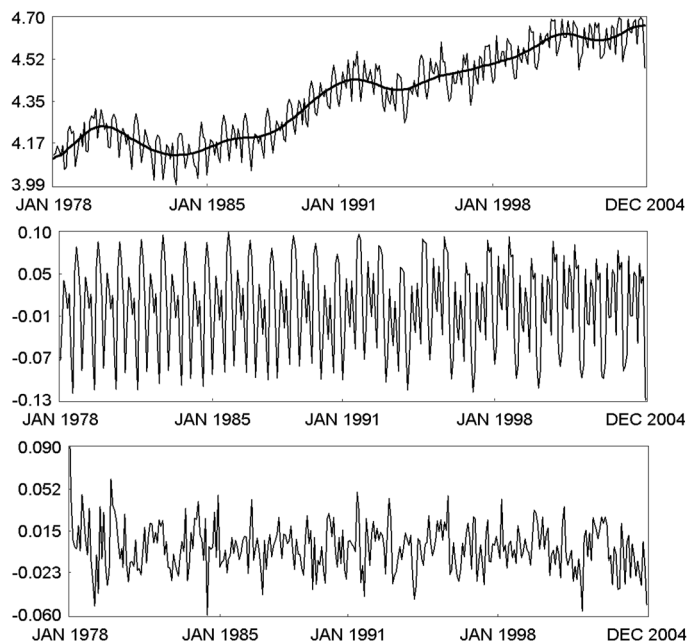

Figure 12. Reconstructed trend (top), harmonic (middle) and noise (bottom).

the series: eigentriple pairs 3–4, 6–7, 8–9, 10–11, 14–15 and 17–18 are related to the harmonics with specific periods (It has been shown that they correspond to the periods of 6, 4, 12, 3, 36 and 2.4 months [12].) Another way of grouping is to examine the matrix of the absolute values of the w-correlations [7]. Fig. 11 shows the w-correlations for the 120 reconstructed components in a 20-grade grey scale from white to black corresponding to the absolute values of correlations from 0 to 1. Based on this information, we select the first 18 eigentriples for the reconstruction of the original series and consider the rest as noise.

Identification of trend, harmonics and noise components

Trend is a slowly varying component of a time series which does not contain oscillatory components. Hence to capture the trend in the series, we should look for slowly varying eigenvectors. Fig. 12 (top) shows the extracted trend which is obtained from the eigentriples 1, 2, 5, and 12–13. It clearly follows the main tendency in the series.

Fig. 12 (middle) represents the selected harmonic components (3,4, 6–11, 14–18) and clearly shows the same pattern of seasonality as in the original series. Thus, we can classify the rest of the eigentriples components (19–120) as noise (for more information see [12]). Fig. 12 (bottom) shows the residuals which are obtained from these eigentriples. The w-correlation between the reconstructed series (the eigentriples 1–18) and the residuals (the eigentriples 19–120) is equal to 0.0006, which can be considered as a confirmation that this grouping is very reasonable.

Comparison of the accuracy of the forecasts

Hassani et al. [12] considered the forecasting performance of the SSA, ARIMA and Holt-Winter techniques at different horizons h, of up to a year for the European Industrial Production. The results were represented for $h = 1, 3, 6$ and 12 (months). They considered in-sample and out-of-sample forecasting performance. The results showed that SSA has produced lower RMSE and gives the highest R^2 for all the series for the in-sample results. The results of out-of-sample forecast reported in [12] shows 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; 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 (at least, in the range of values of h considered in their paper).

9.2 Short time series

An important feature of SSA is that it can be used for analyzing relatively short and non-stationary series [10]. Hassani and Zhigljavsky [10] considered 16 quarterly data sets each containing 68 data points, and 16 yearly data sets which contain 45 observations. They used the series of Iranian GDP (quarterly) as the main data set for illustrating details of the practical application of the SSA methodology for short time series. The forecasting results, for both yearly and quarterly series, confirms that SSA works very well even for short time series.

In parallel, they have performed SSA analysis and forecast for the data taken in the logarithmic scale. They found that taking logarithms of the data does not improve the quality of the SSA forecast (on the opposite, it typically leads to its deterioration). This is related to the fact that the quarterly data have periodic components which are easier to extract when the data are considered in the original scale (taking logarithms produces additional smoothing and makes extraction of periodic components more difficult).

