ISSN 1440-771X

MONASH University

Department of Econometrics and Business Statistics
http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Window Length Selection and Signal–Noise Separation and Reconstruction in Singular Spectrum Analysis

Md Atikur Rahman Khan and D. S. Poskitt

November 2011

Working Paper 23/11

Window Length Selection and Signal–Noise Separation and Reconstruction in Singular Spectrum Analysis

Md Atikur Rahman Khan
&
D. S. Poskitt*
Department of Econometrics and Business Statistics, Monash University, Australia

Abstract

In Singular Spectrum Analysis (SSA) window length is a critical tuning parameter that must be assigned by the practitioner. This paper provides a theoretical analysis of signal-noise separation and reconstruction in SSA that can serve as a guide to optimal window choice. We establish numerical bounds on the mean squared reconstruction error and present their almost sure limits under very general regularity conditions on the underlying data generating mechanism. We also provide asymptotic bounds for the mean squared separation error. Evidence obtained using simulation experiments indicates that the theoretical properties are reflected in observed behaviour, even in relatively small samples, and the results indicate how an optimal choice for the window length can be made.

Keywords: Dimension, Embedding, Mean squared error, Reconstruction, Signal–noise separation, Window length.

JEL Classification: C18, C22, C52

1 Introduction

Singular spectrum analysis (SSA) is a non-parametric technique that is designed to look for nonlinear, non–stationary, and intermittent or transient behaviour in an observed time series. By way of introduction to SSA and in order to set the scene, suppose that $x(t) \equiv x'(\tau)$ is a stochastic process of interest that is observed at a sequence of points $\tau = \tau_{\min} + t\triangle t$, $t = 1, \ldots, N$, in the interval $T = (\tau_{\min}, \tau_{\max}]$ where $\triangle t = (\tau_{\max} - \tau_{\min})/N$, giving rise to an observed time series $\{x(1), x(2), \ldots, x(N)\}$ of

*Correspondence: Don Poskitt, Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia. Tel.:+61-3-9905-9378; fax:+61-3-9905-5474.
E-mail address: Donald.Poskitt@monash.edu

length N.[1] The aim of SSA is to decompose the observed series into the sum of independent and interpretable components, akin to the classical decomposition of a time series into the sum of trend, cyclical, seasonal and noise components, and SSA looks for such structure in an observed series via an eigen–decomposition of the so-called trajectory matrix.

The development of SSA is often attributed to researchers working in the physical sciences, namely Broomhead and King (1986), Vautard and Ghil (1989) and Vautard, Yiou and Ghil (1992), and SSA has gained popularity in such areas as meteorology, bio-mechanics and hydrology. The basic building blocks of SSA where previously outlined by Basilevsky and Hum (1979), however, who argued that in the social sciences standard frequency domain methods based on Fourier decompositions may lack appeal and that a discrete Karhunen-Loève analysis was more suitable, and the application of SSA in economics and finance are now also finding favour.

The general structure of the algorithm underlying SSA can be described in four basic steps:

1. Embedding: For a given window size m the $m \times n$ trajectory matrix is given by $\mathbf{X} = [\mathbf{x}_1 : \ldots : \mathbf{x}_n]$ where $n = N - m + 1$ and $\mathbf{x}_t = (x(t), x(t+1), \ldots, x(t+m-1))'$ for $(t = 1, 2, \ldots, n)$ are known as the m–lagged vectors of $\mathbf{X}$. The parameter m is called the trajectory matrix window length, and following standard practice we will suppose that m is assigned by the practitioner such that $2 < m \leq N/2 \leq n$. We will denote the mapping from $\{x(1), x(2), \ldots, x(N)\}$ to its trajectory matrix $\mathbf{X}$ by $x(t) \overset{T_{(N,m)}}{\longleftrightarrow} \mathbf{X}$. It is straightforward to show that $T_{(N,m)}$ is a linear mapping that defines an isomorphism between $\mathbb{R}^N$ and the vector space of $m \times n$ Hankel matrices.

2. Singular Value Decomposition: Let $\ell_1 \geq \ell_2 \geq \ldots, \geq \ell_m > 0$ denote the eigenvalues of $\mathbf{X}\mathbf{X}'$ arranged in descending order of magnitude, and denote by $\mathbf{u}_1, \ldots, \mathbf{u}_m$ the associated orthonormal system of eigenvectors. Then $\mathbf{X}$ can be expressed as the sum of m rank one projections $\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_m$,wherein $\mathbf{X}_i = \sqrt{\ell_i}\mathbf{u}_i\mathbf{v}_i'$ and $\sqrt{\ell_i}$ is the ith singular value, and $\mathbf{u}_i$ and $\mathbf{v}_i = \mathbf{X}'\mathbf{u}_i/\sqrt{\ell_i}$ are the ith left and right eigenvectors of $\mathbf{X}$.

3. Signal–Noise Separation: It is well known that $\|\mathbf{X}\|^2 = \text{trace}\{\mathbf{X}\mathbf{X}'\} = \sum_{i=1}^m \ell_i$ and $\|\mathbf{X}_i\|^2 = \ell_i$ for $i = 1, \ldots, m$, and $\ell_i / \sum_{i=1}^m \ell_i$ can be interpreted as the proportion of the total variation in $\mathbf{X}$ attributable to $\mathbf{X}_i$. Since not every eigentriple,

[1] To avoid a proliferation of notation we adopt the common practice of not distinguishing between a stochastic process and realized values of that process, relying on the context or some explicit statement to make the meaning clear. For notational simplicity we have supposed that the series is observed on a uniform grid, but extension to observations $\{x(t_1), x(t_2), \ldots, x(t_N)\}$ on a non–uniform grid is straightforward.

$\{\ell_i, \mathbf{u}_i, \mathbf{v}_i\}$, need contribute significantly to the overall variation, the next step is to determine a subset of eigentriples that encompass the dominant variation in $\mathbf{X}$. This amounts to selecting a $k \leq m$ dimensional subspace on which to project $\mathbf{X}$ with the associated projection $\mathbf{S}_k = \mathbf{X}_1 + \cdots + \mathbf{X}_k$ being attributed to the signal and the residual $\mathbf{E}_k = \sum_{i=k+1}^{m} \mathbf{X}_i$ being taken as noise, with k denoting the designated dimension of the signal.

4. Time Series Reconstruction: The purpose of this step is to transform the signal–noise representation $\mathbf{X} = \mathbf{S}_k + \mathbf{E}_k$ into a corresponding reconstruction of the observed series. Noting that $\mathbf{X}$ is Hankel, this is achieved by a process of diagonal averaging or *Hankelization* in which the r, cth element of $\mathbf{S}_k = [s_{rc,k}]$ is replaced by the average over all r and c such that $r + c = t + 1$ where $r = 1, \ldots, m$, $c = 1, \ldots, n$ and $t = 1, \ldots, N$. This operation implicitly defines a time series and an associated trajectory matrix, $\{\widetilde{s}_k(1), \widetilde{s}_k(2), \ldots, \widetilde{s}_k(N)\}$ and $\widetilde{\mathbf{S}}_k = [\widetilde{s}_{rc,k}]$ say, where

$$\widetilde{s}_k(t) = \widetilde{s}_{r(t-r+1),k} = \begin{cases} \frac{1}{t} \sum\sum_{r+c=t+1} s_{rc,k}\,, & \text{when } 1 \leq t \leq m-1; \\ \frac{1}{m} \sum\sum_{r+c=t+1} s_{rc,k}\,, & \text{when } m \leq t \leq n; \\ \frac{1}{N-t+1} \sum\sum_{r+c=t+1} s_{rc,k}\,, & \text{when } n+1 \leq t \leq N. \end{cases}$$

After applying the Hankelization procedure the SSA(m, k) model is given by the specification $\widetilde{\mathbf{S}}_k + \widetilde{\mathbf{E}}_k \overset{T_{(m,N)}}{\longleftrightarrow} \widetilde{s}_k(t) + \widetilde{e}_k(t)$, with $x(t) = \widetilde{s}_k(t) + \widetilde{e}_k(t)$, $t = 1, \ldots, N$, denoting the reconstruction of the original time series.

