MONASH
BUSINESS
SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

Singular Spectrum Analysis of Grenander Processes and Sequential Time Series Reconstruction

D.S. Poskitt

August 2016

Working Paper 15/16

Singular Spectrum Analysis of Grenander Processes and Sequential Time Series Reconstruction

D. S. Poskitt
Department of Econometrics and Business Statistics,
Monash University, VIC 3800
Australia.
Email: Donald.Poskitt@monash.edu.au

31 July 2016

JEL classification: C14, C22, C52

Singular Spectrum Analysis of Grenander Processes and Sequential Time Series Reconstruction

Abstract: This paper provides a detailed analysis of the properties of Singular Spectrum Analysis (SSA) under very general conditions concerning the structure of the observed series. It translates the SSA interpretation of the singular value decomposition of the so called trajectory matrix as a discrete Karhunen-Loève expansion into conventional principle components analysis, and shows how this motivates a consideration of SSA constructed using standardized or re-scaled trajectories (R-SSA). The asymptotic properties of R-SSA are derived assuming that the true data generating process (DGP) satisfies sufficient regularity to ensure that Grenander's conditions are satisfied. The spectral structure of different population ensemble models implicit in the large sample properties so derived is examined and it is shown how the decomposition of the spectrum into discrete and continuous components leads to an application of sequential R-SSA series reconstruction. As part of the latter exercise the paper presents a generalization of Szegö's theorem to fractionally integrated precesses. The operation of the theoretical results is demonstrated via simulation experiments. The latter serve as a vehicle to illustrate the numerical consequences of the results in the context of different processes, and to assess the practical impact of the sequential R-SSA processing methodology.

Keywords: embedding; principle components; re-scaled trajectory matrix; singular value decomposition; spectrum.

1 Introduction

Singular spectrum analysis (SSA) is a non-parametric modeling technique that is designed to accommodate nonlinear, non–stationary, and intermittent or transient behaviour in an observed time series. The basic idea behind SSA is to represent the observed series as the sum of uncorrelated components, a signal-plus-noise decomposition, where the decomposition is obtained via a singular value decomposition (SVD) of the so-called (in the terminology of SSA) *trajectory* matrix. Many of the basic ideas and methods used in SSA were developed in the physical sciences by Broomhead and King (1986) and Vautard and Ghil (1989), and they have subsequently been described in Elsner and Tsonis (1996) and Golyandina et al. (2001), where more detailed accounts of the techniques and their practical application (with several examples) can be found. An early formulation of some of the key ideas in a socio-demographic setting can be found in the work of Basilevsky and Hum (1979), and several of the basic building blocks of SSA can be traced back to Prony (1795). In more recent times SSA has been applied in a diverse range of disciplines, including meteorology (Ghil et al., 2002), bio-mechanics (Alonso et al., 2005) and hydrology (Marques et al., 2006), as well as in economics and finance (Hassani and Thomakos, 2010). See also Jolliffe (2002, Chapters 12.1 and 12.2).

Suppose that $x(t)$ is a time series of interest that is observed on a uniform grid, giving rise to a realization $\{x(1), x(2), \ldots, x(N)\}$ of 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 a SVD of the so called trajectory matrix, an $m \times n$ matrix $\mathbf{X}$ defined by the mapping

[1] To avoid a proliferation of notation we adopt the common practice of not distinguishing between a process and realized values of that process, relying on the context or some explicit statement to make the meaning clear.

$$\begin{bmatrix} x(1) \\ \vdots \\ x(N) \end{bmatrix} \mapsto \mathbf{X} = \left.\begin{bmatrix} x(1) & x(2) & \dots & x(n) \\ x(2) & x(3) & \dots & x(n+1) \\ \vdots & \vdots & & \vdots \\ x(m) & x(m+1) & \dots & x(N) \end{bmatrix} \begin{matrix} 1^{\text{st}} \\ 2^{\text{nd}} \\ \vdots \\ m^{\text{th}} \end{matrix}\right\} \text{trajectory}, \tag{1}$$

where m is a preassigned window length, variously referred to as the trajectory matrix window size or lag length, and $n = N - m + 1$.

The SVD of $\mathbf{X}$ implies that $\mathbf{X}$ can be written as the sum of m rank one orthogonal components,

$$\mathbf{X} = \sum_{i=1}^{m} \mathbf{X}_i = \sum_{i=1}^{m} \sqrt{\ell_i}\mathbf{u}_i\mathbf{v}_i^\top \tag{2}$$

where $\sqrt{\ell_1} \geq \sqrt{\ell_2} \geq \ldots \geq \sqrt{\ell_m} > 0$ denote the the singular values of $\mathbf{X}$ arranged in descending order of magnitude, and $\mathbf{u}_i^\top$ and $\mathbf{v}_i = \mathbf{X}^\top\mathbf{u}_i/\sqrt{\ell_i}$, $i = 1, \ldots, m$, are the corresponding left and right orthonormal eigenvectors. Borrowing the nomenclature of Basilevsky and Hum (1979), the decomposition in (2) will be referred to as a discrete Karhunen-Loève expansion with coefficients $\mathbf{u}_j$ and empirical eigen functions $\mathbf{v}_j$, $j = 1, \ldots, m$. Let $\|\mathbf{X}\| = \sqrt{\text{trace}\{\mathbf{X}\mathbf{X}^\top\}}$ denote the Frobenius norm of $\mathbf{X}$. Since $\|\mathbf{X}\|^2 = \sum_{i=1}^{m} \ell_i$ and $\|\mathbf{X}_i\|^2 = \ell_i$ for $i = 1, \ldots, m$, the ratio $\ell_i / \sum_{i=1}^{m} \ell_i$ represents the proportion of the total variation in $\mathbf{X}$ attributable to $\mathbf{X}_i$. Now suppose that a large proportion of the total variation can be associated with a subset of dominant eigentriples $\{\ell_i, \mathbf{u}_i, \mathbf{v}_i\}$, $i = 1, \ldots, k$. If the designated eigentriples are thought to encompass variation due to the presence of a signal in the original series, then

$$\mathbf{X} = \sum_{i=1}^{k} \mathbf{X}_i + \sum_{i=k+1}^{m} \mathbf{X}_i = \mathbf{X}_{\mathscr{S}} + \mathbf{X}_{\mathscr{N}},$$

where $\mathbf{X}_{\mathscr{S}} = \sum_{i=1}^{k} \mathbf{X}_i$ can be viewed as the component of $\mathbf{X}$ due to the signal, with k being the dimension or rank of the signal, and the remainder $\mathbf{X}_{\mathscr{N}} = \sum_{i=k+1}^{m} \mathbf{X}_i$ can then be interpreted as the component due to noise.

Given the decomposition $\mathbf{X} = \mathbf{X}_{\mathscr{S}} + \mathbf{X}_{\mathscr{N}}$, the transformation into a corresponding signal-plus-noise reconstruction of the original series is achieved by a process of diagonal

averaging or Hankelisation. The Hankel operator $\mathscr{H}(\cdot)$ replaces the r,cth entry of $\mathbf{X}_{\mathscr{G}} = [s_{rc}]_{r=1,\ldots,m,c=1,\ldots,n}$ by the average over all r and c such that $r+c-1=t$ where $r=1,\ldots,m$, $c=1,\ldots,n$ and $t=1,\ldots,N$.[2] By so doing the operator implicitly defines a time series and an associated orthosymmetric trajectory matrix, $\{s(1),s(2),\ldots,s(N)\}$ and $\mathscr{H}(\mathbf{X}_{\mathscr{G}}) = \mathbf{S} = [s(r+c-1)]_{r=1,\ldots,m,c=1,\ldots,n}$ say, where

$$s(t) = \begin{cases} \frac{1}{t}\sum\sum_{r+c=t+1} s_{rc}\,, & \text{when } 1 \le t \le m-1; \\ \frac{1}{m}\sum\sum_{r+c=t+1} s_{rc}\,, & \text{when } m \le t \le n; \\ \frac{1}{N-t+1}\sum\sum_{r+c=t+1} s_{rc}\,, & \text{when } n+1 \le t \le N. \end{cases} \tag{3}$$

After applying diagonal averaging the resulting Hankel matrices $\mathbf{S}$ and $\mathbf{E} = \mathbf{X} - \mathbf{S}$ give the signal–noise decomposition $\mathbf{X} = \mathbf{S} + \mathbf{E}$ of the trajectory matrix, and $x(t) = s(t) + \varepsilon(t)$, $t = 1,\ldots,N$, yields the associated signal–noise reconstruction of the original time series.

The reconstruction obviously depends on the window length, a tuning parameter that must be assigned by the practitioner, and the signal dimension, a modeling parameter that the practitioner must select, and we will denote an SSA model with window length m and signal dimension k by $SSA(m,k)$. In practice the methods outlined in Elsner and Tsonis (1996) and Golyandina et al. (2001) are commonly employed to determine the specification of the $SSA(m,k)$ model. This approach employs pattern recognition techniques and methods similar to those used in conventional principal component analysis – the use of scree-plots and various correlation methods as described in Jolliffe (2002, Chapter 6.) – and supposes that values for m, $2 < m \le N/2 \le n$, and $k < m$ are allocated that ensure that the signal and noise components are *strongly separated*. An alternative approach advanced in Khan and Poskitt (2013, 2015) is to employ appropriate model selection decision rules derived from more formal statistical techniques based on hypothesis testing procedures and description length principles.

[2]Note that diagonal averaging is a minimum norm operation in the sense that for any matrix $\mathbf{X}$ the Hankel matrix $\mathscr{H}(\mathbf{X})$ minimizes the Frobenius norm of the approximation error, i.e. $\|\mathbf{X} - \mathscr{H}(\mathbf{X})\| \le \|\mathbf{X} - \mathbf{H}\|$ for all conformable Hankel matrices $\mathbf{H}$.

Whatever approach is adopted to determine the model specification, the SVD in (2) plays a key role in the model selection. In the following section the SSA interpretation of the representation in (2) as a discrete Karhunen-Loève expansion is translated back into conventional principle components analysis (as in Jolliffe (2002, Chapters 12.1 and 12.2)) and it is shown how this motivates a consideration of SSA constructed using standardized or re-scaled trajectories (R-SSA).

In Section 3 the asymptotic properties of R-SSA are derived under the presumption that the observed series satisfies very general regularity conditions. The assumptions admit series of a rather general type and imply that Grenander's conditions (Grenander, 1954) concerning the structure of the observed series apply. The features of different asymptotic population ensemble models implicit in the large sample properties derived in Section 3 are examined in Section 4.

The exposition in Section 4 exploits some classical time series analysis spectral techniques to be found, for example, in Grenander and Rosenblatt (1957) and Anderson (1971), and presents a generalization of Szegö's theorem (Grenander and Szego, 1958) to fractionally integrated processes. The results in Section 4 motivate the consideration of a two stage sequential R-SSA methodology in which the first stage series reconstruction deals with the discrete component of the process spectrum and the second stage the continuous component.

Section 5 illustrates the workings of the theoretical results in the context of particular examples, and presents numerical results obtained via some related simulations. This section serves as a vehicle to illustrate the manifestation of the results in the context of different processes, and provides a guide to the practical impact of the sequential R-SSA methodology. Section 6 presents a brief conclusion.

2 Centered and Re-scaled SSA

In order to translate the discrete Karhunen-Loève representation $\mathbf{X} = \sum_{i=1}^{m} \sqrt{\ell_i}\mathbf{u}_i\mathbf{v}_i^\top$ of SSA into the terminology of conventional principle component analysis, note that

the values seen in the trajectory matrix in (1) can be viewed as n observations on m variables, the first row of $\mathbf{X}$ corresponding to the first variable, the second row to the second, and so on through to the mth row corresponding to the last variable. We can think of $\mathbf{X}$ (heuristically) as containing data on an m component vector random variable, in which case the ith Karhunen-Loève coefficient vector $\mathbf{u}_i$ corresponds to the ith principle component loading, and the ith empirical eigenfunction $\mathbf{v}_i$ corresponds to the ith principle component.

Principle components are neither location nor scale invariant. In classical multivariate analysis it is common practice to address the lack of location invariance by applying mean correction. Lack of scale invariance is dealt with by either ensuring that all the variables are of the same type and are measured in "natural" units, or by standardizing the variables if their values are recorded using widely different units of measurement. In the former case the principle components are derived from the sample covariance matrix, and in the latter the sample correlation matrix, and the components so obtained are not the same, nor is it possible to pass from one solution to the other by means of a simple transformation. See Jolliffe (2002, Chapter 2.3).

In SSA mean correction is often not applied since a constant mean value is deemed to be inappropriate. Furthermore, the different trajectories are obviously evaluated using the same unit of measurement and standardization is more often than not, therefore, also omitted. However, despite being measured using the same units, it is possible that the values obtained in different trajectories may not be commensurate; for example, if the series contains strong trends or a high degree of heteroscedasticity. In light of this possibility we wish to examine the consequences of employing standardization in SSA.

2.1 Mean Correction: Centered SSA

Basic SSA is implemented via an orthogonal decomposition of the trajectory matrix $\mathbf{X}$ using the eigentriples $\{\ell_i, \mathbf{u}_i, \mathbf{v}_i\}$, $i = 1, \ldots, m$, of $\mathbf{X}$ itself. Centred SSA works in

terms of the eigentriples $\{\bar{\ell}_i, \bar{\mathbf{u}}_i, \bar{\mathbf{v}}_i\}$, $i = 1, \ldots, m$, of the re–centred matrix $\mathbb{X} = \mathbf{X} - \bar{\mathbf{x}}\mathbf{1}_n^\top$ where $\bar{\mathbf{x}} = n^{-1}\mathbf{X}\mathbf{1}_n = (\bar{x}_0, \ldots, \bar{x}_{m-1})^\top$, $\mathbf{1}_n = (1, \ldots, 1)^\top$. Let $\mathbf{v}_0 = \mathbf{1}_n/\sqrt{n}$ and set $\ell_0 = n\|\bar{\mathbf{x}}\|^2$. Direct calculation shows that $\mathbf{u}_0 = \mathbf{X}\mathbf{v}_0/\sqrt{\ell_0} = \bar{\mathbf{x}}/\|\bar{\mathbf{x}}\|$ and $\mathbf{v}_0^\top \mathbf{X}^\top \mathbf{X}\mathbf{v}_0 = \ell_0$; the eigentriple $\{\ell_0, \mathbf{u}_0, \mathbf{v}_0\}$ is often referred to as the *first average triple* of $\mathbf{X}$. Now $\mathbb{X}\mathbf{v}_0 = \mathbf{X}\mathbf{v}_0 - \bar{\mathbf{x}}\mathbf{1}_n^\top \mathbf{v}_0 = \mathbf{0}$ and $\mathbf{v}_0^\top \mathbb{X}^\top \mathbb{X}\bar{\mathbf{v}}_i = \bar{\ell}_i \mathbf{v}_0^\top \bar{\mathbf{v}}_i = 0$, implying that $\mathbf{v}_0$ is orthogonal to $\bar{\mathbf{v}}_i$ for all i such that $\bar{\ell}_i \neq 0$. Thus $\bar{\mathbf{v}}_i$ for $i = 1, 2, \ldots, m$ and $\mathbf{v}_0$ form an orthonormal system and $\mathbf{X} = \mathbb{X} + \bar{\mathbf{x}}\mathbf{1}_n^\top = \sum_{i=1}^m \sqrt{\bar{\ell}_i}\bar{\mathbf{u}}_i\bar{\mathbf{v}}_i^\top + \sqrt{\ell_0}\mathbf{u}_0\mathbf{v}_0^\top$ yields an alternative orthogonal decomposition of $\mathbf{X}$ such that $\|\mathbf{X}\|^2 = \ell_0 + \sum_{i=1}^m \bar{\ell}_i$.

