Estimation methods for stationary Gegenbauer processes

Richard Hunt¹ · Shelton Peiris¹ · Neville Weber¹

Received: 16 July 2021 / Revised: 21 December 2021 / Accepted: 15 January 2022 /

Published online: 14 February 2022

© The Author(s) 2022

Abstract

This paper reviews alternative methods for estimation for stationary Gegenbauer processes specifically, as distinct from the more general long memory models. A short set of Monte Carlo simulations is used to compare the accuracy of these methods. The conclusion found is that a Bayesian technique results in the highest accuracy. The paper is completed with an examination of the SILSO Sunspot Number series as collated by the Royal Observatory of Belgium.

Keywords Gegenbauer · GARMA · Whittle · Bayesian · Semi-parametric · SILSO

Mathematics Subject Classification 60G22

1 Introduction

Long memory models have become useful in a large number of areas including hydrology, astronomy, econometrics and finance. One of the first practitioners to appreciate the significance of these models was Hurst (1951, 1956) who studied the flows of the river Nile, and identified the distinction between short and long memory models; specifically noting that

It is commonly assumed that the frequency distribution for river flows approximates to the Normal or Gaussian curve, and this is usually true, but it is only part of the description of the phenomenon, since there is also a tendency for high or low years to be grouped together.

Thus there is a tendency for phenomena (e.g. floods, high inflation levels etc) to occur in runs. If the Nile were to flood at very high levels for year after year, it is likely that a model based on traditional short-memory estimates (which assume flooding in one year is correlated only to the most recent few years of flood levels) would suggest a

✉ Richard Hunt
richard@huntemail.id.au

¹ School of Mathematics and Statistics, University of Sydney, Camperdown, Australia

dam level too low. The likely result of this would be that the dam would overflow after only a few years of the higher flood levels. The work of Hurst was pivotal in helping hydrologists decide to build the high dam at Aswan much higher than had previously been planned. A summary of Hurst's life and work may be found in Sutcliffe et al. (2016).

This paper will deal primarily with a subclass of long memory models, known as Gegenbauer models, which are formally defined in Sect. 2. Gegenbauer models provide long memory in 'cycles' and this property has enabled them to be successfully applied to a wide range of phenomena, including modeling the El-Niño effect, where Lustig et al. (2017) found evidence of a 2-factor Gegenbauer process driving the long-term memory in this climate phenomena, and also Diongue and Guégan (2008) examined electricity spot prices using a Gegenbauer-GARCH process (introduced formally below).

A short introduction to the area of long memory models in general, including the 'standard' ARFIMA model and also Gegenbauer models in particular was given by Holan et al. (2010), in Section 6, and a more detailed one is provided in Woodward et al. (2017), Ch 11. Other excellent and detailed references for long memory models specifically (but with only a short mention of Gegenbauer models) are Beran et al. (2013) and Palma (2006), both of whom also cover in detail the estimation of long memory ARFIMA processes.

Section 2 will provide a very brief introduction to Gegenbauer processes and also cover a few further variations on these models. Section 3 details a variety of estimation methods for Gegenbauer models. Section 4 illustrates the estimation methods using Monte-Carlo simulations to compare performance and Sect. 5 examines an application of these models to updated SILSO sunspot data. Section 6 concludes the paper.

2 The Gegenbauer seasonal process and variants

2.1 Box–Jenkins seasonality and the general seasonal model

A Box–Jenkins (Box and Jenkins 1976) style of seasonal model has a general form of

$$\Phi(B; s)\phi(B)(X_t - \mu) = \Theta(B; s)\theta(B)\epsilon_t,$$

where $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ is a polynomial of order p in z , $\theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q$ is a polynomial of order q in z , $\Phi(z; s) = 1 - \Phi_1 z^s - \dots - \Phi_P z^{Ps}$ is a polynomial of order P in z^s , and $\Theta(z; s) = 1 - \Theta_1 z^s - \dots - \Theta_Q z^{Qs}$ is a polynomial of order Q in z^s , with $s \in \mathbb{N}$ representing the seasonality or periodicity of the model (this paper will use both terms interchangeably), μ is the mean of the stationary process, B is the backward shift operator defined as $BX_t = X_{t-1}$, and ϵ_t is a set of uncorrelated zero mean innovations with constant variance σ^2 , possibly non-Gaussian.