For more detailed discussions of the techniques underlying SSA and their practical application (with several examples) we refer to Elsner and Tsonis (1996) and Golyandina, Nekrutkin and Zhigljavski (2001).

Signal–noise separation and reconstruction are critical initial steps in SSA that underly any application – to forecasting or the analysis of missing data or change point problems, for example – and from the preceding description it is apparent that these two steps depend upon two basic parameters that must be assigned or chosen by the practitioner, namely, the window length of the embedding (a tuning parameter) and the dimension of the signal (a modeling parameter). Clearly the values chosen for m and k will interact one with another so as to effect performance and it is therefore necessary to ensure that the techniques used for the assignment and selection of m and k will lead to strong separability and minimize (in some sense) reconstruction error. Our purpose in this paper is to present a detailed theoretical analysis of signal–noise separation and time series reconstruction as implemented in SSA and thereby to indicate how the previous goal may be achieved.

Standard practice in SSA is to use a value for m large enough to ensure that the signal and noise components are *strongly separated*. Strong separation is deemed to have been achieved when a weighted correlation between $\widetilde{s}_k(t)$ and $\widetilde{e}_k(t)$, computed once the signal and noise groupings have been assigned, a sufficiently small. The signal–noise groupings are determined via procedures that employ pattern recognition techniques and methods similar to those used in conventional principal component analysis (the use of the scree–plot and various correlation methods as described in Jolliffe, 2002, chap. 6). The difficulty with this approach is that in the absence of clear cut statistical decision rules, and with few guidelines on how to set appropriate thresholds, the modeling involves substantial subjective assessment.

In an attempt to provide a more objective criterion for assigning window length Golyandina (2010) has examined the application of the rule $m = \beta N$ with $\beta \in (0, 0.5)$ and concludes that the use a value of β close to 0.5 will produce optimal SSA signal-noise separation and reconstruction. The recommendation that the window length be chosen close to one-half of the time series length is based primarily on simulation evidence derived from time series constructed using deterministic trigonometric signals, however, and it is not obvious that extrapolation of Golyandina's conclusions to more general stochastic processes is appropriate. More recently Khan and Poskitt (2010) have developed a Minimum Description Length (MDL) criterion that can be employed to identify the dimension of the signal from the data. They show that under appropriate regularity, and given a window length $m = \log(N)^c$, $c > 0$, the MDL criterion will provide a strongly consistent estimate of k. Experimental and empirical results presented in Khan and Poskitt (2010) clearly demonstrate the practical efficacy of using the MDL criterion to determine k, and they also illustrate that setting m too large can have deleterious effects on signal-noise separation and reconstruction.

In this paper we present theoretical results on SSA signal-noise separation and reconstruction that provide a clear guide to the experimental outcomes obtained by Golyandina (2010) and Khan and Poskitt (2010). We obtain bounds on mean squared separation and reconstruction error that can be used to explain the differences in performance characteristics reported in these two papers. The theoretical properties that we develop also support simulation results indicating how an optimal choice of window length can be made.

The plan of the remainder of the paper is as follows. In Section 2 we define mean squared reconstruction error ($MSRE$), obtain finite sample numerical bounds on $MSRE$, and define mean squared separation error ($MSSE$). In Section 3 we outline the signal–plus–noise model that underlies our analysis and use this to establish corresponding asymptotic bounds for both $MSRE$ and $MSSE$ under very general regularity conditions. Section 4 employs simulation experiments to demonstrate the

practical impact of our results in the context of two very different stochastic processes, and illustrates the effects of varying window length on signal–noise separation and reconstruction. Section 5 presents some concluding remarks.

2 Separation and Reconstruction Error

Now suppose that the observed process $x(t) = s(t) + \varepsilon(t)$ where $s(t)$ denotes an underlying signal that is masked by random noise $\varepsilon(t)$ with zero mean and variance $\sigma_\varepsilon^2 < \infty$. If $s(t)$ and $\varepsilon(t)$ are orthogonal and $\varepsilon(t)$ satisfies sufficient regularity to ensure that $N^{-1}\sum_{t=1}^{N}\varepsilon(t)^2$ converges to σ_ε^2 (statistical ergodic theorem) then the signal–noise ratio, defined as the almost sure limit

$$SNR = \lim_{N\to\infty} 10\log_{10}\left(\frac{\sum_{t=1}^{N} s(t)^2}{\sum_{t=1}^{N}\varepsilon(t)^2}\right) \text{ a.s.},$$

provides a natural measure of our ability to disentangle the signal from the noise, large values of SNR indicating that fluctuations in the signal are closely reflected in corresponding variations in the observed process. In practice, of course, $\varepsilon(t)$ and $s(t)$ are unobservable, but an obvious empirical counterpart to SNR can be calculated by evaluating

$$SNR_N = 10\log_{10}\left(\frac{\sum_{t=1}^{N} x(t)^2}{\sum_{t=1}^{N}\widetilde{e}_k(t)^2} - 1\right).$$

The only term in SNR_N that changes as a consequence of using different window lengths and signal dimensions in the signal–noise separation and reconstruction steps is

$$\frac{1}{N}\sum_{t=1}^{N}\widetilde{e}_k(t)^2 = \frac{1}{N}\sum_{t=1}^{N}(x(t) - \widetilde{s}_k(t))^2,$$

which we will designate the mean squared reconstruction error ($MSRE_k$) following the common practice in SSA of referring to $\widetilde{s}_k(t)$ as the reconstructed time series.

Recognizing that $MSRE_k$ depends on the values assigned to m and k, we begin by establishing algebraic bounds on $MSRE_k$ as a function of these two parameters. From the structure of the mapping $\widetilde{\mathbf{E}}_k \overset{T_{(m,N)}}{\longleftrightarrow} \widetilde{e}_k(t)$ we have

$$\begin{aligned}
\frac{m}{N}\sum_{t=1}^{N}\widetilde{e}_k^2(t) &\geq \frac{1}{N}\left(\sum_{t=1}^{m-1} t\widetilde{e}_k^2(t) + m\sum_{t=m}^{n}\widetilde{e}_k^2(t) + \sum_{t=n+1}^{N}(N-t+1)\widetilde{e}_k^2(t)\right) \\
&= \frac{1}{N}\|\widetilde{\mathbf{E}}_k\|^2 \\
&\geq \frac{1}{N}\sum_{j=k+1}^{m}\ell_j,
\end{aligned}$$

where $\|\cdot\|$ denotes the Euclidean norm, the final inequality following because $\|\widetilde{\mathbf{E}}_k\|^2 = \|\mathbf{X} - \widetilde{\mathbf{S}}_k\|^2 \geq \|\mathbf{X} - \mathbf{S}_k\|^2 = \|\mathbf{E}_k\|^2$ (Rao, 1965, Sec. 8g.2). Thus we can conclude that

$$MSRE_k \geq \frac{\sum_{j=k+1}^{m} \ell_j}{mN}\,.$$

This establishes the lower bound exhibited in the following lemma.