Centered SSA now proceeds as for basic SSA by replacing the eigentriples $\{\ell_i, \mathbf{u}_i, \mathbf{v}_i\}$, $i = 1, \ldots, m$, in the decomposition $\mathbf{X} = \mathbf{X}_{\mathscr{S}} + \mathbf{X}_{\mathscr{N}}$, where $\mathbf{X}_{\mathscr{S}} = \sum_{i=1}^k \sqrt{\ell_i}\mathbf{u}_i\mathbf{v}_i^\top$ and $\mathbf{X}_{\mathscr{N}} = \sum_{i=k+1}^m \sqrt{\ell_i}\mathbf{u}_i\mathbf{v}_i^\top$, with $\{\ell_0, \mathbf{u}_0, \mathbf{v}_0\}$ and $\{\bar{\ell}_i, \bar{\mathbf{u}}_i, \bar{\mathbf{v}}_i\}$, $i = 1, \ldots, m$. Thus, in the centered SSA(m,k) model the signal component corresponds to the first average triple plus the eigentriples $\{\bar{\ell}_i, \bar{\mathbf{u}}_i, \bar{\mathbf{v}}_i\}$, $i = 1, \ldots, k$, so that $\mathbf{X}_{\mathscr{S}} = \sqrt{\ell_0}\mathbf{u}_0\mathbf{v}_0^\top + \sum_{i=1}^k \sqrt{\bar{\ell}_i}\bar{\mathbf{u}}_i\bar{\mathbf{v}}_i^\top$ and $\mathbf{X}_{\mathscr{N}} = \sum_{i=k+1}^m \sqrt{\bar{\ell}_i}\bar{\mathbf{u}}_i\bar{\mathbf{v}}_i^\top$. This specification leads to a signal component that is deemed to be a constant plus a process of dimension k, in parallel with the basic SSA(m,k) model. For further details see Golyandina et al. (2001, sec. 6.3).

2.2 Standardization: Re-scaled SSA

Standardization in SSA is accomplished by re-scaling each trajectory by reference to its observed sample second moment. Thus, in the case of basic SSA the raw trajectory matrix is replaced by $\mathbf{Y} = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}$ where

$$\begin{aligned}\mathbf{D} &\equiv \operatorname{diag}\{\mathbf{X}\mathbf{X}^\top\} \\ &= \begin{bmatrix} \sum_{t=1}^n x(t)^2 & 0 & \cdots & \cdots & 0 \\ 0 & \sum_{t=1}^n x(t+1)^2 & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \sum_{t=1}^n x(t+m-1)^2 \end{bmatrix},\end{aligned}$$

and the analysis proceeds on the basis of the eigentriples $\{\ell_i, \mathbf{u}_i, \mathbf{v}_i\}$, $i = 1, \ldots, m$, calculated from $\mathbf{Y}\mathbf{Y}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^\top\mathbf{D}^{-\frac{1}{2}}$.[3] The signal-noise decomposition of $\mathbf{X}$ is constructed from that of $\mathbf{Y}$ as $\mathbf{X}_\mathscr{S} + \mathbf{X}_\mathscr{N} = \mathbf{D}^{\frac{1}{2}}(\mathbf{Y}_\mathscr{S} + \mathbf{Y}_\mathscr{N})$ where $\mathbf{Y}_\mathscr{S} = \sum_{i=1}^{k} \sqrt{\ell_i}\mathbf{u}_i\mathbf{v}_i^\top$ and $\mathbf{Y}_\mathscr{N} = \sum_{i=k+1}^{m} \sqrt{\ell_i}\mathbf{u}_i\mathbf{v}_i^\top$. Similarly, for centered SSA the normalised eigentriples $\{\bar{\ell}_i, \bar{\mathbf{u}}_i, \bar{\mathbf{v}}_i\}$, $i = 1, \ldots, m$, are calculated from $\mathbb{Y}\mathbb{Y}^\top = \mathbb{D}^{-\frac{1}{2}}\mathbb{X}\mathbb{X}^\top\mathbb{D}^{-\frac{1}{2}}$ where, obviously, $\mathbb{Y} = \mathbb{D}^{-\frac{1}{2}}\mathbb{X}$ and $\mathbb{D} = \mathbf{D} - \mathrm{diag}\{n\bar{x}_0^2, \ldots, n\bar{x}_{m-1}^2\}$. The signal-noise decomposition of $\mathbb{Y}$ is then constructed as $\mathbb{Y} = \mathbb{Y}_\mathscr{S} + \mathbb{Y}_\mathscr{N}$ where $\mathbb{Y}_\mathscr{S} = \sum_{i=1}^{k} \sqrt{\bar{\ell}_i}\bar{\mathbf{u}}_i\bar{\mathbf{v}}_i^\top$ and $\mathbb{Y}_\mathscr{N} = \sum_{i=k+1}^{m} \sqrt{\bar{\ell}_i}\bar{\mathbf{u}}_i\bar{\mathbf{v}}_i^\top$, and the signal-noise decomposition of the raw trajectory matrix is given by $\mathbf{X}_\mathscr{S} = \sqrt{\ell_0}\mathbf{u}_0\mathbf{v}_0^\top + \mathbb{D}^{\frac{1}{2}}\mathbb{Y}_\mathscr{S}$ and $\mathbf{X}_\mathscr{N} = \mathbb{D}^{\frac{1}{2}}\mathbb{Y}_\mathscr{N}$. Henceforth we will denote an SSA model calculated from re-scaled trajectories using a window length m and a signal dimension k by *R-SSA*(m, k).

For both basic and re-centered SSA the effect of re-scaling is that the eigentriples used in the signal-noise decomposition are derived from correlation matrices, in keeping with usage established in the principle factor method of factor analysis, as described in Mardia et al. (1979, Chapter 9.2 & 9.3) and Jolliffe (2002, Chapter 7.3). See also Watanabe (1965) for a discussion of the relationship between the Karhunen-Loéve expansion and the factor analysis model. Our purpose here is to examine the consequences of such considerations for SSA.

3 Grenander Processes

In order to examine the statistical properties of SSA we will assume that $x(t)$ is generated by an affine combination of processes (not necessarily stochastic) that satisfy the following ergodicity assumption.

Assumption 1 *The data generating mechanism underlying the observed process can be characterized as* $x(t) = \alpha_1 z_1(t) + \cdots + \alpha_d z_d(t) = \boldsymbol{\alpha}^\top \mathbf{z}(t)$ *where* $0 < \|\boldsymbol{\alpha}\| = \sqrt{\boldsymbol{\alpha}^\top\boldsymbol{\alpha}} < \infty$ *and the process* $\mathbf{z}(t) = (z_1(t), \ldots, z_d(t))^\top$ *satisfies the following conditions almost surely:*

[3] Here and in what follows we will not distinguish between the eigentriples calculated from $\mathbf{Y}$ and those calculated from $\mathbf{X}$. It should be clear from the context whether it is the eigentriples calculated from the re-scaled trajectory matrix or those from the raw trajectory matrix that is meant.

1. $\lim_{N\to\infty} a_{ii}^N(0) = \infty$ *where* $a_{ij}^N(h) = \sum_{t=1}^{N-h} z_i(t+h)z_j(t)$*;*

2. *if* S_N *denotes a subset of* T_N *indices* t*,* $1 \le t \le N$*, such that* $\lim_{N\to\infty} T_N/N = 0$*, then*

$$\lim_{N\to\infty} \frac{\sum_{t\in S_N} z_i^2(t)}{a_{ii}^N(0)} = 0$$

uniformly in T_N/N *for all* $T_N/N < \delta$*; and*

3. *for* $h = 0, 1, \ldots, T_N$*,* $\lim_{N\to\infty} r_{ij}^N(h) = \rho_{ij}(h)$*, where*

$$r_{ij}^N(h) = \frac{a_{ij}^N(h)}{\sqrt{a_{ii}^N(0)a_{jj}^N(0)}}.$$

In addition,

4. *if* $\mathbf{R}(h)$ *denotes the* $d \times d$ *matrix with entries* $\rho_{ij}(h)$*,* $i, j = 1, \ldots, d$*, then* $\mathbf{R}(0)$ *is nonsingular.*

A series that satisfies Assumption 1 will satisfy the conditions introduced by Grenander (1954); and a process that satisfies the conditions of Assumption 1 will therefore be referred to as a Grenander process, and the conditions of Assumption 1 will be called Grenander's conditions. The first condition ensures that the process does not ultimately degenerate and is persistently exciting. The second condition implies that for each $i = 1, \ldots, d$ we have $\lim_{N\to\infty} \sum_{t=1}^n z_i^2(t+r)/a_{ii}^n(0) = 1$ for all $r = 0, \ldots, m-1$. A necessary condition that a sequence $z_i(t)$ satisfy the second condition is that $\lim_{N\to\infty} a_{ii}^N(0)/(\log N)^p = \infty$ for $p > 0$, and $z_i(t)$ must increase more slowly than exponentially. The second and third conditions ensure that

$$\lim_{n\to\infty} \frac{\sum_{t=1}^n z_i(t+r)z_j(t+s)}{\sqrt{\sum_{t=1}^n z_i^2(t+r)\sum_{t=1}^n z_j^2(t+s)}} = \rho_{ij}(r-s), \quad r,s = 0, \ldots, m-1,$$

so that the correlation computed from the subsets $z_i(t+r)$ and $z_j(t+s)$, $t = 1, \ldots, n$, with $r-s = h$ posses the common limit $\rho_{ij}(h)$, provided of course that $\lim_{N\to\infty} m/N = 0$. Condition 4 rules out the possibility that $\mathbf{z}(t)$ contains linearly dependent redundant processes.

It is well known that polynomial trends and trigonometric series obey Assumption 1, see Grenander and Rosenblatt (1957, Chapter 7.5 & 7.6) and Anderson (1971, Chapter 10.2.3). Primitive regularity conditions under which statistical ergodic properties of the type implicit in Assumption 1 are applicable to stochastic processes are also well documented in the literature, see *inter alia* Davidson (1994, Part IV). It should, perhaps, be emphasized that the specification $x(t) = \alpha_1 z_1(t) + \cdots + \alpha_d z_d(t)$ is being used here to characterize the unknown data generating process (DGP), it is akin to a dynamic factor model (Forni and Lippi, 2001), but it does not represent a statistical model that is to be fitted to data.

As an example of the application of Grenander's conditions, suppose that $x(t)$ equates to the solution of a difference equation with polynomial difference operator $d(z) = \prod_{i=1}^{r}(1-\zeta_i^{-1}z)^{m_i}$ where ζ_i denotes a distinct root of $d(z) = 0$ of multiplicity m_i, $i = 1,\ldots,r$. Solving for $x(t)$ gives

$$x(t) = \sum_{i=1}^{r}\Big(\sum_{j=0}^{m_i-1}\alpha_{ij}t^j\Big)\zeta_i^{-t} \tag{4}$$

for $t > \sum_{i=1}^{r} m_i$ where the coefficients α_{ij}, $j = 0,1,\ldots,m_i-1$, $i = 1,\ldots,r$, are obtained from the initial conditions. Since

$$\lim_{t\to\infty} t^j \exp(-t\log|\zeta|) = \begin{cases} 0, & |\zeta| > 1\,; \\ \infty, & |\zeta| < 1\,, \end{cases}$$

for all $j = 1,\ldots,\max\{m_1,\ldots,m_r\}-1$ it follows from (4) that Grenander's conditions will be violated if $\min_{i=1,\ldots,r}\{|\zeta_i|\} > 1$, or $\max_{i=1,\ldots,r}\{|\zeta_i|\} < 1$, or $\min_{i=1,\ldots,r}\{|\zeta_i|\} < 1$ and $\max_{i=1,\ldots,r}\{|\zeta_i|\} \geq 1$. If, however, $\zeta_i = \exp(\imath\theta_i)$ where $-\pi \leq \theta_i \leq \pi$, $i = 1,\ldots,r$, we can write $x(t)$ as a linear combination of

$$\left.\begin{matrix} t^j\cos(\theta_i t) \\ t^j\sin(\theta_i t) \end{matrix}\right\} \quad j = 0,\ldots,m_i-1, \quad i = 1,\ldots,r\,, \tag{5}$$

where of course the sine terms do not occur if $\theta_i = 0$ or $\theta_i = \pm\pi$. Let $\mathbf{z}(t)$ denote the sequence in which these are arranged such that all those with the same frequency occur together; so that $t^j \cos(\theta_i t)$ is followed by $t^j \sin(\theta_i t)$, $j = 0, \ldots, m_i - 1$, for $i = 1, \ldots, r$. Then following the arguments in (Grenander and Rosenblatt, 1957, pages 245-248) or (Anderson, 1971, pages 581-583) we find that $\mathbf{R}(h)$ is block diagonal with ith $m_i \times m_i$ diagonal block $\mathbf{T}_i$ where the matrix

$$\mathbf{T}_i = \left[\frac{\sqrt{(2r-1)(2c-1)}}{(r+c-1)} \right]_{r,c=1,\ldots,m_i}$$

if $\theta_i = 0$ or $\theta_i = \pm\pi$, and $2m_i \times 2m_i$ diagonal block $\mathbf{G}_i \otimes \mathbf{T}_i$ where

$$\mathbf{G}_i = \begin{bmatrix} \cos(\theta_i h) & -\sin(\theta_i h) \\ \sin(\theta_i h) & \cos(\theta_i h) \end{bmatrix},$$

the Givens rotation matrix at frequency $-\theta_i h$, otherwise. This example is relevant to our subsequent discussion, where some specific examples of processes of this type are examined in more detail.

The Whitney embedding theorem (Broomhead and King, 1986, Section 2.3) states that a k-dimensional manifold with $k > 0$ can be smoothly embedded in a Euclidean space of dimension $2k+1$. In order to see the relevance of this result here, suppose that the observed process satisfies a difference equation of order $k = \sum_{i=1}^{r} m_i$. Then we can express each of $x(t+k), \ldots, x(t+m-1)$ as a linear combination of $x(t), \ldots, x(t+k-1)$ where the coefficients of the polynomial difference operator can be determined from the $(k+1)k$ entries in the sub–matrix $\mathbf{X}_{11} = [x(r+c-1)]_{r=1,\ldots,k+1,c=1,\ldots,k}$. The entries of $\mathbf{X}_{11}$ are uniquely defined by the $2k$ values $x(1), \ldots, x(2k)$, however, and the Hankel structure of $\mathbf{X}_{11}$ and $\mathbf{X}$ means that all the entries of $\mathbf{X}$ can be generated recursively from the $2k+1$ values $x(1), \ldots, x(2k+1)$. The upshot of this is that the smallest window length consistent with the reproduction of the observed trajectory matrix in this manner is $m = k+1$.

If asymptotic properties of SSA are to be derived then it is essential to impose Grenander's conditions, or some such restrictions, on the properties of the observed DGP. As we have seen, Assumption 1 admits a broad range of series, and the imposition of Grenander's conditions appears not to exclude any cases of importance in SSA. This implies that any results based upon Assumption 1 are likely to have broad applicability.[4]

Theorem 1 *If the process $x(t)$ satisfies Assumption 1 then for any window length $m \leq M$ where $\lim_{N\to\infty} M/N = 0$ there exists a unique function of bounded variation, $F(\omega)$, $-\pi \leq \omega \leq \pi$, whose increments are non-negative definite, such that*

$$\lim_{N\to\infty} \|\mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^{\top}\mathbf{D}^{-\frac{1}{2}} - \int_{-\pi}^{\pi} \mathbf{e}_m(\omega)\mathbf{e}_m(\omega)^* dF(\omega)\| = 0$$

where $\mathbf{e}_m(\omega)^* = (1, \exp(\imath\omega), \exp(\imath 2\omega), \ldots, \exp(\imath(m-1)\omega))$. The function is given by $F(\omega) = \boldsymbol{\mu}^{\top}\mathbf{N}(\omega)\boldsymbol{\mu}$ where

$$\boldsymbol{\mu} = \frac{\mathbf{R}(0)^{\frac{1}{2}}\boldsymbol{\beta}}{\sqrt{\boldsymbol{\beta}^{\top}\mathbf{R}(0)\boldsymbol{\beta}}}, \quad \boldsymbol{\beta}^{\top} = \left(\alpha_1\sqrt{a_{11}^n(0)}, \ldots, \alpha_d\sqrt{a_{dd}^n(0)}\right),$$

$\mathbf{R}(0)^{\frac{1}{2}}$ is the unique symmetric square root of $\mathbf{R}(0)$, and

$$\mathbf{N}(\omega) = \mathbf{R}(0)^{-\frac{1}{2}}\left(\lim_{N\to\infty}\sum_{h=-N+1}^{N-1}\mathbf{R}(h)\left[\frac{\exp(-\imath\omega h)-1}{-\imath h}\right]\right)\mathbf{R}(0)^{-\frac{1}{2}}.$$

The function $\mathbf{N}(\omega)$ introduced in Theorem 1 is a $d \times d$ matrix valued function whose increments, $\mathbf{N}(\omega_2) - \mathbf{N}(\omega_1)$, $\omega_1 \leq \omega_2$, are Hermitian non-negative definite. It is continuous from the right and null at $-\pi$, and by construction $\mathbf{N}(\omega)$ satisfies

$$\int_{-\pi}^{\pi} d\mathbf{N}(\omega) = \mathbf{I}_d. \tag{6}$$

An immediate corollary of Theorem 1 is that for any choice of window length $m \leq M$ where $M/N \to 0$ as $N \to \infty$ the almost sure limit of the Gramian $\mathbf{Y}\mathbf{Y}^{\top} = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^{\top}\mathbf{D}^{-\frac{1}{2}}$

[4] In what follows we will derive our results for basic SSA. The adaptations necessary to cater for re-centered SSA are straightforward.

from which the re-scaled SSA model will be constructed is given by the Toeplitz matrix

$$\mathbf{\Gamma}_m = \begin{bmatrix} 1 & \varrho(1) & \cdots & \varrho(m-2) & \varrho(m-1) \\ & 1 & \cdots & \cdot & \varrho(m-2) \\ & & \ddots & & \vdots \\ & -\cdots- & & 1 & \varrho(1) \\ & & & & 1 \end{bmatrix},$$

where $\varrho(h) = \int_{-\pi}^{\pi} e^{\iota\omega h} dF(\omega)$. Thus we find that the effect of adopting Grenander's conditions in conjunction with the use of the re-scaled trajectory matrix is to introduce a type of asymptotic covariance stationarity.