The above model may be re-written in a more general form as

$$\Psi(B)\phi(B)(X_t - \mu) = \theta(B)\epsilon_t, \quad (1)$$

where

$$\Psi(B; s, \Phi, \Theta) = \frac{\Phi(B; s)}{\Theta(B; s)}. \quad (2)$$

The Box-Jenkins seasonality model defined by (1) and (2) provides an absolutely summable autocorrelation function (ACF)—equivalently a continuous and bounded spectral density—and is thus a short memory process. Many standard results hold for this model—for example Mann and Wald (1943) provides consistency and central limit theorem results for maximum likelihood-derived parameter estimates and Hannan (1973) provides similar results for the Whittle likelihood parameter estimates.

The form of (1) allows for alternative forms of seasonality. One of those as detailed in Sect. 2.2 is Gegenbauer seasonality. For this reason (1) is referred to as a general seasonal model.

It should be noted that (1) can be extended to include k seasonal factors by repeating the factor (2) k times, with different periodicities $s_1, \dots, s_k$.

Finally note that if $\Phi(B; s) \equiv 1$ then (1) is generally referred to as an ARMA (or autoregressive, moving average) model.

2.2 Gegenbauer seasonality

Adenstedt (1974) originally introduced the concept of processes based on Gegenbauer polynomials in a general form, with the specific details of applying (1) developed further by Gray et al. (1989). These authors introduced a type of seasonal model, defined with just 2 parameters and using the generating function for the Gegenbauer (or hyper-spherical polynomials (Rainville 1960; Szegő 1959)) as

$$\Psi(B) \equiv \Psi(B; u, d) = (1 - 2uB + B^2)^d, \quad (3)$$

where $|u| \leq 1$. For this model, define $\lambda = \arccos(u)$ and the period is thus $s = 2\pi/\lambda$. It is also noted that the periodicity here may be fractional (unlike Box-Jenkins seasonality) and secondly that the periodicity can be estimated directly from the likelihood function without needing to do this as a part of an identification step as in traditional Box-Jenkins models (Box and Jenkins 1976). The Gegenbauer frequency is sometimes written as $f_G = 1/s$.

A full k -factor Gegenbauer model is defined as

$$\phi(B) \prod_{j=1}^k \Psi(B; u_j, d_j)(X_t - \mu) = \phi(B) \prod_{j=1}^k (1 - 2u_j B + B^2)^{d_j} (X_t - \mu) = \theta(B)\epsilon_t, \quad (4)$$

where u_j , $j = 1, \dots, k$ are distinct real numbers with $|u_j| \leq 1$, and d_j are real numbers, which must satisfy $|d_j| < 1/2$ for the process to be stationary and invertible, unless $u_j = 1$ in which case $|d_j| < 1/4$.

As mentioned above, the factor $(1 - 2uB + B^2)^d$ is the generating function for the Gegenbauer orthogonal polynomials, which can be written as

$$(1 - 2uz + z^2)^{-d} = \sum_{j=0}^{\infty} C_j^{(d)}(u)z^j. \quad (5)$$

Recursion formulae exist for calculating these polynomials—refer to Szegő (1959).

The spectral density of model (4) is given by

$$\begin{aligned} f(\omega) &= \frac{\sigma_\epsilon^2}{2\pi} \left| \frac{\theta(e^{i\omega})}{\phi(e^{i\omega})} \right|^2 \prod_{j=1}^k \left[4(\cos(\omega) - u_j)^2 \right]^{-d_j} \\ &= \frac{\sigma_\epsilon^2}{2\pi} \left| \frac{\theta(e^{i\omega})}{\phi(e^{i\omega})} \right|^2 \prod_{j=1}^k \left[4 \sin\left(\frac{\omega - \lambda_j}{2}\right) \sin\left(\frac{\omega + \lambda_j}{2}\right) \right]^{-2d_j}, \end{aligned} \quad (6)$$

where $\lambda_j = \arccos(u_j)$.

When $k = 1$ there is an alternative formulation of (6), which reflects only the behavior of the function around the unbounded peak when $d > 0$

$$f(\omega) \simeq C |\lambda - \omega|^{-2d} \quad \text{as } \omega \rightarrow \lambda. \quad (7)$$

This formulation is often used to define the regularity conditions for semi-parametric estimators of λ and d (refer Sect. 3.4).