9.3 Exchange rate series

The most prominent models used in predicting the exchange rates in the fundamentalist tradition include the purchasing power parity theory [4, 49, 50], sticky-price monetary model [49], the Balassa-Samuleson productivity differential model, the behavioral equilibrium exchange rate model, and the interest rate parity model [51].

The earlier empirical works in the latter strand of exchange rate economics often used linear stochastic models such as ARIMA process, however, recent development in nonlinear dynamical systems theory, methods of time-delay embedding, and phase space reconstruction has opened up the possibility of testing for presence of nonlinear, deterministic structures in the dynamics of the exchange rates [4]. In this context, the SSA technique has been used for forecasting exchange rate series in [20].

Hassani et al. [20] used two series of daily exchange rates: pound/dollar (UK) and Euro/dollar (EU). They performed one-step ahead forecasting based on the most up-to-date information available at the time of the forecast. It is clear that, the correlation between the UK/dollar and the EU/dollar exchange rate is high. This correlation motivated them to use a multivariate version of the SSA technique. They found that the difference between MSSA predictions and RW are significant with respect to all chosen criteria. The results confirmed, with strong evidence, that using MSSA enables us to improve both accuracy and direction of change of the forecasting results. Comparing to univariate SSA, they improved the accuracy of the forecasting results from only 4% (for univariate case) to 20% on average. Therefore, using the EU exchange rate information enabled them to improve the results up to 16% on average.

9.3.1 Connection with martingale theory

Note that finding whether SSA/MSSA forecasting procedures for exchange rate series are superior to the random walk (RW) forecasts or not, may be interpreted from the viewpoint of martingale theory as follows.

A series $\{y_t\}$ is called a martingale (with respect to its own past) if $E_t(y_{t+1}|y_1, \ldots, y_t) = y_t$ for all t, i.e. if the series has a unit root. It is widely believed that many financial time series (including exchange rate series) are martingales in this sense. If a series $\{y_t\}$ is a martingale (with respect to its own past), then it is not possible to improve on the random walk (RW) forecast, where $\hat{y}_{t+1} = y_t$ is used as the forecast for y_{t+1}.

The results in [20] evaluate an important assumption for prediction of the exchange rate series: indeed, we are unable to build a forecasting method for the exchange rate series that is more precise than the RW forecast, if the information available was restricted to the series itself. However, when we allowed ourselves to use additional information (the values of other exchange rate series of up to time t), then we are able to build a forecast that is superior to the RW forecast. This may imply that the exchange rate time series are not martingales with respect to all available information at the markets. Formally, if $\{y_t\}$ is the series we are interested in and $\{x_t\}$ is a multivariate series of all other currency exchange rates, then the results in [20] show that $E_t(y_{t+1}|y_1, \ldots, y_t, x_1, \ldots, x_t) \neq y_t$, which is equivalent to saying that the RW is not the best possible (in the RMSE sense) forecast.

9.3.2 Connection of MSSA results with the EMH

Those financial economists who believe in efficiency of financial markets, however, seriously doubt accurate pre-

dictability of the financial asset prices. Efficient Market Hypothesis (EMH) in its weak form implies that the returns of financial asset prices are white noise processes consisting of independent, identically distributed random variables. The white noise nature of the returns implies that the series at level follows a random walk model and is unpredictable.

In spite of the popularity of EMH, mostly in the academic circles, a vast literature dealing with predictions of the financial asset prices exists. Reviewing the empirical exchange rate economics literature one could discern two strands of research in the field that closely follow fundamentalist and chartist (technical analyst and its rough counterpart in academia time series analysts) that prevails in prediction of equity prices in the stock markets. In the context of exchange rate economics, the fundamentalists believe that the money supply, the price level, national income, interest rates, productivity, and other relevant economic variables determine exchange rates. The empirical results of the present study are instructive in examining the efficient market hypothesis controversy. Accordingly, we first present formal discussions of the martingale games, random walk processes, their relationship with the EMH, and then we elaborate on the implications of our findings for the EMH.