Theorem 2 *Suppose that the observed series $x(t)$ satisfies Assumption 1. Denote the eigenvalue-eigenvector pairs of $\mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^\top\mathbf{D}^{-\frac{1}{2}}$ by $\{\ell_j, \mathbf{u}_j\}$, $j = 1, \ldots, m$, and those of $\mathbf{\Gamma}_m$ by $\{\lambda_j, \boldsymbol{v}_j\}$, $j = 1, \ldots, m$. Then for each fixed $m \leq M$, where $M/N \to 0$ as $N \to \infty$, $\lim_{N\to\infty} |\ell_j - \lambda_j| = 0$ and $\lim_{N\to\infty} \|\varsigma_j \mathbf{u}_j - \boldsymbol{v}_j\| = 0$ where $\varsigma_j = sign(\boldsymbol{v}_j' \mathbf{u}_j)$, $j = 1, \ldots, m$.*

A corollary of Theorem 2 is that the spectral decompositions of $\mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^\top\mathbf{D}^{-\frac{1}{2}}$ and $\mathbf{\Gamma}_m$, namely $\mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^\top\mathbf{D}^{-\frac{1}{2}} = \sum_{i=1}^{m} \ell_i \mathbf{u}_i \mathbf{u}_i^\top$ and $\mathbf{\Gamma}_m = \sum_{i=1}^{m} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^\top$, will converge. This implies that for any process that satisfies sufficient regularity the values calculated from an observed realization will yield consistent estimates of the parameters of a corresponding (asymptotic) population ensemble *R-SSA*(m, k) model.

4 Population Ensemble Properties

In order to describe the corresponding population ensemble model we begin with a result due to Grenander and Rosenblatt (1957). Following Grenander and Rosenblatt, we will designate the set of points ω such that for any interval (ω_1, ω_2) containing ω, $\omega_1 < \omega < \omega_2$, the difference $\mathbf{N}(\omega_2) - \mathbf{N}(\omega_1)$ is nonnegative definite and not the null matrix, the spectrum of $\mathbf{N}(\omega)$.

Theorem 3 *The spectrum of $\mathbf{N}(\omega)$ can be uniquely partitioned into disjoint sets E_j, $j =$*

$1,\ldots,q \leq d$, *that differ only on a set of* trace$\{\mathbf{N}(\omega)\}/d$ *measure zero, such that*

$$\mathbf{P}(E_j) = \int_{E_j} d\mathbf{N}(\omega) > 0\,, \quad j = 1,\ldots,q\,,$$

and

$$\sum_{j=1}^{q} \mathbf{P}(E_j) = \mathbf{I}_d \quad \textit{and} \quad \mathbf{P}(E_i)\mathbf{P}(E_j) = 0\,, i \neq j\,.$$

There is no finer such partition.

The sets E_j, $j = 1,\ldots,q \leq d$ are called the elements of the spectrum and the spectral decomposition of $\mathbf{\Gamma}_m$ is obviously governed by the elements of the spectrum of $\mathbf{N}(\omega)$.

Suppose that the spectrum of $\mathbf{N}(\omega)$ has elements that are q distinct points $\omega_1,\ldots,\omega_q$, $q \leq d$. Then $\mathbf{N}(\omega)$ is made up of a denumerable set of saltuses. Theorem 3 indicates that the elements of the spectrum provide a unique maximal set of mutually annihilating Hermitian idempotents and we can therefore conclude that $\mathbf{N}(\omega)$ can be expressed as

$$\mathbf{N}(\omega) = \mathbf{P}_1 + \mathbf{P}_2 + \cdots + \mathbf{P}_j\,, \quad \omega_j \leq \omega < \omega_{j+1}$$

where the orthogonal projections $\mathbf{P}_j$, $j = 1,\ldots,q$, yield a resolution of the identity. In this case

$$\mathbf{\Gamma}_m = \sum_{j=1}^{q} \mathbf{e}_m(\omega_j)\mathbf{e}_m(\omega_j)^*\boldsymbol{\mu}^\top\mathbf{P}_j\boldsymbol{\mu}$$

and the eigenvalues of $\mathbf{\Gamma}_m$ are

$$\lambda_i = \sum_{j=1}^{q} |\gamma_{ji}|^2\boldsymbol{\mu}^\top\mathbf{P}_j\boldsymbol{\mu}$$

where $\gamma_{ji} = \mathbf{e}_m(\omega_j)^*\boldsymbol{v}_i$, for $i = 1,\ldots,\min\{m,q\}$, plus $\lambda = 0$ with multiplicity $m-q$ when $m > q$, where $\boldsymbol{v}_i$, $i = q+1,\ldots,m$, form a maximal orthonormal set orthogonal to the manifold generated by $\mathbf{e}_m(\omega_j)$, $j = 1,\ldots,q$. Since the eigenvectors $\boldsymbol{v}_i$, $i = 1,\ldots,m$, form an orthonormal basis for $\mathbb{R}^m$ we have $\mathbf{e}_m(\omega_j) = \sum_{i=1}^{m}\gamma_{ji}\boldsymbol{v}_i$ where

the coefficients γ_{ji} are such that $|\gamma_{ji}| \leq \sqrt{m}$ and $\sum_{i=1}^{m} |\gamma_{ji}|^2 = \|\mathbf{e}_m(\omega_j)\|^2 = m$. Consequently,

$$\begin{aligned}\sum_{i=1}^{m} \lambda_i &= \sum_{i=1}^{m}\sum_{j=1}^{q} |\gamma_{ji}|^2 \boldsymbol{\mu}^\top \mathbf{P}_j \boldsymbol{\mu} \\ &= m \sum_{j=1}^{q} \boldsymbol{\mu}^\top \mathbf{P}_j \boldsymbol{\mu} \\ &= m = \text{trace}\{\boldsymbol{\Gamma}_m\}, \quad \text{as it must,}\end{aligned}$$

where the last line follows since $\sum_{j=1}^{q} \mathbf{P}_j = \mathbf{I}_d$ and $\|\boldsymbol{\mu}\| = 1$.

Now, from Theorem 2 it follows that if the window length m is sufficiently large, i.e. $m > q$, then as $N \to \infty$ the SVD of $\mathbf{Y} = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}$ will contain q non-null singular values bounded away from zero and $m-q$ arbitrarily small singular values, reflecting that the elements of the spectrum of $\mathbf{N}(\omega)$ is made up of q distinct points. The corresponding signal-noise decomposition of the re-scaled trajectory matrix will give $\mathbf{Y} = \mathbf{Y}_{\mathscr{S}} + \mathbf{Y}_{\mathscr{N}}$ where

$$\mathbf{Y}_{\mathscr{S}} = \sum_{i=1}^{q} \sqrt{\ell_i}\mathbf{u}_i\mathbf{v}_i^\top = \sum_{i=1}^{q} \sqrt{\lambda_i}\boldsymbol{v}_i\boldsymbol{\nu}_i^\top + o(1) \tag{7}$$

where $\boldsymbol{\nu}_i = \mathbf{Y}^\top \boldsymbol{v}_i/\sqrt{\lambda_i}$, $i = 1, \ldots, q$, and

$$\mathbf{Y}_{\mathscr{N}} = \sum_{i=q+1}^{m} \sqrt{\ell_i}\mathbf{u}_i\mathbf{v}_i^\top = o(1). \tag{8}$$

In this case the population ensemble model implies that the *R-SSA*(m,q) model will reproduce the spectral characteristics of the discrete component of the re-scaled trajectory matrix of the observed series with an ever decreasing error as $N \to \infty$.

The process $x(t)$ consisting of an affine combination of $t^j\cos(\theta_i t)$ and $t^j\sin(\theta_i t)$, $j = 0, \ldots, m_i - 1$, for $i = 1, \ldots, r$, as described following (5) above, provides an example in which the elements of the spectrum of $\mathbf{N}(\omega)$ are distinct points of $(-\pi, \pi]$. The spectrum of $\mathbf{M}(\omega) = R(0)^{\frac{1}{2}}\mathbf{N}(\omega)\mathbf{R}(0)^{\frac{1}{2}}$ consists of a set of points determined by θ_i, $i = 1, \ldots, r$, with an $m_i \times m_i$ block diagonal jump of $\mathbf{T}_i$ at $\omega = \theta_i$ if $\theta_i = 0$ or $\theta_i = \pi$,

and a $2m_i \times 2m_i$ block diagonal jump of

$$\frac{1}{2}\begin{bmatrix} 1 & \imath \\ -\imath & 1 \end{bmatrix} \otimes \mathbf{T}_i \quad \text{and} \quad \frac{1}{2}\begin{bmatrix} 1 & -\imath \\ \imath & 1 \end{bmatrix} \otimes \mathbf{T}_i$$

at $\omega = \pm\theta_i$ otherwise. See Grenander and Rosenblatt (1957, pages 245-248) or Anderson (1971, pages 581-583). These points translate directly into the elements of the spectrum of $\mathbf{N}(\omega)$ in an obvious way.

Now suppose that the spectrum of $\mathbf{N}(\omega)$ has a single element consisting of the set $(-\pi, \pi]$ itself. The increments $\mathbf{N}(\omega_2) - \mathbf{N}(\omega_1)$, $\omega_1 < \omega_2$, are Hermitian positive definite, and via an appeal to Lebesgue's decomposition $\mathbf{N}(\omega) = \mathbf{N}^d(\omega) + \mathbf{N}^c(\omega)$ where $\mathbf{N}^d(\omega)$ is an increasing step function and $\mathbf{N}^c(\omega)$ is an increasing continuous function. Suppose that the saltuses in $\mathbf{N}^d(\omega)$ occur at the values ω_j, $j = 1, 2, \ldots$, and that the jump at ω_j is $\mathbf{N}_j > 0$. Thus,

$$\mathbf{N}^d(\omega) = \sum_{\omega_j < \omega} \mathbf{N}_j$$

where ω_j, $j = 1, 2, \ldots$, designates the discrete spectrum, and $\{(-\pi, \pi] \setminus \{\omega_1, \omega_2, \ldots\}\}$ is the continuous spectrum with (trivially)

$$\mathbf{N}^c(\omega) = \mathbf{N}(\omega) - \sum_{\omega_j < \omega} \mathbf{N}_j \,.$$

We now have $F(\omega) = F^d(\omega) + F^c(\omega) = \boldsymbol{\mu}^\top \mathbf{N}^d(\omega)\boldsymbol{\mu} + \boldsymbol{\mu}^\top \mathbf{N}^c(\omega)\boldsymbol{\mu}$ and

$$\boldsymbol{\Gamma}_m = \sum_{j=1}^{\infty} \mathbf{e}_m(\omega_j)\mathbf{e}_m(\omega_j)^* F_j + \int_{-\pi}^{\pi} \mathbf{e}_m(\omega)\mathbf{e}_m(\omega)^* dF^c(\omega)\,,$$

where, obviously, $F_j = \boldsymbol{\mu}^\top \mathbf{N}_j \boldsymbol{\mu}$. Either of $F^d(\omega)$ or $F^c(\omega)$ can be null.

When $F^c(\omega)$ is null and the spectrum of $\mathbf{N}(\omega)$ is purely discrete and consists of a finite set of q distinct points $\omega_1, \ldots, \omega_q$, $q \leq d$, we have already seen that as $N \to \infty$ a *R-SSA*(m, q) model will reproduce the spectral characteristics of the discrete component

with an ever decreasing error.[5] Let us now suppose that the spectrum is continuous, so that $F^d(\omega)$ is null and only $F^c(\omega)$ remains.

As preparation for the following result set

$$\begin{aligned} F_m^c(\omega) &= \int_{-\pi}^{\omega} \frac{1}{2\pi} \sum_{r=-(m-1)}^{m-1} \left(1-\frac{|r|}{m}\right) \varrho^c(r) \exp(-\imath\theta r) d\theta \\ &= \frac{(\pi+\omega)}{2\pi} + \sum_{\substack{r=-(m-1) \\ r\neq 0}}^{m-1} \left(1-\frac{|r|}{m}\right) \varrho^c(r) \left[\frac{\exp(-\imath\pi r)-\exp(-\imath\omega r)}{\imath 2\pi r}\right] \end{aligned}$$

where

$$\varrho^c(r) = \int_{-\pi}^{\pi} \exp(\imath\omega r) dF^c(\omega). \tag{9}$$

By construction $F_m^c(-\pi) = 0$ and $F_m^c(\pi) = 1$, and because the integrand

$$\begin{aligned} f_m^c(\omega) &= \frac{1}{2\pi} \sum_{r=-(m-1)}^{m-1} \left(1-\frac{|r|}{m}\right) \varrho^c(r) \exp(-\imath\omega r) \\ &= \frac{1}{2\pi m} \sum_{s=1}^{m} \sum_{t=1}^{m} \varrho^c(t-s) \exp(-\imath\omega(t-s)) \\ &= \frac{1}{2\pi m} \int_{-\pi}^{\pi} |\sum_{r=1}^{m} \exp(-\imath(\omega-\theta)r)|^2 dF^c(\theta) \\ &> 0 \end{aligned}$$

is positive for all ω, since the increments in $F^c(\omega) = \boldsymbol{\mu}^\top \mathbf{N}^c(\omega)\boldsymbol{\mu}$ are positive definite, the distribution function $F_m^c(\omega)$ is increasing for each value of m. Furthermore, it is readily verified that

$$\int_{-\pi}^{\pi} \exp(\imath\omega r) dF_m^c(\omega) = \begin{cases} \left(1-\frac{|r|}{m}\right)\varrho^c(r), & |r| \leq m-1, \\ 0, & |r| \geq m, \end{cases} \tag{10}$$

where $\varrho^c(r)$ is defined in (9). Treating the Fourier transforms in (9) and (10) as characteristic functions of distribution functions supported on the interval $[-\pi, \pi]$, it

[5] This feature provides an SSA counterpart to the stochastic process concept of a deterministic series, a time series that is ultimately perfectly predictable. See Anderson (1971, Chapter 7.6) for example.

follows from Helly's theorem and the uniqueness and continuity properties of characteristic functions that $F_m^c(\cdot)$ converges to $F^c(\cdot)$ at each point of continuity.