Lemma 1. *: For all window lengths $m = k+1, \ldots, [N/2]$ the mean squared reconstruction error $MSRE_k$ lies in the interval $[L_R(m,k), U_R(m,k)]$ where*

$$L_R(m,k) = \frac{\sum_{j=k+1}^{m} \ell_j}{mN}$$

and

$$U_R(m,k) = \frac{\sum_{j=k+1}^{m} \ell_j}{N}\left(\frac{N-2(m-1)}{m} + 2\log(m-1) + 2\gamma + \frac{1}{m-1} - \frac{\eta(m)}{(m-1)^2}\right),$$

wherein γ denotes Euler's constant and $\eta(m) \in [0, 0.25]$.

Proof of Lemma 1: It is only necessary for us to establish the stated upper bound. Using the fact that $\widetilde{\mathbf{E}}_k$ in the mapping $\widetilde{\mathbf{E}}_k \overset{T_{(N,m)}}{\longleftrightarrow} \widetilde{e}_k(t)$ is obtained by averaging $\mathbf{E}_k = [e_{rc,k}]$, $r = 1, \ldots, m$, $c = 1, \ldots, n$, along the t-th secondary diagonal, it follows from the Cauchy–Schwartz inequality that

$$\widetilde{e}_k(t)^2 \leq \begin{cases} \frac{1}{t}\sum\sum_{r+c=t+1} (e_{rc,k})^2\,, & \text{for } 1 \leq t \leq m-1; \\ \frac{1}{m}\sum\sum_{r+c=t+1} (e_{rc,k})^2\,, & \text{for } m \leq t \leq n; \\ \frac{1}{N-t+1}\sum\sum_{r+c=t+1} (e_{rc,k})^2\,, & \text{for } n+1 \leq t \leq N. \end{cases} \tag{1}$$

Replacing each $\widetilde{e}_k(t)^2$ in the formula for $MSRE_k$ by it's upper bound in (1) and gathering terms, noting that $\sum_{t=1}^{N}\sum\sum_{r+c=t+1}(e_{rc,k})^2 = \sum_{r=1}^{m}\sum_{c=1}^{n}(e_{rc,k})^2 = \|\mathbf{E}_k\|^2$ now yields the inequality

$$\frac{1}{N}\sum_{t=1}^{N} \widetilde{e}_k^2(t) \leq \frac{1}{N}\left(\sum_{t=1}^{m-1}\frac{1}{t} + \sum_{t=m}^{n}\frac{1}{m} + \sum_{t=n+1}^{N}\frac{1}{N-t+1}\right)\|\mathbf{E}_k\|^2\,. \tag{2}$$

Finally, substituting

$$\sum_{t=1}^{m-1}\frac{1}{t} = \log(m-1) + \gamma + \frac{1}{2(m-1)} - \frac{\eta(m)}{2(m-1)^2}\,,$$

$n - m + 1 = N - 2(m-1)$ and $\|\mathbf{E}_k\|^2 = \|\sum_{j=k+1}^{m}\sqrt{\ell_j}\mathbf{u}_j\mathbf{v}_j'\|^2 = \sum_{j=k+1}^{m}\ell_j$ into (2) gives the formula for $U_R(m,k)$ as stated. □

If we are to use $MSRE_k$ as a guide to assigning appropriate values for the window–length, or as a guide to the consequences of selecting a particular dimension for the signal, we need to investigate the statistical properties of the bounds in Lemma 1 under different scenarios. Before going on to treat such matters, we note here that in her simulation experiments Golyandina (2010) uses

$$\frac{1}{N}\sum_{t=1}^{N}(\widetilde{s}_k(t) - s(t))^2 = \frac{1}{N}\sum_{t=1}^{N}(\widetilde{e}_k(t) - \varepsilon(t))^2 \tag{3}$$

to evaluate the consequences of using different choices of m and k. The expressions in (3) measure the distance between the reconstruction estimates and the actual signal and noise components, and we will therefore call this the mean squared separation error ($MSSE_k$). Although $MSSE_k$ can be evaluated in simulation experiments it cannot be calculated empirically because it depends upon $s(t)$ and $\varepsilon(t)$, which are in practice unobservable. Unlike $MSRE_k$, which can be determined from the data, $MSSE_k$ is an abstract object not available to the practitioner. It follows that the theoretical properties of $MSSE_k$ are of interest if it is to provide a general guide to the consequences of using particular combinations of window length m and signal dimension k. To derive the theoretical behaviour of $MSSE_k$, as well as $MSRE_k$, we must introduce some regularity conditions, needless to say.

3 Asymptotic Bounds

In order for our results to have broad applicability we state our basic assumption in generic form.

Assumption 1: The data generating mechanism underlying the stochastic process $x(t)$ satisfies sufficient conditions to ensure that for any trajectory matrix window length $m = (\log N)^c$, $c < \infty$, there exists a positive definite matrix $\mathbf{\Gamma}_m$ such that $\|n^{-1}\mathbf{X}\mathbf{X}' - \mathbf{\Gamma}_m\| = O(Q_n)$ a.s. as $N \to \infty$ where $Q_n = \sqrt{\log\log n/n}$.

Particular examples of data generating mechanisms that satisfy Assumption 1 are presented below. To show that Assumption 1 is applicable to the so called Karhunen class of processes (Rao, 1985), we follow Khan and Poskitt (2010) and suppose that $\{x(t) : t \in T\}$ is a zero mean stochastic process, defined on a probability space $\mathcal{P} = \{\Omega, \mathcal{B}, P\}$, continuous in mean square, with the continuous covariance kernel $K(t,s) = \mathbb{E}[x(t)x(s)]$ on $T \times T$. By Mercer's theorem $K(t,s) = \sum_{j=1}^{\infty}\lambda_j\phi_j(t)\phi_j(s)$ where the $\{\phi_j\}$ are continuous orthonormal eigenfunctions of K corresponding to the eigenvalues $\{\lambda_j\}$, namely $\int_T K(t,s)\phi_j(s)ds = \lambda_j\phi_j(t)$, and the series converges uniformly and absolutely

on $T \times T$. Let

$$z_j = \int_T x(t)\phi_j(t)dt \qquad j = 1, 2, \ldots . \tag{4}$$

Then the z_j have zero mean and variance λ_j, $z_j \sim (0, \lambda_j)$, $j = 1, 2, \ldots$, and since $K(t,t) = \sum_{j=1}^{\infty} \lambda_j |\phi_j(t)|^2$ converges it follows from the Cauchy criterion that the stochastic series $\sum_{j=1}^{\kappa} z_j \phi_j(t)$ converges uniformly in mean square to $x(t)$ as $\kappa \to \infty$. The limiting expression $x(t) = \sum_{j=1}^{\infty} z_j \phi_j(t)$ is known as the Karhunen-Loève expansion. Now let us suppose that in passage to the limit given by the Karhunen-Loève expansion there exists a value $\kappa = k$ such that the difference $x(t) - \sum_{j=1}^{k} z_j \phi_j(t)$ behaves as a weakly stationary process, so that we may write the observed process as

$$x(t) = \sum_{j=1}^{k} z_j \phi_j(t) + \varepsilon(t)\,, \tag{5}$$

a signal–plus–noise representation of $x(t)$ in which the signal $s(t) = \sum_{j=1}^{k} z_j \phi_j(t)$ and the noise $\varepsilon(t)$ are orthogonal by virtue of the fact that the random coefficients z_j are pairwise uncorrelated. The decomposition in (5) implies that the signal and noise subspaces are strongly separable and that the minimal eigenvalue of the signal is larger than the maximal eigenvalue of the noise.

