

ISSN 1440-771X



MONASH University

Australia

Department of Econometrics and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

**Moment Tests for Window Length
Selection in Singular Spectrum Analysis
of Short- and Long-Memory Processes**

Md Atikur Rahman Khan and D. S. Poskitt

September 2011

Working Paper 22/11

Moment Tests for Window Length Selection in Singular Spectrum Analysis of Short- and Long-Memory Processes

Md Atikur Rahman Khan

&

D. S. Poskitt*

Department of Econometrics and Business Statistics, Monash University

Abstract

In this paper we propose a new methodology for selecting the window length in Singular Spectral Analysis in which the window length is determined from the data prior to the commencement of modeling. The selection procedure is based on statistical tests designed to test the convergence of the autocovariance function. A classical time series portmanteau type statistic and two test statistics derived using a conditional moment principle are considered. The first two are applicable to short-memory processes, and the third is applicable to both short- and long-memory processes. We derive the asymptotic distribution of the statistics under fairly general regularity conditions and show that the criteria will identify true convergence with a finite window length with probability one as the sample size increases. Results obtained using Monte-Carlo simulation indicate the relevance of the asymptotic theory, even in relatively small samples, and that the conditional moment tests will choose a window length consistent with the Whitney embedding theorem. Application to observations on the Southern Oscillation Index shows how observed experimental behaviour can be reflected in features seen with real world data sets.

Keywords: Portmanteau type test, Conditional moment test, Asymptotic distribution, Linear regular process, Singular spectrum analysis, Embedding.

JEL Classification: C12, C22, C52

1 Introduction

Singular spectrum analysis (SSA) is a non-parametric technique that has gained popularity in the analysis of meteorological (Ghil et al. 2002), bio-mechanical (Alonso et al. 2005) and hydrological time series (Marques et al. 2006), and following its successful application in the physical sciences, applications in economics and finance are now also finding favour (Hassani & Zhigljavsky 2009). SSA is designed to look for both persistent and transitory behaviour in an observed time series, and expositions of the basic ideas and methods can be found in the monographs by Elsner & Tsonis (1996) and Golyandina et al. (2001).

*Corresponding address: Don Poskitt, Department of Econometrics and Business Statistics, Monash University, Victoria 3800, Australia. Tel.:+61-3-9905-9378; fax:+61-3-9905-5474.

E-mail address: Donald.Poskitt@monash.edu

In SSA the initial step consists of the construction of the, so called, trajectory matrix. Given an observed time series $\{x_1, x_2, \dots, x_N\}$ of length N , $x_t \in \mathbb{R}$, and a user-specified window length (or lag length) m where $2 \leq m \leq N/2$, the trajectory matrix, \mathbf{X} say, is obtained via the embedding mapping $H : \{x_1, x_2, \dots, x_N\} \mapsto \mathbf{X}$ where

$$\mathbf{X} = [\mathbf{x}_1 : \dots : \mathbf{x}_n] \quad (1)$$

and $\mathbf{x}_i = (x_i, x_{i+1}, \dots, x_{i+m-1})'$. The \mathbf{x}_i for $i = 1, 2, \dots, n$ are known as the lagged vectors of \mathbf{X} , a Hankel matrix of order m by $n = N - m + 1$. Following the embedding three further steps are used to determine a signal-plus-noise type model for the observed series; (i) the singular value decomposition (SVD) of \mathbf{X} , (ii) noise reduction and signal extraction (component grouping), and finally, (iii) time series reconstruction. See [Elsner & Tsonis \(1996\)](#) and [Golyandina et al. \(2001\)](#) for details.

If we let $\ell_1 \geq \ell_2 \geq \dots, \geq \ell_m \geq 0$ denote the ordered eigenvalues of the Gramian $\mathbf{G} = n^{-1}\mathbf{X}\mathbf{X}'$, and denote by $\mathbf{U}_1, \dots, \mathbf{U}_m$ the associated orthonormal system of eigenvectors, then the row space of the trajectory matrix has dimension d where $d = \max\{i : \ell_i > 0\}$ and \mathbf{X} can be expressed exactly as the sum of $d \leq m$ rank one projections;

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d \quad (2)$$

wherein $\mathbf{X}_i = \sqrt{n\ell_i}\mathbf{U}_i\mathbf{V}_i'$ where \mathbf{U}_i and $\mathbf{V}_i = \mathbf{X}'\mathbf{U}_i/\sqrt{n\ell_i}$ are the i th left and right eigenvectors of \mathbf{X} , and $\sqrt{n\ell_i}$ is the i th singular value. The SVD of \mathbf{X} is the fundamental operation underlying SSA and it is apparent from the decomposition in (2) that the window size m will be an important factor in determining the outcome of any further analysis and, indeed, that the choice of window length could be critical in assessing any eigenstructure inherent in the properties of the original time series.

The standard approach to window length selection is to determine a value for m large enough to ensure that the signal and noise components are easily (in the terminology of SSA) *separated*. This is achieved by calculating a weighted correlation between the reconstructed signal and noise components and the window length corresponding to the minimum weighted correlation is selected for further analysis. This approach involves the use of an image plot of the weighted correlation matrix for different window lengths and the selected window length is supposed to provide a 'clear view' of the orthogonality of the components. Given that the components in the SVD are orthogonal by construction, the weighted correlation is likely to be small for various different window lengths, and in the absence of clear cut statistical decision rules and with few guidelines on how to set appropriate thresholds, the associated modeling entails substantial subjective assessment.

Some recent studies have used window lengths proportional to the perceived periodicity of the data, $m = \ell\omega$ where ω is the frequency and ℓ is an integer such that $\ell\omega \leq N/2$ ([Hassani & Zhigljavsky 2009](#), for example). Assuming that ω is known, the value of ℓ , and hence m , is chosen using the pattern recognition techniques as just outlined, but see also [Golyandina \(2010\)](#). For some additional discussion of different aspects of the currently

prevailing methods of window length selection see [Elsner & Tsonis \(1996, Chapter 5\)](#) and [Golyandina et al. \(2001\)](#).

In this paper we propose a methodology in which m is determined from the data prior to the construction of the trajectory matrix and commencement of the SSA modeling. In this we are motivated by the fact that [Khan & Poskitt \(2010\)](#) have developed a description length principle that enables the user to consistently extract signal components (both theoretically and in practice) given a preassigned window length compatible with the Whitney embedding theorem. The techniques that we develop are related to the work of [Tzagkarakis et al. \(2009\)](#), who selected the window length as the point of first crossing of a confidence interval (CI) for the sample autocorrelation function (SACF). Although this might be appropriate for the type of data examined in [Tzagkarakis et al. \(2009\)](#), in general the point of first crossing of a ("white noise" 95%) CI by the SACF does not represent a time interval beyond which there is no memory left in the process. Rather it is the overall profile and convergence properties of the autocovariance function that characterize the properties of the process and it is this feature that the procedures developed in this paper attempt to exploit.

To obtain an appropriate statistical decision rule we need to specify the class of processes to be analyzed, this we do in the following section. Given appropriate regularity, Section 3 shows how some classical results in time series analysis can be used to construct a portmanteau type statistic that can be employed to determine window length. Section 4 then outlines a conditional moment procedure that leads to a test statistic that can also be employed to select m in a similar manner, but which does not suffer from the drawbacks inherent in using the portmanteau type statistic. Section 5 indicates how the conditional moment procedure can be appropriately modified so as to yield a window selection method that can be applied to both short range and long range dependent processes.

2 Regularity

Let x_t denote a stochastic process of interest, and let $\mathcal{X}_{-\infty}^t$ denote the linear manifold determined by x_s , $s \leq t$, and \mathcal{X}_t^{∞} the linear manifold determined by x_s , $s \geq t$.²

Definition 1 *A stationary process x_t is linearly regular (nonsingular) if and only if $\mathcal{X}_{-\infty} = \bigcap_{s=0}^{\infty} \mathcal{X}_{-\infty}^{t-s} = 0$ and linearly singular (deterministic) if and only if $\mathcal{X}_{-\infty} = \mathcal{X} = \bigcup_{s=-\infty}^{\infty} \mathcal{X}_{-\infty}^{t-s}$.*

From the argument presented in [Ibragimov & Linnik \(1971, Section 17.1\)](#) we can deduce the following result.

Theorem 1 *A necessary and sufficient condition for x_t to be a linearly regular, stationary process is that, for all $\zeta \in \mathcal{X}$ with $E[\zeta^2] < \infty$,*

$$\lim_{t \rightarrow -\infty} \sup_{\xi} |E[\xi\zeta] - E[\xi]E[\zeta]| = 0,$$

²Following common practice our notation does not distinguish between a stochastic process and realized values of that process. The required meaning should be readily apparent from the context.

where the supremum is taken over all $\xi \in \mathcal{X}_{-\infty}^t$ with $E[\xi^2] < \infty$.

Theorem 1 conveys the idea that for a linearly regular process events become uncorrelated (orthogonal) with increasing separation, and if we suppose that x_t for $t \in \mathbb{Z}$ is a linearly regular, covariance-stationary process, with mean $E[x_t] = \mu$ and autocovariance function $E[(x_t - \mu)(x_{t+h} - \mu)] = \gamma(h)$, an obvious implication of the theorem is that $|\gamma(k)| \rightarrow 0$ as $k \rightarrow \infty$.