Theorem 4 *Let*

$$\mathbf{\Gamma}_m^c = \int_{-\pi}^{\pi} \mathbf{e}_m(\omega)\mathbf{e}_m(\omega)^* dF^c(\omega)$$

and assume that the Fourier coefficients in (9) decay hyperbolically: $|\varrho^c(r)| \leq C|r|^{2d-1}$ *as* $|r| \to \infty$ *for some parameter* d, $|d| < 1/2$, *and constant* $C < \infty$. *Then the terms in the spectral decomposition* $\mathbf{\Gamma}_m^c = \sum_{i=1}^m \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^\top$ *satisfy*

$$m^{-1}|\lambda_i - 2\pi f_m^c(2\pi j_i/m)| = h(m)$$

and

$$m^{-1}\|\varsigma_j \boldsymbol{v}_i - \mathbf{e}_m(2\pi j_i/m)\| = h(m)$$

for all m *sufficiently large, where for* $i = 1, \ldots, m$; $2\pi j_i/m$, $j_i \in \{0, 1, \ldots, m-1\}$ *denote points of the spectrum such that* $f_m^c(2\pi j_1/m) \geq f_m^c(2\pi j_2/m) \geq \cdots \geq f_m^c(2\pi j_m/m)$, $\varsigma_i = sign(\boldsymbol{v}_i' \mathbf{e}_m(2\pi j_i/m))$, *and*

$$h(m) \sim \begin{cases} C\left(\frac{2(7+12d)}{(4d+1)(4d+2)}\right)^{\frac{1}{2}} m^{2d-1}, & 0 < d < 1/2\,; \\ C(2\log(m))^{\frac{1}{2}} m^{-1}, & d = 0\,; \\ C(2\zeta(1-4d))^{\frac{1}{2}} m^{-1}, & -1/2 < d < 0\,. \end{cases},$$

where $\zeta(\cdot)$ *denotes Riemann's zeta function.*

An immediate corollary of Theorem 4 is that

$$\lim_{m\to\infty} |\frac{1}{m}\sum_{i=1}^m \lambda_i^p - \frac{1}{2\pi}\int_{-\pi}^{\pi} \{2\pi f_m^c(\omega)\}^p d\omega| = 0\,, \quad p = 0, 1, 2, \ldots\,.$$

This follows as a consequence of the inequality

$$\left|\frac{1}{m}\sum_{i=1}^{m}\lambda_i^p - \frac{1}{2\pi}\int_{-\pi}^{\pi}\{2\pi f_m^c(\omega)\}^p d\omega\right| \leq \frac{1}{m}\sum_{i=1}^{m}|\lambda_i^p - \{2\pi f_m^c(2\pi j_i/m)\}^p| + \left|\frac{1}{m}\sum_{j=0}^{m-1}\{2\pi f_m^c(2\pi j/m)\}^p - \frac{1}{2\pi}\int_{-\pi}^{\pi}\{2\pi f_m^c(\omega)\}^p d\omega\right|,$$

where the existence of the limit of $m^{-1}\sum_{j=0}^{m-1}\{2\pi f_m^c(2\pi j/m)\}^p$ as a Riemann integral is guaranteed by the continuity of $f_m^c(\cdot)$.

A great deal is known about the properties of Toeplitz matrices, of course, a classic reference being Grenander and Szego (1958), and arguably the most well known result describing the behavior of the eigenvalues of sequences of $m \times m$ Toeplitz matrices as $m \to \infty$ is Szegö's theorem. Theorem 4 and its corollary provide an adaption of Szegö's theorem that allows for the possibility that the Fourier coefficients are not absolutely summable. Prototypical examples are fractional Gaussian noise, obtained as the increments of self-similar processes, and fractional autoregressive moving average processes (Beran, 1994).

When applied in conjunction with Theorem 2, Theorem 4 indicates that if the series has a purely continuous spectrum, then for any $R\text{-}SSA(m,k)$ model with $k < m$ the signal-noise decomposition constructed from $\mathbf{D}^{\frac{1}{2}}(\mathbf{Y}_{\mathscr{S}} + \mathbf{Y}_{\mathscr{N}})$ with $\mathbf{Y}_{\mathscr{S}} = \sum_{i=1}^{k}\sqrt{\ell_i}\mathbf{u}_i\mathbf{v}_i^\top$ and $\mathbf{Y}_{\mathscr{N}} = \sum_{i=k+1}^{m}\sqrt{\ell_i}\mathbf{u}_i\mathbf{v}_i^\top$ will converge to a population ensemble model that allocates those points of the spectrum that have the greatest power to the signal and relegates the points remaining to the noise.

As an example of a process with a continuous spectrum, consider the series $x(t) = z_1(t) = \varepsilon(t) + \theta\varepsilon(t-1)$ where $\varepsilon(t)$ is a zero mean white noise with variance σ^2. Then it is straightforward to show that in this case $\mathbf{\Gamma}_m$ equals the tri-diagonal Toeplitz matrix

$$\begin{pmatrix} 1 & \varrho & & & & \\ \varrho & 1 & \varrho & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & \varrho & 1 & \varrho \\ & & & & \varrho & 1 \end{pmatrix} \tag{11}$$

where $\varrho = \theta/(1+\theta^2)$. The normalized spectral density of this process is $(2\pi)^{-1}(1 + 2\varrho\cos(\omega))$ and the eigenvalues of the matrix in (11) are $\lambda_i = (1+2\varrho\cos(\omega_{j_i}))$, where $\omega_{j_i} = 2\pi j_i/(m+1)$, $j_i \in \{1,\ldots,m\}$, $i = 1,\ldots,m$. The corresponding eigenvectors are $\boldsymbol{v}_i = \sqrt{\frac{2}{(m+1)}}(\sin(\omega_{j_i}),\ldots,\sin(m\omega_{j_i}))^\top$, $i = 1,\ldots,m$. From this it is clear that if an *R-SSA*(m,k) model with $k < m$ is employed to represent the series the loss incurred will depend critically on the power forgone by assigning the $m-k$ eigenvalue-eigenvector pairs that do not correspond to dominant parts of the power-spectrum to the noise component.

Now assume that the spectrum is mixed, that is, both discrete and continuous components are present, and suppose that in the decomposition $x(t) = \alpha_1 z_1(t) + \cdots + \alpha_d z_d(t)$ the variables $z_i(t)$, $i = 1,\ldots,q$ have a discrete spectrum and $z_i(t)$, $i = q+1,\ldots,d$, have a continuous spectrum – polynomial trends or trigonometric series and regular stationary processes, respectively, for example. Assume also that $z_i(t)$, $i = 1,\ldots,q$ predominate in the sense that there exists a function $a(n)$ that is regularly varying at infinity with a positive index such that; for $i = 1,\ldots,q$, $a_{ii}^n(0)/a(n)$ is regularly varying with a non-negative index, whereas for $i = q+1,\ldots,d$, $a_{ii}^n(0)/a(n)$ is regularly varying with a negative index, implying that for $i = q+1,\ldots,d$, $\lim_{n\to\infty} a_{ii}^n(0)/a(n) = 0$. Then $\boldsymbol{\mu} = \mathbf{R}(0)^{\frac{1}{2}}\boldsymbol{\beta}^d/\sqrt{\boldsymbol{\beta}^{d\top}\mathbf{R}(0)\boldsymbol{\beta}^d}$ where $\boldsymbol{\beta}^d = \boldsymbol{\beta}/\sqrt{a(n)}$, and as n increases $\boldsymbol{\beta}^d$ will collapse on to $(\alpha_1\sqrt{a_{11}^n(0)/a(n)},\ldots,\alpha_q\sqrt{a_{qq}^n(0)/a(n)},0,\ldots,0)^\top$. This indicates that in this case the trajectories of $\mathbf{X}$ will ultimately be driven by $z_i(t)$, $i = 1,\ldots,q$, and as $n \to \infty$ the SVD of $\mathbf{Y} = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}$ will contain q non-null singular values bounded away from zero and $m-q$ arbitrarily small singular values, reflecting the dominance of the discrete spectrum. The corresponding signal-noise decomposition will give

$\mathbf{Y} = \mathbf{Y}_{\mathscr{S}} + \mathbf{Y}_{\mathscr{N}}$ where $\mathbf{Y}_{\mathscr{S}}$ and $\mathbf{Y}_{\mathscr{N}}$ are as in (7) and (8), and the population ensemble *R-SSA*(m,q) model will reproduce the dominant discrete component with an ever decreasing error as $n \to \infty$.

This suggests applying R-SSA sequentially;

- First step, remove the component generated by the q dominating discrete components using a *R-SSA*(m,q) model to construct an approximation $s^d(t)$, to use an obvious notation, to $\alpha_1 z_1(t) + \cdots + \alpha_q z_q(t)$. Filter out the discrete component to give the residuals $x(t) - s^d(t)$, $t = 1, \ldots, N$.

- Second step, analyze the resulting residuals and construct an appropriate *R-SSA*(m,k) model for $x(t) - s^d(t)$ so as to yield an approximation to the continuous component $\alpha_{q+1} z_{q+1}(t) + \cdots + \alpha_d z_d(t)$. Denote the approximation by $s^c(t)$, $t = 1, \ldots, N$.

At the second step the residuals should be analyzed by examining the SVD of the rescaled trajectory $\mathbf{D}^{-\frac{1}{2}}(\mathbf{X} - \mathbf{S}^d)$ where now $\mathbf{D} = \text{diag}\{(\mathbf{X} - \mathbf{S}^d)(\mathbf{X} - \mathbf{S}^d)^\top\}$, and in both the first step and the second step the model can be chosen so as to ensure that the variance ratio $\sum_{i=1}^{k} \ell_i / \sum_{i=k+1}^{m} \ell_i$ is sufficiently large, or by using the description length principle outlined in Khan and Poskitt (2015).

- Third step, amalgamate the two components to give a final R-SSA signal–noise reconstruction $x(t) = s^d(t) + s^c(t) + \varepsilon(t)$, $t = 1, \ldots, N$, for the observed time series.

5 Illustrations

Our purpose in this section is to demonstrate the practical impact of the results presented above. This we will do by examining the details of three hypothetical series that satisfy the regularity conditions of Section 3.

5.1 Polynomial Series

If $z_i(t) = t^{i-1}$, $i = 1, \ldots, d$, then

$$\rho_{rc}(h) = \frac{\sqrt{(2r-1)(2c-1)}}{r+c-1}, \quad r, c = 1, \ldots, d,$$

for all $h = 0, \pm 1, \pm 2, \ldots$, with corresponding spectral distribution function

$$\mathbf{M}(\omega) = \begin{cases} 0, & \omega < 0; \\ \mathbf{R}(0), & \omega \geq 0. \end{cases}$$

The spectrum of $\mathbf{N}(\omega)$ consists of a single element at $\omega = 0$ with a jump of $\mathbf{I}_d$, and

$$F(\omega) = \boldsymbol{\mu}^\top \mathbf{N}(\omega)\boldsymbol{\mu} = \begin{cases} 0, & \omega < 0\,; \\ 1, & \omega \geq 0\,, \end{cases}$$

the Heaviside (unit) step function. This yields a limiting value of $\boldsymbol{\Gamma}_m = \mathbf{1}_m \mathbf{1}_m^\top$ for the Gramian $\mathbf{Y}\mathbf{Y}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^\top\mathbf{D}^{-\frac{1}{2}}$ where $\mathbf{1}_m = (1, \ldots, 1)^\top$. The eigenvalues of $\mathbf{1}_m\mathbf{1}_m^\top$ are $\lambda_1 = m$ with eigenvector $\boldsymbol{v}_1 = \mathbf{1}_m/\sqrt{m}$, and $\lambda_m = 0$ with multiplicity $m-1$ and eigenvectors

$$\begin{aligned} \boldsymbol{v}_2^\top &= (-1, 1, \mathbf{0}_{m-2}^\top)/\sqrt{2} \\ \boldsymbol{v}_3^\top &= (1, 1, -2, \mathbf{0}_{m-3}^\top)/\sqrt{6} \\ \boldsymbol{v}_4^\top &= (-1, -1, -1, 3, \mathbf{0}_{m-4}^\top)/\sqrt{12} \\ &\vdots \\ \boldsymbol{v}_m^\top &= (-1)^{m-1}(1, 1, \ldots, 1, -(m-1))/\sqrt{m(m-1)}. \end{aligned}$$

The upshot of this is that any *R-SSA*(m, k) model applied to the polynomial series $x(t) = \alpha_1 z_1(t) + \cdots + \alpha_d z_d(t)$, with $2 \leq m \leq M$ where $\lim_{N\to\infty} M/N = 0$ as $N \to \infty$ and $1 \leq k < m$, will ultimately lead to a *R-SSA*$(m, 1)$ (asymptotic) population ensemble model in which $\mathbf{Y}_{\mathscr{S}} = \mathbf{1}_m\mathbf{1}_m^\top\mathbf{Y} + o(1)$ and $\mathbf{Y}_{\mathscr{N}} = o(1)$, irrespective of the

value of d.

Tables 1 and 2 present entries in the Gramian matrix $\mathbf{YY}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{XX}^\top\mathbf{D}^{-\frac{1}{2}}$ and values of the eigenvalues ℓ_j, $j = 1,\ldots,m$, when calculated from realizations of the process $x(t) = \alpha_1 z_1(t) + \cdots + \alpha_4 z_4(t)$ where $z_j(t) = t^{j-1}$, $j = 1,3$, and $z_4(t)$ is a generalized autoregressive conditionally heteroscedastic $GARCH(2,1)$ process with parameters $(0.2, 0.1)$ and 0.4, and an unconditional variance of unity. For the linear trend series $(\alpha_1,\ldots,\alpha_4)^\top = (1, 0.1, 0, 2.51)$ and for the quadratic trend series $(\alpha_1,\ldots,\alpha_4)^\top = (1, 2, -0.01, 14.91)$, where the coefficient values are chosen so that the signal-noise ratio of both series is 15.0 dB.[6] The calculations are based on $N = 200$ observations with a window length of $m = \lfloor\sqrt{N}\rfloor + 1 = 15$. Figure 1 graphs the individual realizations that gave rise to the values presented in Tables 1 and 2.

Table 1: *Entries in upper triangle of* $\mathbf{YY}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{XX}^\top\mathbf{D}^{-\frac{1}{2}}$ *and* ℓ_j, $j = 1,\ldots,m$, *calculated from* $N = 200$ *observations on linear trend series with* $m = 15$.

$\mathbf{YY}^\top$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	0.97	0.96	0.96	0.97	0.97	0.97
2	·	1	0.97	0.96	0.97	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
3	·	·	1	0.97	0.96	0.97	0.96	0.97	0.97	0.96	0.97	0.96	0.97	0.97	0.97
4	·	·	·	1	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
5	·	·	·	·	1	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
6	·	·	·	·	·	1	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
7	·	·	·	·	·	·	1	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
8	·	·	·	·	·	·	·	1	0.97	0.97	0.97	0.97	0.97	0.97	0.97
9	·	·	·	·	·	·	·	·	1	0.97	0.97	0.97	0.97	0.97	0.97
10	·	·	·	·	·	·	·	·	·	1	0.97	0.97	0.97	0.97	0.97
11	·	·	·	·	·	·	·	·	·	·	1	0.98	0.97	0.97	0.97
12	·	·	·	·	·	·	·	·	·	·	·	1	0.98	0.97	0.97
13	·	·	·	·	·	·	·	·	·	·	·	·	1	0.98	0.97
14	·	·	·	·	·	·	·	·	·	·	·	·	·	1	0.98
15	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1
ℓ_j	14.57	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.01

Visual inspection of Figure 1 clearly indicates that over the time frame considered the two series generate a range of values that differ by an order of magnitude, and that they exhibit very different trend behaviour. Nevertheless, from Tables 1 and 2 we find that both series yield values of $\mathbf{YY}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{XX}^\top\mathbf{D}^{-\frac{1}{2}}$ that closely approximate an $m \times m$

[6] Designating the deterministic component as the signal, $\mathfrak{s}(t) = \alpha_1 z_1(t) + \alpha_2 z_2(t) + \alpha_3 z_3(t)$, and the random component as noise, $\mathfrak{n}(t) = \alpha_4 z_4(t)$, the signal-noise ratio $\sum_{t=1}^{T}\mathfrak{s}(t)^2 / \sum_{t=1}^{T}\mathfrak{n}(t)^2 \sim (\alpha_{p+1}/\alpha_4)^2 N^{2p}/(p+1)$, with $p = 1$ for the linear trend series and $p = 2$ for quadratic trend series.

Table 2: *Entries in upper triangle of* $\mathbf{YY}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{XX}^\top\mathbf{D}^{-\frac{1}{2}}$ *and* ℓ_j, $j = 1, \ldots, m$, *calculated from* $N = 200$ *observations on quadratic trend series with* $m = 15$.