To relate SSA to the Karhunen-Loève expansion note that if the model in (5) obtains the m-lagged vectors of the trajectory matrix can be written as

$$\mathbf{x}_t = \sum_{j=1}^{k} z_j \boldsymbol{\phi}_j(t) + \boldsymbol{\varepsilon}_t\,, \tag{6}$$

where

$$\boldsymbol{\phi}_j(t) = \begin{bmatrix} \phi_j(t) \\ \vdots \\ \phi_j(t+m-1) \end{bmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon(t) \\ \vdots \\ \varepsilon(t+m-1) \end{bmatrix}.$$

Now suppose that ϕ_j, $j = 1, \ldots, k$, satisfy the Lipschitz condition $|\phi_j(t) - \phi_j(t-1)| \leq M\triangle t$ – smoothness of the dominant eigenfunctions is commonly supposed in SSA. Then $\|\boldsymbol{\phi}_j(t) - \boldsymbol{\phi}_j(t-1)\| \leq \sqrt{m} M \triangle t$. Let $\boldsymbol{\varphi}_j$ be a point on the line segment joining $\boldsymbol{\phi}_j(t)$ to $\boldsymbol{\phi}_j(t-1)$ and set $\zeta_{jt} = (\boldsymbol{\varphi}_j'\boldsymbol{\varphi}_j)^{-1}\boldsymbol{\varphi}_j'\boldsymbol{\phi}_j(t) z_j$. Then $\zeta_{jt}\boldsymbol{\varphi}_j = z_j \boldsymbol{\phi}_j(t)$ and (6) can be reexpressed in matrix–vector form as

$$\mathbf{x}_t = \boldsymbol{\Phi}\mathbf{z}_t + \boldsymbol{\varepsilon}_t \tag{7}$$

where $\mathbf{z}_t = (\zeta_{1t}, \ldots, \zeta_{kt})'$ and $\boldsymbol{\Phi} = [\boldsymbol{\varphi}_1 : \cdots : \boldsymbol{\varphi}_k]$ is an $m \times k$ matrix of functional values. Furthermore, $|1 - (\boldsymbol{\varphi}_j'\boldsymbol{\varphi}_j)^{-1}\boldsymbol{\varphi}_j'\boldsymbol{\phi}_j(t)| \leq \|\boldsymbol{\varphi}_j - \boldsymbol{\phi}_j(t)\|/\|\boldsymbol{\varphi}_j\|^2$ and $\|\boldsymbol{\varphi}_j - \boldsymbol{\phi}_j(t)\| \leq \sqrt{m} M \triangle t$. Hence, as $N \to \infty$ and $\triangle t \to 0$ the representation in (6) will be equivalent

to the model in (7) where, with a slight abuse of notation, $\mathbf{z}_t \sim (\mathbf{0}, \mathbf{\Lambda})$ with $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_k\}$ and is orthogonal to $\boldsymbol{\varepsilon}_t \sim (\mathbf{0}, \mathbf{\Sigma}_m)$ where $\mathbf{\Sigma}_m = \mathbb{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t']$.

The specification in (7) generates a combined functional–structural relationship for $\mathbf{x}_t$ (Kendal and Stuart, 1979, chap. 29) and given the nature of the approximation inherent in (7) it does not seem unreasonable to suppose that the stochastic properties of $\mathbf{z}_t$ are characterized by those of a near epoch dependent (mixing) processes. Pötscher and Prucha (1997, Chapt. 5–10) present a detailed discussion of stochastic approximation and near epoch dependence, with extensive references. If we also assume that $\|n^{-1}\sum_{i=1}^{n} \mathbf{z}_t \mathbf{z}_t' - \mathbf{\Lambda}\| = O(Q_n)$, $\|n^{-1}\sum_{t=1}^{n} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' - \mathbf{\Sigma}_m\| = O(Q_n)$ and $\|n^{-1}\sum_{t=1}^{n} \mathbf{z}_t \boldsymbol{\varepsilon}_t'\| = O(Q_n)$, which we will christen Assumption 1', then it is straightforward to show that $x(t)$ will satisfy Assumption 1 with $\mathbf{\Gamma}_m = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}' + \mathbf{\Sigma}_m$. We will not indicate more primitive conditions under which an $O(Q_n)$ convergence rate will hold since a speed of convergence governed by the law of the iterated logarithm is not critical to our subsequent analysis. Suffice it to say that none of the requirements of Assumption 1 and Assumption 1' seems onerous, indeed, our results will still hold for any Q_n such that $Q_n \to 0$ as $N \to \infty$.

In SSA the window length is simply assumed to satisfy $2 \leq m \leq \frac{N}{2}$, but for our theoretical development we have supposed that $m = (\log N)^c$, $c < \infty$. The latter reflects that the kth order signal–noise representation $\mathbf{X} = \mathbf{S}_k + \mathbf{E}_k$ is a function of $k(m+1) - \frac{1}{2}k(k+1)$ freely varying parameters with $k < m$, namely, k singular values plus mk eigenvector elements minus their $\frac{1}{2}k(k+1)$ orthonormalization constraints. Although we can allow m to approach infinity with N it is obvious that we require $m/N \to 0$ as $N \to \infty$ if the effective sample size $n = N - m + 1$ is to grow faster than the number of parameters. Hence the imposition the condition that $m = (\log N)^c$, $c < \infty$. We will return to a consideration of the choice of c when using the rule $m = (\log N)^c$ to assign window length in Section 3.

For convenience we state the following convergence property taken from Khan and Poskitt (2010), where the proof can be found.

Lemma 2. *: Suppose that $x(t)$ satisfies Assumption 1 and let $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_m > 0$ denote the ordered eigenvalues of $\mathbf{\Gamma}_m$. Then $\max_{j=1,\ldots,m} |\gamma_j - \ell_j/n| = O(Q_n)$.*

From Lemma 2 we can readily deduce that

$$\left| \frac{\sum_{i=k+1}^{m} \ell_i}{n \sum_{i=k+1}^{m} \gamma_i} - 1 \right| \leq \frac{1}{(m-k)\gamma_m} \left| \sum_{i=k+1}^{m} \left(\frac{\ell_i}{n} - \gamma_i \right) \right| = O(Q_n) \tag{8}$$

uniformly in $k < m$. Employing this result in conjunction with Lemma 1 leads to the following theorem.