In order to illustrate the significance of linear regularity for SSA consider the structure of the trajectory matrix. Theorem 1 indicates that values of the process separated by more than m time periods will be (roughly) uncorrelated when m is sufficiently large. This means that the first and m th row of \mathbf{X} will be almost orthogonal,

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_n \\ x_2 & x_3 & x_4 & \dots & x_{n+1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ x_m & x_{m+1} & x_{m+2} & \dots & x_N \end{bmatrix} \left. \begin{array}{l} \Leftarrow \\ \\ \\ \Leftarrow \end{array} \right\} \text{Orthogonal,}$$

and we can think of these two rows (heuristically) as uncorrelated realizations of the process, each of length n . Now, supposing for simplicity that $E[x_t] = 0$, for any choice of window length $m \in [2, N/2]$ we have $\mathbf{\Gamma} = E[\mathbf{G}] = n^{-1} \sum_{i=1}^n E[\mathbf{x}_i \mathbf{x}_i']$, which by ergodicity yields the Toeplitz matrix

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(m-2) & \gamma(m-1) \\ & \gamma(0) & \dots & \cdot & \gamma(m-2) \\ & & \ddots & & \vdots \\ - \dots - & & & \gamma(0) & \gamma(1) \\ & & & & \gamma(0) \end{bmatrix}$$

as the almost sure limit of the Gramian from which the SVD of \mathbf{X} is constructed. Again, this suggests selecting the value of m such that $|\gamma(k)| < \delta$ for all $k > m$ where δ small.

In practice $\gamma(k)$ will not be known and must be estimated from the data, x_1, x_2, \dots, x_N , which is now, of course, assumed to be a realization of the process x_t of length N . Set $\bar{x} = N^{-1} \sum_{t=1}^N x_t$ and denote the SACF by $\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}$, $k = 0, 1, \dots, N-1$, where the k th sample autocovariance

$$\hat{\gamma}(k) = \frac{1}{N} \sum_{t=k+1}^N (x_t - \bar{x})(x_{t-k} - \bar{x}), \quad k = 0, 1, \dots, N-1. \quad (3)$$

If we are to base the choice of lag window length on the SACF we must obviously allow for sampling variability.

More often than not, the starting point for deriving the sampling properties of $\hat{\gamma}(k)$ and $\hat{\rho}(k)$ is Wold's representation theorem, namely, that if x_t is a linearly regular, covariance-stationary process then x_t can be expressed as

$$x_t = \mu + \sum_{j=0}^{\infty} \kappa(j) \varepsilon_{t-j} \quad (4)$$

where ε_t is a zero mean white noise (innovation) process with variance σ^2 . The coefficients of the transfer function $k(z) = \sum_{j \geq 0} \kappa(j) z^j$ satisfy the conditions $\kappa(0) = 1$ and $\sum_{j \geq 0} \kappa(j)^2 < \infty$. If $0 < \sum_{j \geq 0} |\kappa(j)| < \infty$ then x_t is said to be a short memory processes, whereas if $\sum_{j \geq 0} |\kappa(j)| = \infty$ then x_t is said to exhibit long memory, see [Beran \(1994\)](#) or [Palma \(2007\)](#). This division of linearly regular processes into short and long memory series according to the speed of decay of their impulse response coefficients proves to be of crucial importance in our subsequent analysis.

3 Short Memory Processes and a Portmanteau Type Test

To begin we state a classical theorem concerning the sampling properties of the sample autocovariances.

Theorem 2 *If x_t is a linear regular process with short memory, driven by independent and identically distributed innovations ε_t with zero mean, variance σ^2 and finite fourth moment $E[\varepsilon_t^4] = \eta\sigma^4 < \infty$, then for any non-negative integer k*

$$\sqrt{N}(\hat{\gamma}(0) - \gamma(0), \dots, \hat{\gamma}(k) - \gamma(k)) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{\Omega})$$

where (Bartlett's formula)

$$\mathbf{\Omega} = \left[\sum_{s=-\infty}^{\infty} \{ \gamma(s+p)\gamma(s+q) + \gamma(s-q)\gamma(s+p) \} + (\eta - 3)\gamma(p)\gamma(q) \right]_{p,q=0,\dots,k}.$$

Theorems of this type can be traced back to the pioneering work of [Bartlett \(1946\)](#), for a proof see [Brockwell & Davis \(1991, Proposition 7.3.4\)](#). In what follows we will continue to assume that the innovations are independent and identically distributed with zero mean, variance σ^2 and finite fourth moment $\eta\sigma^4$ {henceforth abbreviated to i.i.d.(0, σ^2 , η)}. As pointed out in [Hannan & Heyde \(1972\)](#), however, if independence is replaced by the imposition of a classical martingale difference structure – $E[\varepsilon_t | \mathcal{E}_{t-1}] = 0$ and $E[\varepsilon_t^2 | \mathcal{E}_{t-1}] = \sigma^2$ where \mathcal{E}_t denotes the σ -algebra of events determined by ε_s , $s \leq t$ – then subject to only mild additional conditions the classical theory goes through (See also [Fuller 1996, Chapter 6, Exercises 23 & 24](#)).

The quintessential example of short memory is an M -dependent process, for which $\mathcal{X}_{t+s}^{\infty}$ and $\mathcal{X}_{-\infty}^t$ are uncorrelated (orthogonal) for all $s > M$, and the process can be expressed as $x_t = \mu + \sum_{j=0}^M \kappa(j) \varepsilon_{t-j}$. This means that $\gamma(k) = 0$ for all $k > M$ and the autocorrelation function $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$ is null for $|k| > M$. Specializing [Theorem 2](#) and Bartlett's formula to

the M -dependent case it is easy to show the following result. (c.f. Brockwell & Davis 1991, Theorem 7.2.1, Remark 1.)

Theorem 3 *If x_t is an M -dependent process with i.i.d. $(0, \sigma^2, \eta)$ innovations then for all $k > M$ and any non-negative integer h*

$$\sqrt{N}(\hat{\rho}(k), \dots, \hat{\rho}(k+h)) \xrightarrow{D} (\zeta_1, \dots, \zeta_h)$$

where $\zeta_j = \sum_{r=-M}^M \rho(r) Z_{j-r}$ and the Z_s are independent standard normal, i.i.d. $\mathcal{N}(0, 1)$.

To formulate a statistical decision rule suitable for window length selection in SSA, let us now consider testing the null hypothesis H_0 : x_t is an m -dependent process.

Proposition 1 *Let $\hat{\rho}' = (\hat{\rho}(m+1), \dots, \hat{\rho}(m+h))$ and set $PTT_m(h) = N\hat{\rho}'\hat{\Omega}^{-1}\hat{\rho}$ where $\hat{\Omega} = [\hat{\omega}_{pq}]_{p,q=1,\dots,h}$*

$$\begin{aligned} \hat{\omega}_{pq} &= \sum_{s=-m}^{m-|p-q|} \hat{\rho}(s)\hat{\rho}(s+|p-q|), & |p-q| \leq 2m \\ &= 0, & |p-q| > 2m. \end{aligned}$$

Suppose that x_t is a linear regular process with i.i.d. $(0, \sigma^2, \eta)$ innovations. Then under H_0 the statistic $PTT_m(h) \xrightarrow{D} \chi_h^2$ where χ_h^2 denotes the χ^2 -distribution with h degrees of freedom. The critical region $\{PTT_m(h) \geq \chi_h^2(1-p)\}$, where $\chi_h^2(1-p)$ denotes the χ_h^2 quantile function, has asymptotic size p and yields a consistent test of H_0 against the alternative H_1 that $\rho = (\rho(m+1), \dots, \rho(m+h)) \neq \mathbf{0}$.

PROOF: The null distribution of $PTT_m(h)$ follows as a direct consequence of Theorem 3. That $\{PTT_m(h) \geq \chi_h^2(1-p)\}$ provides an asymptotic critical region of size p is then obvious. Consistency follows by observing that $\text{plim}\hat{\rho} = \rho$ and $\hat{\Omega}$ converges to a positive definite matrix, Ω say. If $\rho \neq \mathbf{0}$ then $\rho'\Omega^{-1}\rho > 0$ and for any $\delta > 0$

$$\lim_{N \rightarrow \infty} Pr \left[\left| \frac{PTT_m(h)}{N} - \rho'\Omega^{-1}\rho \right| < \delta \right] = 1.$$

Setting $0 < \delta < \rho'\Omega^{-1}\rho$ it follows that for all $N > \chi_h^2(1-p)/(\rho'\Omega^{-1}\rho - \delta)$ we have

$$\begin{aligned} Pr [PTT_m(h) > \chi_h^2(1-p)] &= Pr \left[\frac{PTT_m(h)}{N} > \frac{\chi_h^2(1-p)}{N} \right] \\ &\geq Pr \left[\frac{PTT_m(h)}{N} > \rho'\Omega^{-1}\rho - \delta \right] \\ &\geq Pr \left[\left| \frac{PTT_m(h)}{N} - \rho'\Omega^{-1}\rho \right| < \delta \right], \end{aligned}$$

and, as we have just seen, the later converges to one as N increases. We can therefore conclude that $\lim_{N \rightarrow \infty} Pr [PTT_m(h) > \chi_h^2(1-p)] = 1$ under H_1 . \blacksquare

We can now contemplate calculating $PTT_m(h)$ for a sequence of values $m \leq N/2$ and choosing for m a value beyond which H_0 is regularly deemed to be acceptable.

REMARK 1: The hypothesis H_0 may seem over-restrictive in its current guise, but by the Weierstrass (trigonometric) approximation theorem we know that if x_t is a linear regular process, then for any $\epsilon > 0$ there exists an M -dependent process with transfer function $k_M(e^{i\omega})$ such that $||k(e^{i\omega})|^2 - |k_M(e^{i\omega})|^2| < \epsilon$, $\omega \in [-\pi, \pi]$. If $\gamma_M(k)$ denotes the autocovariance of such an M -dependent process then this implies that $|\gamma(k) - \gamma_M(k)| < \sigma^2\epsilon$. Hence H_0 is equivalent to the hypothesis that x_t can be approximated arbitrarily closely by an M -dependent process and that $|\gamma(k)| < \delta = \sigma^2\epsilon$ for all $k > M = m$.

REMARK 2: Note that when $m = 0$, implying that $x_t = \varepsilon_t$, the covariance matrix $\hat{\Omega}$ collapses to the identity and $PTT_0(h)$ is equivalent to the [Box & Pierce \(1970\)](#) portmanteau statistic $N \sum_{r=1}^h \hat{\rho}(r)^2$.