$\mathbf{YY}^\top$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	0.97	0.96	0.96	0.96	0.96	0.95
2	·	1	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.97	0.96	0.96	0.96	0.96
3	·	·	1	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	0.96	0.96	0.96
4	·	·	·	1	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	0.96	0.96
5	·	·	·	·	1	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	0.96
6	·	·	·	·	·	1	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
7	·	·	·	·	·	·	1	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
8	·	·	·	·	·	·	·	1	0.97	0.97	0.97	0.97	0.97	0.97	0.97
9	·	·	·	·	·	·	·	·	1	0.97	0.97	0.97	0.97	0.97	0.97
10	·	·	·	·	·	·	·	·	·	1	0.97	0.97	0.97	0.97	0.97
11	·	·	·	·	·	·	·	·	·	·	1	0.98	0.97	0.97	0.97
12	·	·	·	·	·	·	·	·	·	·	·	1	0.98	0.97	0.97
13	·	·	·	·	·	·	·	·	·	·	·	·	1	0.98	0.97
14	·	·	·	·	·	·	·	·	·	·	·	·	·	1	0.98
15	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1
ℓ_j	14.56	0.08	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.01

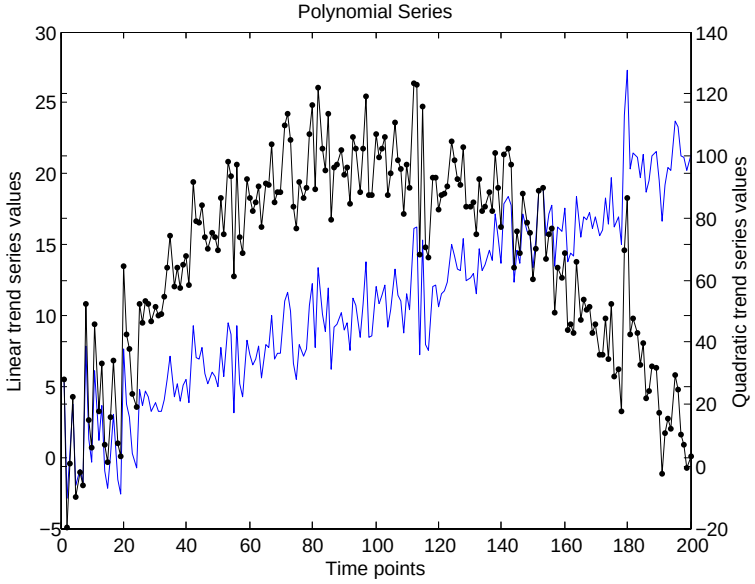

Figure 1: *Realizations of linear and quadratic trend series with* $N = 200$.

equi-correlation matrix of the form

$$\begin{pmatrix} 1 & \varrho & \varrho & \cdots & \cdots & \varrho \\ \varrho & 1 & \varrho & \cdots & \cdots & \varrho \\ \cdot & \cdots & \ddots & \cdots & \cdots & \cdot \\ \cdot & \cdots & \cdots & \ddots & \cdots & \cdot \\ \varrho & \cdots & \cdots & \varrho & 1 & \varrho \\ \varrho & \cdots & \cdots & \varrho & \varrho & 1 \end{pmatrix}, \tag{12}$$

with eigenvalues $1+(m-1)\varrho$ and $1-\varrho$ with multiplicity $m-1$, where $\varrho \approx 0.97$. Thus we find that at this signal-noise ratio a sample of size $N = 200$ is sufficient for the finite sample value of $\mathbf{Y}\mathbf{Y}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^\top\mathbf{D}^{-\frac{1}{2}}$ to be in close accord with the limiting value of $\mathbf{\Gamma}_m = \mathbf{1}_m\mathbf{1}_m^\top$, and the $R\text{-}SSA(15,1)$ signal-noise reconstructions for both series will therefore be close to those based upon the corresponding population ensemble model.

In Figure 2 we graph for each series the true signal $\mathfrak{s}(t) = \alpha_1 z_1(t) + \alpha_2 z_2(t) + \alpha_3 z_3(t)$ and noise process $\mathfrak{n}(t) = \alpha_4 z_4(t)$ together with the reconstructions $s^d(t)$ and $s^c(t)$ derived from $R\text{-}SSA(15,1)$ and $R\text{-}SSA(15,13)$ models respectively. Both models were chosen by assigning the signal dimension to the smallest value of k such that $100(\sum_{i=1}^{k}\ell_i/m) \geq 95\%$. To ascertain the precision of the reconstructions we have calculated the correlation coefficients between the true signal and its reconstruction, $r_{\mathfrak{s},s^d}$, the true noise and its reconstruction, $r_{\mathfrak{n},s^c}$, and the original series and its reconstruction, r_{x,s^d+s^c}. These provide simple scale invariant measures that can be used to directly compare the performance of different SSA models when applied to different series.

For the linear trend series we have $r_{\mathfrak{s},s^d} = 0.9957$ and for the quadratic trend series $r_{\mathfrak{s},s^d} = 0.9906$, indicating that the level of accuracy achieved by the $R\text{-}SSA(15,1)$ model for each of the two series is on a par. Hence we find that, despite the values of $\mathbf{Y}\mathbf{Y}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^\top\mathbf{D}^{-\frac{1}{2}}$ for each series being very close to each other and close to their common theoretical limiting value, the discrete reconstructions accurately reproduce the different trend components of each series. Consequently, apart from the change in scale necessary to achieve the common signal-noise ratio, the noise reconstructions are almost identical and precisely capture the evolution of the noise component. For the linear trend series $r_{\mathfrak{n},s^c} = 0.9526$, and for the quadratic trend series $r_{\mathfrak{n},s^c} = 0.9417$. The overall effect is to produce R-SSA series reconstructions with correlations of $r_{x,s^d+s^c} = 0.9990$ and $r_{x,s^d+s^c} = 0.9986$ for the linear and quadratic trend series respectively.

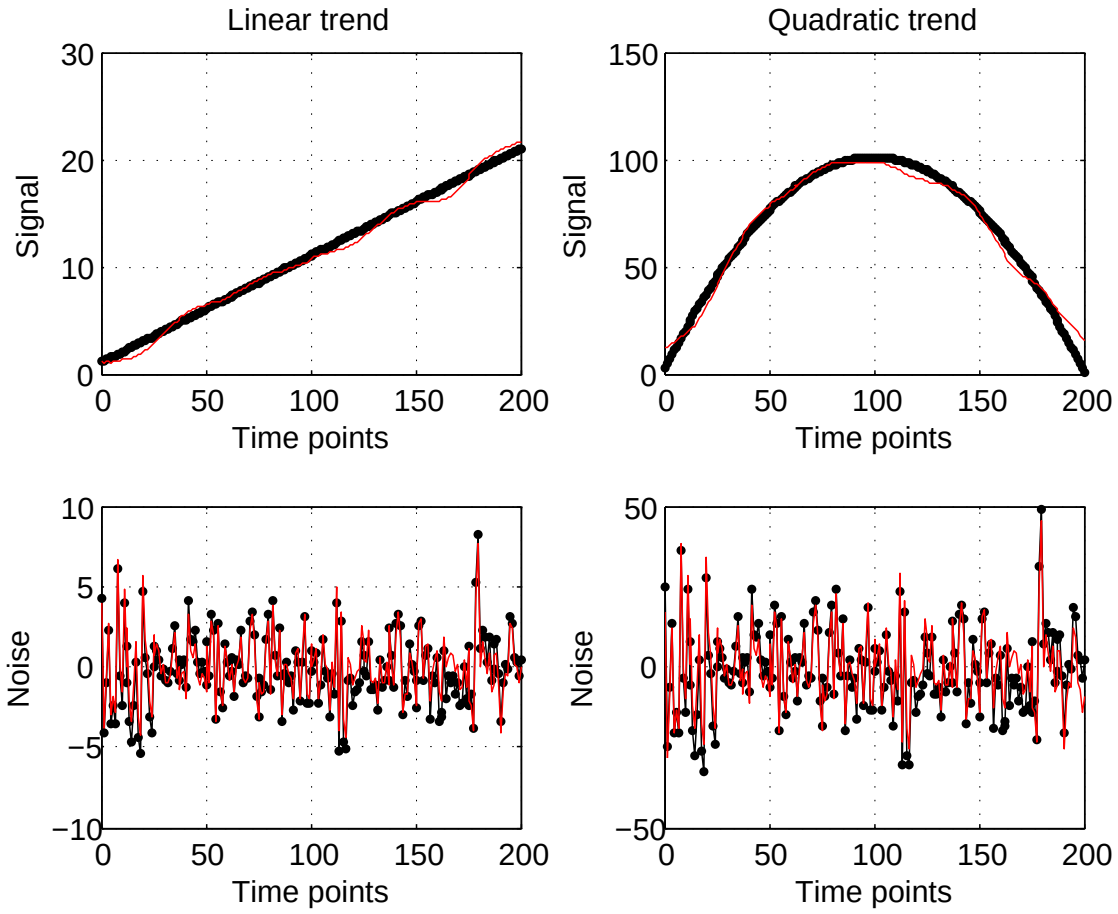

Figure 2: *Realized signal and noise components of linear and quadratic trend series overlaid with R-SSA first stage signal reconstruction $s^d(t)$ and second stage noise reconstruction $s^c(t)$.*

In summary, the comparison of the values given by the linear and quadratic trend series in Tables 1 and 2, when combined with an examination of Figures 1 and 2 and the associated reconstruction correlations, provides clear evidence of how detailed structure embedded within a series that is lost in the characterization of $\mathbf{YY}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{XX}^\top\mathbf{D}^{-\frac{1}{2}}$ via $\mathbf{N}(\omega)$ is recovered in SSA by projecting back into the time domain through the empirical eigenfunctions.

5.2 Random Walk Process

Now consider an observed process $x(t)$ such that

$$x(t) = \alpha_1 z_1(t) + \alpha_2 z_2(t) = \sum_{\tau=0}^{t-1} \eta(t-\tau) + \nu(t) \tag{13}$$

where $\eta(t)$ is a zero mean i.i.d. Gaussian white noise processes with a unit variance, and $\nu(t)$ is an independent zero mean stationary and ergodic process.[7] For reasons that will become apparent below, we will label the first component as the signal, $\mathfrak{s}(t) = \alpha_1 z_1(t) = \sum_{\tau=0}^{t-1} \eta(t-\tau)$, and the second component as noise, $\mathfrak{n}(t) = \alpha_2 z_2(t) = \nu(t)$. See Thomakos (2008*a*,*b*) for a number of results concerning the application of SSA to random walk processes.

To derive the re-scaled SSA population ensemble model, let us express $\mathbf{X}\mathbf{X}^\top$ as $\sum_{t=1}^{n} \mathbf{x}(t)\mathbf{x}(t)^\top$ where

$$\mathbf{x}(t) = z_1(t-1)\mathbf{1}_m + \begin{bmatrix} \eta(t) \\ \eta(t)+\eta(t+1) \\ \eta(t)+\eta(t+1)+\eta(t+2) \\ \vdots \\ \eta(t)+\cdots\cdots+\eta(t+m-1) \end{bmatrix} + \begin{bmatrix} \nu(t) \\ \nu(t+1) \\ \nu(t+2) \\ \vdots \\ \nu(t+m-1) \end{bmatrix}. \tag{14}$$

From the strong Markov property of the random walk process it follows that the three components on the right hand side of (14) are mutually orthogonal. Since $\mathbb{E}[z_1(t-1)^2] = t-1$ and $\mathbb{E}[\sum_{\tau=0}^{r} \eta(t+\tau) \sum_{\tau=0}^{s} \eta(t+\tau)] = \min(r+1, s+1)$, we obtain $\mathbb{E}(\mathbf{x}_t \mathbf{x}_t^\top) = (t-1)\mathbf{1}_m \mathbf{1}_m^\top + \mathbf{\Psi} + \mathbf{\Upsilon}$ where $\mathbf{\Upsilon} = [\mathbb{E}(\nu(t+r-c)\nu(t))]_{r,c=1,\ldots,m}$ and $\mathbf{\Psi} = [\min(r,c)]_{r,c=1,\ldots,m}$. For the raw trajectory matrix $\mathbf{X}$ we therefore have

$$\frac{1}{n}\mathbb{E}[\mathbf{X}\mathbf{X}^\top] = \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}[\mathbf{x}_t \mathbf{x}_t^\top] = \frac{n-1}{2}\mathbf{1}_m \mathbf{1}_m^\top + \mathbf{\Psi} + \mathbf{\Upsilon}.$$

[7] The first component is a random walk, of course, a non–stationary process of orthogonal increments. In the structural times series literature the process in (13) is referred to as the local level model when $\nu(t)$ is white noise. The assumption of Gaussianity is adopted for convenience, it can be replaced by appropriate martingale or mixing conditions but such a level of generality is not required here.

Moreover, applying Donsker's theorem and the fact that $n^{-3/2}\sum_{t=1}^{n} z_1(t-1)\eta(t+s) = O(\sqrt{\log\log n})$, $s = 0,\ldots,m-1$, (Poskitt, 2000, Lemma A.1.(ii)) we can also deduce that $\|n^{-1}\mathbf{X}\mathbf{X}' - \mathbf{\Sigma}_m\| = O(\sqrt{\log\log n/n})$ where $\mathbf{\Sigma}_m = n\beta_n^2\mathbf{1}_m\mathbf{1}_m^\top + \mathbf{\Psi} + \mathbf{\Upsilon}$ and

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

where $\mathbb{B}(\omega)$ denotes standard Brownian motion. We are therefore lead to the conclusion that $\mathbf{Y}\mathbf{Y}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^\top\mathbf{D}^{-\frac{1}{2}}$ converges to $\mathbf{\Delta}_m^{-\frac{1}{2}}\mathbf{\Sigma}_m\mathbf{\Delta}_m^{-\frac{1}{2}}$ where $\mathbf{\Delta}_m = \mathrm{diag}\{n\beta_n^2 + 1 + \sigma^2, n\beta_n^2 + 2 + \sigma^2, \ldots, n\beta_n^2 + m + \sigma^2\}$, $\sigma^2 = \mathbb{E}(v(t)^2)$, and hence that $\mathbf{\Gamma}_m = \mathbf{1}_m\mathbf{1}_m^\top + o(1)$. This reflects that the random walk component eventually dominates the behaviour of the entries in $\mathbf{Y}\mathbf{Y}^\top$ and asymptotically the contribution of the stationary component is smothered.[8]

Figure 3 graphs a realization of a random walk process as specified in (13). Here

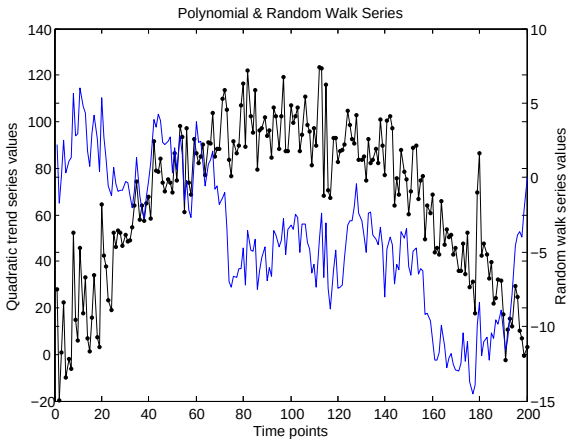

Figure 3: *Realizations of random walk process and quadratic trend series with* $N = 200$.

the noise component is a generalized autoregressive conditionally heteroscedastic $GARCH(2,1)$ process with parameters $(0.2, 0.1)$ and 0.4, and an unconditional variance set so that the signal-noise ratio $\sum_{t=1}^{N}\mathfrak{s}(t)^2/\sum_{t=1}^{N}\mathfrak{n}(t)^2$ is 15.0 dB, the same as

[8] The limiting value of $\mathbf{\Gamma}_m$ obtained for this process is the same as that produced in the polynomial series case, and likewise, this will ultimately be manifest in a R-SSA model in which the signal component $\mathbf{Y}_{\mathscr{S}} = \mathbf{1}_m\mathbf{1}_m^\top\mathbf{Y} + o(1)$. An explicit algebraic representation of the series given by $\mathscr{H}(\mathbf{1}_m\mathbf{1}_m^\top\mathbf{Y})$ is presented in Thomakos (2008*a*, Section 3.1.)

that for the polynomial trend series examined previously. To provide a basis for comparison, the plot of the random walk process is superimposed upon the quadratic trend series.