Theorem 1. *: Suppose that $x(t)$ satisfies Assumption 1. Then for all window lengths $m = k+1, \ldots, [N/2]$ the bounds on $MSRE_k$ can be reformulated as*

$$L_R(m,k) = \frac{n}{mN}\left(\sum_{j=k+1}^{m}\gamma_j\right)(1+O(Q_n))$$

and

$$U_R(m,k) = \frac{n\nu_{N,m}}{N}\left(\sum_{j=k+1}^{m}\gamma_j\right)(1+O(Q_n))$$

where

$$\nu_{N,m} = \frac{N-m}{N}\left(\frac{N-2(m-1)}{m} + 2\log(m-1) + 2\gamma + \frac{1}{m-1} - \frac{\eta(m)}{(m-1)^2}\right).$$

Proof of Theorem 1: The formula for the lower bound follows from equation (8) upon substituting $n\sum_{i=k+1}^{m}\gamma_i$ for $\sum_{i=k+1}^{m}\ell_i$ in Lemma 1. Let $S_N = N^{-1}\sum_{t=1}^{N}\widetilde{e}_k^2(t)$ and $C_N = N^{-2}\sum_{t=1}^{N}(N-t)\widetilde{e}_k^2(t)$. From the properties of Césaro sums we have $\lim_{N\to\infty}|S_N - C_N| = 0$, and for N sufficiently large $C_N < N^{-2}(N-m)\sum_{t=1}^{N}\widetilde{e}_k^2(t) = N^{-1}(N-m)S_N$ since by assumption $m/N \to 0$ as $N \to \infty$. Applying Lemma 1 to S_N, and once again substituting $n\sum_{i=k+1}^{m}\gamma_i$ for $\sum_{i=k+1}^{m}\ell_i$ and appealing to Lemma 2 via equation (8) in similar manner to the derivation of the lower bound, now gives the upper bound. □

The corresponding theorem for $MSSE_k$ is as follows.

Theorem 2. *: Suppose that $x(t)$ satisfies Assumption 1'. Then for all window lengths $m = k+1, \ldots, [N/2]$ the inequalities $0 \le MSSE_k \le U_S(m,k)$ obtain where*

$$U_S(m,k) = \frac{n\nu_{N,m}}{N}\left(\sum_{j=k+1}^{m}\gamma_j + \sum_{j=1}^{m}\sigma_j - 2\sum_{j=k+1}^{m}\sigma_j\right)(1+O(Q_n))$$

and $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_m > 0$ denote the ordered eigenvalues of $\mathbf{\Sigma}_m$.

Proof of Theorem 2: That $MSSE_k \ge 0$ is obvious. Let $\mathbb{E} = [\varepsilon(r+c-1)]$, $r = 1, \ldots, m$, $c = 1, \ldots, n$. It is readily verified that $\widetilde{e}_k(t) - \varepsilon(t) \overset{T_{(N,m)}}{\longleftrightarrow} \widetilde{\mathbf{E}}_k - \mathbb{E}$ where $\widetilde{\mathbf{E}}_k - \mathbb{E}$ is obtained by Hankelizing $\mathbf{E}_k - \mathbb{E}$. It follows that

$$(\widetilde{e}_k(t) - \varepsilon(t))^2 = \left(\frac{1}{t}\sum\sum_{r+c=t+1} e_{rc,k} - \varepsilon(t)\right)^2 \le \frac{1}{t}\sum\sum_{r+c=t+1}(e_{rc,k} - \varepsilon(t))^2,$$

when $1 \le t \le m-1$,

$$(\widetilde{e}_k(t) - \varepsilon(t))^2 = \left(\frac{1}{m}\sum\sum_{r+c=t+1} e_{rc,k} - \varepsilon(t)\right)^2 \le \frac{1}{m}\sum\sum_{r+c=t+1}(e_{rc,k} - \varepsilon(t))^2,$$

when $m \leq t \leq n$, and

$$(\widetilde{e}_k(t) - \varepsilon(t))^2 = \left(\frac{1}{N-t+1} \sum\sum_{r+c=t+1} e_{rc,k} - \varepsilon(t) \right)^2 \leq \frac{1}{N-t+1} \sum\sum_{r+c=t+1} (e_{rc,k} - \varepsilon(t))^2$$

when $n+1 \leq t \leq N$. Replacing each $(\widetilde{e}_k(t) - \varepsilon(t))^2$ in the formula for $MSSE_k$ by the corresponding mean squared difference and proceeding as in the proof of Theorem 1 leads us to the inequality

$$MSSE_k \leq \frac{\nu_{N,m}}{N} \|\mathbf{E}_k - \mathbb{E}\|^2 .$$

Now, $\|\mathbf{E}_k - \mathbb{E}\|^2 = \|\mathbf{E}_k\|^2 + \|\mathbb{E}\|^2 - 2\mathrm{tr}(\mathbf{E}_k \mathbb{E}')$ and by Lemma 2 it follows that

$$n^{-1}\|\mathbf{E}_k\|^2 = n^{-1} \sum_{j=k+1}^{m} \ell_j = \left(\sum_{j=k+1}^{m} \gamma_j \right) (1 + O(Q_n))$$

and

$$n^{-1}\|\mathbb{E}\|^2 = \mathrm{tr}\left(n^{-1} \sum_{t=1}^{n} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \right) = \left(\sum_{j=1}^{m} \sigma_j \right) (1 + O(Q_n)) .$$

To evaluate $n^{-1}\mathrm{tr}\,(\mathbf{E}_k \mathbb{E}')$ substitute $\mathbf{E}_k = \sum_{j=k+1}^{m} \sqrt{\ell_j}\mathbf{u}_j \mathbf{v}_j' = \sum_{j=k+1}^{m} \mathbf{u}_j \mathbf{u}_j' \mathbf{X}$ and note from (7) that $\mathbf{X} = [\boldsymbol{\Phi}\mathbf{z}_1 + \boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\Phi}\mathbf{z}_n + \boldsymbol{\varepsilon}_n]$. It follows from the orthogonality between the signal and the noise that $n^{-1} \sum_{t=1}^{n} (\boldsymbol{\Phi}\mathbf{z}_t + \boldsymbol{\varepsilon}_t)\boldsymbol{\varepsilon}_t' = \boldsymbol{\Sigma}_m + O(Q_n)$ and hence that

$$\begin{aligned} n^{-1}\mathrm{tr}\,(\mathbf{E}_k \mathbb{E}') &= n^{-1} \sum_{j=k+1}^{m} \mathrm{tr}\left(\mathbf{u}_j \mathbf{u}_j' \sum_{t=1}^{n} (\boldsymbol{\Phi}\mathbf{z}_t + \boldsymbol{\varepsilon}_t)\boldsymbol{\varepsilon}_t' \right) \\ &= \sum_{j=k+1}^{m} \mathbf{u}_j' (\boldsymbol{\Sigma}_m + O(Q_n)) \mathbf{u}_j . \end{aligned}$$

From Poincaré's separation theorem we can therefore deduce that $n^{-1}\mathrm{tr}\,(\mathbf{E}_k \mathbb{E}') \geq (\sum_{j=k+1}^{m} \sigma_j)(1 + O(Q_n))$. Collecting the limiting expressions for $n^{-1}\|\mathbf{E}_k\|^2$, $n^{-1}\|\mathbb{E}\|^2$ and $n^{-1}\mathrm{tr}(\mathbf{E}_k \mathbb{E}')$ together now gives the required result. □