Both (6) and (7) illustrate that the spectral density of a process with Gegenbauer seasonality will in general have k values where there is a discontinuity in the spectral density, and where the function becomes unbounded in a neighborhood of that peak. The peaks correspond to the Gegenbauer frequencies.

In the remainder of this paper, when $k = 1$, the subscript will be dropped from the parameters. Note that it is usual to refer to a $k = 1$ process of the form of (4) as a GARMA(p,d,q;u) process, and this format is used when referring to specific models in Sect. 4.

Finally we define an ARFIMA process from (4) with $k = 1$, $u = 1$, and $\delta = 2d$, giving

$$\phi(B)(1 - B)^\delta X_t = \theta(B)\epsilon_t. \quad (8)$$

2.3 Comparing theoretical ACF of seasonal models

Figure 1 illustrates the ACF for the Box-Jenkins seasonal component defined by (2) (with $P = 1$, $\Phi_1 = 0.8$) and a Gegenbauer seasonal component defined by (3) (with $d = 0.4$). The ACF for the Gegenbauer seasonal component was generated using the approximation formula of Gray et al. (1989) Theorem 3(c). For clarity of interpretation the seasonal period s for both has been set to 24.

Referring to Fig. 1 it is apparent that the Box-Jenkins ACF provides correlation only at the seasonal lags, whereas the Gegenbauer ACF provides some correlation at adjacent lags as well, since it represents a damped sine wave. This might be appropriate for instance when modeling daily data with a (calendar) monthly periodicity. However

Fig. 1 Theoretical ACF for Box Jenkins and Gegenbauer seasonality components

in different circumstances and with different datasets, it is expected that either form might prove to be useful.

A second key feature is perhaps less obvious, but can be seen by observing the height of the ACF at lags 24 and 192—for the Box-Jenkins model the ACF decays quite rapidly, with ACF at lag 192 being only 21% of the value at lag 24. However the Gegenbauer ACF at lag 192 is still 66% of the value at lag 24.

This illustrates the ‘long memory’ component of the Gegenbauer seasonal model. Observations quite far apart in time are still quite highly correlated with each other.

2.4 Variations on Gegenbauer seasonality

One variation on model (4) was introduced by Giraitis and Leipus (1995) which was a fractional ARUMA process. This has 2 unit roots with differing degrees of fractional differencing, and is defined as

$$\phi(B)(1-B)^{d_0} \prod_{j=1}^k (1-2u_j B + B^2)^{d_j} (1+B)^{d_{k+1}} (X_t - \mu) = \theta(B)\epsilon_t. \quad (9)$$

Giraitis and Leipus (1995) showed consistency of the Whittle estimates of the long memory components and the Gegenbauer frequencies of this process. Setting $d_0 = d_{k+1} = 0$ shows this result includes (4).

Let

$$Y_t = \prod_{j=1}^k (1-2u_j B + B^2)^{d_j} (X_t - \mu). \quad (10)$$

Then $f_Y(\omega) = \frac{\sigma_\epsilon^2}{2\pi} \left| \frac{\theta(e^{i\omega})}{\phi(e^{i\omega})} \right|^2$, $-\pi < \omega < \pi$ represents the spectral density of the short memory component of the process (with Y_t representing the short memory process)—also known as the underlying ARMA process, and thus

$$\begin{aligned} f(\omega) &= f_Y(\omega) \prod_{j=1}^k \left[4 (\cos(\omega) - u_j)^2 \right]^{-d_j} \\ &= f_Y(\omega) \prod_{j=1}^k \left[4 \sin\left(\frac{\omega - \lambda_j}{2}\right) \sin\left(\frac{\omega + \lambda_j}{2}\right) \right]^{-2d_j}. \end{aligned} \quad (11)$$

