On perturbation stability of SSA and MSSA forecasts and the supportiveness of time series

Maria Vronskaya-Robl *[1] and Karl Michael Schmidt†[1]

[1]School of Mathematics, Cardiff University, Cardiff CF24 4AG, Wales, UK

August 7, 2015

Abstract

Studying the stability of Singular Spectrum Analysis (SSA) and Multivariate Singular Spectrum Analysis (MSSA) forecasts under random perturbations of the input time series, we make the empirical observation that the reconstruction kernel of SSA as a convolution filter and the forecast recurrence vector are remarkably stable both under generated Gaussian and natural non-Gaussian noise. Assuming that these elements of the forecast procedure are noise-independent, we derive concise formulae for the variance under perturbations of SSA and MSSA forecasts. We suggest a criterion of supportiveness based on the behaviour of these proxy variances under scaling of the support series. Finally, we remark on a problem of lacking scaling invariance of MSSA.

1 Introduction

In the present study, we analyse the stability of the SSA and MSSA forecasts, i.e. the properties of the forecast as a random variable when the original time series is perturbed by added noise, with and without the inclusion of a second time series in the analysis. This question was motivated by the analogy between the recurrence system (4) for a recurrent MSSA forecast of a pair of time series and the stationary vector autoregressive process

$$\begin{array}{l} X_t = \sum_{j=1}^{m} a_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon_t \\ Y_t = \sum_{j=1}^{m} c_j X_{t-j} + \sum_{j=1}^{m} d_j Y_{t-j} + \xi_t \end{array} \tag{1}$$

used by C. J. Granger to implement his concept of causality, as expressed in the words: "We say that Y_t is causing X_t if we are better able to predict X_t using all available information than if the information apart from Y_t had been used" [6, p.428], i.e. Y is causing X if it improves the quality of the forecast of series X.

*vronskayam@cardiff.ac.uk, vronskaya.maria@gmail.com

†schmidtkm@cardiff.ac.uk

Note that there are essential differences between the models: (1) is a *stochastic* process modeling the *noise only*, and indeed the definition of Granger's causality coherence requires noise to be present in both time series, whereas (4) is a *deterministic* process (although the coefficients are derived from random variables in practice) modeling the *signal*, with noise added afterwards, so that in this case noise does not propagate and no specific noise model is assumed.

Nevertheless, the analogy between (4) and (1) could provide a basis for the study of a causal or *support* relationship between time series (without the assumptions of stationarity, zero mean or autoregressive structure) by comparing the quality of the MSSA forecast of time series x using the support series y with that of the SSA forecast of x alone, taking an improved forecast as an indicator of supportiveness. What exactly counts as an improvement is open to interpretation: it can refer both to *accuracy* of the forecast, i.e. how close the predicted value is to the real value, and to the *stability* of the forecast, i.e. the variance of the forecast as a random variable. In practice, actual future values are subject to random fluctuation and hence introduce further variance into a measurement of accuracy. Therefore, we focus on working with the stability of the forecast value in the present study. This is effectively done in practical use of Granger causality tests, which are closely related to the F-test and essentially compare goodness of fit for regression with and without the support series [7, 14]. There have been earlier attempts to estimate causality by different means than autoregressive modeling, including SSA [12, 2]. For example, in [2] the forecast improvement is estimated from the point of view of accuracy, i.e. predicted values in a forecast interval are compared to actual future values. In contrast, we study the stability of time series forecasts under random perturbations., considering the following question: *Does inclusion of the support series in MSSA make the forecast more stable under perturbations?*

Put simply, the SSA procedure can be visualised as a black box (see diagram below) which takes the initial series x and its perturbation ($\sigma\varepsilon$) as an input and outputs the forecast. For the bivariate MSSA, an extra time series y (support series) is used as input. The resultant outcome is the forecast point, calculated with the appropriate SSA/MSSA linear recurrence. This will be a random variable; in the diagram below, ξ has mean 0 and variance 1.

$$x + \sigma\varepsilon \longrightarrow \boxed{\text{SSA}} \longrightarrow \hat{x}_{\text{SSA}} + \hat{\sigma}_{\text{SSA}}\xi$$

$$\left.\begin{matrix} x + \sigma\varepsilon \\ y \end{matrix}\right\} \longrightarrow \boxed{\text{MSSA}} \longrightarrow \hat{x}_{\text{MSSA}} + \hat{\sigma}_{\text{MSSA}}\xi$$

We are interested in the relation of input and outcome variances with the main focus on the ratio between outcome variances of SSA $\hat{\sigma}_{SSA}$ and MSSA $\hat{\sigma}_{MSSA}$. We do not perturb the support series, as this would only add variance to the MSSA output in a way which has no comparable analogue in SSA.

After noting, in Section 2, that the first-order perturbation theory of the spectral decomposition, which is a core element of SSA, does not give a sufficiently transparent link between the variances of input perturbation and forecast, we make the discovery, in Section 3, that the central part of the time series, excluding the first and the last SSA window, is remarkably stable under perturbations, allowing the use of the unperturbed SSA projector as a proxy for the actual perturbed one. To a lesser extent, this flattening effect carries over

to MSSA. Note that for this observation to be useful, the time series needs to be longer than twice the chosen window length for SSA, to leave a sufficiently indicative central part of the series. This discovery motivates a simplifying assumption which we can use to derive very convenient and practical formulae (47), (54) for the SSA and MSSA forecast variances in Section 4. When comparing the performance of these formulae against empirical forecast variances calculated in random trials for pairs of time series with and without expected supportiveness, in Section 5, we find that they appear to reflect the qualitative relationship between the SSA and MSSA variances very well and are also robust against distorting effects caused by random eigenvalue crossings in the pseudorandom empirical trials.

Accepting these formulae as indicative proxies for the actual forecast variances, and with Granger's guiding idea in mind, we suggest the following practical computational method for establishing or rejecting supportiveness of a time series y_n for a time series x_n. Calculate the two fundamental objects, the MSSA convolution kernel (44) and the MSSA forecast vectors R_{11}, R_{12} (5) from MSSA of x_n and ρy_n, with a scaling parameter $\rho > 0$, and hence compute the convolution norm (54). If this norm tends to 0 for large ρ, the series y_n is supportive for x_n, if the norm approaches a positive level for large ρ, then y_n is not supportive for x_n. Note that the application of this criterion does not require any calculation of (pseudo)random perturbations, but uses only MSSA data for the unperturbed time series.

The above method makes crucial use of the fact that the MSSA forecast is not homogeneous in the scaling factor ρ. However, in some other applications of MSSA, this may be rather problematic. We briefly touch upon this scaling problem and suggest a partial remedy in Section 6.

2 First-order perturbation theory of SSA

We start with a look at the information on the forecast provided by first-order perturbation theory [13], which essentially corresponds to linearisation of the singular-value (spectral) decomposition underlying SSA and MSSA and of the recurrent forecast formula. As a result, we shall see that this linearisation in itself is not sufficient to give a sufficiently simple overview of the connection between the forecast variance and the variance of the input perturbation; but it shows the different steps of error variance propagation in the SSA process and pinpoints the crucial elements of this process for our subsequent considerations. We remark that, beyond the simple first-order perturbation analysis below, the full perturbation series has been studied extensively in [11]; however, that study is mainly concerned with the limit of infinitely long time series, whereas we consider fixed-length time series and small perturbation variance.

We shall use the following notation for SSA and bivariate MSSA analysis and recurrent forecast.

Consider a time series $(x_1, ..., x_N)$. Choosing a window length $1 < L < N$,

we set up a *trajectory matrix*

$$\mathbb{X} = \begin{pmatrix} x_1 & x_2 & \dots & x_{N-L+1} \\ x_2 & & \dots & x_{N-L+2} \\ \vdots & & \dots & \vdots \\ x_L & & \dots & x_N \end{pmatrix}$$

and consider the spectral decomposition of the lag-covariance matrix $\mathbb{X}\mathbb{X}^T = \sum_{j=1}^{L} \lambda_j \eta_j \eta_j^T$, where η_j are orthonormal eigenvectors of $\mathbb{X}\mathbb{X}^T \in \mathbb{R}^{L\times L}$ with corresponding eigenvalues λ_j, enumerated in non-increasing order, i.e. $\lambda_j \geq \lambda_{j+1}$. Choosing $r < L$, we use the orthogonal spectral projector of rank r, $\mathbb{P} = \sum_{i=1}^{r} \eta_j \eta_j^T$, to extract a lower rank matrix $\tilde{\mathbb{X}} = \mathbb{P}\mathbb{X}$. Hankelization, i.e. diagonal averaging of $\tilde{\mathbb{X}}$ and of $\mathbb{X} - \tilde{\mathbb{X}} = (1-\mathbb{P})\mathbb{X}$, gives a decomposition of the initial time series $x_i = \tilde{x}_i + \varepsilon_i$ into reconstructed time series $\tilde{x}_i$, which can be interpreted as a signal, and residuals ε_i which are treated as noise. For the discussion of the optimal choice of window length L and number of components r see [4]. Based on this representation, it is possible to forecast the initial time series by extending the Hankelized output matrix $\tilde{\mathbb{X}}_H$ in such a way that the next added column $(\tilde{x}_{N-L+2}, ..., \tilde{x}_N, \hat{x}_{N+1})^T$ has minimal distance to the projection subspace $\mathbb{P}\mathbb{R}^L$, giving the linear L-term recurrence

$$\hat{x}_{n+1} = (\tilde{x}_{n-L+2}, ..., \tilde{x}_n)R, \qquad (n \geq N) \tag{2}$$

with the recurrence vector

$$R = \frac{\sum_{k=1}^{r} \eta_{k,L} \eta_k^{\nabla}}{1 - \sum_{k=1}^{r} \eta_{k,L}^2} \in \mathbb{R}^{L-1}, \quad \eta_k = \begin{pmatrix} \eta_k^{\nabla} \\ \eta_{k,L} \end{pmatrix} \in \mathbb{R}^L. \tag{3}$$

The forecast can be further extended using the same recurrence formula.