To examine the performance of $PTT_m(h)$ we generated observations from an M -dependent process using the coefficients from Spencer's 15 point moving average smoothing filter. The data generating mechanism was taken as the (two-sided) $MA(q)$ process

$$x_t = \varepsilon_t - \sum_{i=-q}^q \theta_i \varepsilon_{t-i}$$

where $\theta_i = \theta_{-i}$ and ε_t is i.i.d. $\mathcal{N}(0, 1)$, the coefficients for Spencer's 15 point moving average smoothing filter being

$$\{\theta_0, \theta_1, \dots, \theta_7\} = \{0.231, 0.209, 0.144, 0.066, 0.009, -0.016, -0.019, -0.009\} .$$

Figure 1 plots two typical realizations of this process with sample size $N = 500$. The

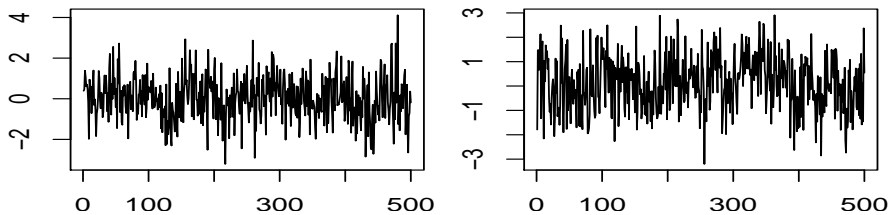


Figure 1: Typical realizations of Spencer's 15 point moving average process

operation of Slutsky's effect ([Slutzky 1937](#)) is apparent despite the fact that the coefficients are small (the smallest root of $\sum_{i=0}^q \theta_i z^i$ has modulus $|z| = 1.5707$).

Figure 2 depicts the average value of $\hat{\rho}(m)$ and the average value of $p_m = Prob(\chi_h^2 > PTT_m(h))$, the average p -value evaluated from $PTT_m(h)$, with $h = \sqrt{N}$, computed from 10000 realizations of the process with $N = 500$. To provide points of reference the theoretical value $\rho(m)$, the white noise 95% CI $\pm 1.96/\sqrt{N}$, and $p = 0.5$, are also plotted in Figure 2.

The inappropriateness of the point of first crossing of the white noise CI as a choice for m in this case is obvious. Under H_0 the p -values should be uniformly distributed on the unit interval and we therefore expect the values of p_m to equal 0.5 for all $m \geq 15$ and to be less than 0.5 otherwise. Examination of the p -values generated by $PTT_m(h)$ reveals that at this sample size this criterion will most likely select a window length $m = 10$. Unfortunately, however, the

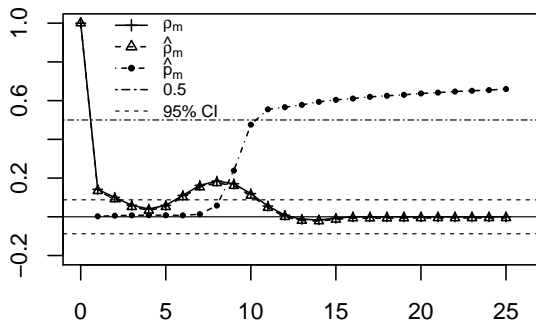


Figure 2: Average p -value $PTT_m(h)$, 10000 replications of Spencer's 15-point moving average processes with $N = 500$.

average value of p_m for $m > 10$ increases monotonically, indicating that $Prob(p_m \leq p) < p$ and $PTT_m(h)$ is undersized.³ The tendency for portmanteau type tests to exhibit low power is well known and this has lead investigators to examine various modifications (Newbold 1981, Section 3). Rather than seeking to modify $PTT_m(h)$, in the following section we will consider a different approach that enables us to consider further reaching adaptations that facilitate application to long range dependent processes.

4 A Conditional Moment Test

Suppose that features of a process x_t are characterized by a parameter $\theta_1 = (\theta_{11}, \dots, \theta_{1p})'$, and we can specify p moment conditions $\mu_1(\theta_1)$ with $E[\mu_1(\theta_1)] = \mathbf{0}$ that exactly identify θ_1 . A set of q auxiliary moment conditions $\mu_2(\theta_1, \theta_2)$ where $\theta_2 = (\theta_{21}, \dots, \theta_{1q})'$ that we wish to use to test other features of the process are also available. The auxiliary moment conditions are constructed so that $E[\mu_2(\theta_1, \theta_2)] = \mathbf{0}$ and the augmented set of moments

$$\mu^a(\theta) = \begin{bmatrix} \mu_1(\theta_1) \\ \mu_2(\theta_1, \theta_2) \end{bmatrix},$$

where $\theta = (\theta_1', \theta_2')'$, are such that $\partial\mu^a(\theta)/\partial\theta'$ is continuous in θ and $E[\partial\mu^a(\theta)/\partial\theta']$ has full column rank. We will also suppose that there exists a diagonal matrix

$$\mathbf{D}_N = \text{diag}(\mathbf{D}_{1N}, \mathbf{D}_{2N}) = \text{diag}(d_{1N}, \dots, d_{pN}, d_{(p+1)N}, \dots, d_{(p+q)N}),$$

where d_{iN} , $i = 1, \dots, p+q$ are monotonically increasing sequences in N , such that $\mathbf{D}_N\mu^a(\theta) = \Sigma^{\frac{1}{2}}\mathbf{Z}_N + o_p(1)$ where Σ is positive definite and $\mathbf{Z}_N \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I})$. We will refer these conditions collectively as Assumption *MC*.

³The use of Henderson's 13 point moving average smoothing filter with coefficients $\{\theta_0, \theta_1, \dots, \theta_6\} = \{0.240, 0.214, 0.147, 0.066, 0, -0.028, -0.019\}$ produced qualitatively identical outcomes.

Theorem 4 Suppose that there exists a set of sample moments

$$\mathbf{m}^a(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{m}_1(\boldsymbol{\theta}_1) \\ \mathbf{m}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \end{bmatrix}$$

such that $\|\mathbf{m}^a(\boldsymbol{\theta}) - \boldsymbol{\mu}^a(\boldsymbol{\theta})\| \rightarrow 0$ and $\|\partial(\mathbf{m}^a(\boldsymbol{\theta}) - \boldsymbol{\mu}^a(\boldsymbol{\theta}))/\partial\boldsymbol{\theta}'\| \rightarrow 0$ uniformly in $\boldsymbol{\theta}$ as $N \rightarrow \infty$ and $\|\mathbf{D}_N(\mathbf{m}^a(\boldsymbol{\theta}) - \boldsymbol{\mu}^a(\boldsymbol{\theta}))\| = o_p(1)$ where $\boldsymbol{\mu}^a(\boldsymbol{\theta})$ obeys Assumption MC. Set $\boldsymbol{\Sigma}_{2|1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ and let $\widehat{\boldsymbol{\theta}}_1$ denote the method of moments estimator obtained by solving $\mathbf{m}_1(\widehat{\boldsymbol{\theta}}_1) = \mathbf{0}$. Then under the null hypothesis $H_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{20}$

$$QCM = \mathbf{m}_2(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{20})' \mathbf{D}_{2N} \boldsymbol{\Sigma}_{2|1}^{-1} \mathbf{D}_{2N} \mathbf{m}_2(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{20}) \xrightarrow{D} \chi_q^2, \quad (5)$$

whilst under the sequence of local alternative hypothesis $H_{1N} : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{20} + \mathbf{D}_{2N}^{-1} \boldsymbol{\delta}$, $0 < \|\boldsymbol{\delta}\| < \infty$, the distribution of QCM is contiguous to a noncentral Chi-squared distribution with degrees of freedom q and noncentrality parameter λ_N , $\chi_q^2\{\lambda_N\}$, where

$$\lambda_N = \boldsymbol{\delta}' \mathbf{D}_{2N}^{-1} \frac{\partial \mathbf{m}_2'(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2} \mathbf{D}_{2N} \boldsymbol{\Sigma}_{2|1}^{-1} \mathbf{D}_{2N} \frac{\partial \mathbf{m}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2'} \mathbf{D}_{2N}^{-1} \boldsymbol{\delta}.$$

PROOF: Let

$$\boldsymbol{\mu}_{2|1}(\boldsymbol{\theta}) = \boldsymbol{\mu}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) - \mathbf{D}_{2N}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \mathbf{D}_{1N} \boldsymbol{\mu}_1(\boldsymbol{\theta}_1). \quad (6)$$

By the continuous mapping theorem $\mathbf{D}_{2N} \boldsymbol{\mu}_{2|1}(\boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{2|1})$ and the corresponding quadratic form $\boldsymbol{\mu}_{2|1}(\boldsymbol{\theta})' \mathbf{D}_{2N} \boldsymbol{\Sigma}_{2|1}^{-1} \mathbf{D}_{2N} \boldsymbol{\mu}_{2|1}(\boldsymbol{\theta})$ converges to a Chi-squared random variable with q degrees of freedom. Replacing $\boldsymbol{\mu}_1(\boldsymbol{\theta}_1)$ by $\mathbf{m}_1(\boldsymbol{\theta}_1)$ and $\boldsymbol{\mu}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ by $\mathbf{m}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ in (6) gives us

$$\mathbf{m}_{2|1}(\boldsymbol{\theta}) = \mathbf{m}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) - \mathbf{D}_{2N}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \mathbf{D}_{1N} \mathbf{m}_1(\boldsymbol{\theta}_1) \quad (7)$$

and we can infer via Slutsky's theorem that

$$\mathbf{m}_{2|1}(\boldsymbol{\theta})' \mathbf{D}_{2N} \boldsymbol{\Sigma}_{2|1}^{-1} \mathbf{D}_{2N} \mathbf{m}_{2|1}(\boldsymbol{\theta}) \xrightarrow{D} \chi_q^2 \quad (8)$$

since $\|\mathbf{D}_{2N}(\mathbf{m}_{2|1}(\boldsymbol{\theta}) - \boldsymbol{\mu}_{2|1}(\boldsymbol{\theta}))\| = o_p(1)$.