A stochastic process x_t follows a martingale if

$$(27) \qquad E_t(y_{t+1}|\Omega_t) = y_t$$

where Ω_t is the information set at time t that includes y_t also. Equation (27) implies that if y_t follows a martingale the best forecast of y_{t+1} is y_t, given the information set Ω_t.

The implication of the fair game model (27) in financial economics is that the returns of the asset price y_t are unpredictable, given the information set Ω_t. Accordingly, the information set Ω_t is fully reflected in the asset price, and this is known as the EMH[3].

Note that one may restrict the information set Ω_t only to the asset's past price history, making alternative representation of (27) as

$$(28) \qquad E(y_{t+1}|y_t, y_{t-1}, \ldots) = y_t$$

What are the implications of the MSSA forecasting results for the EMH? Based on the results of random walk predictions, which are based only on the past price history, we conclude that the currency markets are efficient. However, the results based on MSSA which are obtained by including other information, i.e. EU/dollar exchange rate, clearly point to inadequacy of the random walk in modeling exchange rate for predictions. Moreover, the superior results obtained from the direction of change criterion, also provide additional support for the view that currency markets may not be efficient in the sense discussed above (for further information see [20]).

---

[3]We are using EMH in a generic sense, to avoid further discussion of the types of efficient market hypothesis which is not germane to the issue here.

9.4 Inflation rate series

In recent years a number of comparative studies of inflation forecasting methods resulted in two major insights about inflation forecasting methods and inflation rate in the United States. First, the studies are inconclusive about the superiority of the competing forecasting methods. For example, Stock and Watson [52] documents that Phillips curve-based models tend to have the most accurate forecast of the inflation in the United States up to 1996. While Atkeson and Ohanian [53] contradicts the conclusion about the relative forecasting accuracy of the Phillips curve-based models and shows that a naive random walk model has a superior predictive capability. Given that the dynamics of the U.S. economy has gone through many variations due to policy and structural changes during the time period under consideration, one needs to make certain that the method of prediction is not sensitive to the dynamical variations. Therefore, SSA can be considered as a suitable technique for the forecasting US inflation rate.

9.4.1 Forecasting US inflation rate based on the CPI-all and CPI-core series

Let us now consider several U.S. price indexes in out-of-sample, h-step-ahead moving prediction exercises. These indices including consumer price index with and without highly volatile food and energy items, CPI-all and CPI-core, respectively as well as real-time quarterly chain-weighted GDP price index. Specifically, we used monthly CPI-all and CPI-core data for the period JAN 1986 – DEC 2006. We do not report inflation forecasting based on GDP price index here (the results can be found in [16]). We use moving h-step-ahead prediction, which means that we include all available information for the predictions.

The MSSA has been used for forecasting inflation rate based on the CPI-all and CPI-core series over the period Jan-1986 to Dec-1996 that was used as the training set data [16]. The forecasting results indicate that MSSA outperforms the random walk predictions in both one and 3-step ahead forecasts. Additionally, the test results for the null hypothesis of whether the percentages of the direction of changes are greater than the pure chance (50%) shows that all results are statistically significant and higher than a 50% chance. The important result is that MSSA predicts direction of change for 3-step as accurately as it can predict 1-step ahead. This confirms that we are able to capture the hidden dynamic in the inflation series using MSSA which is not usually captured by other classical methods. The results obtained below can be considered as a confirmation for this conclusion.

9.4.2 Comparison with the other methods

Comparative study is somewhat difficult, since data, methods, forecasting horizons and error criteria are not uni-

Table 5. RMSE of MSSA forecast with other models for 3-month ahead forecast for 3-month moving averages of inflation rate based on the CPI-all and CPI-core series