The above procedure can be extended to the simultaneous analysis of two time series by stacked bivariate MSSA. Given a second time series $y = (y_1, ..., y_N)$, we stack trajectory matrices, both with the same window length L, in the following manner $\begin{pmatrix} \mathbb{X} \\ \mathbb{Y} \end{pmatrix}$ and use the spectral decomposition of

$$\begin{pmatrix} \mathbb{X} \\ \mathbb{Y} \end{pmatrix} \begin{pmatrix} \mathbb{X} \\ \mathbb{Y} \end{pmatrix}^T = \begin{pmatrix} \mathbb{X}\mathbb{X}^T & \mathbb{X}\mathbb{Y}^T \\ \mathbb{Y}\mathbb{X}^T & \mathbb{Y}\mathbb{Y}^T \end{pmatrix} = \sum_{j=1}^{2L} \lambda_j^M \eta_j^M \eta_j^{M^T};$$

the spectral projector will be $\mathbb{P}_M = \sum_{j=1}^{r} \eta_j^M \eta_j^{M^T}$. Separate Hankelization of $\tilde{\mathbb{X}}$ and $\tilde{\mathbb{Y}}$, defined by $\begin{pmatrix} \tilde{\mathbb{X}} \\ \tilde{\mathbb{Y}} \end{pmatrix} = \mathbb{P}_M \begin{pmatrix} \mathbb{X} \\ \mathbb{Y} \end{pmatrix}$, gives MSSA reconstructions $\tilde{x}$ and $\tilde{y}$. For the forecasting, following the same principle as before, we get a bivariate L-term linear recurrence

$$\begin{aligned} \hat{x}_{N+1} &= (\tilde{x}_{N-L+2}, ..., \tilde{x}_N)R_{11} + (\tilde{y}_{N-L+2}, ..., \tilde{y}_N)R_{12} \\ \hat{y}_{N+1} &= (\tilde{x}_{N-L+2}, ..., \tilde{x}_N)R_{21} + (\tilde{y}_{N-L+2}, ..., \tilde{y}_N)R_{22} \end{aligned} \tag{4}$$

with the recurrence vectors ($j \in \{1, 2\}$)

$$R_{1j} = \frac{1}{\det A}\big((1 - \sum_{k=1}^{r} \eta_{k,2L}^2) \sum_{k=1}^{r} \eta_k^{(j)} \eta_{k,L} + \sum_{k=1}^{r} \eta_{k,2L}\eta_{k,L} \sum_{k=1}^{r} \eta_k^{(j)} \eta_{k,2L}\big), \quad (5)$$

$$R_{2j} = \frac{1}{\det A}\big((1 - \sum_{k=1}^{r} \eta_{k,L}^2) \sum_{k=1}^{r} \eta_k^{(j)} \eta_{k,2L} + \sum_{k=1}^{r} \eta_{k,L}\eta_{k,2L} \sum_{k=1}^{r} \eta_k^{(j)} \eta_{k,L}\big), \quad (6)$$

where

$$A = \begin{pmatrix} 1 - \sum_{k=1}^{r} \eta_{k,L}^2 & -\sum_{k=1}^{r} \eta_{k,2L}\eta_{k,L} \\ -\sum_{k=1}^{r} \eta_{k,L}\eta_{k,2L} & 1 - \sum_{k=1}^{r} \eta_{k,2L}^2 \end{pmatrix} \quad \text{and} \quad \eta_k = \begin{pmatrix} \eta_k^{(1)} \\ \eta_{k,L} \\ \eta_k^{(2)} \\ \eta_{k,2L} \end{pmatrix} \in \mathbb{R}^{2L}.$$

Note that in addition to the above recurrent forecasting method, which will be the basis of our considerations in this paper, there are other SSA- and MSSA-based forecasting methods, see [15, 3]; our MSSA forecasting method corresponds to MSSA-K in [15] and recurrent row MSSA forecasting in [3].

First-order perturbation theory of the SSA process and forecast gives the following result.

Proposition 2.1. *Let $\mathbb{X}$ be the trajectory matrix of an unperturbed time series x and L the chosen SSA window length. Assume that the eigenvalues $\lambda_1, ..., \lambda_L$ of $\mathbb{X}\mathbb{X}^\intercal$ are all simple.*

Let R be the SSA recurrence vector (3) obtained from the unperturbed time series x_n. Then the SSA recurrence vector $R(\sigma)$ obtained from the randomly perturbed time series $x_n + \sigma\varepsilon_n$, where $\varepsilon \sim N(0, 1)$ is i.i.d., with the same SSA parameters L and r is, to first order,

$$R(\sigma) = R + \sigma\left(\mathbf{c}R + \widetilde{R}\right) + O(\sigma^2) \quad (\sigma \to 0), \quad (7)$$

where

$$\mathbf{c} = \frac{2\sum_{k=1}^{r}\sum_{i=r+1}^{L} \alpha_{i,k}\eta_{i,L}\eta_{k,L}}{1 - \sum_{k=1}^{r} \eta_{k,L}^2} \quad (8)$$

and

$$\widetilde{R} = \frac{\sum_{k=1}^{r}\sum_{i=r+1}^{L} \alpha_{i,k}(\eta_{i,L}\eta_k^{\nabla} + \eta_{k,L}\eta_i^{\nabla})}{1 - \sum_{k=1}^{r} \eta_{k,L}^2}. \quad (9)$$

Proof. Let N be the trajectory matrix of the perturbation time series ε, and

$$\mathrm{Z} = \mathbb{X}\mathrm{N}^T + \mathrm{N}\mathbb{X}^T. \quad (10)$$

The perturbed matrix

$$(\mathbb{X} + \sigma\mathrm{N})(\mathbb{X} + \sigma\mathrm{N})^T = (\mathbb{X}\mathbb{X}^T + \sigma\mathrm{Z}) + O(\sigma^2)$$

has orthonormal eigenvectors

$$\gamma_{\sigma,k} = \eta_k + \sigma\nu_{1,k} + \sigma^2\nu_{2,k} + \dots \quad (11)$$

and eigenvalues

$$\lambda_{\sigma,k} = \lambda_k + \sigma\mu_{1,k} + \sigma^2\mu_{2,k} + \ldots, \tag{12}$$

$k \in \{1, ..., L\}$, which are analytical in σ [10, Chapter 2,§1]. Substitution of these power series into the eigenvalue equation gives, in order σ,

$$\mathbb{X}\mathbb{X}^T\nu_{1,k} + \mathrm{Z}\eta_k = \lambda_k\nu_{1,k} + \mu_{1,k}\eta_k, \tag{13}$$

and hence $\mu_{1,k} = \eta_k^T\mathrm{Z}\eta_k$. Hence, writing $\nu_{1,k}$ in terms of basis eigenvectors η_i,

$$\nu_{1,k} = \sum_{i=1}^{L} \alpha_{i,k}\eta_i,$$

gives, for $j \neq k$,

$$\alpha_{j,k} = \frac{\eta_j^T(\mu_{1,k} - \mathrm{Z})\eta_k}{\lambda_j - \lambda_k} = -\frac{\eta_j^T\mathrm{Z}\eta_k}{\lambda_j - \lambda_k}.$$

For $j = k$, note that

$$\begin{aligned} 1 = \gamma_{\sigma,k}^T\gamma_{\sigma,k} = &\quad \eta_k^T\eta_k + \sigma(\eta_k^T\nu_{1,k} + \nu_{1,k}^T\eta_k) + O(\sigma^2) \\ = &\quad 1 + 2\sigma\alpha_{k,k} + O(\sigma^2), \end{aligned}$$

so $\alpha_{k,k} = 0$. Furthermore, the fact that Z is symmetric implies the antisymmetry $\alpha_{j,k} = -\alpha_{k,j}$. Now the recurrence vector for the SSA forecast is

$$\begin{aligned} R(\sigma) &= \frac{\sum_{k=1}^{r} \gamma_{k,\sigma,L}\gamma_{k,\sigma}^{\triangledown}}{1 - \sum_{k=1}^{r} \gamma_{k,\sigma,L}^2} \\ &= \frac{\sum_{k=1}^{r} \left(\eta_{k,L}\eta_k^{\triangledown} + \sigma \sum_{i=1,i\neq k}^{L} \alpha_{i,k}(\eta_{i,L}\eta_k^{\triangledown} + \eta_{k,L}\eta_i^{\triangledown})\right) + O(\sigma^2)}{1 - \sum_{k=1}^{r}(\eta_{k,L}^2 + 2\sigma \sum_{i=1,i\neq k}^{r} \alpha_{i,k}\eta_{i,L}\eta_{k,L}) + O(\sigma^2)}. \end{aligned} \tag{14}$$

Due to antisymmetry, we have

$$\alpha_{i,k}(\eta_{i,L}\eta_k^{\triangledown} + \eta_{k,L}\eta_i^{\triangledown}) + \alpha_{k,i}(\eta_{k,L}\eta_i^{\triangledown} + \eta_{i,L}\eta_k^{\triangledown}) = 0 \tag{15}$$

in the numerator of (14), and similarly in the denominator of (14),

$$\alpha_{i,k}\eta_{i,L}\eta_{k,L} + \alpha_{k,i}\eta_{k,L}\eta_{i,L} = \alpha_{i,k}\eta_{i,L}\eta_{k,L} - \alpha_{i,k}\eta_{k,L}\eta_{i,L} = 0. \tag{16}$$

Therefore

$$R(\sigma) = \frac{\sum_{k=1}^{r} \eta_{k,L}\eta_k^{\triangledown} + \sigma(\sum_{k=1}^{r}\sum_{i=r+1}^{L} \alpha_{i,k}(\eta_{i,L}\eta_k^{\triangledown} + \eta_{k,L}\eta_i^{\triangledown})) + O(\sigma^2)}{1 - \sum_{k=1}^{r} \eta_{k,L}^2 - 2\sigma \sum_{k=1}^{r}\sum_{i=r+1}^{L} \alpha_{i,k}\eta_{i,L}\eta_{k,L} + O(\sigma^2)}. \tag{17}$$

Equations (7),(8),(9) now follow by observing that

$$\begin{aligned} &\left(1 - \frac{2\sigma \sum_{k=1}^{r}\sum_{i=r+1}^{L} \alpha_{i,k}\eta_{i,L}\eta_{k,L}}{1 - \sum_{k=1}^{r} \eta_{k,L}^2} + O(\sigma^2)\right)^{-1} \\ &= 1 + \frac{2\sigma \sum_{k=1}^{r}\sum_{i=r+1}^{L} \alpha_{i,k}\eta_{i,L}\eta_{k,L}}{1 - \sum_{k=1}^{r} \eta_{k,L}^2} + O(\sigma^2). \end{aligned}$$

□

Note that, to first order, the noise from the perturbation enters the recurrence vector $R(\sigma)$ only through the coefficients $\alpha_{j,k}$.