As an alternative to an underlying ARMA process, Bloomfield (1973) introduced the exponential model for short memory processes with a rational spectrum defined by $f(\omega) = \frac{\sigma_\epsilon^2}{2\pi} \frac{|\theta(e^{i\omega})|^2}{|\phi(e^{i\omega})|^2} = \frac{\sigma_\epsilon^2}{2\pi} k(\omega)$, say. Bloomfield noted that $\log k(\omega) = \log \left(\frac{|\theta(e^{i\omega})|^2}{|\phi(e^{i\omega})|^2} \right)$ is “often found to be a well behaved function” and this might then be approximated by a truncated Fourier series $\log \hat{k}(\omega) = \sum_{r=1}^{m_e} \xi_r \cos(r\omega)$, with the intention of using a small value for m_e , resulting in the spectral density for a new class of models defined by

$$f(\omega) = \frac{\sigma_\epsilon^2}{2\pi} k(\omega) = \frac{\sigma_\epsilon^2}{2\pi} \exp \left\{ \sum_{r=1}^{m_e} \xi_r \cos(r\omega) \right\},$$

where the ξ_r are Fourier coefficients.

Beran (1993) extended this formulation to long memory models, labeling the result as the Fractional EXponential or FEXP model, and this was further extended to Gegenbauer models by Hsu and Tsai (2009) (for the $k = 1$ model) and McElroy and Holan (2012) for the general $k > 1$ model. This model is also defined through its spectral density as

$$f(\omega) = \frac{\sigma_\epsilon^2}{2\pi} \prod_{j=1}^k \left[4 (\cos(\omega) - u_j)^2 \right]^{-d_j} \exp \left\{ \sum_{r=1}^{m_e} \xi_r \cos(r\omega) \right\}, \quad (12)$$

where m_e indicates the truncation of the approximation to the short memory process, and ξ_r , $r = 1, \dots, m_e$ are the short memory parameters to be estimated.

ARFIMA models defined by (8) have been extended to include conditional heteroscedasticity. The fractionally integrated (long memory) GARCH (FIGARCH) model was introduced by Baillie et al. (1996) and Bollerslev and Mikkelsen (1996). Ling and Li (1997) extended these models to cover the GARCH(r, s) error structure where the ϵ_t in (8) are modeled as

$$\begin{aligned} \epsilon_t &= e_t \sqrt{h_t} \\ h_t &= \alpha_0 + \sum_{l=1}^r \alpha_l \epsilon_{t-l}^2 + \sum_{j=1}^s \beta_j h_{t-j} \\ \alpha_0 &> 0, \alpha_l \geq 0 \quad \forall l = 1, \dots, r, \beta_j \geq 0 \quad \forall j = 1, \dots, s, \end{aligned} \quad (13)$$

with $e_t | \mathcal{F}_{t-1} \sim N(0, 1)$, where $\mathcal{F}_t$ denotes the sigma-field spanned by $(X_1, \dots, X_{t-1}, X_t)$. Note that some authors assume a simpler model for the GARCH($r, 0$) process, labeled as ARCH(r) when $s = 0$. This is distinct from a stochastic volatility model as defined by Phillip et al. (2018) (for instance) following from Bos et al. (2014), who introduced the GARFIMA-SV model, where the error structure is given as

$$\begin{aligned} \epsilon_t &\sim N(0, e^{h_t}) \\ h_t &= \alpha + \beta(h_{t-1} - \alpha) + \eta_t, \quad \eta_t \sim N(0, \sigma^2). \end{aligned} \quad (14)$$

This defines $h_t | \mathcal{F}_{t-1}$ (and thus the magnitude/scaling of ϵ_t) as stochastic, which is not the case with a GARCH model.

Dissanayake and Peiris (2011); Dissanayake et al. (2014) and later Boubaker (2015) introduced the GARMA-GARCH variation, which applies the disturbance term defined by (13) to (4). Dissanayake et al. (2014) used a truncated maximum likelihood technique to estimate these models, and Boubaker (2015) used a wavelet based method to estimate the Whittle likelihood (refer Sect. 3.2).