Year	N	MSSA/AO		MSSA/AR	
		CPI-all	CPI-core	CPI-all	CPI-core
Jan 97–Dec 98	24	0.54**	0.34**	0.57**	0.27**
Jan 97–Dec 00	48	0.73**	0.37**	0.79**	0.31**
Jan 97–Dec 02	72	0.68**	0.45**	0.73**	0.39**
Jan 97–Dec 04	96	0.70**	0.42**	0.70**	0.40**
Jan 97–Aug 06	116	0.75**	0.39**	0.76**	0.36**
Year	N	MSSA/DFM88		MSSA/DFM158	
		CPI-all	CPI-core	CPI-all	CPI-core
Jan 97–Dec 98	24	0.48**	0.68**	0.58**	0.26**
Jan 97–Dec 00	48	0.75**	0.36**	0.78**	0.35**
Jan 97–Dec 02	72	0.75**	0.38**	0.77**	0.34**
Jan 97–Dec 04	96	0.73**	0.40**	0.74**	0.38**
Jan 97–Aug 06	116	0.79**	0.37**	0.81*	0.36**

Note to Table 5. AO = Atkeson and Ohanian; AR = Autoregressive; DFM88 = Dynamic factor model based on 88 variable; DFM158 = Dynamic factor model based on 158 variables.

form. Nevertheless, we compare the results based on the MSSA method with the results obtained from other inflation prediction methods. Table 5 presents RMSEs of MSSA prediction results and the forecasting results obtained using the other models considered in this study for 3-step ahead predictions. The results clearly indicate that MSSA outperforms other classical methods considered here.

The results for the direction of change can be found in [16]. The results, not reported here, indicate that MSSA method correctly predicts the direction of the change of the CPI-Core consistently higher than the competing models. This is particularly true for longer prediction horizons.

9.5 MSSA with different series length

One of the main advantages of MSSA compared to classical multivariate time series methods is that we can use several series with different series length in SSA, whilst this is not the cast for classical time series (the series should have the same series length). The Index of industrial production (IIP) is a typical example of such usage. The IIP series as a leading economic indicator, is the most widely scrutinised and intensively studied of all economic time series. However, this index is revised several times before a final figure is published. The time between the publication of the initial and final data is a lengthy one and thus forecasting the final vintage of data on IIP is an important issue. Let us now consider the structure of these series.

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

Table 6. Forecasting results over period Jan 2001–March 2006

	RMSE			Bias		DC	
V	P-V	SSA	RRMSE	P-V	SSA	P-V	SSA
1	1.71	1.53	0.90**	−0.75	−0.55	0.75**	0.79**
2	1.60	1.43	0.90**	−0.71	−0.52	0.76**	0.81**
3	1.48	1.33	0.90**	−0.63	−0.46	0.76**	0.84**
4	1.37	1.23	0.90**	−0.56	−0.42	0.73**	0.79**
5	1.28	1.16	0.91**	−0.49	−0.35	0.81**	0.87**
6	1.14	1.06	0.92**	−0.40	−0.29	0.86**	0.87**
7	1.02	0.97	0.95**	−0.33	−0.25	0.89**	0.92**
8	0.92	0.86	0.94**	−0.27	−0.20	0.92**	0.89**
9	0.79	0.76	0.96**	−0.21	−0.16	0.92**	0.90**
10	0.66	0.65	0.98*	−0.16	−0.12	0.94**	0.94**
11	0.54	0.52	0.98*	−0.11	−0.08	0.97**	0.97**

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

$$(29) \quad \begin{pmatrix} y_1^1 & y_1^2 & y_1^3 & \cdots & y_1^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{t-m}^1 & y_{t-m}^2 & y_{t-m}^3 & \cdots & y_{t-m}^m \\ \vdots & \vdots & \vdots & \ddots & \\ y_{t-2}^1 & y_{t-2}^2 & y_{t-2}^3 & & \\ y_{t-1}^1 & y_{t-1}^2 & & & \\ y_t^1 & & & & \end{pmatrix}$$

Thus, publication from a particular issue of MDS trace back a diagonal of this data matrix which is a composite of data of different vintages. The problem here is to forecast h step ahead ($h = 1, \ldots, 11$) of the final vintage, $v = m$, given the information at time t. The forecast could be obtained using classical univariate time series methods. However, the forecasts are not optimal since other information (vintages) available at time t are not used. For example, in forecasting y_{t-m+1}^m we also have available information of y_{t-m+1}^v for $v = 1, \ldots, m-1$, each of which could itself be regarded as a forecast of y_{t-m+1}^m. This matter motivates us to use a multivariate method for forecasting h step ahead of y_t^m. For example, to obtain the final vintage value at time t, y_t^m, we can use the information for the first vintage data $y_1^1, \ldots, y_t^1$ and the final vintage data $y_1^m, \ldots, y_{t-m}^m$.