Now, the sample moments $\mathbf{m}_1(\boldsymbol{\theta}_1)$ exactly identify $\boldsymbol{\theta}_1$ and solving $\mathbf{m}_1(\widehat{\boldsymbol{\theta}}_1) = \mathbf{0}$ yields a consistent estimate $\widehat{\boldsymbol{\theta}}_1$ of $\boldsymbol{\theta}_1$. Thus, under the null hypothesis $H_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{20}$ the vector $\widehat{\boldsymbol{\theta}}_0 = (\widehat{\boldsymbol{\theta}}_1', \boldsymbol{\theta}_{20}')'$ provides a consistent estimate of $\boldsymbol{\theta}$. Substituting $\widehat{\boldsymbol{\theta}}_0$ for $\boldsymbol{\theta}$ in (7) and (8) we are therefore lead to the conclusion that under H_0 the quadratic form QCM $\xrightarrow{D} \chi_q^2$.

Applying the mean value theorem (Apostol 1960, Section 6.8) to $\mathbf{m}_2(\widehat{\boldsymbol{\theta}}_1, \cdot)$ gives

$$\mathbf{m}_2(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{20}) = \mathbf{m}_2(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2) + \frac{\partial \mathbf{m}_2(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2^*)}{\partial \boldsymbol{\theta}_2'} (\boldsymbol{\theta}_{20} - \boldsymbol{\theta}_2)$$

where $\|\boldsymbol{\theta}_2^* - \boldsymbol{\theta}_2\| \leq \|\boldsymbol{\theta}_{20} - \boldsymbol{\theta}_2\|$. Substituting $(\widehat{\boldsymbol{\theta}}_1', \boldsymbol{\theta}_2')'$ for $\boldsymbol{\theta}$ in (7) and (8), and noting that $\|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1\| = o_p(1)$ and $\|\boldsymbol{\theta}_{20} - \boldsymbol{\theta}_2\| = \|\mathbf{D}_{2N}^{-1} \boldsymbol{\delta}\| \rightarrow 0$ as $N \rightarrow \infty$ under H_{1N} , we can now infer that

$$\nabla \mathbf{m}_2(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{20})' \mathbf{D}_{2N} \boldsymbol{\Sigma}_{2|1}^{-1} \mathbf{D}_{2N} \nabla \mathbf{m}_2(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{20}) \xrightarrow{D} \chi_q^2,$$

where $\nabla \mathbf{m}_2(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{20}) = \mathbf{m}_2(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{20}) - (\partial \mathbf{m}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) / \partial \boldsymbol{\theta}'_2) \mathbf{D}_{2N}^{-1} \boldsymbol{\delta}$. It follows that under H_{1N} the distribution of QCM is contiguous to

$$\chi_q^2 \left\{ \boldsymbol{\delta}' \mathbf{D}_{2N}^{-1} \frac{\partial \mathbf{m}'_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2} \mathbf{D}_{2N} \boldsymbol{\Sigma}_{2|1}^{-1} \mathbf{D}_{2N} \frac{\partial \mathbf{m}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}'_2} \mathbf{D}_{2N}^{-1} \boldsymbol{\delta} \right\}. \quad \blacksquare$$

REMARK 3: The null hypothesis H_0 is here being implicitly tested against H_{1N} subject to the constraints embedded in the moment conditions $\mathbf{m}_1(\boldsymbol{\theta}_1)$ that define the estimator $\widehat{\boldsymbol{\theta}}_1$ being maintained. It is for this reason that we employ the nomenclature conditional moment test.

REMARK 4: As an example of the need to consider non-standard normalization suppose, at the risk of getting slightly ahead of ourselves, that x_t is a fractionally integrated long memory process with mean μ and index $d \in (0.0, 0.5)$. If $\theta_{11} = \mu$ and the first component of the augmented moments is $\mu_{11}(\boldsymbol{\theta}_1) = N^{-1} \sum_{t=1}^N (x_t - \mu) = \bar{x} - \mu$ then in order for Assumption MC to hold it will be necessary to set $d_{1N} = N^{\frac{1}{2}-d}$ because the re-normalized average $N^{\frac{1}{2}-d}(\bar{x} - \mu) \xrightarrow{D} \mathcal{N}(0, \omega^2)$ where $\omega = \{\sigma\phi(1)\}^2 \Gamma(1-2d)/(1+2d)\Gamma(1+d)\Gamma(1-d)$, see [Hosking \(1996, Theorem 8\)](#).

To construct an explicit formula for the test statistic assume the auxiliary moments are structured such that $\mathbf{D}_{2N} = N^{\frac{1}{2}} \mathbf{I}_q$ and, for ease of notation, denote $\mathbf{m}_2(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{20})$ by $\mathbf{m}_2(\widehat{\boldsymbol{\theta}}_0)$. From Theorem 4 we can conclude that under H_0 the statistic

$$CMS = N \mathbf{m}_2(\widehat{\boldsymbol{\theta}}_0)' [\widehat{\boldsymbol{\Sigma}}_{22} - \widehat{\boldsymbol{\Sigma}}_{21} \widehat{\boldsymbol{\Sigma}}_{11}^{-1} \widehat{\boldsymbol{\Sigma}}_{12}]^{-1} \mathbf{m}_2(\widehat{\boldsymbol{\theta}}_0) \xrightarrow{D} \chi_q^2 \quad (9)$$

for any consistent estimate $\widehat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$. To complete the specification of CMS let us rewrite the moment conditions in the form

$$\mathbf{m}_1(\boldsymbol{\theta}_1) = \frac{1}{N} \sum_{t=1}^N \mathbf{z}_{1t}(\boldsymbol{\theta}_1) \quad \text{and} \quad \mathbf{m}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{1}{N} \sum_{t=1}^N \mathbf{z}_{2t}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$$

where $\mathbf{z}_{1t}(\boldsymbol{\theta}_1)$ and $\mathbf{z}_{2t}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ are suitably defined functions of the parameters and the data. Set $\mathbf{i} = (1, 1, \dots, 1)'$, the $N \times 1$ sum vector, and let

$$\widehat{\mathbf{Z}}'_1 = [\mathbf{z}_{11}(\widehat{\boldsymbol{\theta}}_1), \dots, \mathbf{z}_{1N}(\widehat{\boldsymbol{\theta}}_1)] \quad \text{and} \quad \widehat{\mathbf{Z}}'_{20} = [\mathbf{z}_{21}(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{20}), \dots, \mathbf{z}_{2N}(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{20})].$$

Then by construction

$$\mathbf{m}^a(\widehat{\boldsymbol{\theta}}_0) = \begin{bmatrix} \mathbf{m}_1(\widehat{\boldsymbol{\theta}}_1) \\ \mathbf{m}_2(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_{20}) \end{bmatrix} = \frac{1}{N} \begin{bmatrix} \widehat{\mathbf{Z}}'_1 \mathbf{i} \\ \widehat{\mathbf{Z}}'_{20} \mathbf{i} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{m}_2(\widehat{\boldsymbol{\theta}}_0) \end{bmatrix}.$$

If under H_0 the $\mathbf{z}_t^a(\boldsymbol{\theta}) = [\mathbf{z}_{1t}(\boldsymbol{\theta}_1)', \mathbf{z}_{2t}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)']'$ form a serially uncorrelated (orthogonal) sequence then

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} \widehat{\boldsymbol{\Sigma}}_{11} & \widehat{\boldsymbol{\Sigma}}_{21} \\ \widehat{\boldsymbol{\Sigma}}_{21} & \widehat{\boldsymbol{\Sigma}}_{22} \end{bmatrix} = \frac{1}{N} \sum_{t=1}^N \mathbf{z}_t^a(\widehat{\boldsymbol{\theta}}_0) \mathbf{z}_t^a(\widehat{\boldsymbol{\theta}}_0)' = \frac{1}{N} \begin{bmatrix} \widehat{\mathbf{Z}}'_1 \widehat{\mathbf{Z}}_1 & \widehat{\mathbf{Z}}'_1 \widehat{\mathbf{Z}}_{20} \\ \widehat{\mathbf{Z}}'_{20} \widehat{\mathbf{Z}}_1 & \widehat{\mathbf{Z}}'_{20} \widehat{\mathbf{Z}}_{20} \end{bmatrix}$$

yields a consistent estimate of Σ and in terms of the zed matrices the test statistic in (9) becomes

$$CMS = \mathbf{i}' \widehat{\mathbf{Z}}_{20} [\widehat{\mathbf{Z}}'_{20} \widehat{\mathbf{Z}}_{20} - \widehat{\mathbf{Z}}'_{20} \widehat{\mathbf{Z}}_1 (\widehat{\mathbf{Z}}'_1 \widehat{\mathbf{Z}}_1)^{-1} \widehat{\mathbf{Z}}_1' \widehat{\mathbf{Z}}_{20}]^{-1} \widehat{\mathbf{Z}}'_{20} \mathbf{i}. \quad (10)$$

In this form the statistic is ideally suited for numerical implementation because we can re-write (10) as

$$CMS = \begin{bmatrix} \mathbf{i}' \widehat{\mathbf{Z}}_1, \mathbf{i}' \widehat{\mathbf{Z}}_{20} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{Z}}'_1 \widehat{\mathbf{Z}}_1 & \widehat{\mathbf{Z}}'_1 \widehat{\mathbf{Z}}_{20} \\ \widehat{\mathbf{Z}}'_{20} \widehat{\mathbf{Z}}_1 & \widehat{\mathbf{Z}}'_{20} \widehat{\mathbf{Z}}_{20} \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\mathbf{Z}}'_1 \mathbf{i} \\ \widehat{\mathbf{Z}}'_{20} \mathbf{i} \end{bmatrix},$$

which is N times the coefficient of determination from the regression of \mathbf{i} on $\widehat{\mathbf{Z}}_1$ and $\widehat{\mathbf{Z}}_{20}$.