Other variations on the basic process (4), mentioned here but otherwise not followed up include:

1. Artech (1998) introduced the asymmetric Gegenbauer processes. The interested reader may consult Artech (1998) and also Artech (2007) the latter of which describes an application of this model to a monthly Spanish inflation process.
2. Bordignon et al. (2007) introduced a variation on (13) so that the process variance (rather than the underlying process) follows a GARMA definition. The idea is that since a Gegenbauer process can model seasonal long memory behavior, so there can be a minimal set of frequencies to be included in the model. Bordignon et al. (2007) give an example of the log returns for the Euro-Dollar exchange rate, with $k = 4$.
3. Lu and Guegan (2011) introduced a version of (4) which allowed for a time varying fractional differencing parameter $d_j = d_j(t/n)$ which can be considered as locally stationary. Lu and Guegan (2011) introduced a conditional method for estimating the parameters of these models.
4. Espejo et al. (2014) presented a 2-dimensional spatial version of the Gegenbauer process—a Gegenbauer random field, although without a moving average short-memory component. These can be used for instance for spatial modeling of climate phenomena—for example, see Rypdal et al. (2015) for a pure long memory example. Further, Espejo et al. (2015) presents a class of minimum contrast estimators for these Gegenbauer random fields and establishes consistency and asymptotic normality results for those estimators.
5. Li et al. (2020) introduced long memory functional time series, including a ARFIMA process as a special case. A number of key theorems were provided as well as some examples.
6. Wu and Peiris (2018) introduced a vector (dimension m , $k = 1$) cointegrated Gegenbauer process, which they labeled as VEGARMA, which allowed cointegration only for the short memory components of the model. They derived a

quasi-maximum likelihood estimation procedure for this model. They do not examine any theoretical properties of these estimators. This was used to model a vector process consisting of 669 consumer price index (CPI) figures from France and the USA.

3 Parameter estimation

Whilst a natural approach for many practitioners might be to evaluate a maximum-likelihood function to obtain parameter estimates, in the case of long memory and Gegenbauer models, this can amount to a considerable investment in time and computer resources. It is also true that the assumptions behind such a model may not be satisfied. Note that the maximum likelihood estimation of just 1 of the simulations from Sect. 4 (below) using a Kalman filter to evaluate the likelihood as documented in Palma and Chan (2005) took more than 11 hours on the same hardware as the other simulations were run typically taking less than 1 second. Thus alternative methods to estimate the values of the parameters have become popular.

There are 2 main approaches taken to the estimation problem. One approach is first to establish the order of the model (the model identification step—to determine appropriate values of k , p , q), and then estimate the parameters based on a range of techniques like maximum-likelihood, approximate maximum-likelihood, quasi maximum-likelihood, conditional sum-of-squares or Whittle.

The second method is to estimate the long memory parameters first. This technique was originally used to allow the original series X_t to be transformed to Y_t via the relation (10)—which is an approximate process only—and then use techniques such as those described in Box and Jenkins (1976) to estimate the remainder of the parameters (refer for instance to Geweke and Porter-Hudak (1983)). Whilst this technique provides conditional parameter estimates, they are fast to compute and further can often be used as starting values in a numerical optimization process for other methods.

More practically, these latter techniques have been used to estimate the long memory components without specific knowledge of the structure of the short memory components.

For instance if a single unbounded peak can be identified in the spectrum of a given process, then these techniques can be used to identify the location (u) and sharpness (d) of that peak, allowing (10) to be used to generate an ARMA process which can then be estimated via traditional means—see for example Box and Jenkins (1976).

Define ν to represent the vector of parameters to be estimated, which may vary from occurrence to occurrence based on the underlying model structure. For a given sample of n observations of the process X_t available, $t = 1, \dots, n$, define $\omega_j = \frac{2\pi j}{n}$, and $w_n(\omega) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n X_t e^{it\omega}$, where the mean is omitted since $\sum_{t=1}^n e^{it\omega_j} = 0$, for all $j = 1, \dots, n$.

The periodogram of (4) is $I(\omega) = |w_n(\omega)|^2$. Further, define $\tilde{n} = [n/2]$ where the square brackets indicate the largest integer which is less than or equal to the argument. $\psi(x)$ denotes the di-gamma function.

3.1 Likelihood-related methods

The Gaussian log-likelihood function for (4) is

$$\mathcal{L}_G(X; v) = -1/2 \log |\Sigma(v)| - 1/2 (X - \mu 1_n)^T \Sigma^{-1}(v) (X - \mu 1_n), \quad (15)$$

where X represents the vector of realizations and 1_n represents an $n \times 1$ vector of 1's, and the covariance matrix of the process X_t is $\Sigma \equiv \Sigma(v)$ which is assumed to be a strictly positive definite, $n \times n$ matrix, and where v represents the vector of parameters. Further define $\hat{X}_t$ as the fitted value of the model and $e_t = X_t - \hat{X}_t$ as the residuals.