Casual perusal of Figure 3 might lead an unwary practitioner who has no prior knowledge of the true DGPs to suggest that Figure 3 provides a counterpart to Figure 1, save that the quadratic trend series is matched with a series containing a downward sloping rather than an upward sloping linear trend: an erroneous conclusion that could be reinforced by noting that the Gramian $\mathbf{YY}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{XX}^\top\mathbf{D}^{-\frac{1}{2}}$ (given in Table 3) loosely

Table 3: *Entries in upper triangle of* $\mathbf{YY}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{XX}^\top\mathbf{D}^{-\frac{1}{2}}$ *and* ℓ_j, $j = 1, \ldots, m$, *calculated from* $N = 200$ *observations on random walk process with* $m = 15$.

$\mathbf{YY}^\top$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0.95	0.94	0.93	0.92	0.91	0.91	0.90	0.88	0.87	0.85	0.84	0.83	0.81	0.80
2	·	1	0.96	0.94	0.92	0.92	0.91	0.90	0.90	0.87	0.87	0.85	0.84	0.83	0.80
3	·	·	1	0.96	0.94	0.93	0.92	0.91	0.91	0.89	0.87	0.87	0.86	0.84	0.82
4	·	·	·	1	0.96	0.94	0.93	0.92	0.92	0.90	0.89	0.88	0.87	0.85	0.83
5	·	·	·	·	1	0.96	0.94	0.93	0.92	0.92	0.90	0.89	0.88	0.87	0.85
6	·	·	·	·	·	1	0.96	0.94	0.93	0.92	0.91	0.90	0.89	0.87	0.86
7	·	·	·	·	·	·	1	0.96	0.94	0.93	0.91	0.91	0.90	0.88	0.86
8	·	·	·	·	·	·	·	1	0.96	0.94	0.93	0.92	0.91	0.90	0.88
9	·	·	·	·	·	·	·	·	1	0.96	0.94	0.93	0.92	0.91	0.89
10	·	·	·	·	·	·	·	·	·	1	0.96	0.94	0.93	0.92	0.91
11	·	·	·	·	·	·	·	·	·	·	1	0.96	0.94	0.93	0.92
12	·	·	·	·	·	·	·	·	·	·	·	1	0.96	0.94	0.93
13	·	·	·	·	·	·	·	·	·	·	·	·	1	0.96	0.94
14	·	·	·	·	·	·	·	·	·	·	·	·	·	1	0.96
15	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1
ℓ_j	13.68	0.56	0.16	0.10	0.08	0.07	0.07	0.05	0.05	0.04	0.03	0.03	0.03	0.02	0.01

approximates to an $m \times m$ equi-correlation matrix as in (12) with $\varrho \approx 0.91$, as might be observed with a polynomial series.

In Figure 6 we graph for each series the true signal and noise components together with the first step $R\text{-}SSA(15, 1)$ signal reconstruction and second step $R\text{-}SSA(15, 13)$ noise reconstruction. Both models were chosen by assigning the signal dimension to the smallest value of k such that $100(\sum_{i=1}^{k} \ell_i/m) \geq 95\%$. The ability of the re-scaled SSA signal reconstruction to track the underlying true signal component of each series is apparent, the fact that one signal is continuously differentiable whilst the other approximates a process that is nowhere differentiable, except possibly on a set of probability measure zero, notwithstanding. The $R\text{-}SSA(15, 1)$ signal reconstructions clearly show that neither set of observations are derived from a linear trend series. And

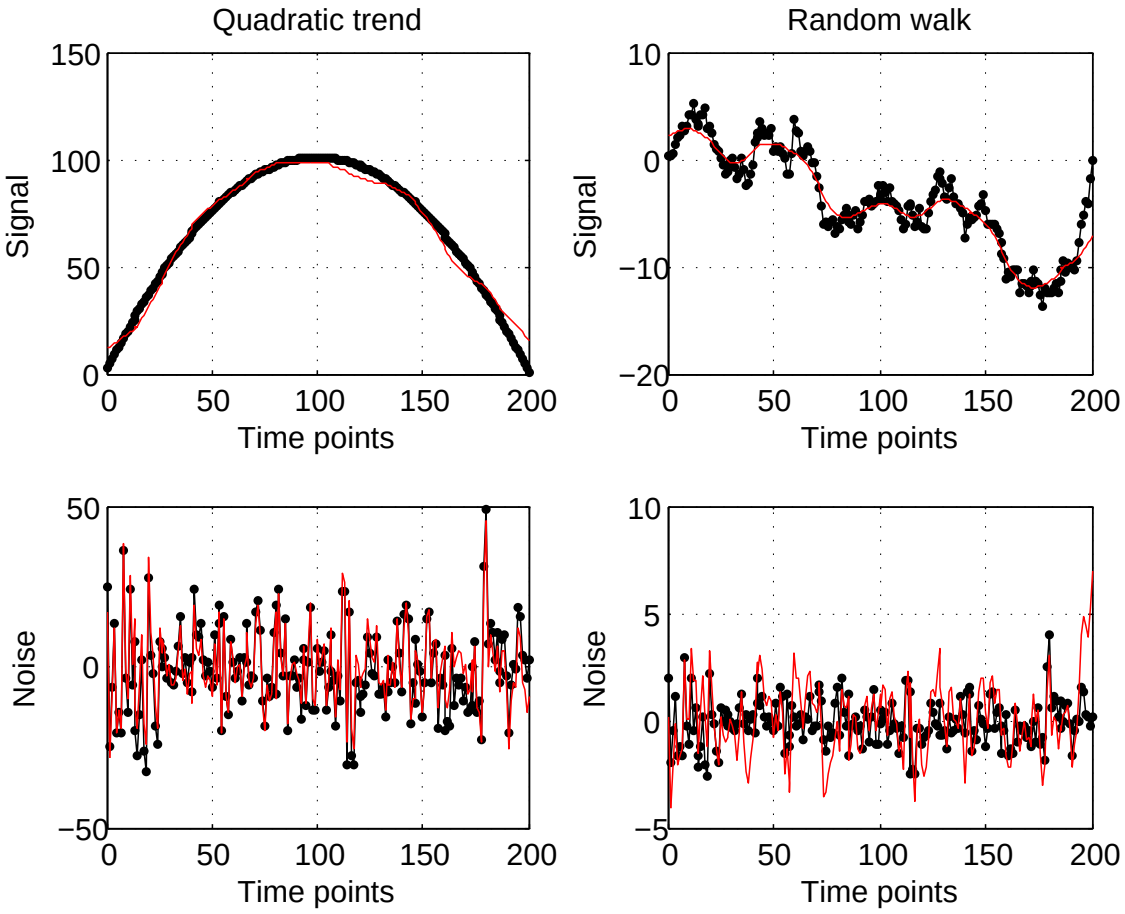

Figure 4: *Realized signal and noise components of quadratic trend series and random walk process overlaid with R-SSA first stage signal reconstruction $s^d(t)$ and second stage noise reconstruction $s^c(t)$.*

once again, apart from the change in scale necessary to achieve the common signal-noise ratio, the noise reconstructions behave similarly and capture the evolution of the true noise component reasonably accurately. Given the relatively erratic behaviour of random walks, however, it is perhaps not too surprising to find that for the random walk process we have $r_{\mathfrak{s},s^d} = 0.9524$ and $r_{\mathfrak{n},s^c} = 0.5952$, compared to $r_{\mathfrak{s},s^d} = 0.9906$ and $r_{\mathfrak{n},s^c} = 0.9417$ for the quadratic trend series. Interestingly enough, the lack of precision in the second stage noise reconstruction implicit in the decline in $r_{\mathfrak{n},s^c}$ for the random walk process is recovered in the overall series reconstruction, and both perform similarity with correlations of $r_{x,s^d+s^c} = 0.9986$ and $r_{x,s^d+s^c} = 0.9987$ for the quadratic trend series and random walk process respectively.

The outcomes reported above clearly illustrate that different DGPs that are not equivalent can give rise to the same asymptotic structure for $\mathbf{Y}\mathbf{Y}^{\top} = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^{\top}\mathbf{D}^{-\frac{1}{2}}$, and hence the same spectral distribution function and spectrum. Nevertheless, although the description of the DGP through $\mathbf{N}(\omega)$ is poor, as all fine detail is lost, and SSA will ultimately handle such series identically due to their common "frequency domain" features as a consequence, by projecting back into the time domain through the empirical eigenfunctions SSA is able to retrieve the individual structure embedded within each series.

5.3 A Random Walk with Drift and Autocorrelated Increments

To illustrate a situation involving a more complex mixed spectral structure, suppose that $x(t) = \boldsymbol{\alpha}^{\top}\mathbf{z}(t)$ where $\mathbf{z}(t) - \mathbf{z}(t-1) = \boldsymbol{\delta} + \mathbf{u}(t)$ and $\mathbf{u}(t)$ is a stationary process with Wold representation $\sum_{s=0}^{\infty} \boldsymbol{\Phi}_s \boldsymbol{\eta}(t-s)$; where the process $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{dt})'$ constitutes an i.i.d. sequence of zero mean Gaussian variables with covariance matrix $^{\circ} > 0$, and the $d \times d$ coefficient matrices of the transfer function satisfy the conditions $\sum_{s\geq 0} j||\boldsymbol{\Phi}_s|| < \infty$ and $\sum_{s=0}^{\infty} \boldsymbol{\Phi}_s = \overline{\boldsymbol{\Phi}} \neq \mathbf{0}$. Re-expressing $\mathbf{u}(t)$ using a Beveridge-Nelson decomposition (Phillips and Solo, 1992) gives $\sum_{s=0}^{\infty} \boldsymbol{\Phi}_s \boldsymbol{\eta}(t-s) = \overline{\boldsymbol{\Phi}}\boldsymbol{\eta}(t) + \boldsymbol{\nu}(t) - \boldsymbol{\nu}(t-1)$ where $\boldsymbol{\nu}(t) = \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s \boldsymbol{\eta}(t-s)$, $\boldsymbol{\Psi}_s = -(\boldsymbol{\Phi}_{s+1} + \boldsymbol{\Phi}_{s+2} + \cdots)$, and from this it follows that

$$\begin{aligned} \mathbf{z}(t) &= \mathbf{z}(0) + \boldsymbol{\delta}\, t + \sum_{\tau=1}^{t} \mathbf{u}(\tau) \\ &= \mathbf{z}(0) + \boldsymbol{\delta}\, t + \overline{\boldsymbol{\Phi}} \sum_{\tau=1}^{t} \boldsymbol{\eta}(\tau) + \boldsymbol{\nu}(t) - \boldsymbol{\nu}(0), \end{aligned}$$

where the initial values $\mathbf{z}(0)$ and $\nu(0)$ may be taken to be equal to fixed constants. The upshot of all this is that

$$x(t) = \boldsymbol{\alpha}^{\top}(\mathbf{z}(0) - \boldsymbol{\nu}(0)) + \boldsymbol{\alpha}^{\top}\boldsymbol{\delta}\, t + \boldsymbol{\alpha}^{\top}\overline{\boldsymbol{\Phi}} \sum_{\tau=1}^{t} \boldsymbol{\eta}(\tau) + \boldsymbol{\alpha}^{\top}\boldsymbol{\nu}(t),$$

which we can rewrite (for notational convenience) as

$$x(t) = x^{d_P}(t) + x^{d_{RW}}(t) + x^{c}(t)$$

where $x^{d_P}(t) = \boldsymbol{\alpha}^\top(\mathbf{z}(0) - \boldsymbol{\nu}(0) + \boldsymbol{\delta} t)$, $x^{d_{RW}}(t) = \boldsymbol{\alpha}^\top \overline{\boldsymbol{\Phi}} \sum_{\tau=1}^{t} \boldsymbol{\eta}(\tau)$ and $x^{c}(t) = \boldsymbol{\alpha}^\top \boldsymbol{\nu}(t)$. This decomposition of $x(t)$ into the sum of a polynomial series (the drift term), a random walk process and a stationary process obviously governs the limiting properties of the trajectory matrix, and in order that $\mathbf{N}(\omega)$ have a mixed spectrum we require that at least one of $\boldsymbol{\alpha}^\top\boldsymbol{\delta}$ and $\boldsymbol{\alpha}^\top\overline{\boldsymbol{\Phi}}$ be non-zero, to ensure the existence of a discrete component, and that $\boldsymbol{\alpha}^\top\boldsymbol{\nu}(t)$ is not identically zero, to guarantee the existence of the continuous component.

Taken together, the previously stated conditions allow appeal to be made to established results on linear processes and the asymptotic convergence properties of associated partial sum processes, see Phillips and Solo (1992) and Davidson (1994, Part VI) for example. In particular we have that for all $m = 1, \ldots, M$ where $M/N \to 0$ as $N \to \infty$;

1. $n^{-3}\sum_{t=1}^{n} x^{d_P}(t+r)^2 \sim (\boldsymbol{\alpha}^\top\boldsymbol{\delta})^2/3$,
2. $n^{-2}\sum_{t=1}^{n} x^{d_{RW}}(t+r)^2 \sim \boldsymbol{\alpha}^\top\overline{\boldsymbol{\Phi}}\boldsymbol{\Sigma}^{\frac{1}{2}}\int_0^1 \mathbb{B}(\omega)\mathbb{B}(\omega)^\top d\omega \boldsymbol{\Sigma}^{\frac{1}{2}}\overline{\boldsymbol{\Phi}}^\top\boldsymbol{\alpha}$
3. $n^{-1}\sum_{t=1}^{n} x^{c}(t+r)^2 \sim \sum_{s=0}^{\infty}\boldsymbol{\alpha}^\top\boldsymbol{\Phi}_s\boldsymbol{\Sigma}\boldsymbol{\Phi}_s^\top\boldsymbol{\alpha}$
4. $n^{-5/2}\sum_{t=1}^{n} x^{d_P}(t+r)x^{d_{RW}}(t+r) \sim \boldsymbol{\alpha}^\top\boldsymbol{\delta}\int_0^1 \omega\mathbb{B}(\omega)^\top d\omega\boldsymbol{\Sigma}^{\frac{1}{2}}\overline{\boldsymbol{\Phi}}^\top\boldsymbol{\alpha}$
5. $n^{-3/2}\sum_{t=1}^{n} x^{d_P}(t+r)x^{c}(t+r) \sim \boldsymbol{\alpha}^\top\boldsymbol{\delta}\int_0^1 \omega d\mathbb{B}(\omega)^\top d\omega\boldsymbol{\Sigma}^{\frac{1}{2}}\overline{\boldsymbol{\Phi}}^\top\boldsymbol{\alpha}$
6. $n^{-3/2}\sum_{t=1}^{n} x^{d_{RW}}(t+r)x^{c}(t+r) \sim$

$$\boldsymbol{\alpha}^\top\left(\overline{\boldsymbol{\Phi}}\boldsymbol{\Sigma}^{\frac{1}{2}}\int_0^1 \mathbb{B}(\omega)d\mathbb{B}(\omega)^\top d\omega\boldsymbol{\Sigma}^{\frac{1}{2}}\overline{\boldsymbol{\Phi}}^\top + \sum_{\tau=1}^{\infty}\sum_{s=0}^{\infty}\boldsymbol{\Phi}_s\boldsymbol{\Sigma}\boldsymbol{\Phi}_{s+\tau}^\top\right)\boldsymbol{\alpha},$$

uniformly in $r = 0, \ldots, m-1$, where $\mathbb{B}(\omega)$ denotes d dimensional standard Brownian motion

From the orders of magnitude implicit in relationships 1 through 6 it follows that

$\sum_{t=1}^{n} x(t+r)^2$ will be dominated by $\sum_{t=1}^{n} \mathfrak{s}(t+r)^2$ where $\mathfrak{s}(t) = x^{d_P}(t) + x^{d_{RW}}(t)$, and the size of $\sum_{t=1}^{n} \mathfrak{s}(t+r)^2$ will in turn be determined by the relative magnitudes of $x^{d_P}(t)$ and $x^{d_{RW}}(t)$ over the time interval $t = 1, \ldots, N$. Whichever term dominates $\mathfrak{s}(t)$, be it $x^{d_P}(t)$ or $x^{d_{RW}}(t)$, as $N \to \infty$ the Gramian $\mathbf{Y}\mathbf{Y}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^\top\mathbf{D}^{-\frac{1}{2}}$ will converge to $\mathbf{\Gamma}_m = \mathbf{1}_m\mathbf{1}_m^\top + o(1)$.