The perturbation of the forecast is determined not only by the perturbation of the forecast recurrence vector (7), but also by the perturbation of the reconstructed time series on which gives the initial values for the forecast recurrence. The signal of the unperturbed time series is reconstructed, by diagonal averaging, from the sum of the first r elementary matrices $\mathbf{X}_k(0) = \eta_k \eta_k^T \mathbb{X}$ of the decomposition of the trajectory matrix $\mathbb{X}$

$$\mathbf{X} = \sum_{k=1}^{r} \mathbf{X}_k(0). \tag{18}$$

Correspondingly, for the first order reconstruction for the perturbed time series, the elementary matrices are calculated from the perturbed eigenvectors $\gamma_{\sigma,i}$,

$$\begin{aligned}
\mathbf{X}_k(\sigma) &= \gamma_{\sigma,k}\gamma_{\sigma,k}^T(\mathbb{X} + \sigma\mathbb{N}) \\
&= (\eta_k + \sigma \sum_{i=1, i\neq k}^{L} \alpha_{i,k}\eta_i + O(\sigma^2))(\eta_k + \sigma \sum_{i=1, i\neq k}^{L} \alpha_{i,k}\eta_i + O(\sigma^2))^T(\mathbb{X} + \sigma\mathbb{N}) \\
&= \mathbf{X}_k(0) + \sigma \left(\sum_{i=1, i\neq k}^{L} \alpha_{i,k}(\eta_i\eta_k^T + \eta_k\eta_i^T)\mathbb{X} + \eta_k\eta_k^T\mathbb{N} \right) + O(\sigma^2).
\end{aligned}$$

Again, to first order the added noise enters through the coefficients $\alpha_{j,k}$ only. Then

$$\mathbf{X}(\sigma) = \sum_{i=1}^{r} \mathbf{X}_i(\sigma)$$

gives rise, after diagonal averaging, to the reconstructed time series $\widetilde{x}_n(\sigma)$. To calculate the forecast, we substitute the vector (14) into the linear recurrence formula and use the reconstruction series

$$\widetilde{x}_n(\sigma) = \widetilde{x}_n(0) + \sigma\widetilde{\varepsilon}_n + O(\sigma^2),$$

where $\sigma\widetilde{\varepsilon}_n$ arises from perturbation terms in elementary matrices $\mathbf{X}_k(\sigma)$. The forecast for the $(N+1)^{st}$ point is calculated from the linear recurrence formula (2) as

$$\begin{aligned}
\widetilde{x}_{N+1}(\sigma) =& \sum_{k=1}^{L-1} \big(a_k + \sigma(\mathbf{c}a_k + b_k) + O(\sigma^2))(\widetilde{x}_{N-k+1}(0) + \sigma\widetilde{\varepsilon}_{N-k+1} + O(\sigma^2)\big) \\
=& \sum_{k=1}^{L-1} a_k\widetilde{x}_{N-k+1}(0) + \sigma \sum_{k=1}^{L-1} \big((\mathbf{c}a_k + b_k)\widetilde{x}_{N-k+1}(0) + a_k\widetilde{\varepsilon}_{N-k+1}\big) + O(\sigma^2) \\
=& \widetilde{x}_{N+1}(0) + \sigma \sum_{k=1}^{L-1} \big((\mathbf{c}a_k + b_k)\widetilde{x}_{N-k+1}(0) + a_k\widetilde{\varepsilon}_{N-k+1}\big) + O(\sigma^2).
\end{aligned}$$

where we have written $R = (a_{L-1}, ..., a_1)^T$ and $\tilde{R} = (b_{L-1}, ..., b_1)^T$ for the unperturbed vector R (cf. (3)) and the vector $\tilde{R}$ (cf. (9)) respectively, and $\tilde{x}_{N+1}(0)$ is the unperturbed forecast.

Clearly,

$$\frac{var(\widetilde{x}_{N+1}(\sigma))}{\sigma^2} = var\left(\sum_{k=1}^{L-1}\left((\mathbf{c}a_k + b_k)\widetilde{x}_{N-k+1}(0) + a_k\widetilde{\varepsilon}_{N-k+1}\right)\right).$$

Note that coefficients b_k ($k \in \{1, ..., L-1\}$) and $\mathbf{c}$ are random variables due to the noise, which complicates the understanding of the output variance and makes the above formula rather inconclusive. Therefore we need to analyse the process of noise propagation in more detail in the following.

3 The flattening effect

In order to reach a better understanding of how the forecast variance depends on the variance of the input noise ε, we study the effect of the perturbation at three stages separately. These stages are: projector construction, time series reconstruction and forecast. Firstly, we deal with the noise propagation at the stage of constructing the projector, obtained from perturbed eigenvector components. Second stage is reconstruction, where the noise comes through the Hankel matrix and perturbed projector. And finally, the forecast, where the noise comes in through the perturbed recurrence vectors and through reconstruction. On each stage we assess the effect size of noise and see if any of these effects are dominant, so that others could be neglected in comparison.

Univariate case

The perturbation of the time series affects the reconstruction of the time series x_n directly and linearly through the Hankel matrix σN, and indirectly and nonlinearly by way of the perturbed eigenvectors γ_σ in the spectral projection of the Hankel matrix $\mathbb{X} + \sigma$N.
The reconstruction resulting from elementary matrices

$$\eta_i\eta_i^T\mathbb{X}, \quad i \in \{1, ..., r\}, \tag{19}$$

is the signal reconstruction $\widetilde{x}^{(1)}$ with no perturbation either in the time series or in the eigenvectors η_i, i.e. the result of SSA of the unperturbed time series.

The reconstruction $\widetilde{x}^{(2)}$ resulting from the elementary matrices

$$\gamma_{\sigma,i}\gamma_{\sigma,i}^T(\mathbb{X} + \sigma\mathrm{N}), \quad i \in \{1, ..., r\}, \tag{20}$$

is the reconstruction with the double perturbation effect in eigenvectors and projected Hankel matrix, i.e. this is the result of SSA of the perturbed time series. Both $\widetilde{x}^{(1)}$ and $\widetilde{x}^{(2)}$ are standard SSA reconstructions of time series x_n and $x_n + \sigma\varepsilon_n$, respectively.

The following two constructions are hybrids we use to study the influence of the perturbation. The reconstruction series $\widetilde{x}^{(3)}$ is based on the effect of the direct Hankel matrix perturbation only, but using the unperturbed eigenvectors,

$$\eta_i\eta_i^T(\mathbb{X} + \sigma\mathrm{N}), \quad i \in \{1, ..., r\}, \tag{21}$$

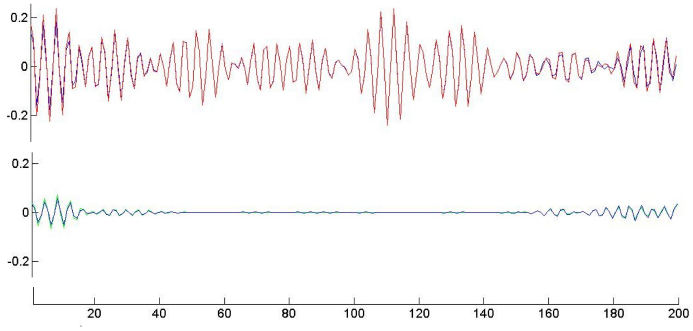

Figure 1: Reconstructions differences for the model (23) with perturbation $\sigma\varepsilon_n \sim N(0, 0.25)$: $\widetilde{x}^{(1)} - \widetilde{x}^{(3)}$, $\widetilde{x}^{(2)} - \widetilde{x}^{(4)}$ (top); $\widetilde{x}^{(2)} - \widetilde{x}^{(4)}$, $\widetilde{x}^{(2)} - \widetilde{x}^{(3)}$ (bottom).

and the reconstructed series $\widetilde{x}^{(4)}$ uses elementary matrices resulting from perturbed eigenvectors, but applied to the unperturbed time series

$$\gamma_{\sigma,i}\gamma_{\sigma,i}^T\mathbb{X}. \tag{22}$$

For a first experiment, we use the generated time series

$$x_n = \sin(\frac{3\pi}{2}n) + \sin(\frac{\pi}{2}n) \qquad n \in \{1, ..., 200\}, \tag{23}$$

perturbing it with Gaussian i.i.d. noise $\sigma\varepsilon_n \sim N(0, 0.25)$, and performing SSA with $L = 50$ and $r = 4$.

Figure 1 illustrates that

$$\widetilde{x}_n^{(2)} - \widetilde{x}_n^{(3)} \approx 0 \text{ for } n \in \{L+1, ..., N-L\}, \tag{24}$$

i.e. the difference is negligibly small through the whole reconstruction, apart from the first and last length L interval. Therefore, we see that the difference of the sum of elementary matrices $\sum_{i=1}^r \gamma_i\gamma_i^T(\mathbb{X} + \sigma\mathbb{N}) - \sum_{i=1}^r \eta_i\eta_i^T(\mathbb{X} + \sigma\mathbb{N})$ is approximately zero after Hankelization. That suggests that the difference in the eigenvectors is small and may be neglected in comparison with the effect of the perturbation coming through the Hankel matrix $\mathbb{X} + \sigma\mathbb{N}$ directly.

Similar behaviour is observed for the difference

$$\widetilde{x}_n^{(4)} - \widetilde{x}_n^{(1)} \approx 0 \text{ for } n \in \{L+1, ..., N-L\}, \tag{25}$$

which is approximately zero through all series, apart from the first and last L interval.

The observation indicates that the change in perturbed eigenvectors γ_σ is not crucial in the main (central) part of time series. One could think that this holds because the time series in this example had a simple structure and was perturbed with generated Gaussian white noise. However, we also find this flattening effect with real data.