More generally, the elements of the vector $\mathbf{z}_t^a(\boldsymbol{\theta})$ will not be serially uncorrelated. In this case the covariance matrix of $\sqrt{N} \mathbf{m}^a(\boldsymbol{\theta})$ is

$$\begin{aligned} \Sigma &= \frac{1}{N} \sum_{t=1}^N \sum_{s=1}^N E[\mathbf{z}_t^a(\boldsymbol{\theta}) \mathbf{z}_s^a(\boldsymbol{\theta})'] \\ &= \frac{1}{N} \left(\sum_{t=1}^N E[\mathbf{z}_t^a(\boldsymbol{\theta}) \mathbf{z}_t^a(\boldsymbol{\theta})'] + \sum_{r=1}^{N-1} \sum_{t=r+1}^N E[\mathbf{z}_t^a(\boldsymbol{\theta}) \mathbf{z}_{t-r}^a(\boldsymbol{\theta})' + \mathbf{z}_{t-r}^a(\boldsymbol{\theta}) \mathbf{z}_t^a(\boldsymbol{\theta})'] \right), \end{aligned}$$

which we can estimate using the heteroskedasticity and autocorrelation consistent (*HAC*) estimator

$$\widehat{\Sigma}_{HAC} = \sum_{r=1-N}^{N-1} k\left(\frac{r}{B_N}\right) \widehat{\Gamma}_0^a(r),$$

where

$$\widehat{\Gamma}_0^a(r) = \frac{1}{N} \sum_{t=r+1}^N \mathbf{z}_t^a(\widehat{\boldsymbol{\theta}}_0) \mathbf{z}_{t-r}^a(\widehat{\boldsymbol{\theta}}_0)' \quad r = 0, 1, \dots, N-1,$$

$k(\cdot)$ is a positive semidefinite kernel, such as the Parzen or quadratic spectral kernel, and the bandwidth parameter B_N increases with N such that $B_N/N \rightarrow 0$ as $N \rightarrow \infty$. For a detailed discussion of the properties of *HAC* estimators see [Andrews \(1991\)](#) and [Hansen \(1992\)](#).

By way of illustration, suppose that x_t is a linear regular, short memory process. Let $\boldsymbol{\theta}_1 = (\mu, \gamma(0))'$, $p = 2$, and $\boldsymbol{\theta}_2 = (\rho(m+1), \dots, \rho(m+h))'$, $q = h$. Set

$$\boldsymbol{\mu}_1(\boldsymbol{\theta}_1) = \begin{bmatrix} N^{-1} \sum_{t=1}^N (x_t - \mu) \\ N^{-1} \sum_{t=1}^N (x_t - \mu)^2 - \gamma(0) \end{bmatrix} \quad (11)$$

and

$$\boldsymbol{\mu}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \begin{bmatrix} \frac{\sum_{t=1}^N (x_t - \mu)(x_{t-m-1} - \mu)}{N\gamma(0)} - \rho(m+1) \\ \vdots \\ \frac{\sum_{t=1}^N (x_t - \mu)(x_{t-m-h} - \mu)}{N\gamma(0)} - \rho(m+h) \end{bmatrix}.$$

Then $\boldsymbol{\mu}^a(\boldsymbol{\theta})$ satisfies Assumption *MC*, the required asymptotic normality following from Theorem 2 with $d_{iN} = N^{\frac{1}{2}}$, $i = 1, \dots, p+q$.

Let $\mathbf{m}_1(\boldsymbol{\theta}_1) = \boldsymbol{\mu}_1(\boldsymbol{\theta}_1)$ where $\boldsymbol{\mu}_1(\boldsymbol{\theta}_1)$ is defined as in (11), and set

$$\mathbf{m}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \left[\frac{\sum_{t=m+s+1}^N (x_t - \mu)(x_{t-m-s} - \mu)}{N\gamma(0)} - \rho(m+s) \right]_{s=1, \dots, h}.$$

Then $\widehat{\boldsymbol{\theta}}_1 = (\bar{x}, s_x^2)$ where $\bar{x} = N^{-1} \sum_{t=1}^N x_t$ and $s_x^2 = N^{-1} \sum_{t=1}^N (x_t - \bar{x})^2$, and under the null hypothesis $H_0 : \boldsymbol{\theta}_2 = \mathbf{0}$, i.e. that $\rho(m+s) = 0$ for $s = 1, \dots, h$,

$$\mathbf{m}_2(\widehat{\boldsymbol{\theta}}_0) = \left[\frac{\sum_{t=m+s+1}^N (x_t - \bar{x})(x_{t-m-s} - \bar{x})}{Ns_x^2} \right]_{s=1, \dots, h}.$$

For the zed variables we have $\mathbf{m}_1(\boldsymbol{\theta}_1) = N^{-1} \sum_{t=1}^N \mathbf{z}_{1t}(\boldsymbol{\theta}_1)$ where $\mathbf{z}_{1t}(\boldsymbol{\theta}_1)' = [(x_t - \mu), (x_t - \mu)^2 - \gamma(0)]$, and $\mathbf{m}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = N^{-1} \sum_{t=1}^N \mathbf{z}_{2t}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ where

$$\mathbf{z}_{2t}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)' = \left[\frac{(x_t - \mu)(x_{t-m-1} - \mu)}{\gamma(0)} - \rho(m+1), \dots, \frac{(x_t - \mu)(x_{t-m-h} - \mu)}{\gamma(0)} - \rho(m+h) \right]$$

and it is understood that any element of $\mathbf{z}_{2t}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ containing a time subscript that is not positive is replaced by zero.

A simple manipulation indicates that with this choice of moments the statistic in (9) can be re-written as $N\hat{\boldsymbol{\rho}}'\widehat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\rho}}$ and is equivalent to a portmanteau type statistic similar to $PTT_m(h)$ that employs a different covariance estimate. When CMS is evaluated using $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}_{HAC}$ calculated from $\mathbf{z}_{1t}(\widehat{\boldsymbol{\theta}}_1)$ and $\mathbf{z}_{2t}(\widehat{\boldsymbol{\theta}}_1, \mathbf{0})$ as specified immediately above the statistic will therefore be designated $PM_{HAC}T_m(h)$.

To illustrate the differences in the performance of $PM_{HAC}T_m(h)$ and $PTT_m(h)$ Figure 3 depicts the average value of $\hat{\rho}(m)$ and the average p -value, $p_m = Prob(\chi_h^2 > PM_{HAC}T_m(h))$ with $h = \sqrt{N}$, $N = 500$, computed from the same 10000 realizations of Spencer's 15 point moving average smoothing filter process as used in the construction of Figure 2. For $PM_{HAC}T_m(h)$ we used the Tukey-Hanning kernel with the optimal mean squared error bandwidth $B_N = (2/3)N^{1/5}$. As in Figure 2 $\rho(m)$, $\pm 1.96/\sqrt{N}$ and $p = 0.5$ are also plotted to provide points of reference. Comparison of the p -values in Figures 2 and 3 indicates

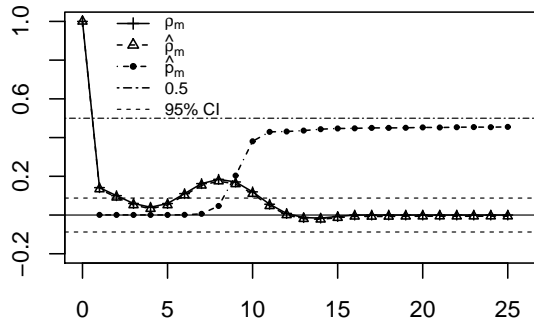


Figure 3: Average p -value $PM_{HAC}T_m(h)$, 10000 replications of Spencer's 15-point moving average processes with $N = 500$.

that $PM_{HAC}T_m(h)$ behaves similarly to $PTT_m(h)$, but $PM_{HAC}T_m(h)$ seems more likely to

choose a value of $m = 11$ rather than 10, and the average value of p_m for $m > 14$ is very stable around 0.5, indicating that $PM_{HAC}T_m(h)$, unlike $PTT_m(h)$, is correctly sized.

5 Long Range Dependence and Testing for Convergence

The class of fractionally integrated $FI(d)$ processes can be characterized by the specification $x_t = \mu + \sum_{j \geq 0} \kappa(j) \varepsilon_{t-j}$ wherein

$$k(z) = \frac{\phi(z)}{(1-z)^d}$$

and $\phi(z) = \sum_{j \geq 0} \varphi(j) z^j$. If $\sum_{j \geq 0} |\varphi(j)| < \infty$ then it can be shown that

$$\kappa(j) \sim \frac{\phi(1)}{\Gamma(d)} j^{d-1} \quad \text{as } j \rightarrow \infty. \quad (12)$$

If $|d| < 0.5$ then $\sum_{j \geq 0} |\kappa(j)|^2 < \infty$ and x_t is well-defined as the limit in mean square of a linearly regular, covariance-stationary process. We note that; $\sum_{j \geq 0} |\kappa(j)| < \infty$ when $d \leq 0$ and x_t is a short memory process with $\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$, $\sum_{j \geq 0} |\kappa(j)| = \infty$ when $d > 0$ and x_t is a long range dependent process with $|\gamma(k)| \sim Ck^{2d-1}$ as $k \rightarrow \infty$. For a more detailed examination of the properties of $FI(d)$ processes see [Beran \(1994\)](#) and [Palma \(2007\)](#).

Long range dependence presents us with a problem in the context of window selection in SSA for two basic reasons: First, $PTT_m(h)$ and $PM_{HAC}T_m(h)$ both test the significance of the sample autocorrelations $\hat{\rho}(k)$ for $k > m$. As pointed out in [Hosking \(1996\)](#), for long memory processes the bias of the sample autocorrelations can be substantial – a feature that is clearly illustrated in [Figure 5](#) below. This bias decays very slowly as the sample size increases and is likely to impact adversely on the performance of both test statistics even in moderate to large samples; Second, and more critically, if x_t is long range dependent then [Theorem 2](#) does not hold, nor do the conditions for application of [Theorem 4](#) when $d > 0.25$, and neither theorem can be simply reinstated. In particular, the usual standardization of $\hat{\gamma}(k) - \gamma(k)$ by $N^{1/2}$ to achieve asymptotic normality does not work for every $d \in (0.0, 0.25]$, and when $d \in (0.25, 0.5)$ the cumulants of $N^{1-2d}(\hat{\gamma}(k) - \gamma(k))$ convergence to those of a Rosenblatt distribution ([Hosking 1996](#), [Theorem 4](#)). So the asymptotic theory upon which the previous tests are based breaks down irretrievably.