Figure 5 graphs two realizations of a random walk process with drift and autocorrelated errors where the Beveridge-Nelson decomposition gives rise to the characterization $x^{d_P}(t) = 1 + 0.1t$, $x^{d_{RW}}(t) = 2(1+\theta)\sum_{\tau=0}^{t-1}\eta(t-\tau)$ and $x^{d_C}(t) = -2\theta\eta(t)$, $|\theta| < 1$. For the first realization $\theta = -0.9$ and for the second $\theta = 0.9$, and for this process $\sum_{t=1}^{N} x^{d_P}(t)^2 \sim 0.003N^3$, $\sum_{t=1}^{N} x^{d_{RW}}(t)^2 \sim 4(1+\theta)^2N^2\int_0^1 \mathbb{B}^2(\omega)d\omega$ and $\sum_{t=1}^{N} x^{d_P}(t)x^{d_{RW}}(t) \sim 0.2(1+\theta)N^{5/2}\int_0^1 \omega\mathbb{B}(\omega)d\omega$, and the contribution of the two discrete components $x^{d_P}(t)$ and $x^{d_{RW}}(t)$ to the signal-noise ratio gives 17.83 dB when $\theta = -0.9$ and 22.46 dB when $\theta = 0.9$. Both processes exhibit a clearly perceptible drift, with the first behaving very much like the linear trend series observed previously. For both specifications we find that the Gramian $\mathbf{Y}\mathbf{Y}^\top = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^\top\mathbf{D}^{-\frac{1}{2}}$ approximates to

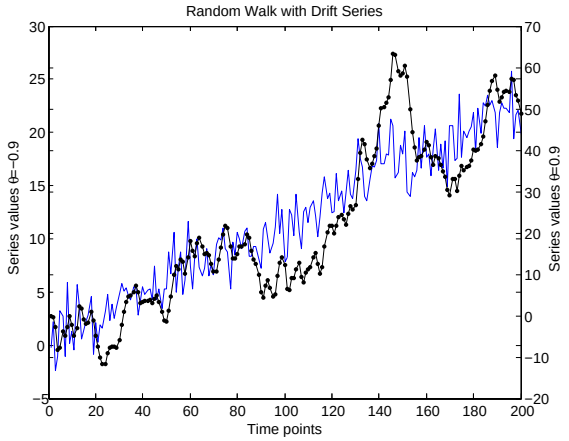

Figure 5: *Realizations of random walk process with drift and autocorrelated increments.*

an $m \times m$ equi-correlation matrix as in (12) with $\varrho \approx 0.9817$ when $\theta = -0.9$ and $\varrho \approx 0.9676$ when $\theta = 0.9$, and the first eigenvalue accounts for 98.29% and 96.97% of $m = \text{trace}\mathbf{Y}\mathbf{Y}^\top = \|\mathbf{D}^{-\frac{1}{2}}\mathbf{X}\|^2$, respectively.

In Figure 6 we graph for each series the true signal together with the first stage *R-SSA*$(15,1)$ signal reconstruction $s^d(t)$, and the true noise together with the second stage noise reconstruction $s^c(t)$, the latter derived from models chosen by assigning the signal dimension to the smallest value of k such that $100(\sum_{i=1}^{k}\ell_i/m) \geq 95\%$, *R-SSA*$(15,14)$ and *R-SSA*$(15,7)$ respectively. The difference in the specifications for the second stage models is consistent with the properties outlined in Theorem 4. The different specifications arise because the spectral distribution function of the noise $\mathfrak{n}(t)$ is $F^c(\omega) = 4\theta^2(\omega+\pi)/2\pi$ and, as is shown below, when $\theta = -0.9$ $s^d(t) \approx \mathfrak{s}(t)$ so the first stage residuals $x(t) - s^d(t) = \mathfrak{s}(t) - s^d(t) + \mathfrak{n}(t)$ closely approximate $\mathfrak{n}(t)$, when $\theta = 0.9$, on the other hand, $s^d(t)$ is much smoother than $\mathfrak{s}(t)$ so that $\mathfrak{s}(t) - s^d(t)$ introduces additional power into the first stage residuals that is not evenly spread over the interval $[-\pi,\pi)$ and the first stage residuals therefore have a power spectrum distribution that is no longer uniform. The upshot is that we have $r_{\mathfrak{s},s^d} = 0.9985$ and $r_{\mathfrak{n},s^c} = 0.9712$ for the first specification, compared to $r_{\mathfrak{s},s^d} = 0.9777$ and $r_{\mathfrak{n},s^c} = -0.0486$ for the second.

These seemingly anomalous results arise because when $\theta = -0.9$ the magnitude of $x^{d_{RW}}(t)$ is damped down (and approaches zero as $\theta \to -1$) so that $\mathfrak{s}(t) = x^{d_P}(t) + x^{d_{RW}}(t)$ closely approximates a linear trend; whereas when $\theta = 0.9$ the fluctuations in $x^{d_{RW}}(t)$ amplify the variation in the signal around the linear trend so as to increase the signal-noise ratio. But these additional fluctuations in $\mathfrak{s}(t)$ are precisely the feature that the *R-SSA*$(15,1)$ model evaluated in the first step finds difficulty in tracking. Thus, although the first stage residuals of both specifications are not too dissimilar, more extreme deviations are seen with the second specification because when $\theta = 0.9$ the first stage *R-SSA*$(15,1)$ model smooths out the sharp turning points in $\mathfrak{s}(t)$ induced by the presence of a significant random walk component. This manifests itself in the fact that although the contemporaneous correlation of the reconstructed noise with the true noise is very small $\theta = 0.9$, the cross-autocorrelation function, $r_{\mathfrak{n},s^c}(\tau)$, $\tau = 0, \pm1, \pm2, \ldots, \pm14$, (plotted in Figure 7) exhibits significant correlations at non-zero lags. The cross-autocorrelation function obtained when $\theta = -0.9$ behaves like that of two strongly correlated white noise processes.

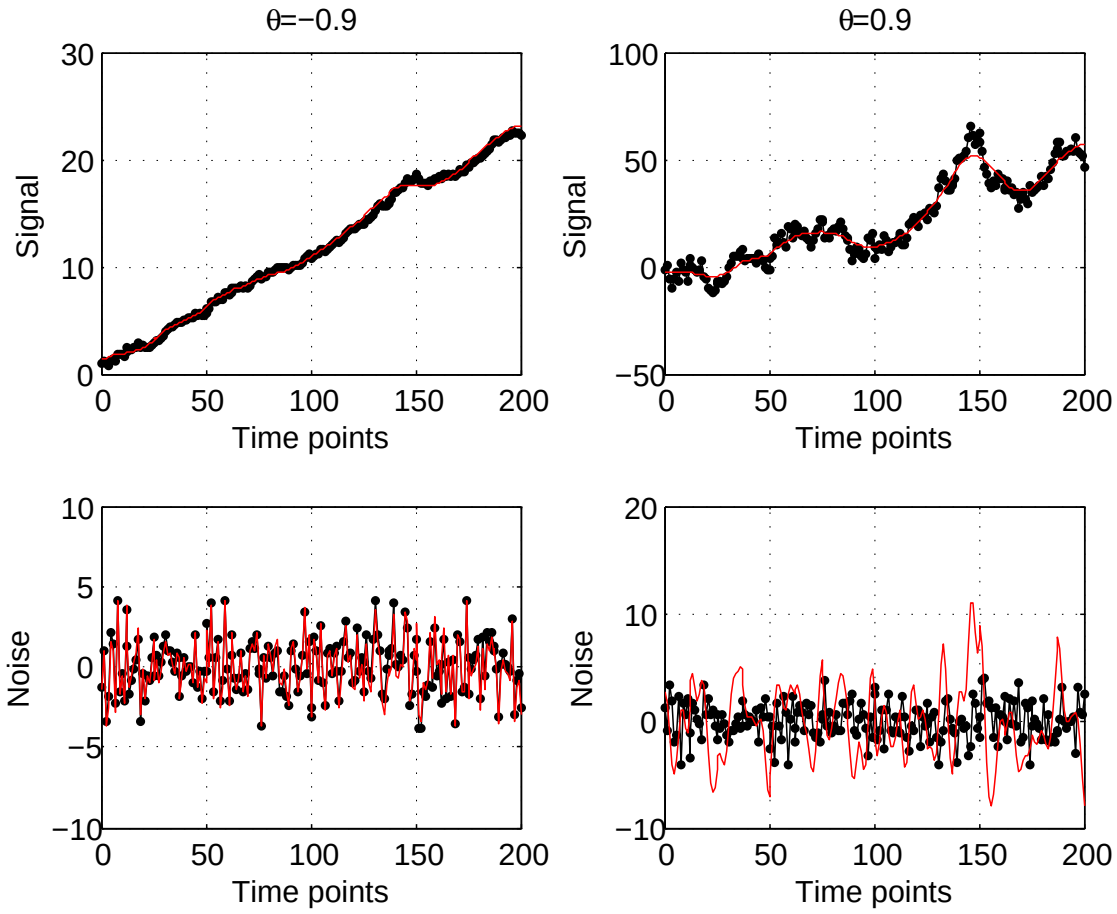

Figure 6: *Realized signal and noise components of random walk process with drift and autocorrelated increments overlaid with R-SSA first stage signal reconstruction reconstruction $s^d(t)$ and second stage noise reconstruction $s^c(t)$.*

These cross-autocorrelation values reflect that when $\theta = -0.9$ the first stage residual $x(t) - s^d(t) \approx \mathfrak{n}(t)$ because $s^d(t) \approx \mathfrak{s}(t)$, but when $\theta = 0.9$ the incorporation of the non-trivial difference $\mathfrak{s}(t)$—$s^d(t)$ into the first stage residuals produces both an increase in amplitude and a quasi-periodic phase shift effect that arises as a result of the over-smoothing that occurs with the first stage $R\text{-}SSA(15, 1)$ reconstruction. Nevertheless, properties of the signal that escape characterization in the first step are recaptured at in second step, and the overall series reconstructions $s^d(t) + s^c(t)$ perform similarly with correlations of $r_{x,s^d+s^c} = 0.9998$ and $r_{x,s^d+s^c} = 0.9993$ respectively. The ability of the two stage reconstruction to closely track the true signal-plus-noise decomposition is apparent even in this more complicated case.

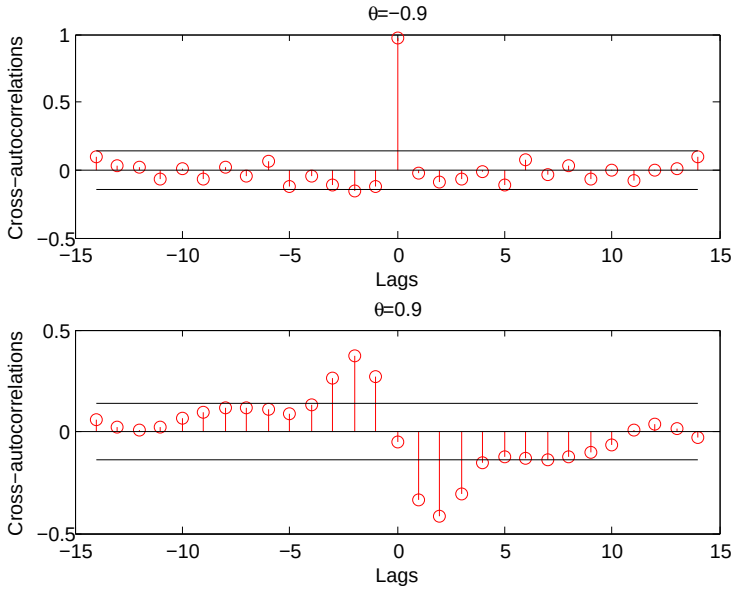

Figure 7: *Cross-autocorrelations between true noise* $\mathfrak{n}(t)$ *and second stage noise reconstruction* $s^c(t)$ *with approximate 95% white noise error bands: random walk process with drift and autocorrelated increments.*

6 Conclusions

In this paper we have demonstrated that if the observed series is a realization of a DGP characterized by an affine combination of Grenander processes then R-SSA – that is, SSA based upon the re-scaled trajectory matrix – will converge to a population ensemble model almost surely as the sample size increases. Provided that window lengths that are commensurate with the Whitney embedding theorem are employed, the R-SSA modelling will characterise the discrete and continuous components of the spectrum of the process. Numerical results have demonstrated that different DGPs that are not equivalent can give rise to the same asymptotic structure and hence the same spectral distribution function and spectrum; nevertheless, by projecting back into the time domain through the empirical eigenfunctions R-SSA is able to retrieve the individual structure embedded within such series. The numerical results obtained using simulated data series suggest that the theoretical properties outlined in the paper will be manifest in practice, and that the application of sequential R-SSA series reconstruction provides a viable alternative methodology to current practice.

Acknowledgement: I am grateful to Dimitrios Thomakos for helpful discussions and insightful comments on a previous version of this paper.

References

Alonso, F. J., Castillo, J. M., and Pintado, P. (2005), "Application of singular spectrum analysis to the smoothing of raw kinematic signals," *Journal of Biomechanics*, 38(5), 1085–1092.

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, New York: J. Wiley.

Basilevsky, A., and Hum, D. P. J. (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.

Beran, J. (1994), "Statistics for long-memory processes, volume 61 of," *Monographs on Statistics and Applied Probability*, .

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

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford: Oxford University Press.

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

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

Gentle, G. E. (2007), *Matrix Algebra: Theory, Computation, and Applications in Statistics* New York: Springer.

Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P. (2002), "Advanced spectral methods for climatic time series," *Reviews of Geophysics*, 40(1), 1003.

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

Grenander, U. (1954), "On the Estimation of Regression Coefficients in the Case of an Autocorrelated Disturbance," *Annals of Mathematical Statistics*, 25, 252–272.

Grenander, U., and Rosenblatt, M. (1957), *Statistical Analysis of Stationary Times Series*, New York: J. Wiley.

Grenander, U., and Szego, G. (1958), *Toeplitz Forms and Their Application*, Berkeley: University of California Press.

Hassani, H., and Thomakos, D. (2010), "A review on singular spectrum analysis for economic and financial time series," *Stat. Interface*, 3, 377–397.

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

Khan, M. A. R., and Poskitt, D. S. (2013), "A note on window length selection in singular spectrum analysis," *Australian & New Zealand Journal of Statistics*, 55(2), 87–108.

Khan, M. A. R., and Poskitt, D. S. (2015), "Signal Identification In Singular Spectrum Analysis: A Description Length Approach," *Australian and New Zealand Journal of Statistics*, . Forthcoming.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis* Academic Press.

Marques, C. A. F., Ferreira, J. A., Rocha, A., Castanheira, J. M., Melo-Gonçalves, P., Vaz, N., and Dias, J. M. (2006), "Singular spectrum analysis and forecasting of hydrological time series," *Physics and Chemistry of the Earth*, 31(18), 1172–1179.

Phillips, P. C. B., and Solo, V. (1992), "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

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

Prony, G. S. (1795), "Essai Experimental et Analytique: Sur les Lois de la Dilatabilitie de Fluides Elasttique et sur Celles de la Force Expansive de la Vapeur de L'Alkool, a Differentes Temperatures," *Journal de l'Ecole Polytechnique*, 1, 24–76.

Thomakos, D. (2008*a*), Optimal Linear Filtering, Smoothing and Trend Extraction for Processes with Unit Roots and Cointegration,, Working paper, University of Pelopeonnese. Available at Social Science Research Network.

Thomakos, D. (2008*b*), Optimal Linear Filtering, Smoothing and Trend Extraction of M-Period Differences of Processes with a Unit Root,, Working paper, University of Pelopeonnese. Available at Social Science Research Network.

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

Watanabe, S. (1965), Karhunen-Loéve Expansion and Factor Analysis: Theoretical Remarks and Applications,, in *Transactions of the Fourth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, Prague: Czechoslovak Academy of Sciences, pp. 635–660.

A Proofs

In what follows we will make use of the following result.