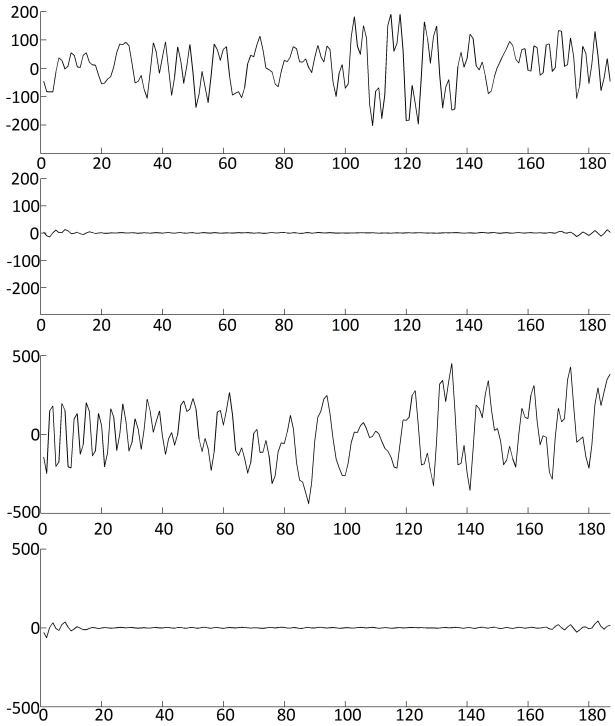

Figure 2: Australian dry wine sales, $L = 24$, $r = 5$: $(x^{(2)} - x^{(1)})$ (above), $(x^{(4)} - x^{(1)})$ (below). Noise is Gaussian N(0,4000) (top two graphs) or permuted residuals (bottom two graphs).

The real data presented here is monthly sales of dry Australian wine for the period 1980 - 1994 [9]. The analysis of the dry wine time series is done as in [4, Chapter II], where the same time series was used. The length of the series is $N = 187$ and the natural period is equal to one year, i.e. 12 months (12 data points). It is therefore natural to choose a multiple of 12 for the window length. According to the book [4, p. 138], the optimal window length to obtain the structure of the time series is $L = 24$ and the number of eigentriples which correspond to the trend and main periodics is $r = 5$.

Here we consider the unperturbed signal time series to be based on the first 5 eigentriples, and the added noise term is either generated Gaussian noise or randomly permuted residuals from the SSA reconstruction (so statistically independent, but not Gaussian).

Figure 3 illustrates the differences of reconstructions for both experiments, where the difference between $x^{(1)}$ and $x^{(4)}$ is approximately zero in both cases in the central part of the reconstruction and is oscillating in the first and last window length interval.

We also investigated the effect of the perturbed eigenvectors on the recurrence vector used for forecasting. In the example shown in Figure 3, the unperturbed signal time series is based on the first 7 eigentriples ($L = 60$) of the red wine sales time series ([9]), and the added noise term is either generated Gaussian noise or randomly permuted residuals from the SSA. We generated random

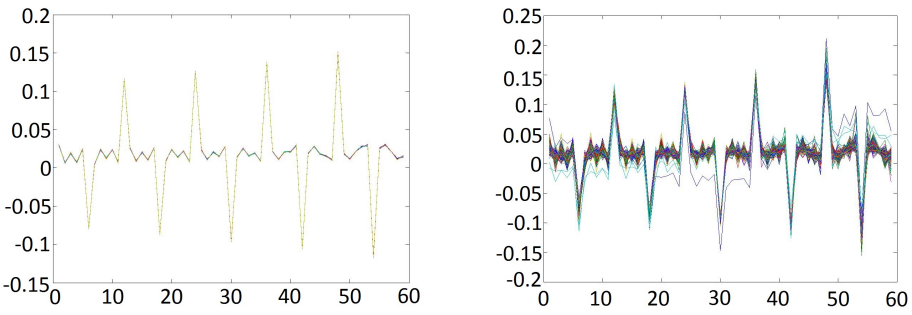

Figure 3: Australian red wine sales, $L = 60$, $r = 7$: Recurrence vector with generated noise $\varepsilon \sim N(0, 100)$ (left) vs. permuted residuals (right); 250 random instances in each graph.

(Gaussian or permuted residuals) noise 250 times to study the stability of the recurrence vector under perturbations. In the case of Gaussian white noise (left in Figure 3), the recurrence vector is remarkably stable and can be considered as practically independent on noise; the permuted natural noise (right Figure 3) gives greater variation of the recurrence vector, but the positions and heights of the characteristic spikes are very stable.

Bivariate case

In the bivariate case the main time series x_n is perturbed with $\sigma\varepsilon_n$, and y_n is a support series. Performing bivariate MSSA with parameters L, r, we study two signal reconstructions of the main series x_n: the reconstruction $\tilde{x}_n^{(1),MSSA}$ from diagonal averaging of grouped elementary matrices of

$$\eta^{MSSA}\left(\eta^{MSSA}\right)^T \begin{pmatrix} \mathbb{X} \\ \mathbb{Y} \end{pmatrix}$$

and $\tilde{x}_n^{(2),MSSA}$ from the perturbed series

$$\gamma^{MSSA}(\sigma)\left(\gamma^{MSSA}(\sigma)\right)^T \begin{pmatrix} \mathbb{X}+\sigma\mathbb{N} \\ \mathbb{Y} \end{pmatrix},$$

along with two hybrid cases, $\tilde{x}_n^{(3),MSSA}$ from $\eta^{MSSA}\left(\eta^{MSSA}\right)^T \begin{pmatrix} \mathbb{X}+\sigma\mathbb{N} \\ \mathbb{Y} \end{pmatrix}$ and $\tilde{x}_n^{(4),MSSA}$ from $\gamma^{MSSA}(\sigma)\left(\gamma^{MSSA}(\sigma)\right)^T \begin{pmatrix} \mathbb{X} \\ \mathbb{Y} \end{pmatrix}$. Considering a simple example with time series

$$x_n = y_n = \sin(\pi n\omega), \tag{26}$$

where $n \in \{1, ..., 200\}$, $\omega = 0.48$ and i.i.d. perturbation $0.5\varepsilon_n \sim N(0, 0.25)$, we calculated MSSA reconstructions and hybrid cases $\widetilde{x}_n^{(1),MSSA}, \widetilde{x}_n^{(2),MSSA}$, $\widetilde{x}_n^{(3),MSSA}, \widetilde{x}_n^{(4),MSSA}$ with $L = 10$, $r = 2$. Figure 4 illustrates that the difference

$$\widetilde{x}_n^{(4),MSSA} - \widetilde{x}_n^{(1),MSSA}$$

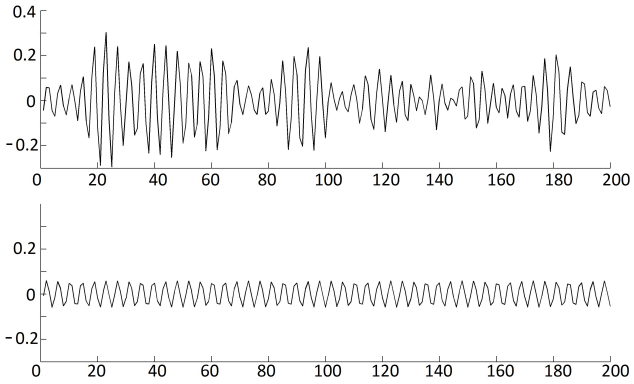

Figure 4: Reconstructions differences $\widetilde{x}_n^{(4),MSSA} - \widetilde{x}_n^{(1),MSSA}$ (top) of perturbed time series (26), in comparison to $\widetilde{x}_n^{(2),MSSA} - \widetilde{x}_n^{(1),MSSA}$, $L = 10$, $r = 2$

is not negligibly small as for the generated example for univariate case (see Figure 1), but is considerably smaller than $\widetilde{x}_n^{(3),MSSA} - \widetilde{x}_n^{(1),MSSA}$ and $\widetilde{x}_n^{(2),MSSA} - \widetilde{x}_n^{(1),MSSA}$.

4 A simplified model of forecast variance

We now proceed to derive formulae for the variance of SSA and MSSA forecasts of the randomly perturbed time series, calculated from the SSA and MSSA data of the *unperturbed* time series. They will give a simple and transparent model for the forecast variance. However, they will rely on the following two assumptions. Firstly, we make the simplifying assumption that the flattening effect observed in the preceding section occurs exactly, so that the SVD eigenvectors and the forecast recurrence vectors of the unperturbed time series can be used as proxies for those of the perturbed time series in the central part of the time series, i.e. omitting the first and last L entries, where L is the window length. Note that this requires L to be substantially smaller than the total length N of the time series, so the choice $L = N/2$, which is not uncommonly used in SSA, will be unsuitable. However, if the time series comprises several period intervals, then the period length or a multiple of it will be a natural choice for L, allowing to leave a sufficient central part of the series. Secondly, because of the restriction to the central part of the series and for reasons which will become apparent in the following, the formulae will apply not to a forecast at the end of the series, i.e. a future forecast, but to the forecast at the end of the central part of the series. This will not normally be a forecast of interest in itself; however, we here use the forecast as a tool to assess the supportiveness of a second time series, and for this specific purpose a forecast from an interior part of the time series will be suitable.

The formulae (47), (54) for the SSA and MSSA forecast variance will be based on the following reformulation of SSA reconstruction as a (mid-point) linear filter; for this aspect of SSA, see also [1, 8] and [5], Section 3.9.

Proposition 4.1. *Let $f = (f_1, ..., f_N)$ be a time series and $\widetilde{f} = (\widetilde{f}_1, ..., \widetilde{f}_N)$ its SSA signal reconstruction for a suitable choice of parameters L and r. Then*

$$\widetilde{f_n} = (q \star f)_n = \sum_{m=-L+1}^{L-1} q_m f_{n+m}, \qquad n \in \{L+1, ..., N-L\}, \tag{27}$$

where the reconstruction kernel $q : \{-L+1, ..., L-1\} \to \mathbb{R}$ has the symmetry property

$$q_{-m} = q_m, \qquad m \in \{-L+1, ..., L-1\}. \tag{28}$$

The convolution formula (27) gives the exact same result as the standard SSA reconstruction, except in the first and last L terms. In fact, the convolution concept is more natural to doubly-infinite time series $f = (f_j)_{j \in \mathbb{Z}}$. Our approach to the proof is to derive the convolution formula for the doubly-infinite case; the fact that the series f is finite (and extended to be doubly-infinite by padding zeros) is invisible from the central part due to the finite support of q. However, for points in the first and last window length, the convolution will in general be different from the result of standard SSA, as it gives the SSA of the time series extended by 0.