The performance of the MSSA has been assessed by applying it to the first twelve vintages of the index of industrial production in [19]. The results shows that the preliminary vintage can not be viewed as an accurate forecast of the final vintage and using SSA enables us to obtain more accurate forecasts in terms of reducing the mean square errors. It has been found that SSA performs well in predicting the direction of change. Table 6 shows the results in details. As appears from Table 6, there are gains to using SSA throughout the revision process, these being between 8% and 11% for vintage up to $v = 5$, reducing to 5% or slightly less for latter vintages for basic SSA. Table 6 also shows the mean

Table 7. The value of $F_{v^m|v^i}^{(h,m-i)}$ and $D_{v^m|v^i}^{(h,m-i)}$ in forecasting of ith vintage of the index of industrial production series

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

prediction error for each vintage $v = 1, \ldots, 11$, using the SSA and initial estimates. As the table shows, the negative bias present in just using preliminary vintage has also been decreased. Table 6 provides the percentage of forecasts that correctly predict the direction of change for each vintage. It also shows whether they are significantly greater than the pure chance ($p = 0.50$). The symbols $**$ in the table indicate the 1% level of significance. The percentage of correct signs produced by SSA are all higher than those given by the preliminary vintages and they are all significant at the 1% level. The results obtained by MSSA are also much better than VAR model in both forecasting accuracy and predictability of direction of change (for more information see [19].

We expect that there is a SSA causal relationship between preliminary vintage (v^{th} vintage) and the final vintage (m^{th} vintage). To answer this, we need to forecast h step ahead ($h = 1, \ldots, 11$) of the final vintage, $v = m$, given the information at time t. If the results of h step ahead forecast MSSA are better than SSA, e.g. $F_{v^m|v^i}^{(h,m-i)} < 1$ and $D_{v^m|v^i}^{(h,m-i)} < 1$, we then conclude that there is a SSA causal relationship of order h between i^{th} vintage and final vintage. The value of $F_{v^m|v^i}^{(h,m-i)}$ and $D_{v^m|v^i}^{(h,m-i)}$ ($i = 1, \ldots, 11$) for each vintage and relative to single SSA is given in Table 7. The results, without exception, confirm that there exists a SSA causal relationship between each vintage and the final vintage. In fact the results strongly indicate that there is SSA causality between i^{th} vintage and final vintage is of order $m - i$. It should be noted that here i is equal to h step ahead forecast which is the time lag difference between i^{th} vintage and final vintage. Here, as the results show, SSA causality holds for lower lag order such as in the case of the exchange rate series. This confirms that SSA causality of order $m - i$ has consequences for other orders of causality. Note that here the problem of interest is one-side causality as the final vintage is forecasted. The results of Granger causality tests, also showed that there is a Granger

causal relationship between these series. This is not surprising as each column of the data matrix is a revised version of the previous column and therefore they are high correlated. It should be noted that the results of VAR model in forecasting these series are worse than the MSSA results [19].

10. CONCLUSION

Singular Spectrum Analysis (SSA) has emerged over the past 20 years as a powerful tool for analyzing a variety of time series. Although its origins lie in the natural sciences, and the series associated from such processes, it has and can be applied in several different fields. Economic and financial time series have several characteristics that make them amenable to SSA-based analysis, in smoothing, trend extraction and forecasting. In this paper we provide a thorough review of SSA in the context of economic and financial time series. Besides a concise overview and references of the SSA methodology, we present: (a) theoretical results on SSA applied to time series with unit-root non-stationarity & connect them to filtering methods popular in economics (b) a potential method for selecting the embedding dimension; (c) two new methods for performing SSA based on minimum variance and perturbation theory concepts; (d) the relationship between SSA and causality testing and a large variety of empirical results based on simulated and real-world time series.