Before examining the practical implications of Theorems 1 and 2 it is of interest to note that Mercer's theorem implies that $\lambda_k \to 0$ as $k \to \infty$ and hence that the noise component, $\varepsilon(t) = \sum_{j=k+1}^{\infty} z_j \phi_j(t)$, in the signal–plus–noise representation of $x(t)$ will deviate from zero with arbitrarily small probability as k increases. This intimates that it is appropriate for us to examine scenarios where there is no noise. In this case the standard SSA concept of signal–noise separation breaks down since $x(t) \equiv s(t)$ and $\varepsilon(t) \equiv 0$. Nevertheless, from Rao (1965, Sections 8g.1–8g.2) we know that the minimum mean squared error projection of $\mathbf{X}$ into $\mathbb{R}^k$ is achieved by the first k "principle components" $\mathbf{S}_k = \sum_{j=1}^{k} \sqrt{\ell_j}\mathbf{u}_j \mathbf{v}_j'$, with a residual mean square equal to $\sum_{j=k+1}^{m} \ell_j$, and we can interpret the employment of the specification $\widetilde{\mathbf{S}}_k +$

$\widetilde{\mathbf{E}}_k \overset{T_{(m,N)}}{\longleftrightarrow} \widetilde{s}_k(t) + \widetilde{e}_k(t)$ for the time series reconstruction as the use of a k dimensional approximation to the truly infinite dimensional process $x(t)$. Moreover, $MSRE_k$ can still be calculated and Theorem 1 used to evaluate the reconstruction and hence assess the quality of the approximation.

4 Numerical Illustrations

Our purpose in this section of the paper is to examine via simulation experimentation the extent to which the asymptotic properties presented in Theorems 1 and 2 are reflected in finite sample behaviour. Theorems 1 and 2 can be used to assess the possible consequences of employing different combinations of m and k when modeling different types of time series, and inferences drawn from them will be devoid of vagaries that might be associated with basing conclusions exclusively on experimental simulation outcomes.

4.1 Example I

Consider a process $x(t)$ such that

$$x(t) = \sum_{r=1}^{p} A_r cos(\lambda_r t + \theta_r) + \varepsilon(t)$$

where A_r is the amplitude, λ_r the frequency (in cycles per unit time), and θ_r the phase shift of the rth cosinusoidal element of the signal, and $\varepsilon(t)$ is a zero mean white noise process with variance σ^2. If the θ_r are independent and uniformly distributed over the interval $(-\pi, \pi)$ it is straightforward to show that $x(t)$ is a zero mean stationary process with covariance kernel

$$E[x(t)x(s)] = \begin{cases} \frac{1}{2}\sum_{r=1}^{p} A_r^2 cos(\lambda_r(t-s)) & \text{if } t \neq s; \\ \frac{1}{2}\sum_{r=1}^{p} A_r^2 + \sigma^2 & \text{if } t = s. \end{cases}$$

Moreover, $x(t)$ will satisfy Assumption 1 with

$$\mathbf{\Gamma}_m = \frac{1}{2}\sum_{r=1}^{p} A_r^2 \left[\mathbf{c}_r \mathbf{c}_r' + \mathbf{s}_r \mathbf{s}_r'\right] + \sigma^2 \mathbf{I}_m$$

where $\mathbf{c}_r = [1, \cos(\lambda_r), \cdots, \cos((m-1)\lambda_r)]'$ and $\mathbf{s}_r = [1, \sin(\lambda_r), \cdots, \sin((m-1)\lambda_r)]'$. It is straightforward to verify that the ordered eigenvalues of $\mathbf{\Gamma}_m$ are $\gamma_i = \upsilon_i + \sigma^2$, $i = 1, \ldots, k$, and $\gamma_i = \sigma^2$, $i = k+1, \ldots, m$, where $\upsilon_1 > \upsilon_2 > \cdots > \upsilon_k > 0$ denote the

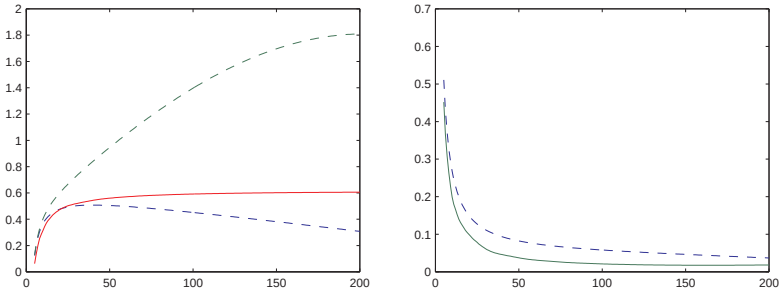
ordered, nonzero eigenvalues of $\frac{1}{2}\sum_{r=1}^{p} A_r^2 \left[\mathbf{c}_r\mathbf{c}_r' + \mathbf{s}_r\mathbf{s}_r'\right]$ and $k = 2p$, the dimension of the cosinusoidal signal component.

The signal–to–noise ratio for this process is

$$SNR = 10\log_{10}\left(\frac{\frac{1}{2}\sum_{r=1}^{p} A_r^2}{\sigma^2}\right) dB\,,$$

and for known amplitudes $A_1, \ldots, A_p$ simulated realizations from processes with different pre–assigned signal–to–noise ratios can be generated by setting the noise variance $\sigma^2 = 10^{\log_{10}(\sum_{r=1}^{p} A_r^2) - \log_{10}(0.1SNR)}$. In our experiments we employed $p = 2$ with $A_1 = 1.0$, $A_2 = 0.5$ and $\lambda_1 = 2\pi/7$, $\lambda_2 = 2\pi/10$. The noise process was i.i.d. Gaussian with variance chosen such that SNR ranged from about $5\,dB$ to $-4\,dB$, and we examined sample sizes $N = 200, 400, 600, 1000$ and 1600.

Figure 1 displays the average value of $MSRE_4$ and $MSSE_4$ evaluated across 10000 replications when $SNR = 0$ and $N = 400$. Figure 1 also plots approximate lower



(a) $MSRE_4$ (b) $MSSE_4$

Figure 1: Simulated mean squared error (solid lines) and asymptotic bounds (dashed lines) as a function of window length – 4 dimensional trigonometric signal.

and upper bounds derived from Theorems 1 and 2, namely, $\sigma^2 n(m-k)/mN$ and $\sigma^2 n(m-k)\nu_{N,m}/N$ in Figure 1a, and $\sigma^2 nk\nu_{N,m}/N$ in Figure 1b. Each is graphed as a function of $m = k+1, \ldots, [N/2]$, any of which can correspond to a window length assigned by a practitioner. From Figure 1a we see that the average value of $MSRE_4$ is increasing in m, but for window lengths with $m \geq 50$ it is apparent that $MSRE_4$ becomes more or less stable and is close to $\sigma^2 = 0.625$. Whereas $MSRE_4$ is minimized for small values of m the opposite is obviously true of $MSSE_4$. From Figure 1b we see that the average value of $MSSE_4$ is decreasing in m, but for window lengths with $m \geq 100$ the values of $MSSE_k$ become more or less stable and are close to their lower bound. Figure 1 clearly demonstrates that $MSRE$ and $MSSE$ are in conflict and that the choice of a single window length that simultaneously optimizes both is not possible. A rule that assigns a value of m between, say 50 and 100, will however achieve a compromise where the least desirable outcomes are avoided for both $MSRE$ and $MSSE$, and both are close to optimal.