To overcome these problems we propose selecting a value of m by devising a stochastic convergence criterion that does not test the significance of the sample autocorrelations. Interestingly enough, [Hosking](#) has shown that by considering differences in the autocovariances asymptotic Normality can be accomplished with the conventional $N^{1/2}$ standardization.

Theorem 5 ([Hosking 1996](#), [Theorem 5](#)): *Suppose that x_t is a linear regular $FI(d)$ process where $|d| < 0.5$. Let $v(\tau) = \gamma(\tau) - \gamma(0)$ and set $\hat{v}(\tau) = \hat{\gamma}(\tau) - \hat{\gamma}(0)$. Then $\sqrt{N}\{\hat{v}(\tau) - v(\tau)\}$, for $\tau = 1, \dots, h$, have a non-degenerate multivariate Normal limiting distribution with mean*

zero and covariance matrix

$$\mathbf{\Omega} = \left[\frac{1}{2} \sum_{s=-\infty}^{\infty} (\gamma(s) - \gamma(s-k) - \gamma(s+l) + \gamma(s-k+l))^2 + (\eta-3)v(k)v(l) \right]_{k,l=1,\dots,h}$$

Theorem 5 indicates that in order to obtain statistics that are \sqrt{N} -consistent and asymptotically Normal for all d with $|d| < 0.5$ we must use appropriate functions of $v(\tau)$ and $\hat{v}(\tau)$.

Now consider testing that the autocovariance function $\gamma(k)$ has converged. Cauchy's convergence criterion states that for every $\delta > 0$ there exists an m such that $|\gamma(k) - \gamma(l)| < \delta$ for all $k, l > m$. This implies that for all m such that $|\gamma(k)| < 0.5\gamma(0)$, $k > m$, the ratio $v(k)/v(l)$ will be arbitrarily close to one, for

$$\frac{v(k)}{v(l)} = 1 + \frac{\gamma(k) - \gamma(l)}{v(l)}$$

and

$$\left| \frac{\gamma(k) - \gamma(l)}{v(l)} \right| < \frac{2\delta}{\gamma(0)}$$

where δ can be taken arbitrarily small for m sufficiently large. This suggests that we can test the hypothesis $H_0 : |\gamma(k)| < \delta$ for all $k > m$ by examining the deviations of $\hat{v}(m+k)/\hat{v}(m+h+1)$, $k = 1, \dots, h$, from unity, because the ratios $\hat{v}(m+k)/\hat{v}(m+h+1)$ are continuous functions of the $\hat{v}(\tau)$ of Theorem 5 and are therefore asymptotically Normal with mean $v(m+k)/v(m+h+1)$ and an asymptotic variance-covariance that can be evaluated by the delta method, and as we have just shown under H_0 we have $v(m+k)/v(m+h+1) \approx 1$.

To apply the conditional moment principle set

$$\boldsymbol{\mu}^a(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\mu}_1(\boldsymbol{\theta}_1) \\ \boldsymbol{\mu}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \end{bmatrix} = \frac{1}{N} \sum_{t=1}^N \begin{bmatrix} x_t - \mu \\ (x_t - \mu)(x_{t-m-h-1} - x_t) - v(m+h+1) \\ \frac{(x_t - \mu)(x_{t-m-1} - x_t)}{v(m+h+1)} - \frac{v(m+1)}{v(m+h+1)} \\ \vdots \\ \frac{(x_t - \mu)(x_{t-m-h} - x_t)}{v(m+h+1)} - \frac{v(m+h)}{v(m+h+1)} \end{bmatrix}$$

wherein $\boldsymbol{\theta}_1 = (\mu, v(m+h+1))'$, $p = 2$, and $\boldsymbol{\theta}_2 = v(m+h+1)^{-1}(v(m+1), \dots, v(m+h))$ with $q = h$. Then $\boldsymbol{\mu}^a(\boldsymbol{\theta})$ satisfies Assumption *MC* with $d_{1N} = N^{\frac{1}{2}-d}$ and $d_{iN} = N^{\frac{1}{2}}$, $i = 2, \dots, p+q$ (Hosking 1996, Theorems 5 and 8). The corresponding sample moments are

$$\mathbf{m}^a(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{m}_1(\boldsymbol{\theta}_1) \\ \mathbf{m}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \end{bmatrix} = \frac{1}{N} \sum_{t=1}^N \begin{bmatrix} \mathbf{z}_{1t}(\boldsymbol{\theta}_1) \\ \mathbf{z}_{2t}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \end{bmatrix}$$

where

$$\mathbf{z}_{1t}(\boldsymbol{\theta}_1)' = [x_t - \mu, (x_t - \mu)(x_{t-m-h-1} - x_t) - v(m+h+1)] \quad (13)$$

and

$$\mathbf{z}_{2t}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)' = \left[\frac{(x_t - \mu)(x_{t-m-1} - x_t)}{v(m+h+1)} - \frac{v(m+1)}{v(m+h+1)}, \dots, \frac{(x_t - \mu)(x_{t-m-h} - x_t)}{v(m+h+1)} - \frac{v(m+h)}{v(m+h+1)} \right], \quad (14)$$

with $x_{t-s} - x_t = \mu - x_t$ whenever $t \leq s$. Let $\hat{\Delta} = \hat{\Sigma}_{22} - \hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}$ where $\hat{\Sigma} = \hat{\Sigma}_{HAC}$ is calculated using $\mathbf{z}_t^a(\hat{\theta}_0) = [\mathbf{z}_{1t}(\hat{\theta}_1)', \mathbf{z}_{2t}(\hat{\theta}_1, \theta_{20})]'$ with $\mathbf{z}_{1t}(\theta_1)$ and $\mathbf{z}_{2t}(\theta_1, \theta_2)$ defined as in (13) and (14), and $\hat{\theta}_1 = (\bar{x}, \hat{v}(m+h+1))'$ and $\theta_{20} = \mathbf{i}' = (1, \dots, 1)$. Appeal to Theorem 4 now yields the following result.

Proposition 2 Let $\hat{\psi}'_m = (\hat{\psi}_m(1), \dots, \hat{\psi}_m(h))$ where $\hat{\psi}_m(k) = \hat{v}(m+k)/\hat{v}(m+h+1) - 1$ and set $CM_{HAC}T_m(h) = N\hat{\psi}'_m\hat{\Delta}^{-1}\hat{\psi}_m$. Suppose that x_t is a linear regular $FI(d)$ process with $|d| < 0.5$ that is driven by i.i.d. $(0, \sigma^2, \eta)$ innovations. Then under the null hypothesis $H_0 : \theta_2 = \mathbf{i}$ the statistic $CM_{HAC}T_m(h) \xrightarrow{D} \chi_h^2$, and under the Pitman sequence $H_{1N} : \theta_2 = \mathbf{i}' + N^{-\frac{1}{2}}\delta$ $CM_{HAC}T_m(h) \xrightarrow{D} \chi_h^2\{\delta'\Delta^{-1}\delta\}$ where $\Delta = \text{plim}_{N \rightarrow \infty} \hat{\Delta}$.

As previously, we can now contemplate calculating $CM_{HAC}T_m(h)$ for a sequence of values $m < N/2$ and choosing for our window length the first value of m for which H_0 is not rejected.

Since $CM_{HAC}T_m(h)$ is applicable to both short and long memory processes we present in Figure 4 the counterpart to Figures 2 and 3, with $CM_{HAC}T_m(h)$ calculated using the Tukey–Hanning kernel as previously. The behaviour of $CM_{HAC}T_m(h)$ seen here closely resembles

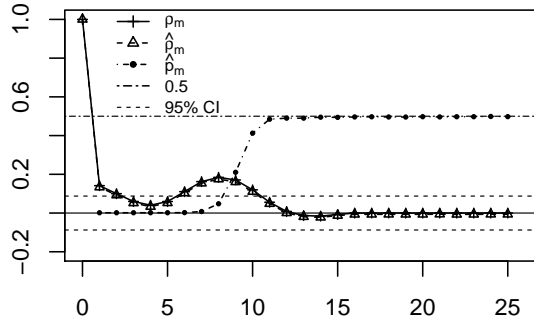


Figure 4: Average p -value $CM_{HAC}T_m(h)$, 10000 replications of Spencer’s 15-point moving average processes with $N = 500$.

that of $PM_{HAC}T_m(h)$ as seen in Figure 3. The interesting feature of this graph is that the observed p -values are in close accord with Proposition 2, they indicate that $p_m = \text{Prob}(\chi_h^2 > CM_{HAC}T_m(h))$ is such that under H_{1N} $\text{Prob}(p_m \leq p) \gg p$ when $m \leq 8$, and as the value of $\delta'\Delta^{-1}\delta$ falls as m increases $\text{Prob}(p_m \leq p)$ decreases until $\text{Prob}(p_m \leq p) = p$ for all $m \geq 15$ under H_0 .

To further explore the behaviour of $CM_{HAC}T_m(h)$ data from an $ARFIMA(1, d, 0)$ process was also examined. The data generating mechanism was

$$x_t - \alpha x_{t-1} = \nu_t = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)\varepsilon_{t-j}}{\Gamma(j+1)\Gamma(d)} \quad (15)$$

where ε_t is i.i.d $\mathcal{N}(0, 1)$. Three specifications for the fractional noise ν_t are considered with the differencing parameter set to $d = 0, 0.2, 0.4$, and for each specification α was set equal to 0.0, 0.2, 0.4, 0.6 and 0.8. Figure 5 depicts $\rho(m)$, the average value of $\hat{\rho}(m)$ and the average

p -value, $p_m = \text{Prob}(\chi_h^2 > CM_{HAC}T_m(h))$, when $h = \sqrt{N}$, the averages being computed from 10000 realizations of the process in (15) with $N = 500$.