Both the theoretical developments and the results presented strongly support the use and potential of SSA in economic and financial time series. As a non-parametric method for smoothing and trend extraction we believe that SSA should be part of the practitioners toolbox as it can give results that are similar or better than those obtained by current methods. As a forecasting method in economic and financial time series SSA has given us some very promising results but has not revealed yet its full potential: issues like optimizing the embedding dimension, the construction of accurate predictive intervals, of forecasting trend components (e.g. potential output, fair prices etc.) are still open and can be pursued in future research.

Received 17 March 2010

REFERENCES

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

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

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

[4] 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.

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

[6] SCHEINKMAN, J. and LEBARON, B. (1989). Nonlinear Dynamics and Stock Returns. *Journal of Business* **62** 311–337.

[7] GOLYANDINA, N., NEKRUTKIN, V., and ZHIGLJAVSKY, A. (2001). *Analysis of Time Series Structure: SSA and related techniques*, Chapman & Hall/CRC, New York - London.

[8] BROOMHEAD, D. S. and KING, G. (1986). Extracting qualitative dynamics from experimental data *Physica D* **20** 217–236.

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

[10] HASSANI, H. and ZHIGLJAVSKY, A. (2009). Singular Spectrum Analysis: Methodology and Application to Economics Data. *Journal of System Science and Complexity* **22** 372–394.

[11] 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** 319–352.

[12] HASSANI, H., HERAVI, S., and ZHIGLJAVSKY, A. (2009). Forecasting European Industrial Production with Singular Spectrum Analysis. *International Journal of Forecasting* **25** 103–118.

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

[14] BENEKI, C., EECKELS, B., and LEON, C. (2009). Signal Extraction and Forecasting of the UK Tourism Income Time Series. A Singular Spectrum Analysis Approach. *http://mpra.ub.uni-muenchen.de/18354/*

[15] HASSANI, H. (2009). Singular Spectrum Analysis based on the Minimum Variance Estimator. *Nonlinear Analysis: Real World Application* **11** 2065–2077.

[16] HASSANI, H. (2009). Development of the theoretical and methodological aspects of the singular spectrum analysis and its application for analysis and forecasting of economics data, PHD Thesis, Cardiff University.

[17] 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** 492–502.

[18] 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.

[19] PATTERSON, K., HASSANI, H., HERAVI, S., and ZHIGLJAVSKY, A. (2010). Forecasting the final vintage of the industrial production series. *Journal of Applied Statistics*, Forthcoming.

[20] HASSANI, H., SOOFI, A., and ZHIGLJAVSKY, A. (2009). Predicting Daily Exchange Rate with Singular Spectrum Analysis. *Nonlinear Analysis: Real World Applications* **11** 2023–2034.

[21] LISI, F. and MEDIO A. (1997). Is a random walk the best exchange rate predictor? *International Journal of Forecasting* **13** 255–267.

[22] THOMAKOS, D. D. (2008). Optimal Linear Filtering, Smoothing and Trend Extraction for Processes with Unit Roots and Cointegration. working paper, available at at SSRN: *http://ssrn.com/abstract=1113331*.

[23] FULLER, W. (1995). *Time Series Analysis*, 2nd edition, New York: John Wiley.

[24] THOMAKOS, D. D. (2008). Optimal Linear Filtering, Smoothing and Trend Extraction of m-Period Differences of Processes with a Unit Root, working paper, available at *SSRN: http://ssrn.com/abstract=1295982*

[25] PHILLIPS, P. C. B. (1996). Spurious Regression Unmasked, CFDP 1135.

[26] PHILLIPS, P. C. B. (1998). Econometric Analysis of Fisher's Equation, CFDP 1180.

[27] PHILLIPS, P. C. B. (2005). Challenges of Trending Time Series Econometrics. *Mathematics and Computers in Simulation* **68** 401–416.

[28] HODRICK, R. and PRESCOTT E. C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking* **29** 1–16.

[29] KING, R. G. and REBELO, S. T. (1993). Low Frequency Filtering and Business Cycles. *Journal of Economic Dynamics and Control* **17** 207–232.

[30] BAXTER, M. and KING, R. G. (1999). Measuring Business-cycles: Approximate Band-pass Filters for Economic Time Series". *Review of Economics and Statistics* **81** 575–593.