Lemma A.1 *Let* $\mathbf{A}_m$ *and* $\mathbf{B}_m$ *denote* $m \times m$ *Hermitian matrices with eigenvalue-eigenvector pairs of* $\{\mu_j, \boldsymbol{\psi}_j\}$, $j = 1, \ldots, m$, *and* $\{\lambda_j, \boldsymbol{\upsilon}_j\}$, $j = 1, \ldots, m$, *respectively. If* $\|\mathbf{A}_m - \mathbf{B}_m\| = h(m)$ *where* $h(m) \to 0$ *as* $m \to \infty$, *then* $|\mu_j - \lambda_j| \leq h(m)$ *and* $\|\varsigma_j \boldsymbol{\psi}_j - \boldsymbol{\upsilon}_j\| = o(h(m))$ *where* $\varsigma_j = sign(\boldsymbol{\upsilon}_j' \boldsymbol{\psi}_j)$, $j = 1, \ldots, m$.

Proof: By the Hoffman-Wielandt Theorem (Gentle, 2007, Page 271) we have

$$\sum_{j=1}^{m} (\mu_j - \lambda_j)^2 \leq \|\mathbf{A}_m - \mathbf{B}_m\|^2,$$

from which it follows that $(\mu_j - \lambda_j)^2 \leq \|\mathbf{A}_m - \mathbf{B}_m\|^2$ for any $j \in \{1, \ldots, m\}$. This implies that $|\mu_j - \lambda_j| \leq h(m)$, $j = 1, \ldots, m$.

Since the eigenvectors are orthonormal and span $\mathbb{R}^m$ we may set $\boldsymbol{\psi}_k = \sum_{j=1}^m c_j \boldsymbol{v}_j$ where the coefficients $c_j = \boldsymbol{v}_j' \boldsymbol{\psi}_k$ are such that $|c_j| \leq 1$ and $\sum_{j=1}^m c_j^2 = 1$. It follows that

$$(\mathbf{A}_m - \mathbf{B}_m + \mathbf{B}_m) \sum_{j=1}^m c_j \boldsymbol{v}_j = (\mu_k - \lambda_k + \lambda_k) \sum_{j=1}^m c_j \boldsymbol{v}_j \,,$$

which can be re-expressed as

$$\mathbf{B}_m \sum_{j=1}^m c_j \boldsymbol{v}_j = \lambda_k \sum_{j=1}^m c_j \boldsymbol{v}_j + o(h(m))$$

because $\|\mathbf{A}_m - \mathbf{B}_m\| = h(m)$ and $|\mu_k - \lambda_k| \leq h(m)$. It follows that $\sum_{j=1}^m c_j^2 (\lambda_j - \lambda_k)^2 = o(h(m)^2)$. Thus we can conclude that $c_j = o(h(m))$ whenever $\lambda_j \neq \lambda_k$ and hence that $|c_k| = 1 + o(h(m))$. Multiplying $\boldsymbol{\psi}_k$ by $\operatorname{sgn}(c_k)$ we obtain

$$\operatorname{sgn}(c_k) \boldsymbol{\psi}_k = \operatorname{sgn}(c_k) \sum_{j=1}^m c_j \boldsymbol{v}_j = |c_k| \boldsymbol{v}_k + \operatorname{sgn}(c_k) \sum_{\substack{j=1 \\ j \neq k}}^m c_j \boldsymbol{v}_j \,,$$

and subtracting $\boldsymbol{v}_k$ from either side and substituting $c_j = h(m)$ $j = 1, \ldots, m$, $j \neq k$, and $|c_k| = 1 + o(h(m))$ into the resulting equation we have

$$\operatorname{sgn}(c_k) \boldsymbol{\psi}_k - \boldsymbol{v}_k = (|c_k| - 1) \boldsymbol{v}_k + \operatorname{sgn}(c_k) \sum_{\substack{j=1 \\ j \neq k}}^m c_j \boldsymbol{v}_j = o(h(m)) \,.$$

Thus we find that the orthonormal eigenvectors of $\mathbf{A}_m$ differ from the orthonormal eigenvectors of $\mathbf{B}_m$ by a term of order $o(h(m))$, modulo a change in sign, since for the kth eigenvector we have that $\|\varsigma_k \boldsymbol{\psi}_k - \boldsymbol{v}_k\| = o(h(m))$ where $\varsigma_k = \operatorname{sgn}(c_k)$. ∎

Proof of Theorem 1: To begin, note that the rth diagonal entry of $\mathbf{D}^{-\frac{1}{2}}$ is the reciprocal of $\{\sum_{t=1}^{n} x^2(t+r-1)\}^{\frac{1}{2}}$, $r=1,\ldots,m$, and

$$\begin{aligned}\sum_{t=1}^{n} x^2(t+r-1) &= \sum_{t=1}^{n}\sum_{i=1}^{d}\sum_{j=1}^{d}\alpha_i\alpha_j z_i(t+r-1)z_j(t+r-1)\\ &= \sum_{i=1}^{d}\sum_{j=1}^{d}\alpha_i\alpha_j\sum_{t=1}^{n} z_i(t+r-1)z_j(t+r-1).\end{aligned}$$

From the definition of $r_{ij}^n(0)$ we have

$$\left|\frac{\sum_{t=1}^{n} z_i(t+r-1)z_j(t+r-1)}{\sqrt{a_{ii}^n(0)a_{jj}^n(0)}} - r_{ij}^n(0)\right| \le T_1 + T_2$$

where, via the Cauchy-Schwartz inequality, T_1 and T_2 are bounded by

$$\sqrt{\frac{\sum_{t=1}^{r-1} z_i^2(t)\sum_{t=1}^{r-1} z_j^2(t)}{a_{ii}^n(0)a_{jj}^n(0)}} \quad \text{and} \quad \sqrt{\frac{\sum_{t=n+1}^{n+r-1} z_i^2(t)\sum_{t=n+1}^{n+r-1} z_j^2(t)}{a_{ii}^n(0)a_{jj}^n(0)}},$$

respectively. Let $\boldsymbol{\beta}^\top = (\alpha_1\sqrt{a_{11}^n(0)},\ldots,\alpha_m\sqrt{a_{mm}^n(0)})$. Then by parts 2 and 3 of Assumption 1 it follows that for each $r=1,\ldots,m$ the absolute relative difference

$$\left|\frac{\sum_{t=1}^{n} x^2(t+r-1)}{\boldsymbol{\beta}^\top\mathbf{R}(0)\boldsymbol{\beta}} - 1\right|,$$

will converge to zero as $n=(N-m+1)\ge(N-M+1)\to\infty$ as $N\to\infty$.

The rcth entry in $\mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^\top\mathbf{D}^{-\frac{1}{2}}$, $r,c=1,\ldots,m$, is

$$\sum_{t=1}^{n} x(t+r-1)x(t+c-1) = \sum_{i=1}^{d}\sum_{j=1}^{d}\alpha_i\alpha_j\sum_{t=1}^{n} z_i(t+r-1)z_j(t+c-1)$$

divided by $\sqrt{\sum_{t=1}^{n} x^2(t+r-1)\sum_{t=1}^{n} x^2(t+c-1)}$, and

$$\sum_{t=1}^{n} z_i(t+r-1)z_j(t+c-1) = \sum_{\tau=1}^{n} z_i(\tau+r-c)z_j(\tau) - S_1 + S_2$$

where

$$S_1 = \sum_{\tau=1}^{c-1} z_i(\tau + r - c) z_j(\tau) \quad \text{and} \quad S_2 = \sum_{\tau=n+1}^{n+c-1} z_i(\tau + r - c) z_j(\tau).$$

Bounding the truncation effects S_1 and S_2 using the Cauchy-Schwartz inequality gives

$$|S_1|^2 \leq \sum_{\tau=1}^{c-1} z_i^2(\tau + r - c) \sum_{\tau=1}^{c-1} z_j^2(\tau) \quad \text{and} \quad |S_2|^2 = \sum_{\tau=n+1}^{n+c-1} z_i^2(\tau + r - c) \sum_{\tau=n+1}^{n+c-1} z_j^2(\tau),$$

from which we can conclude via parts 2 and 3 of Assumption 1 that

$$\lim_{n\to\infty} \left| \frac{\sum_{t=1}^{n} z_i(t + r - 1) z_j(t + c - 1)}{\sqrt{a_{ii}^n(0) a_{jj}^n(0)}} - r_{ij}^n(h) \right| = 0$$

for all r and c such that $r - c = h$, where $r, c = 1, \ldots, m$ and $h = 0, 1, \ldots, m - 1$. It therefore follows that

$$\lim_{n\to\infty} |\sum_{t=1}^{n} x(t + r - 1) x(t + c - 1) - \boldsymbol{\beta}^\top \mathbf{R}(h) \boldsymbol{\beta}| = 0$$

where $h = r - c$.

Collecting the previous properties together we find that for all $r, c = 1, \ldots, m$, the rcth entry in $\mathbf{D}^{-\frac{1}{2}} \mathbf{X}\mathbf{X}^\top \mathbf{D}^{-\frac{1}{2}}$ converges to $\boldsymbol{\beta}^\top \mathbf{R}(r - c) \boldsymbol{\beta} / \boldsymbol{\beta}^\top \mathbf{R}(0) \boldsymbol{\beta}$. Now, by an extension of Herglotz's lemma due to Cramér, $\mathbf{R}(h) = \int_{-\pi}^{\pi} e^{-\imath \omega h} d\mathbf{M}(\omega)$ where the $d \times d$ matrix valued function

$$\mathbf{M}(\omega) = \lim_{N\to\infty} \sum_{h=-N+1}^{N-1} \mathbf{R}(h) \left[\frac{\exp(-\imath \omega h) - 1}{-\imath h} \right], \quad -\pi \leq \omega \leq \pi,$$

has entries of bounded variation, and $\mathbf{M}(\omega_2) - \mathbf{M}(\omega_1) = (\mathbf{M}(\omega_2) - \mathbf{M}(\omega_1))^* \geq 0$, $\omega_1 \leq \omega_2$. Recalling from part 4 of Assumption 1 that $\mathbf{R}(0)$ is nonsingular, and therefore possesses a unique symmetric square root $\mathbf{R}(0)^{\frac{1}{2}}$, the result in the theorem now follows directly. ∎

Proof of Theorem 2: From Theorem 1 it follows that $\lim_{N\to\infty} \|\mathbf{D}^{-\frac{1}{2}} \mathbf{X}\mathbf{X}^\top \mathbf{D}^{-\frac{1}{2}} - \boldsymbol{\Gamma}_m\| = 0$. Applying Lemma A.1 with $\mathbf{A}_m = \mathbf{D}^{-\frac{1}{2}} \mathbf{X}\mathbf{X}^\top \mathbf{D}^{-\frac{1}{2}}$, $\mathbf{B}_m = \boldsymbol{\Gamma}_m$ and $h(m) = o(1)$, we can

therefore conclude that $\lim_{N\to\infty}|\ell_j-\lambda_j|=0$ and $\lim_{N\to\infty}\|\varsigma_j\mathbf{u}_j-\boldsymbol{v}_j\|=0$ where $\varsigma_k=\operatorname{sgn}(\mathbf{u}_j^\top\boldsymbol{v}_j)$, for $j=1,\ldots,m$. ∎

Proof of Theorem 3: See Grenander and Rosenblatt (1957, Chapter 7.4).

Proof of Theorem 4: Set $\mathbf{T}_m$ equal to the $m\times m$ Toeplitz matrix with first row $(1,(1-1/m)\varrho^c(1),\ldots,(1/m)\varrho^c(m-1))$. Then

$$\begin{aligned}\frac{1}{m^{1+2d}}\|\boldsymbol{\Gamma}_m^c-\mathbf{T}_m\|^2 &= \frac{1}{m^{2d}}\sum_{r=1-m}^{m-1}\left(1-\frac{|r|}{m}\right)\left|\varrho^c(r)\frac{|r|}{m}\right|^2\\ &\le 2C^2\sum_{r=1}^{m-1}\left(1-\frac{r}{m}\right)\frac{r^{4d}}{m^{2(1+d)}}\\ &\sim\begin{cases}2C^2m^{2d-1}/(4d+1)(4d+2), & -1/4<d<1/2;\\ 2C^2\log(m)m^{-3/2}, & d=-1/4;\\ 2C^2\zeta(-4d)m^{-2(1+d)}, & -1/2<d<-1/4.\end{cases}.\end{aligned}$$

Now let $\mathbf{C}_m$ be the $m\times m$ Toeplitz matrix whose first row is given by the vector $(c_m(0),c_m(1),\ldots,c_m(m-1))$ where

$$c_m(r)=\frac{2\pi}{m}\sum_{j=0}^{m-1}f_m^c(2\pi j/m)\exp(\imath 2\pi jr/m).$$

From the relationship

$$\begin{aligned}c_m(r) &= \frac{1}{m}\sum_{j=0}^{m-1}\left\{\sum_{s=-(m-1)}^{m-1}\left(1-\frac{|s|}{m}\right)\varrho^c(s)\exp(-\imath 2\pi js/m)\right\}\exp(\imath 2\pi jr/m)\\ &= \sum_{s=-(m-1)}^{m-1}\left(1-\frac{|s|}{m}\right)\varrho^c(s)\left\{\frac{1}{m}\sum_{j=0}^{m-1}\exp(\imath 2\pi j(r-s)/m)\right\} \qquad (A.1)\end{aligned}$$

we find that

$$\begin{aligned} c_m(0) &= \varrho^c(0) = 1 \quad \text{and} \\ c_m(r) &= \left(1-\frac{|r|}{m}\right)\varrho^c(r) + \left(1-\frac{|m-r|}{m}\right)\varrho^c(m-r), \end{aligned} \tag{A.2}$$

for $r = \pm 1, \ldots, \pm(m-1)$, where in (A.1) we have used the orthogonality of the complex exponentials:

$$\frac{1}{m}\sum_{j=0}^{m-1}\exp(\imath 2\pi jk/m) = \begin{cases} 1, & k = 0 \bmod m; \\ 0, & \text{otherwise.} \end{cases}.$$

From (A.2) it follows that $\mathbf{C}_m$ is also a circulant matrix. The eigenvectors of $\mathbf{C}_m$ are therefore $\mathbf{e}_m(2\pi j/m)$, $j = 0, 1, \ldots, (m-1)$, and the eigenvalues equal

$$\begin{aligned} \sum_{r=0}^{m-1} c_m(r)\exp(-\imath 2\pi kr/m) &= \sum_{r=0}^{m-1}\left\{\frac{2\pi}{m}\sum_{j=0}^{m-1} f_m^c(2\pi j/m)\exp(\imath 2\pi jr/m)\right\}\exp(-\imath 2\pi kr/m) \\ &= 2\pi\sum_{j=0}^{m-1} f_m^c(2\pi j/m)\left\{\frac{1}{m}\sum_{r=0}^{m-1}\exp(\imath 2\pi(j-k)r/m)\right\} \\ &= 2\pi f_m^c(2\pi k/m), \quad k = 0, 1, \ldots, (m-1). \end{aligned}$$

Substituting (A.2) for the entries in $\mathbf{C}_m$ we find that

$$\begin{aligned} \frac{1}{m^{1+2d}}\|\mathbf{T}_m - \mathbf{C}_m\|^2 &= \frac{1}{m^{2d}}\sum_{r=1-m}^{m-1}\left(1-\frac{|r|}{m}\right)\left|\left(1-\frac{|m-r|}{m}\right)\varrho^c(m-r)\right|^2 \\ &\le 2C^2\sum_{r=1}^{m-1}\left(1-\frac{r}{m}\right)\frac{r^2}{m^{2(1+d)}(m-r)^{2-4d}} \\ &= 2C^2\sum_{s=1}^{m-1} s^{4d-1}\frac{(m-s)^2}{m^{3+2d}} \\ &\sim \begin{cases} 2C^2 m^{2d-1}(7+12d)/(4d+1)(4d+2), & 0 < d < 1/2; \\ 2C^2\log(m)m^{-1}, & d = 0; \\ 2C^2\zeta(1-4d)m^{-(1+2d)}, & -1/2 < d < 0. \end{cases} \end{aligned}$$

The result stated in the theorem now follows from an application of the triangle inequality $\|\mathbf{\Gamma}_m^c - \mathbf{C}_m\| \le \|\mathbf{\Gamma}_m^c - \mathbf{T}_m\| + \|\mathbf{T}_m - \mathbf{C}_m\|$ and Lemma A.1. ∎