Proof. Consider a doubly infinite time series $f = (f_j)_{j \in \mathbb{Z}}$ and the right shift operator $S : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$, $(Sf)_t = f_{t+1} (t \in \mathbb{Z})$; then the trajectory matrix of infinite time series f with window length L is

$$\mathbb{X} = \begin{pmatrix} f \\ Sf \\ \vdots \\ S^{L-1} f \end{pmatrix} \in \mathbb{R}^{L \times \mathbb{Z}}, \tag{29}$$

and the lag-covariance matrix takes the form

$$\mathbb{X}\mathbb{X}^T = \begin{pmatrix} ff^T & f(Sf)^T & \dots & f(S^{(L-1)}f)^T \\ (Sf)f^T & Sf(Sf)^T & \dots & Sf(S^{(L-1)}f)^T \\ \vdots & & & \\ (S^{(L-1)}f)f^T & S^{(L-1)}f(Sf)^T & \dots & S^{(L-1)}f(S^{(L-1)}f)^T \end{pmatrix} \in \mathbb{R}^{L \times L}, \tag{30}$$

assuming that the inner products are finite. In fact, the matrix $\mathbb{X}\mathbb{X}^T$ is a Toeplitz matrix, as

$$(S^j f)(S^k f)^T = (S^{(j-k)} f) f^T \qquad (j, k \in \mathbb{Z}). \tag{31}$$

The SSA projector P calculated from the first r orthonormal eigenvectors $\eta_1, ..., \eta_r$ of $\mathbb{X}\mathbb{X}^T$ has matrix entries

$$p_{k,j} = \sum_{i=1}^{r} \eta_{i,k} \eta_{i,j} \quad (k, j \in \{1, ..., L\}), \tag{32}$$

where $\eta_{i,j}$ is the j^{th} entry of η_i. Note that

$$p_{k,j} = \sum_{i=1}^{r} \eta_{i,k} \eta_{i,j} = \sum_{i=1}^{r} \eta_{i,j} \eta_{i,k} = p_{j,k}. \tag{33}$$

Therefore, applying the SSA projector to the Hankel matrix, we get

$$P\mathbb{X} = \begin{pmatrix} \sum_{j=0}^{L-1} p_{0,j} S^j f \\ \sum_{j=0}^{L-1} p_{1,j} S^j f \\ \vdots \\ \sum_{j=0}^{L-1} p_{L-1,j} S^j f \end{pmatrix}, \tag{34}$$

The Hankelization of this matrix of L rows and infinitely many columns takes the simple form of a linear operator.

$$\mathbb{H} : \mathbb{R}^{L\times\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}, \qquad \mathbb{H}\begin{pmatrix} y_0 \\ \vdots \\ y_{L-1} \end{pmatrix} = \frac{1}{L}\sum_{k=0}^{L-1} S^{-k} y_k. \tag{35}$$

Hence, the reconstructed signal is

$$\widetilde{f} = \mathbb{H}P\mathbb{X} = \frac{1}{L}\sum_{k=0}^{L-1} P\mathbb{X} = \frac{1}{L}\sum_{k=0}^{L-1}\sum_{j=0}^{L-1} p_{k,j} S^{j-k} f = \sum_{m=-L}^{L} q_m S^m f, \tag{36}$$

i.e.

$$\widetilde{f}_n = \sum_{m=-L+1}^{L-1} q_m f_{n+m} \qquad (n \in \mathbb{Z}),$$

where

$$q_m = \begin{cases} \dfrac{1}{L}\displaystyle\sum_{i=0}^{L-1} p_{i,i}, \text{ if } m = 0, \\ \dfrac{1}{L}\displaystyle\sum_{i=m}^{L-1} p_{i-m,i}, \text{ if } m \in (1, ..., L-1), \\ \dfrac{1}{L}\displaystyle\sum_{i=-m}^{L-1} p_{i,m+i}, \text{ if } m \in (-L+1, ..., -1). \end{cases}$$

As the calculation of convolution coefficients q_m is based on summing projector elements (32), which have the symmetry (33), the q_m are symmetric as well,

$$q_m = q_{-m} \quad (m \in \{-L+1, ..., L-1, L\}). \tag{37}$$

Setting $q_m = 0$ if $|m| > L$, the expression (36) can be rewritten as a convolution

$$\widetilde{f}_n = \sum_{m\in\mathbb{Z}} q_{n-m} f_m = (q \star f)_n.$$

□

Proposition 4.2. *Let $f = (f_1, ..., f_N)$, $g = (g_1, ..., g_N)$ be time series and $\widetilde{f} = (\widetilde{f}_1, ..., \widetilde{f}_N)$, $\widetilde{g} = (\widetilde{g}_1, ..., \widetilde{g}_N)$ their MSSA signal reconstructions, respectively, for a suitable choice of window length L and r eigentriples. Then there exists a convolution representation for the central part of the reconstructed time series*

$$\begin{pmatrix} \widetilde{f} \\ \widetilde{g} \end{pmatrix} = \begin{pmatrix} q^1 \star f + q^2 \star g \\ q^3 \star f + q^4 \star g \end{pmatrix}, \tag{38}$$

where the reconstruction kernels $q^i : \{-L+1, ..., L-1\} \to \mathbb{R}$ for $i \in \{1, 2, 3, 4\}$ depend on the MSSA parameters L, r.

Proof. Extend the given time series by 0 to doubly-infinite time series $(f_n)_{n\in\mathbb{Z}}$, $(g_n)_{n\in\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$.

The Hankel matrix for the bivariate MSSA is a block matrix consisting of the Hankel matrix $\mathbb{X}$ of f_n stacked on top of the Hankel matrix $\mathbb{Y}$ of g_n,

$$\begin{pmatrix} \mathbb{X} \\ \mathbb{Y} \end{pmatrix} = \begin{pmatrix} f \\ Sf \\ \vdots \\ S^{L-1}f \\ g \\ Sg \\ \vdots \\ S^{L-1}g \end{pmatrix} \in \mathbb{R}^{2L\times\mathbb{Z}}. \tag{39}$$

The MSSA projector P is obtained from the first r eigenvectors $\eta_1, ..., \eta_n$ of the $2L \times 2L$ lag-covariance matrix

$$\begin{pmatrix} \mathbb{X} \\ \mathbb{Y} \end{pmatrix} \begin{pmatrix} \mathbb{X} \\ \mathbb{Y} \end{pmatrix}^T = \begin{pmatrix} \left((S^{i-1}f)(S^{j-1}f)\right)_{i,j=1}^{L} & \left((S^{i-1}f)(S^{j-1}g)\right)_{i,j=1}^{L} \\ \left((S^{i-1}g)(S^{j-1}f)\right)_{i,j=1}^{L} & \left((S^{i-1}g)(S^{j-1}g)\right)_{i,j=1}^{L} \end{pmatrix}. \tag{40}$$

The projector P consists of 4 blocks

$$P = \begin{pmatrix} p^1 & p^2 \\ p^3 & p^4 \end{pmatrix},$$

where

$$\begin{aligned} p^1_{k,j} &= \sum_{i=1}^{r} \eta_{i,k}\eta_{i,j} \quad (k, j \in \{0, ..., L-1\}), \\ p^2_{k,j} &= \sum_{i=1}^{r} \eta_{i,k}\eta_{i,j} \quad (k \in \{0, ..., L-1\}, j \in \{L, ..., 2L-1\}), \\ p^3_{k,j} &= \sum_{i=1}^{r} \eta_{i,k}\eta_{i,j} \quad (k \in \{L, ..., 2L-1\}, j \in \{0, ..., L-1\}), \\ p^4_{k,j} &= \sum_{i=1}^{r} \eta_{i,k}\eta_{i,j} \quad (k, j \in \{L, ..., 2L-1\}). \end{aligned} \tag{41}$$

Similarly to the univariate case, we now can write down the expression for the signal reconstruction of f_n and g_n, using the Hankelization operator $\mathbb{H}$ of

(35)

$$\begin{pmatrix}\widetilde{f}\\ \widetilde{g}\end{pmatrix} = \begin{pmatrix}\mathbb{H}\left(\sum_{k=0}^{L-1}\sum_{k=0}^{L-1} p^1_{k,j}S^{j-k}f + \sum_{k=0}^{L-1}\sum_{k=0}^{L-1} p^2_{k,j}S^{j-k}g\right)\\ \mathbb{H}\left(\sum_{k=0}^{L-1}\sum_{k=0}^{L-1} p^3_{k,j}S^{j-k}f + \sum_{k=0}^{L-1}\sum_{k=0}^{L-1} p^4_{k,j}S^{j-k}g\right)\end{pmatrix} \tag{42}$$

$$= \begin{pmatrix}\sum_{m=-L}^{L} q^1_m S^m f + \sum_{m=-L}^{L} q^2_m S^m g\\ \sum_{m=-L}^{L} q^3_m S^m f + \sum_{m=-L}^{L} q^4_m S^m g\end{pmatrix}, \tag{43}$$

where

$$q^j_m = \begin{cases}\frac{1}{L}\sum_{i=0}^{L-1} p^j_{i,i}, \text{ if } m = 0,\\ \frac{1}{L}\sum_{i=m}^{L-1} p^j_{i-m,i}, \text{ if } m \in (1,...,L-1),\\ \frac{1}{L}\sum_{i=-m}^{L-1} p^j_{i,m+i}, \text{ if } m \in (-L+1,...,-1),\end{cases} \tag{44}$$

and $q^j_m = 0$ otherwise. Thus

$$\begin{pmatrix}\widetilde{f_n}\\ \widetilde{g_n}\end{pmatrix} = \begin{pmatrix}\sum_{m\in\mathbb{Z}} q^1_{n-m}f_m + \sum_{m\in\mathbb{Z}} q^2_{n-m}g_m\\ \sum_{m\in\mathbb{Z}} q^3_{n-m}f_m + \sum_{m\in\mathbb{Z}} q^4_{n-m}g_m\end{pmatrix} = \begin{pmatrix}(q^1\star f)_n + (q^2\star g)_n\\ (q^3\star f)_n + (q^4\star g)_n\end{pmatrix} \ (n\in\mathbb{Z}) \tag{45}$$

or briefly

$$\begin{pmatrix}\widetilde{f}\\ \widetilde{g}\end{pmatrix} = \begin{pmatrix}q^1\star f + q^2\star g\\ q^3\star f + q^4\star g\end{pmatrix}. \tag{46}$$

The expression (46) gives the reconstructions $\widetilde{f}_n$, $\widetilde{g}_n$ of the main series f_n and support series g_n for $n \in \{L+1,...,N-L\}$. This expression is identical to the MSSA reconstructions, provided that the parameters of both procedures are the same. □

It is worth mentioning that convolution vectors q^1 and q^4 have the reflection symmetry (37), which q has in the univariate case, but q^2, q^3 do not. Instead, they are related via $q^2_m = q^3_{-m} (m \in \mathbb{Z})$.