[31] CHRISTIANO, L. and FITZGERALD, T. (2003). The Band Pass Filter. *International Economic Review* **44** 435–465.

[32] POLLOCK, D. S. G. (2000). Trend Estimation and De-trending via Rational Square-wave Filters. *Journal of Econometrics* **99** 317–334.

[33] THOMAKOS, D., WANG, T., and WILLE, L. (2002). Modeling daily realized futures volatility using singular spectrum analysis. *Physica A* **312** 505–519.

[34] DROGOBETSKII, S. (2005). Singular Spectrum Analysis Of Price Movement In Forex. *Stocks & Commodities* **23** 24–27.

[35] DROGOBETSKII, S. (2005). Singular Spectrum Analysis Of Price Movement In Forex (Part 2). *Stocks & Commodities* **23** 44–48.

[36] DROGOBETSKII, S. and SMOLYNSKY, V. (2008). Forecasting Singular Spectrum Analysis. *Stocks & Commodities* **26** 38–40.

[37] BUCHSTABER, V. M. (1994). Time Series Analysis and Grassmanians, in S. Gindikin (Ed.). *Applied Problems of Radon Transforms* AMS TransactionsSeries 2, **162** AMS, Providence, RI.

[38] THOMAKOS, D. D. (2009). Median-Unbiased Optimal Smoothing and Trend Extraction. *Journal of Modern Applied Statistical Methods*, Forthcoming.

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

[40] BRITANAK, V., YIP, P. C., and RAO, K. R. (2006). *Discrete Cosine and Sine Transforms: General Properties, Fast Algorithms and Integer Approximations*, Academic Press.

[41] MCELROY, T. (2005). Matrix Formulas for Nonstationary Signal Extraction, SRD Research Report No. RRS2005-04, U.S. Census Bureau.

[42] SCHLICHT, E. (2005). Estimating the Smoothing Parameter in the So-called Hodrick-Prescott Filter. *Journal of the Japan Statistical Society* **35** 99–119.

[43] DERMOUNE, A., DJEHICHE, B., and RAHMANIA, N. (2008). A Consistent Estimator of the Smoothing Parameter in the Hodrick-Prescott Filter. *Journal of the Japan Statistical Society* **38** 225–241.

[44] GRANGER, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37** 424–438.

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

[46] ZHIGLJAVSKY, A., HASSANI, H., and HERAVI, H. (2009). Forecasting European Industrial Production with Multivariate Singular Spectrum Analysis, International Institute of Forecasters, 2008–2009 SAS/IIF Grant, http://forecasters.org/pdfs/SAS-Report.pdf

[47] VAN HUFFEL, S. (1993). Enhanced resolution based on minimum variance estimation and exponential data modeling. *Signal Processing* **33** 333–355.

[48] DE MOOR, B. (1993). The singular value decomposition and long and short spaces on noisy matrices. *IEEE Transaction on Signal Processing* **41** 2826–2838.

[49] FRANKEL, J. A. (1979). On the mark: a theory of floating exchange rates based on real interest differentials. *American Economic review* **69** 610–622.

[50] CORBAE, D. and OULIARIS, S. (1988). Cointegration and tests of purchasing power parity. *Review of Economics and Statistics* **70** 508–511.

[51] CHINN, M. D. (1997). Paper pushers or paper money? Empirical assessment of fiscal and monetary models of exchange rate determination. *Journal of Policy Modeling* **19** 51–78

[52] STOCK, J. and WATSON, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics* **44** 293–335.

[53] Atkeson, A. and Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation?, Federal Reserve Bank of Minneapolis. *Quarterly Review* **25** 2–11.

[54] Gavin, W. T. and Kliesen, K. L. (2006). Forecasting inflation and output: Comparing data-rich models with simple rules. *Research Division, Federal Reserve Bank of St. Louis, Working Paper Series, 2006–054B.*

Hossein Hassani
Cardiff School of Mathematics
Cardiff University
Cardiff, UK, CF24 4AG
E-mail address: hassanih@cf.ac.uk

Dimitrios Thomakos
Department of Economics
School of Management and Economics
University of Peloponnese Tripolis
Greece
E-mail address: thomakos@uop.gr