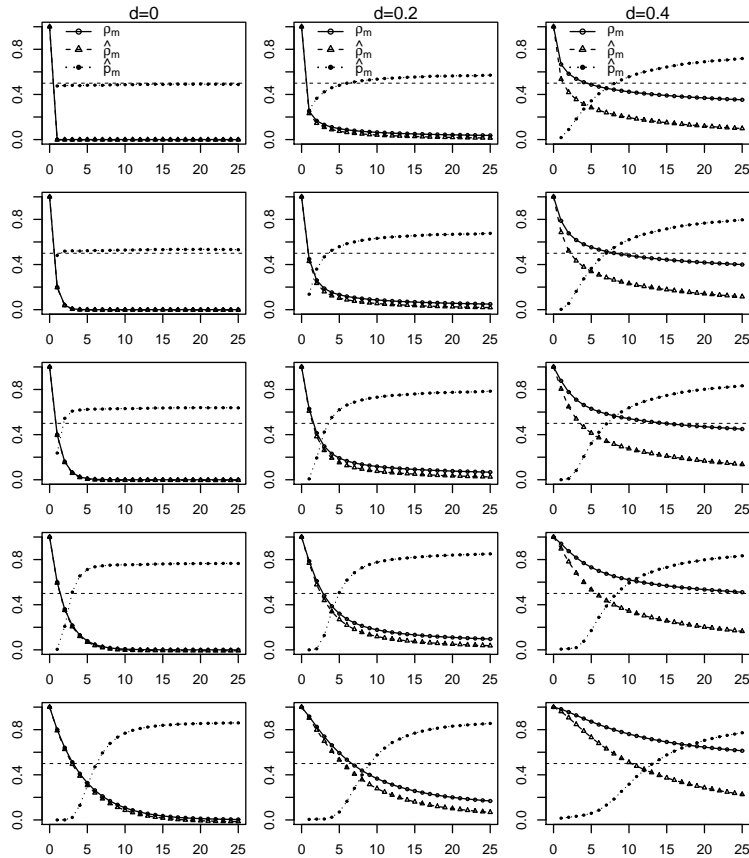


Figure 5: Average p -value of $CM_{HAC}T_m(h)$ from $ARFIMA(1, d, 0)$ process with $\alpha = 0.0, 0.2, 0.4, 0.6, 0.8$ (from top to bottom row) and $d = 0.0, 0.2, 0.4$ (from left to right).

A significant feature that emerges from the panels in this figure is that the finite sample properties of the statistic are not uniform in either d or α . When $d = \alpha = 0$ the average p -value remains close to 0.5 for all m , as it should if $\text{Prob}(p_m \leq p) = p$. When $d \neq 0$ and/or $\alpha \neq 0$, however, the average p -value suggests that $\text{Prob}(p_m \leq p) \gg p$ when m is small and $\text{Prob}(p_m \leq p) \ll p$ when m is large, the transition between the two through the point where $\text{Prob}(p_m \leq p) = p$ being dependent upon the values of d and α . Such behaviour parallels that observed in conventional hypothesis testing and in general we can anticipate that; when m is small and convergence has not yet taken place, i.e. H_0 is false, the distribution of p_m will be positively skewed and p -values will more likely to be small and close to zero; whereas, when m is sufficiently large and H_0 approximately holds p_m will be nearly uniformly distributed in the unit interval; and values near one will become more likely than values near zero as the distribution of p_m becomes negatively skewed as m increases and moves deeper into the set where H_0 is true.

Evidence for the latter is given in Figure 6 which graphs the observed distribution of the p -values computed from $CM_{HAC}T_m(h)$ when $d = 0$ and $\alpha = 0.8$. The distribution is positively skewed at $m = 3$, symmetrically folded around the uniform reference distribution when $m = 6$, and negatively skewed at $m = 12$. That the distribution exhibits such extreme

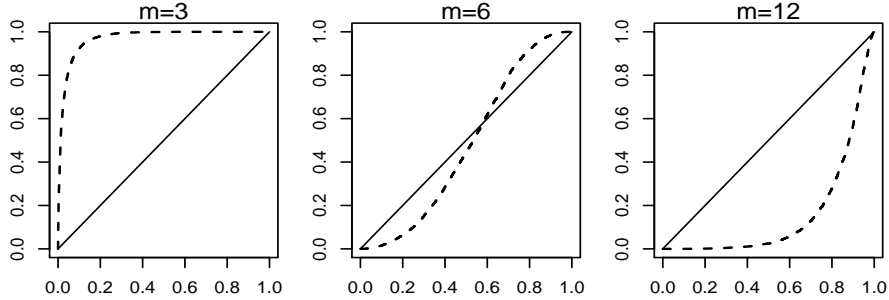


Figure 6: Distributions of $p_m = \text{Prob}(\chi_h^2 > CM_{HAC}T_m(h))$ obtained from AR(1) process $x_t = \alpha x_{t-1} + \epsilon_t$ with $\alpha = 0.8$, for $h = \sqrt{N}$, $N = 500$ and $m = 3, 6$ and 12 .

negative skewness for such a large α is perhaps a somewhat counter intuitive feature given that $\gamma(k) \rightarrow 0$ as $k \rightarrow \infty$ at a rate that decreases as α increases. An examination of the properties of $\hat{\psi}_m(k)$ presented in the Appendix indicates, however, that the expected value is very close to zero for increasing m uniformly in α , but the variance is increasing in both m and α , features that tend to deflate the value of $\hat{\psi}_m' \hat{\Delta}^{-1} \hat{\psi}_m$ and thereby skew the distribution of the statistic.

The Whitney embedding theorem states that any smooth k -dimensional manifold with $k > 0$ (that is also Hausdorff and second-countable) can be smoothly embedded in \mathbb{R}^{2k} . For a short-memory stationary AR(h) process this corresponds to the fact that knowledge of μ and $\gamma(0), \dots, \gamma(2h - 1)$ is sufficient, via the Yule-Walker equations, to determine the autoregressive coefficients and the innovation variance and thus completely characterize the structure of the process. Thus when $d = 0$ and $\alpha \neq 0$ the p -values presented in Figure 5 suggest that $CM_{HAC}T_m(h)$ will select $m \geq 2$, a window length consistent with the Whitney embedding theorem.

When $d \neq 0$ the arguments of the previous two paragraphs do not apply directly, but extension to the fractional case follows by analogy. In particular, if we consider AR(h) approximations to x_t then we find that the window length selected by $CM_{HAC}T_m(h)$ satisfies the condition that $m \geq 2h$ where h is such that the mean squared error of the AR(h) approximation exceeds the innovation variance by less than 3%. So once again $CM_{HAC}T_m(h)$ is selecting a window length in line with the Whitney embedding theorem.¹

6 Empirical Application

To illustrate how the convergence test might be applied in practice we examined monthly observations on the Southern Oscillation Index (SOI) for the period January 1876 to December 2010 inclusive. The data, constructed by the Australian Meteorological Office, can be downloaded from <http://www.bom.gov.au/climate/current/soihtm1.shtml>. No obvious patterns emerge from simple visual inspection of the series, the values appear to fluctuate more or less randomly around zero in a manner not too dissimilar from that seen in Figure 1. We would like to be able to determine if this apparently erratic behaviour disguises a

more systematic signal and for the purposes of SSA a division into yearly observations on annual cycles seems natural, because year-to-year variations in the level of SOI are thought to be very influential in determining annual weather patterns – El Niño (drought) years and La Niña (precipitate) years – and these changes in weather patterns may reflect fluctuations in the level of the underlying signal. This implies a window length of 12 and an obvious question to ask is if such a choice is supported by the data.

Figure 7 graphs the SACF, $\hat{\rho}(m)$, the empirical p -values computed from $CM_{HAC}T_m(h)$, \hat{p}_m , and corresponding bootstrap p -values, $\hat{p}_m^{(B)}$. The bootstrap p -values were calculated as the average value of $p_m = Prob(\chi_h^2 > CM_{HAC}T_m(h))$ computed from 10000 bootstrap realizations of the process using the sieve bootstrap method as described in Poskitt (2008). The bootstrap p -values $\hat{p}_m^{(B)}$ stand instead of simulated p -values since, of course, the true data generating mechanism is now unknown.

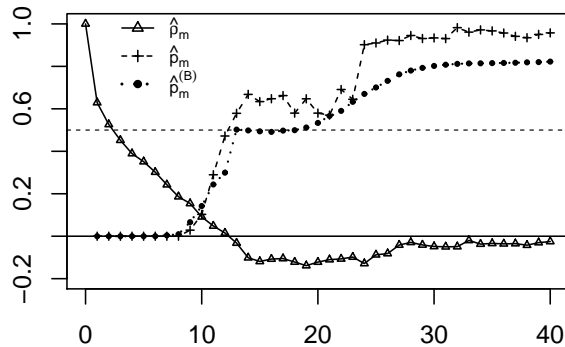


Figure 7: SACF and, observed and bootstrap p -values of $CM_{HAC}T_m(h)$ computed for SOI data.

From Figure 7 we can see that $\hat{\rho}(m)$ and \hat{p}_m mimic the SACF and p -values seen in Figures 4 and 5, the distinctive feature here being the obvious sampling variability apparent in \hat{p}_m for $m > 12$ and the jump that occurs at $m = 24$. The bootstrap p -values $\hat{p}_m^{(B)}$ also seem to behave like a blend of the simulated p -values seen in Figures 4 and 5. Such outcomes have ambiguous implications about the underlying stochastic structure of the series. Nevertheless, both \hat{p}_m and $\hat{p}_m^{(B)}$ clearly indicate that the choice of $m = 12$ for the window length is reasonable, although the data also points to the possible use of $m = 24$, implying division into biennial rather than annual observations.