After these preparations, we are ready to study the forecast for x_{N-L+1} based on the SSA reconstruction for $\{\tilde{x}_{L+1},...,\tilde{x}_{N-L}\}$ from the convolution formula. The expression (27) is conveniently linear and concise; however, both convolution kernel $q_i(\sigma)$ and the recurrence vector $R(\sigma)$ will, strictly speaking,

depend on noise of the initial time series, which complicates the process of estimating the variance of the noise in the reconstruction and in the forecast. Therefore we make the simplifying assumption that these can be replaced with the convolution kernel and recurrence vector of the unperturbed series, as indicated by our observations in Section 4. Then we obtain the following result in the univariate (SSA) case.

Proposition 4.3. *Let R, q be the recurrence vector and convolution kernel, respectively, of SSA performed on a time series $(x_n)_{n\in\{1,\dots,N\}}$ with parameters L, r.*

Moreover, assume that $(\varepsilon_n)_{n\in\{1,\dots,N\}}$ is a random time series such that the ε_n are independent and have variance 1. Let $\hat{x}_{N-L+1}(\sigma)$ be the SSA forecast of the time series $(x_n+\sigma\varepsilon_n)_{n\in\{1,\dots,N\}}$ calculated from the reconstructed $(\tilde{x}_{N-2L+2}(\sigma),\dots,\tilde{x}_{N-L}(\sigma))$.

Then, assuming the recurrence vector $R(\sigma)$ and convolution kernel $q(\sigma)$ are equal to the unperturbed R, q,

$$var(\tilde{x}_{N-L+1}(\sigma)) = \sigma^2\|R\star q\|_2^2. \tag{47}$$

Here $\|y\|_2=\sqrt{\sum_j |y_j|^2}$ is the ℓ^2-norm.

Proof. Using the convolution expression (27) for the SSA reconstruction of the perturbed time series $x_n+\sigma\varepsilon_n$, we find

$$\begin{aligned}\widetilde{x}_n(\sigma) &= \sum_{m\in\mathbb{Z}} q_{n-m}(\sigma)(x_m+\sigma\varepsilon_m)\\ &= \sum_{m\in\mathbb{Z}} q_{n-m}(\sigma)x_m + \sigma\sum_{m\in\mathbb{Z}} q_{n-m}(\sigma)\varepsilon_m.\end{aligned} \tag{48}$$

Applying the LRF (2) to the reconstruction (48), the expression for the forecast $\hat{x}_{N-L+1}(\sigma)$ may be written, using $R(\sigma)=(a_{L-1}(\sigma),\dots,a_1(\sigma))^T$ for the perturbed recurrence vector, as

$$\hat{x}_{N-L+1}(\sigma) = \sum_{i=1}^{L-1} a_i(\sigma)\left(\sum_{m\in\mathbb{Z}} q_{N+1-i-m}(\sigma)x_m + \sigma\sum_{m\in\mathbb{Z}} q_{N+1-i-m}(\sigma)\varepsilon_m\right) \tag{49}$$

$$= \sum_{i=1}^{L-1} a_i(\sigma)\sum_{m\in\mathbb{Z}} q_{N+1-i-m}(\sigma)x_m + \sigma\sum_{i=1}^{L-1} a_i(\sigma)\sum_{m\in\mathbb{Z}} q_{N+1-i-m}(\sigma)\varepsilon_m. \tag{50}$$

Under the simplifying assumptions, (50) turns into

$$\hat{x}_{N-L+1}(\sigma) = \sum_{i=1}^{L-1} a_i\sum_{m\in\mathbb{Z}} q_{N-L+1-i-m}x_m + \sigma\sum_{i=1}^{L-1} a_i\sum_{m\in\mathbb{Z}} q_{N-L+1-i-m}\varepsilon_m \tag{51}$$

$$= \hat{x}_{N-L+1}(0) + \sigma\sum_{i=1}^{L-1} a_i\sum_{m\in\mathbb{Z}} q_{N-L+1-i-m}\varepsilon_m \tag{52}$$

$$= \hat{x}_{N-L+1}(0) + \sigma\sum_{m\in\mathbb{Z}}\Big(\sum_{i=1}^{L-1} a_i q_{N-L+1-i-m}\Big)\varepsilon_m. \tag{53}$$

As ε_m are i.i.d. with variance 1, we hence conclude that

$$var(\hat{x}_{N-L+1}) = \sigma^2 \sum_{m\in\mathbb{Z}} \left| \sum_{i=1}^{L-1} a_i q_{N-L+1-i-m} \right|^2 = \sigma^2 \sum_{k\in\mathbb{Z}} \left| \sum_{i=1}^{L-1} a_i q_{k-i} \right|^2$$
$$= \sigma^2 \sum_{k\in\mathbb{Z}} |(R \star q)_i|^2 = \sigma^2 \|R \star q\|_2^2.$$

□

Similarly, we have the following result in the bivariate (MSSA) case, including the additional support series y_n.

Proposition 4.4. *Let R_{11}, R_{12} and q^i, $i \in \{1,2,3,4\}$ be the recurrence vectors and convolution kernels, respectively, of MSSA performed on $(x_n)_{n\in\{1,...,N\}}$, $(y_n)_{n\in\{1,...,N\}}$ with parameters L, r.*

Assume that $(\varepsilon_n)_{n\in\{1,...,N\}}$ is a random time series such that the ε_n are independent and have variance 1. Let $\hat{x}_{N-L+1}(\sigma)$ be the MSSA forecast of the time series $(x_n + \sigma\varepsilon_n)_{n\in\{1,...,N\}}$ calculated from the reconstructed $(\tilde{x}_{N-2L+2}(\sigma), ..., \tilde{x}_{N-L}(\sigma))$ and $(\tilde{y}_{N-2L+2}(\sigma), ..., \tilde{y}_{N-L}(\sigma))$.

Then, assuming the recurrence vectors $R_{11}(\sigma)$, $R_{12}(\sigma)$ and convolution kernels $(q^i(\sigma), i \in \{1,2,3,4\}$ are equal to the unperturbed R_{11}, R_{12} and q^i, $i \in \{1,2,3,4\}$,

$$var(\hat{x}_{N-L+1}(\sigma)) = \sigma^2 \|R_{11} \star q^1 + R_{12} \star q^3\|_2^2. \tag{54}$$

Proof. Recalling the MSSA LRF for the forecast (4) and its recurrence vectors R_{11}, R_{12} (5), we derive the forecast $\hat{x}_{N-L+1}(\sigma)$ from the following formula

$$\hat{x}_{N-L+1}(\sigma) = \sum_{i=1}^{L-1} a_{1,i} \widetilde{x}_{N-L+1-i}(\sigma) + \sum_{i=1}^{L-1} a_{2,i} \widetilde{y}_{N-L+1-i}(\sigma), \tag{55}$$

where $R_{11} = (a_{L-1,i}, ..., a_{1,1})^T$ and $R_{12} = (a_{2,L-1}, ..., a_{1,1})^T$.

Note that we are operating under the simplifying assumptions that recurrence vectors R_{11} and R_{12} are obtained from the unperturbed time series x_n, i.e. fixed.

Substituting reconstructions (45) into (55) we get, using $x_m(\sigma) = x_m + \sigma\varepsilon_m$,

$$\begin{aligned}\hat{x}_{N-L+1}(\sigma) &= \sum_{i=1}^{L-1} a_{1,i} \Big(\sum_{m\in\mathbb{Z}} q^1_{N-L+1-i-m} x_m(\sigma) + \sum_{m\in\mathbb{Z}} q^2_{N-L+1-i-m} y_m \Big) \\ &\quad + \sum_{i=1}^{L-1} b_{1,i} \Big(\sum_{m\in\mathbb{Z}} q^3_{N-L+1-i-m} x_m(\sigma) + \sum_{m\in\mathbb{Z}} q^4_{N-L+1-i-m} y_m \Big) \\ &= \hat{x}_{N-L+1}(0) + \sigma \sum_{m\in\mathbb{Z}} \sum_{i=1}^{L-1} (a_{1,i} q^1_{N-L+1-i-m} + b_{1,i} q^3_{N-L+1-i-m}) \varepsilon_m.\end{aligned}$$

Similarly to the univariate case, we can now calculate the variance of the fore-

cast,

$$\begin{aligned} var(\hat{x}_{N-L+1}(\sigma)) &= \sigma^2 \sum_{m\in\mathbb{Z}} \left| \sum_{i=1}^{L-1} (a_{1,i} q^1_{N-L+1-i-m} + b_{1,i} q^3_{N-L+1-i-m}) \right|^2 \\ &= \sigma^2 \sum_{k\in\mathbb{Z}} \left| \sum_{i=1}^{L-1} (a_{1,i} q^1_{k-i} + b_{1,i} q^3_{k-i}) \right|^2 \\ &= \sigma^2 \sum_{k\in\mathbb{Z}} \left| R_{11} \star q^1 + R_{12} \star q^3 \right| \\ &= \sigma^2 \| R_{11} \star q^1 + R_{12} \star q^3 \|_2^2. \end{aligned}$$

□

5 Towards a measure of supportiveness

In this section, we aim to explore the possibility of using the forecast variance as a tool to establish a relationship between two time series in the sense that the second (support) series stabilises the forecast of the primary series. This will roughly correspond to Granger's concept of causality. In particular, we address the question of what role the convolution formula (54) based on the simplifying assumption of constant reconstruction kernel and forecast vectors can play as a predictor for the empirical variance one would observe in random trials of perturbed time series.