Varying the values of SNR and N we find that the qualitative nature of the results do not change. As might be expected, in general $MSRE$ and $MSSE$ vary inversely with SNR, and as N increases they approach their natural limiting values of σ^2 and zero, respectively, for all but the very smallest values of m. In general the results indicate that from the perspective of $MSSE$ small values of m are to be avoided, whereas large values of m have a detrimental effect on $MSRE$. Expressing $MSRE$ and $MSSE$ as functions of $c = \log(m)/\log\log(N)$, overall the outcomes suggest that a simple practical rule consistent with the theoretical requirement that $m/N \to 0$ as $N \to \infty$ is to assign a window length equal to $(\log N)^c$ with $c \in (1.5, 3.0)$. This rule gives $m = 12 \approx 0.06N$ when $c = 1.5$ and $N = 200$, and $m = 401 \approx 0.25N$ when $c = 3.0$ and $N = 1600$, moderately sized window lengths that produce values of $MSRE$ and $MSSE$ that are close to optimal. Such window lengths coincide with those found to maximize the probability of correct model determination in Khan and Poskitt (2010), but they are noticeably smaller than those recommended in Golyandina (2010).

REMARK 1: Golyandina (2010) examines the window selection rule $m = \beta N$ and on the basis of $MSSE$ recommends choosing m "close to one-half of the time series length". The assignment $m = \beta N$ obviously does not meet the theoretical requirements of this paper, but setting $m = k + 1, \ldots, [N/2]$ we can investigate the relationship between $\beta = m/N$ and $MSSE$. To demonstrate that our results are in accord with those of Golyandina (2010) we summarise in Figure 2 the outcomes based on 10000 replications of the processes $x(t) = cos(0.2\pi t) + \varepsilon(t)$ where $\sigma^2 = 0.01$ $(SNR = 10\log_{10}(50) \approx 17\,dB)$ and $N = 100$ – Golyandina's process t.s.(4). In Figure 2 we plot the function

$$U(\beta) = \sqrt{\frac{n\nu_{N,\beta N}}{N} k\sigma^2} \quad \beta \in (0.03, 0.5)$$

for $k = 2$, $\sigma^2 = 0.01$ and $N = 100$, and the average value of $\sqrt{MSSE_2}$. We can

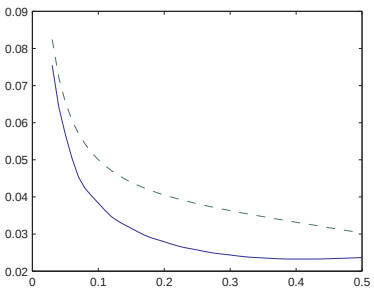


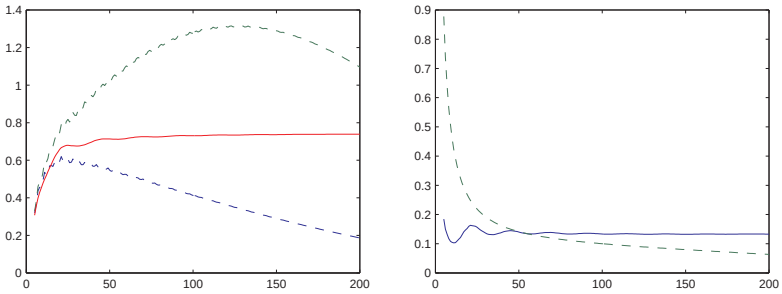
Figure 2: Root mean squared separation error $\sqrt{MSSE_2}$ as function of β – Golyandina's process t.s.(4)

see that the profiles of the approximate asymptotic bound $U(\beta)$ and the observed root mean squared separation error, when reflected about $\beta = 0.5$, closely mirror the

symmetric U–shape of the root mean squared error curve exhibited in Golyandina (*cf.* Golyandina, 2010, Figure 8).

REMARK 2: Our experiments indicate that for the four dimensional trigonometric signal $MSSE_4$ will approach zero for all but the smallest values of m as $N \to \infty$. This finding is consistent with the results of Forni and Lippi (2001). Working in the context of dynamic factor models Forni and Lippi (2001) show that if the observed series is a covariance stationary process composed of incoherent signal and noise components with absolutely continuous spectral distributions, then the two components will be identified asymptotically if the number of factors is known and the number of elements in the series increases with N. If we interpret the window length m as the number of variables in a multivariate time series and the dimension of the signal as the number of factors, then their results imply that, provided k is known, SSA reconstruction will recover the true signal and achieve strong separation if $m \to \infty$ as $N \to \infty$ and $x(t)$ is a nonsingular stationary process. Such regularity for $x(t)$ is too restrictive for our purposes here, and Forni and Lippi (2001) do not specify a rate of increase for m, nevertheless, our results suggest that the assignment rule $m = (\log N)^c$ with $c \in (1.5, 3.0)$ yields a window length sufficiently large to identify the true signal and achieve strong separation.

The correspondence between the simulated outcomes and the asymptotic bounds seen in Figure 1 is partly a consequence of the fact that the true value of k has been employed when evaluating $MSRE_k$ and $MSSE_k$. Allowing k to deviate from the true value we find that $MSRE_k$ continues to behave as previously, but the behaviour of $MSSE_k$ changes quite dramatically. Figure 3 graphs $MSRE_2$ and $MSSE_2$ for the same process as led to Figure 1. From Figure 3a we can see that, apart from its limiting



(a) $MSRE_2$ (b) $MSSE_2$

Figure 3: Simulated mean squared error (solid lines) and asymptotic bounds (dashed lines) as a function of window length – 4 dimensional trigonometric signal.

value exceeding σ^2 due to the under specification of k, $MSRE_2$ behaves similarly to $MSRE_4$ in Figure 1a and the bounds are still applicable. $MSSE_2$, however, no longer declines monotonically and the asymptotic bound is no longer operative, as is seen in

the eventual flattening out of the $MSSE$ curve as m increases. The latter is explicable because the use of the incorrect value of k implies that the results cited in Remark 2 are no longer applicable and the lower bound of $MSSE_2$ is, of course, no longer zero due to the confounding of the signal with the noise implicit in using $k = 2$.

4.2 Example II

As a second example consider an observed process $x(t)$ such that

$$x(t) = \sum_{\tau=0}^{t-1} \eta(t-\tau)$$

where $\eta(t)$ is an i.i.d. Gaussian white noise processes with a variance of one. Here, of course, $x(t)$ is a random walk, $x(t) = x(t-1) + \eta(t)$. Exploiting the strong Markov property of the random walk we can express the ith m–lagged vector as $\mathbf{x}_i = x(i-1)\mathbf{1}_m + \mathbf{\Phi}\mathbf{z}_i$ where: $\mathbf{z}_i = (\zeta_1, \ldots, \zeta_m)' \sim N(\mathbf{0}, \mathbf{\Lambda})$ with $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_m\}$, $\lambda_k = \text{cosec}^2(\theta_k)/4$, $\theta_k = \pi(2k-1)/(4m+2)$, $k = 1, \ldots, m$; $\mathbf{\Phi} = [\boldsymbol{\varphi}_1 : \cdots : \boldsymbol{\varphi}_m]$ is an $m \times m$ matrix with kth column $\boldsymbol{\varphi}_k = \sqrt{4/(2m+1)}(\sin(\theta_k), \ldots, \sin(\theta_k m))'$. For this process $\|n^{-1}\mathbf{X}\mathbf{X}' - \mathbf{\Gamma}\| = O(Q_n)$ where $\mathbf{\Gamma} = n\beta_n^2 \mathbf{1}_m \mathbf{1}_m' + \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}'$ and, via an application of Donsker's theorem and the fact that $n^{-3/2}\sum_{t=1}^{n} x(t-1)\eta(t) = O(\sqrt{\log\log n})$ (Poskitt, 2000, Lemma A.1.(ii)),

$$\beta_n^2 = \frac{1}{n^2}\sum_{t=1}^{n} x(t-1)^2 + O(Q_n) \xrightarrow{D} \int_0^1 \mathbb{B}^2(\omega)d\omega$$

where $Q_n = \sqrt{\log\log n/n}$ and $\mathbb{B}(\omega)$ denotes standard Brownian motion.