Rewriting the decomposition in (2) as $\mathbf{X} = \mathbf{S} + \mathbf{N}$ where the signal component $\mathbf{S} = \mathbf{X}_1 + \dots + \mathbf{X}_\kappa$, and employing the description length principle of Khan & Poskitt (2010) to determine κ , leads to the use of $\kappa = 6$ when the window length $m = 12$ and $\kappa = 10$ when the window length $m = 24$. Denoting the associated signal by s_t^κ , the “reconstructed series” in the terminology of SSA (Golyandina et al. 2001), then the root mean squared reconstruction errors, $\sqrt{T^{-1} \sum_{t=1}^T (x_t - s_t^\kappa)^2}$, are 4.01 and 4.48 respectively.

The two signals, SSA(12,6) and SSA(24,10), are graphed in Figure 8a for the period 1980 to 1999.

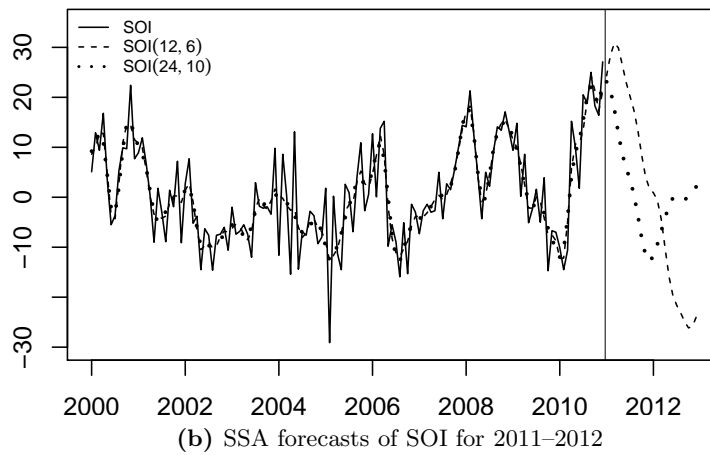
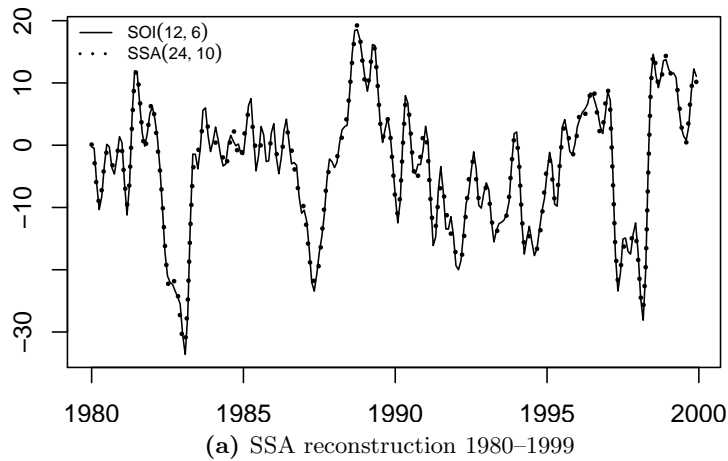


Figure 8: SSA reconstruction & Original Series and Forecasts of SOI data

Treating sustained positive values above +8 as indicative of La Niña events, values between +8 and -8 as indicating generally neutral conditions, and sustained negative values below -8 as being indicative of El Niño periods, we see that historically both signals portend climatic events experienced in Australia in the last century; the major droughts of 1982, 1987, 1991–1994 and 1997, and the record rainfall years of 1988–89 and 1998–1999.

These outcomes are consistent with the idea that aberrant or extreme values of the SOI in particular years are prescient of significant weather events, suggesting that predicting so called "g-phases" (Stone et al. 2000) will be a useful tool in forecasting future El Niño/La Niña effects and their associated weather patterns. Figure 8b graphs the observed series, SSA(12,6) and SSA(24,10) for the period 2000 to 2010, and the forecasts given by SSA(12,6) and SSA(24,10) for 2011 and 2012, constructed using the SSA vector forecasting algorithm of Golyandina et al. (2001, Chapter 2.).

The ability of both signals to track the major fluctuations in the actual series is clear: That Australia recorded its second highest national average rainfall in 2000, the major drought from the early 2000's to around 2008–2009, and the subsequent wet period in 2010, are all events clearly indicated by both SSA(12,6) and SSA(24,10). Both forecasts also give an

indication of the record rainfall and unprecedented floods that actually occurred in Australia at the beginning of 2011. But SSA(12,6) predicts a persistent return to extreme drought conditions by the end of 2012, whereas SSA(24,10) suggests that a brief dry spell in the spring of 2011 will be followed by a return to more neutral conditions in 2012. It is not our purpose to construct a definitive SSA forecasting model at this point, suffice it to say that the sensitivity of the SSA forecast profiles to the choice of m is readily apparent, indicating the importance of employing objectively defined techniques of window length selection of the type considered here.

7 Conclusion

In this paper we presented a new methodology for selecting the window length in SSA based on the use of statistical tests designed to ascertain convergence of the autocovariance function of the observed process. A classical time series portmanteau type statistic and two test statistics derived using a conditional moment principle were considered. We derived the asymptotic distribution of the statistics under fairly general regularity conditions and showed that the criteria will identify true convergence with a finite window length with probability one as the sample size increases. Results obtained using Monte–Carlo simulation indicate that the asymptotic theory is reflected in observed behaviour, even in relatively small samples, and that the conditional moment tests will choose a window length consistent with the Whitney embedding theorem. Finally, the practical relevance of our results were illustrated via a real world data set.

Note:

1. Overall our results suggest that the criteria will favour the selection of window lengths that are orders of magnitude smaller than N . This appears to run counter to the findings of Golyandina (2010), who recommends choosing m close to one half of the series length. These two results are not necessarily incompatible however. Golyandina bases her recommendation upon an investigation of series composed of a deterministic (trigonometric) signal, whereas the results presented here are derived from an analysis of short- and long-memory nonsingular stationary processes.

Acknowledgement

A previous version of this paper was presented at the *International Workshop on SSA and its Applications*, School of Mathematics, University of Cardiff, September 2010. We are grateful to workshop participants, and Dr. Hossein Hassani in particular, for helpful comments.

References

- Alonso, F. J., Castillo, J. M. & Pintado, P. (2005), ‘Application of singular spectrum analysis to the smoothing of raw kinematic signals’, *Journal of Biomechanics* **38**(5), 1085–1092.
- Andrews, D. W. K. (1991), ‘Heteroskedasticity and autocorrelation consistent covariance matrix estimation’, *Econometrica* pp. 817–858.
- Apostol, T. M. (1960), *Mathematical Analysis*, Addison-Wesley, Reading.
- Bartlett, M. (1946), ‘On the theoretical specification and sampling properties of autocorrelated time-series’, *Supplement: Journal of the Royal Statistical Society* **8**(1), 27–41.
- Beran, J. (1994), ‘Statistics for long-memory processes, volume 61 of’, *Monographs on Statistics and Applied Probability* .
- Box, G. E. P. & Pierce, D. A. (1970), ‘Distribution of residual autocorrelations in autoregressive-integrated moving average time series models’, *Journal of the American Statistical Association* **65**(332), 1509–1526.
- Brockwell, P. & Davis, R. A. (1991), *Time Series: Theory and Methods*, 2 edn, Springer-Verlag, New York.
- Elsner, J. B. & Tsonis, A. A. (1996), *Singular Spectrum Analysis: A New Tool in Time Series Analysis*, Plenum Press, New York.
- Fuller, W. A. (1996), *Introduction to Statistical Time Series*, 2 edn, Wiley-Interscience, New Jersey.
- Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F. & Yiou, P. (2002), ‘Advanced spectral methods for climatic time series’, *Reviews of Geophysics* **40**(1), 1003.
- Golyandina, N. (2010), ‘On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods’, *Statistics and Its Interface* **3**(3), 259–279.
- Golyandina, N., Nekrutkin, V. V. & Zhigljavski, A. A. (2001), *Analysis of Time Series Structure: SSA and Related Techniques*, CRC Press.
- Hannan, E. J. & Heyde, C. C. (1972), ‘On limit theorems for quadratic functions of discrete time series’, *The Annals of Mathematical Statistics* **43**, 2058–2066.
- Hansen, B. E. (1992), ‘Consistent covariance matrix estimation for dependent heterogeneous processes’, *Econometrica* **60**(4), 967–972.
- Hassani, H. & Zhigljavsky, A. (2009), ‘Singular Spectrum Analysis: methodology and application to economics data’, *Journal of Systems Science and Complexity* **22**(3), 372–394.
- Hosking, J. R. M. (1996), ‘Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series’, *Journal of Econometrics* **73**(1), 261–284.

- Ibragimov, I. A. & Linnik, Y. V. (1971), *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff Publishing, Groningen.
- Khan, M. A. R. & Poskitt, D. S. (2010), ‘Description length based signal detection in singular spectrum analysis’, *Monash Econometrics and Business Statistics Working Papers* **13/10**.
- Marques, C. A. F., Ferreira, J. A., Rocha, A., Castanheira, J. M., Melo-Gonçalves, P., Vaz, N. & Dias, J. M. (2006), ‘Singular spectrum analysis and forecasting of hydrological time series’, *Physics and Chemistry of the Earth* **31**(18), 1172–1179.
- Newbold, P. (1981), ‘Some recent developments in time series analysis’, *International Statistical Review* **49**(1), 53–66.
- Palma, W. (2007), *Long-memory Time Series: Theory and Methods*, Wiley-Blackwell.
- Poskitt, D. (2008), ‘Properties of the sieve bootstrap for fractionally integrated and non-invertible processes’, *Journal of Time Series Analysis* **29**(2), 224–250.
- Slutzky, E. (1937), ‘The summation of random causes as the source of cyclic processes’, *Econometrica* **5**(2), 105–146.
- Stone, R., Smith, I. & McIntosh, P. (2000), *Statistical methods for deriving seasonal climate forecasts from GCM’s*, chap. 10, In, *Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems: The Australian Experience*, Kluwer Academic, Netherlands.
- Tzagkarakis, G., Papadopouli, M. & Tsakalides, P. (2009), ‘Trend forecasting based on singular spectrum analysis of traffic workload in a large-scale wireless LAN’, *Performance Evaluation* **66**(3-5), 173–190.