The comparison of variances bears some similarity to the statistical F-test; however, our question does not directly fit into the scheme of that test, as here the sample size (the number of random trials) is arbitrary and will not strongly affect the variance. Hence the significance of the test cannot be reasonably calculated, and a single number will not be very indicative. Therefore we consider the *change* in the variance as the relative weight of the support series in bivariate MSSA is increased. More precisely, we consider the pair of time series $(x_n)_{n\in\{1,\dots,N\}}$ (main series) and $(\rho y_n)_{n\in\{1,\dots,N\}}$ (support series) with varying *support series multiplier* ρ and study the dependence of the variance on ρ.

Given this pair of time series, we compare the value for the predicted forecast variance calculated from the convolution formula (54) for the original series $(x_n)_{n\in\{1,\dots,N\}}$, $(\rho y_n)_{n\in\{1,\dots,N\}}$, with the empirical variance from 1000 random trials of the MSSA forecast for the time series $(x_n + \epsilon_n)_{n\in\{1,\dots,N\}}$ and $(\rho y_n)_{n\in\{1,\dots,N\}}$, where $\epsilon_n \sim N(0,\sigma^2)$ is independent pseudo-random generated noise in each trial. In order to assess the impact of our simplifying assumptions in more detail, we also consider the empirical variance of the MSSA forecast where the forecast vector $R_{11}(\sigma)$ is, in each trial, replaced with the fixed unperturbed forecast vector R_{11} (values $\mathcal{A}$), and where both forecast vectors $R_{11}(\sigma)$ and $R_{12}(\sigma)$ are replaced with the fixed unperturbed forecast vectors R_{11}, R_{12} (values $\mathcal{B}$). Note, however, that in these cases the convolution kernel will not be kept fixed, but will vary with the random perturbations.

We analysed two examples, with time series taken from the Australian wine datasets [9]. The main series in both examples is based on the red wine sales time series, the support series in the first example is based on the sparkling wine sales time series, in the second example it is made up of unrelated generated

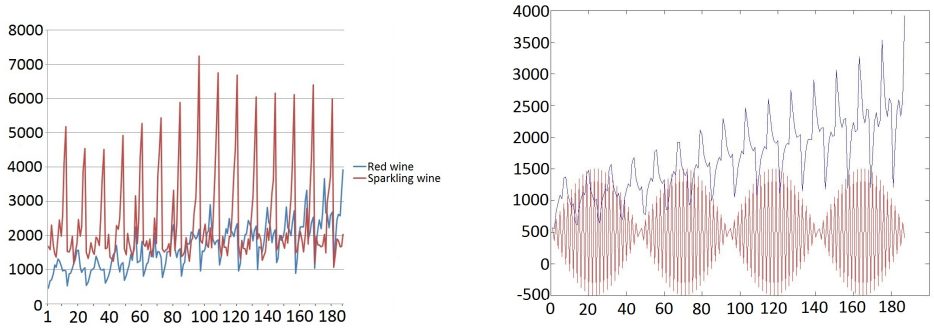

Figure 5: Red wine vs sparkling wine and generated series, 1980-1994.

data. The length of all time series is $N = 187$. The red wine and sparkling wine time series clearly exhibit some structural similarities as well as characteristic differences, see Figure 5 (left).

We constructed time series $(x_n)_{n\in\{1,\dots,187\}}$, $(y_n)_{n\in\{1,\dots,187\}}$ by preconditioning the raw red and sparkling wine time series, respectively, performing SSA with $L = 60$ and $r = 7$ separately on both series; these are the optimal SSA parameters for these standard series (see [4] pp. 138–139). The variance for the perturbative Gaussian white noise added to $(x_n)_{n\in\{1,\dots,187\}}$ in the trials was set to $\sigma^2 = 100$. For the further analyses, we kept the (M)SSA parameters at $L = 60$, $r = 7$. For a fair comparison with the convolution formula, we forecast the value starting the last window in the time series, i.e. $\hat{x}_{128}$, from the reconstructed $\tilde{x}_{68}, \dots, \tilde{x}_{127}$ and $\tilde{y}_{68}, \dots, \tilde{y}_{127}$.

To put the absolute variances calculated for the MSSA forecasts into perspective, we computed the perturbative variance of the SSA forecast of the red wine series $(x_n)_{n\in\{1,\dots,187\}}$ alone. The empirical variance from 1000 trials is $var(\hat{x}_{128}^{SSA}) = 7.9991$. When the perturbed recurrence vector $R(\sigma)$ is replaced with the unperturbed constant R in each trial, the empirical variance is 7.2037. The convolution formula (47) gives the value 6.7355, showing that a large part (although not all) the observed variance can be explained even making the simplifying assumptions.

The results for the first example (red wine sales with sparkling wine sales as support) are shown in Table 1. The empirical variance of forecast values $\hat{x}_{128}$ appears largely stable, apart from the conspicuously large value for $\rho = 0.18$, which we shall discuss below. The variance of forecasts with fixed R_{11} (values $\mathcal{A}$) behaves in a generally similar manner. On the other hand, the variance of forecasts with both R_{11} and R_{12} fixed (values $\mathcal{B}$) shows are marked and sustained decrease with increasing ρ. This is fully paralleled by the prediction from the convolution formula (54).

Thus it appears that the instability under perturbation of $R_{12}(\sigma)$ has a pronounced effect on the result variance, while R_{11} and the reconstruction convolution kernel can be assumed to be unperturbed without qualitative difference in the results.

It is striking in Table 1 that all values obtained from simulated noise trials show very unstable behaviour near $\rho = 0.18$ (values in bold). This instability is

Table 1: MSSA Red Wine Forecast Measures

ρ	$var(\hat{x}_{128})$	$var(\mathcal{A})$	$var(\mathcal{B})$	$\sigma^2\lvert R^s_{11} * q^1 + R^s_{12} * q^3\rvert$
0.1	7.4768	6.6278	6.5749	5.6975
0.17	8.1574	**114.5941**	7.1851	4.5769
0.18	**3898.9**	**3146.2**	**997.9766**	2.6169
0.19	7.3854	**194.1092**	3.6954	2.5575
0.2	6.7346	**30.4935**	3.2459	2.4359
0.22	5.9248	9.5182	3.1407	2.1989
0.3	4.7625	5.1692	1.9477	1.5222
0.4	4.6748	4.5712	1.2182	1.0386
0.5	4.7254	4.371	0.7938	0.7341
0.6	4.8085	4.3115	0.537	0.5253
0.7	4.8949	4.3248	0.3738	0.3792
0.8	4.978	4.3794	0.2663	0.2764
0.9	5.0574	4.4572	0.1934	0.2039
1	5.1319	4.5463	0.1429	0.1523
1.1	5.1994	4.6383	0.1073	0.1152

caused by an uncertainty in the relative position, within the pair of the 7th and 8th, of the eigenvalues corresponding to the continuous branches of eigenvectors. For ρ near 0.18, the 7th and 8th eigenvalues happen to lie very closely together. As the MSSA cut-off was fixed between the 7th and 8th eigenvalue, a small change due to the perturbation can lead to a swap of relative position of these eigenvalues and hence to a large change in the elementary MSSA matrix formed from a completely different nearly orthogonal eigenvector. The instability is thus easily explained as a meaningless artefact of the support series scaling and can therefore be disregarded. Note that the values from the convolution formula do not have this problem, as they are calculated from unperturbed series.

For the second example, we used the same main series as before, but the unrelated generated support series

$$y_n = 500 + 1000 \sin 2\pi \frac{277}{566} n \qquad (n \in \{1, \ldots, 187\}).$$

Figure 5 (right) shows both time series.

The resulting variances are shown in Table 2. The empirical variance of the forecast $\hat{x}_{128}$ is very stable and shows no essential increase or decrease. The variances for fixed R_{11} (values $\mathcal{A}$) even have an increasing trend. Here also the empirical variances for fixed R_{11} and R_{12} (values $\mathcal{B}$) are very stable, showing only very faint decrease, especially for $\rho > 0.5$, even if the multiplier ρ is pushed to extremely high values. This also holds for the values calculated from the convolution formula, which here are excellent proxies for the values $\mathcal{B}$.

In summary, we see that the convolution formula (54) gives very good predictions of the values and, more importantly, qualitative behaviour for varying ρ, of the empirical variances where the recurrence vectors are fixed. However, these predictions do not in general reflect the empirical variance of the forecast of the perturbed time series very well.

On the one hand, this can be taken as an indication that our simplified model based on the assumption that the recurrence vectors do not essentially

Table 2: MSSA Red Wine Forecast Measures with support series

ρ	$var(\hat{x}_{128})$	$var(\mathcal{A})$	$var(\mathcal{B})$	$\sigma^2\lvert R^s_{11} * q^1 + R^s_{12} * q^3\rvert$
0.1	8.6397	8.4545	7.8798	7.8661
0.2	8.1232	7.952	7.5153	7.1994
0.3	7.7428	8.8551	7.1967	6.7942
0.4	7.5411	11.6653	6.9916	6.6424
0.5	7.4265	15.1699	6.8687	6.5881
0.6	7.3044	18.2412	6.793	6.5639
0.7	9.9216	22.2217	6.7272	6.5341
0.8	8.2548	22.6167	6.7453	6.5747
0.9	8.3286	24.0218	6.7349	6.5809
1	8.3883	25.0824	6.7274	6.5853
1.1	8.4366	25.8982	6.7218	6.5888
10	8.7161	30.0821	6.6967	6.6064
100	8.7201	30.1362	6.6964	6.6067

vary with the perturbation is oversimplified and unrealistic, and it is certainly not a suitable tool for a precise estimate of the actual forecast variance.