Under the current scenario $x(t)$ is in truth infinite dimensional and there is no noise. As observed above, the use of a finite k amounts to employing a minimum mean squared error approximation to the process, and the SSA concept of separation breaks down and $MSSE$ is not available. In Figure 4 we have graphed $MSRE_k$ for $k = 18$ since this value of k gives a ratio $\sum_{r=1}^{k}\lambda_r / \sum_{r=1}^{m}\lambda_r$ that just exceeds 0.99, indicating that an approximation that explains a little over 99% of the individual variation in each of the m–lagged vectors of $\mathbf{X}$ is being used. Since β_n^2 converges to a random variable as n increases the eigenvalues of $\mathbf{\Gamma}$ are random and we cannot use Theorem 1 to calculate fixed asymptotic bounds. In Figure 4 we have therefore only plotted the average value of $MSRE_{18}$ evaluated across 10000 replications.

The gradual increase in $MSRE_k$ as m increases beyond k is typical of what we observe in the random walk case. This confirms our previous finding, namely that from the perspective of $MSRE$ smaller values of m, i.e. shorter window lengths, are to be preferred since they generate smaller values of $MSRE$.

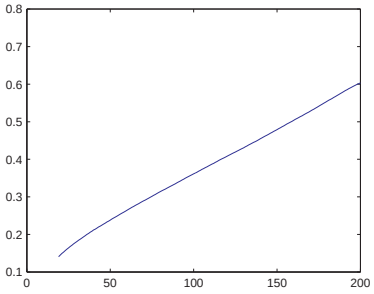


Figure 4: Simulated mean squared reconstruction error $MSRE_{18}$ as a function of window length – random walk process.

5 Concluding Remarks

In this paper we have presented a theoretical analysis of SSA signal–noise separation and reconstruction, a critical initial step that underlies any subsequent use of SSA for other purposes, such as forecasting or the analysis of missing data or change point problems. We established bounds on $MSRE$ and $MSSE$ under very general regularity conditions and our simulation results showed that the theoretical characteristics are reflected in observed behaviour.

The data generating mechanisms considered in our simulations represent two extreme cases. The first consisting of a deterministic signal corrupted by noise, and the second an uncontaminated random walk. If we think of a chaotic series as being one where initially close values diverge so that all predictability is lost as the system evolves in time, then the strong Markov property of a random walk makes this the quintessential example of a chaotic, statistically self similar stochastic process, the random fractal *par excellence*. In both cases our results show that the use of the rule $m = (\log N)^c$ with $c \in (1.5, 3.0)$ to assign window length will yield near optimal performance.

Golyandina (2010) advocates setting m close to one-half of the time series length. Our results indicate that this recommendation is not to be followed in all cases. We should emphasize that we are not suggesting that the analysis conducted in Golyandina (2010) is incorrect, indeed our own results are in accord with those found therein. We would argue, however, that Golyandina's counsel is based upon an analysis of a limited class of processes and is founded on the sole use of $MSSE$ as a guide to performance. Examination of more general processes and consideration of the alternative measure $MSRE$ suggests that Golyandina's conclusion is overstated.

Faced with alternative specifications the practitioner is required to make a choice and a preference for the specification $m = (\log N)^c$ with $c \in (1.5, 3.0)$ might be justified on three grounds. First, the $m = (\log N)^c$ rule yields a sensible compromise between $MSRE$ and $MSSE$ that will avoid the worst choices and provide near optimal performance for both. Second, whereas the optimality of setting m near $N/2$ for

$MSSE$ seems to be contingent on using the correct signal dimension, the properties of $MSRE$ appear to be invariant to the choice of k and $MSRE$ is minimized by using smaller rather than larger values of m. Third, since in practice the calculation of $MSSE$ is infeasible $MSRE$ has more empirical relevance and, as we have just observed, $MSRE$ is minimized by using smaller rather than larger values for the window length.[2]

References

Basilevsky, A. and D. P. J. Hum (1979) Karhunen-Loève analysis of historical time series with an application to plantation births in Jamaica, *Journal of the American Statistical Association*, **74**(366), 284–290.

Broomhead, D. and G. King (1986) Extracting qualitative dynamics from experimental data, *Physica D: Nonlinear Phenomena*, **20**(2-3), 217–236.

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

Forni, M. and M. Lippi (2001) The generalized dynamic factor model: representation theory, *Econometric Theory*, **17**(6), 1113–1142.

Golyandina, N. (2010) On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods, *Statistics and Its Interface*, **3**(3), 259–279.

Golyandina, N., V. V. Nekrutkin and A. A. Zhigljavski (2001) *Analysis of Time Series Structure: SSA and Related Techniques*, CRC Press, Boca Raton.

Jolliffe, I. T. (2002) *Principal Component Analysis*, Springer, Heidelberg.

Kendal, M. and A. S. Stuart (1979) *The Advanced Theory of Statistics*, Griffin, London.

Khan, M. A. R. and D. S. Poskitt (2010) Description length based signal detection in singular spectrum analysis, *Monash Econometrics and Business Statistics Working Papers*, **13/10**.

Poskitt, D. S. (2000) Strongly consistent determination of cointegrating rank via canonical correlations, *Journal of Business and Economic Statistics*, **18**, 77–90.

Pötscher, B. M. and I. R. Prucha (1997) *Dynamic Ninlinear Econometric Models: Asymptotic Theory*, Springer, New York.

[2]For example, for monthly rose wine sales (thousand of litres) in Australia from July 1980 to June 1994 (`www.gistatgroup.com/cat/book2/bookdata.html`) Golyandina et al. (2001) used $m = 0.5N = 84$ with $k = 14$, giving $MSRE_{14} = 173.8386$, using $m = (logN)^{2.2} = 36$ gives $MSRE_{14} = 104.8034$.

Rao, C. R. (1965) *Linear Statistical Inference and its Applications*, John Wiley, New York.

Rao, M. M. (1985) *Harmonizable, Cramér, and Karhunen classes of processes*, in: Hannan, E.J., Krishnaiah, P.R. and Rao, M.M. (Eds.), Time Series in the Time Domain, Handbook of Statistics 5, pp. 279-310, North-Holland, Amsterdam.

Vautard, R. and M. Ghil (1989) Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D: Nonlinear Phenomena*, **35**(3), 395–424.

Vautard, R., P. Yiou and M. Ghil (1992) Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, *Physica D: Nonlinear Phenomena*, **58**(1-4), 95–126.