On the other hand, bearing in mind that we are using the forecast variance just as an experimental tool for assessing to what extent the support series helps improve the forecast of the main series, giving an indicator of supportiveness, our results suggest that the convolution formula (54), when considered for varying support series multiplier ρ, could well be used as such a tool. As it is calculated from the original unperturbed time series only, it is very quick and inexpensive to compute and does not suffer the cut-off instability observed in the empirical results. In short, we suggest the criterion for supportiveness that the convolution norm (54) becomes small for large support series multiplier ρ, compared to its value for small ρ (or the SSA convolution norm (47)), whereas supportiveness is rejected if it settles at a positive level for large ρ.

To explore this concept further, we consider the sparkling wine time series, using a simple cosine with 1 year period,

$$y_n = 3500 + 500 \cos \frac{2\pi}{12} n \qquad (n \in \{1, \dots, 187\}),$$

as a support series ($L = 60$, $r = 7$). The convolution norms (54), shown in Table 3 (a), settle at a positive level after a brief initial decrease, indicating lack of supportiveness.

For comparison, we then take the sparkling wine time series itself, without added noise, as a support series for the same main series and use the same MSSA parameters as above. In this case, the convolution norms tend to zero, see Table 3 (b). When this support series is shifted cyclically by 17 months, to become

$$(y_{18}, \dots, y_{187}, y_1, \dots, y_{17}),$$

the same level of supportiveness appears, see Table 3 (c).

We observed the same effect in the case of a simple sine example, with main series

$$x_n = \sin \frac{2\pi}{40} n + \varepsilon_n \qquad (n \in \{1, \dots, 200\}),$$

Table 3: Convolution norms (54) for Australian sparkling wine sales, support (a) cosine, (b) self, (c) shifted self

ρ	(a)	(b)	(c)
0.1	55678.41	54919.27	53727.07
0.2	54741.06	51796.55	47554.96
0.3	53438.47	47153.5	39250.44
0.4	52003.64	41634.33	30705.39
0.5	50593.95	35854.82	23246.07
0.6	49282.33	30289.33	17424.38
0.7	48086.06	25234.52	13204.97
0.8	46997.06	20829.55	10268.55
0.9	46000.84	17100.56	8243.08
1	45084.41	14005.79	6818.03
10	35443.87	5.49	6.59
100	35522.25	0.00056	0.00069

$\varepsilon_n \sim N(0,1)$ i.i.d., and support series

$$y_n = \sin 2\pi(\frac{n}{40} + \alpha) \qquad (n \in \{1, \ldots, 200\})$$

with offsets $\alpha \in \{0.05, 0.1, 0.15, 0.25\}$, taking $L = 50$, $r = 2$. In all cases, the convolution norms (54) became small with growing ρ, at a rate roughly independent of α.

This indicates that the suggested criterion for supportiveness tests for structural compatibility between the two time series rather than a simple point-by-point conformity.

6 The scaling problem and linearised MSSA

Consider two time series, one (x) giving prices (in units of \$), the other (y) quantities of a commodity (in units of metric tonnes). Then, in the MSSA recurrence formula (4), the recurrence vectors R_{11}, R_{22} will be dimension-free, but the entries of R_{12} will be in units of \$/t and those of R_{21} in units of t/\$.

Now consider the same data, but expressing the quantities of the commodity in units of kilogrammes, i.e. as a time series $\tilde{y} = 1000y$. Then, in order to get the *same* forecast as before, the recurrence vectors must be adjusted by the same conversion factor, i.e. $\tilde{R}_{12} = R_{12}/1000$ and $\tilde{R}_{21} = 1000R_{21}$. However, this is not what MSSA of the new pair of time series x, $\tilde{y}$ will give; instead the result will be completely different forecast vectors, due to the non-linearity of the spectral analysis of the combined (stacked) Hankel matrices. Thus MSSA has the intrinsic problem of lacking scaling invariance in the separate input time series. The practical expedient of only using normalised time series (e.g. mean 0, variance 1) in MSSA hides rather than solves this problem. In situations where the effect of different scalings of a support time series is to be explicitly studied, such as in Section 5 above, normalisation will not be applicable in any case.

A partial solution of the scaling problem is achieved by the following proposed *linearisation* of MSSA around SSA, which will also give a quick MSSA-

type forecast on the basis of the SSA of the main (first) time series alone.

Consider the stacked Hankel matrix $\begin{pmatrix}\mathbb{X}\\ \alpha\mathbb{Y}\end{pmatrix}$ with small α, then

$$\begin{pmatrix}\mathbb{X}\\ \alpha\mathbb{Y}\end{pmatrix}\begin{pmatrix}\mathbb{X}\\ \alpha\mathbb{Y}\end{pmatrix}^T = \begin{pmatrix}\mathbb{X}\mathbb{X}^T & 0\\ 0 & 0\end{pmatrix} + \alpha\begin{pmatrix}0 & \mathbb{X}\mathbb{Y}^T\\ \mathbb{Y}\mathbb{X}^T & \end{pmatrix} + O(\alpha^2).$$

Denoting, as before, the (SSA) eigenvectors and eigenvalues of $\mathcal{X}\mathcal{X}^T$ by η_k, λ_k, respectively, the (MSSA) eigenvectors of $\begin{pmatrix}\mathbb{X}\\ \alpha\mathbb{Y}\end{pmatrix}\begin{pmatrix}\mathbb{X}\\ \alpha\mathbb{Y}\end{pmatrix}^T$ have the form

$$\begin{pmatrix}\eta_k\\ \frac{\alpha}{\lambda_k}\,\mathbb{Y}\mathbb{X}^T\eta_k\end{pmatrix} + O(\alpha^2),$$

and further the MSSA recurrence vectors are

$$\begin{aligned}
R_{11}(\alpha) &= \frac{1}{(1-\sum_k \eta_{k,L}^2)}\sum_k \eta_{k,L}\eta_k^\nabla,\\
R_{12}(\alpha) &= \frac{1}{(1-\sum_k \eta_{k,L}^2)}\alpha\sum_k \eta_{k,L}\gamma_k^\nabla,\\
R_{21}(\alpha) &= \frac{1}{(1-\sum_k \eta_{k,L}^2)}\alpha\sum_k \big(\gamma_{k,L}(1-\sum_l \eta_{l,L}^2) + \eta_{k,L}\sum_l \eta_{l,L}\gamma_{l,L}\big)\eta_k^\nabla,\\
R_{22}(\alpha) &= \frac{1}{(1-\sum_k \eta_{k,L}^2)}\alpha^2\sum_k \big(\gamma_{k,L}(1-\sum_l \eta_{l,L}^2) + \eta_{k,L}\sum_l \eta_{l,L}\gamma_{l,L}\big)\gamma_k^\nabla,
\end{aligned}$$

where $\gamma_k = \frac{1}{\lambda}\mathbb{Y}\mathbb{X}^T\eta_k$. Note that R_{11} is the α-independent SSA recurrence vector R for the series x. Moreover, R_{12} and R_{21} are linear in α, whereas R_{22} is of higher order and hence negligible in the linearisation.

The resulting forecast will be properly homogeneous with respect to scaling of the support series y and is calculated from the SSA of x only. However, it is asymmetric between the two series as it remains non-linear (in particular, not scaling covariant) in x and does not give a usable forecast for the support series y. Also, this *linearised MSSA* cannot directly replace full (non-linear) MSSA in the supportiveness analysis proposed in Section 5.

Nevertheless, linearised MSSA may be a suitable approximation in many situations where MSSA is now used, and its general relationship with the supportiveness question remains to be explored.

Acknowledgements. The authors like to thank the referees for their detailed and valuable comments.

References

[1] E. Bozzo, R. Carniel, and D. Fasino. Relationship between Singular Spectrum Analysis and Fourier Analysis: Theory and Application to the Monitoring of Volcanic Activity. *Comput. Math. Appl.*, 50:812–20, 2010.

[2] F. X. Diebold and R. Mariano. Comparing Predictive Accuracy. *Journal for Business and Economic Statistics*, 13(2-3):253–265, 1995.

[3] N. Golyandina, A. Korobeynikov, A. Shlemov, and K. Usevish. Multivariate and 2D Extensions of Singular Spectrum Analysis with the Rssa Package. arXiv:1309.5050, Sept 2014.

[4] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky. *Analysis of Time Series Structure, SSA and Related Techniques.* Monographs on Statistics and Applied Probability, vol. 90. Chapman & Hall/CRC, Boca Raton, FL, 2001.

[5] N. Golyandina and A. Zhigljavsky. *Singular Spectrum Analysis for Time Series.* Springer, 2013.

[6] C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424–38, July 1969.

[7] D. K. Guilkey and M. K. Salemi. Small Sample Properties of Three Tests for Granger-Causal Ordering in a Bivariate Stochastic System. *The Review of Economics and Statistics*, 64(4):668–80, 1982.

[8] T. Harris and H. Yan. Filtering and Frequency Interpretations of Singular Spectrum Analysis. *Physica D*, 239:1958–67, 2010.

[9] R. J. Hyndman. Time Series Online Data library. Monthly Australian Wine Sales.

[10] T. Kato. *Perturbation Theory for Linear Operators.* Die Grundlehren der mathematischen Wissenschaften, Band 132. Springer-Verlag New York, Inc., New York, 1966.

[11] V. Nekrutkin. Perturbation Expansions of Signal Subspaces for Long Signals. *Stat. Interface*, 3(3):297–319, 2010.

[12] K. Patterson, H. Hassani, S. Heravi, and A. Zhigljavsky. Multivariate Singular Spectrum Analysis for Forecasting Revisions to Real-time Data. *J. Appl. Stat.*, 38(10):2183–2211, 2011.

[13] L.I. Schiff. *Quantum Mechanics.* McGraw-Hill Book Company, third edition, 1986.

[14] C. Sims. Money, Income, and Causality. *American Economic Review*, 62(4):540–52, Sept 1972.

[15] D. Stepanov and N. Golyandina. SSA-based Approaches to Analysis and Forecast of Multidimensional Time Series. In: *Proceedings of the 5th St. Petersburg Workshop on Simulation, June26–July2, 2005*, St. Petersburg State University, St. Petersburg, pp. 293–8.