A key issue with strict likelihood based estimation techniques is the need to invert the covariance matrix $\Sigma(v)$. There have been many proposals to identify methods to avoid or approximate this.

The original papers by Gray et al. (1989, 1994) include suggested details on estimation—a grid search on an approximate log-likelihood—actually a conditional sum-of-squares (CSS) approximation—to identify the long memory parameters d and u and then transform the original process to a short memory ARMA process via (10).

Chung (1996a) provided a central limit theorem for a Gegenbauer $k = 1$ white noise process estimated via a CSS method, and Chung (1996b) examined a $k = 1$ Gegenbauer process defined by (4), including short-memory ARMA parameters as well as the long memory parameters. The CSS method may be illustrated by considering Brockwell and Davis (1991), equations 8.6.6 and 8.6.7 which show that, if $v_t = E(e_{t+1}^2) = E(X_{t+1} - \hat{X}_{t+1})^2$ then

$$\begin{aligned} (X - \mu 1_n)^T \Sigma(v)^{-1} (X - \mu 1_n) &= \sum_{t=1}^n e_t^2 / v_t \\ \det(\Sigma(v)) &= \prod_{t=1}^n v_t. \end{aligned}$$

If we write $\bar{v} = (\prod_{t=1}^n v_t)^{1/n}$, then the resulting log-likelihood approximation is

$$\mathcal{L}_{CSS}(\omega; v) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log \bar{v} - \frac{1}{2} \sum_{t=1}^n e_t^2. \quad (16)$$

Whilst Chung (1996a, b) derives distributional characteristics for the estimators obtained by minimizing this, these arguments were refuted by Giraitis et al. (2001), in part because the proof of consistency was not adequate.

Robinson (2006) provided a consistency result for the $k = 1$ CSS estimates.

Smallwood and Beaumont (2004) examined the full k -factor model defined by (4) and derived results very similar to those of Chung (1996b). Further work was presented in Beaumont and Smallwood (2019), establishing the asymptotic independence of the u_j , $j = 1, \dots, k$ from the other parameters in the model; the asymptotic Gaussian distribution for the parameters excepting the u_j , and a complex asymptotic distribution involving independent Brownian motions for the u_j . However the literature does not

appear to provide a result which establishes consistency of these estimators in the $k > 1$ case.

In a related paper, Beaumont and Smallwood (2020) examined a variety of methods for testing a null hypothesis of no cyclic behavior ($u = 1$) against the alternative that there is cyclic behavior. They examined the methods provided by Chung (1996b) and Hidalgo (2005) (mentioned further in Sect. 3.4 below) and found that the CSS method identified by Chung (1996b) was rarely able to reject the null hypothesis, whereas the method identified by Hidalgo (2005) was much better formed but more difficult to use. Beaumont and Smallwood (2020) proceed to provide advice on the selection of bandwidth parameters which may assist the practitioner.

A technique related to CSS estimation—but not directly based on a likelihood—was provided by Kouamé and Hili (2008) who looked at the process (4) and identified the consistency and asymptotic normality of an estimate of the autocorrelation function of the disturbance term, and then considered an estimator based on minimizing the ‘distance’ between the actual and theoretical autocorrelation functions of the disturbance term, technically defined as

$$\mathcal{L}_{KH}(v) = \sum_{j=1}^{m_f} \hat{\rho}_{e(v),j}^2,$$

where m_f is a fixed integer, greater than the number of parameters to be estimated, with

$\mu = E(X_t) = 0$, $\hat{\rho}_{e(v),j} = \frac{\sum_{t=1}^{n-j} e_t(v)e_{t+j}(v)}{\sum_{t=1}^n e_t^2(v)}$, and $e_t(v) = \sum_{j=0}^{n-1} \alpha_j(v)X_{t-j}$ are the

residuals of the model. The $\alpha_j(v)$ are defined by the relation $\epsilon_t = \sum_{j=0}^{\infty} \alpha_j(v)X_{t-j}$ which converges when the process is invertible and causal. The minimization process can thus be seen as ensuring the autocorrelation function of the residuals is as close to a white noise process as possible, up to a fixed lag.

Kouamé and Hili (2008) find that if the parameter space is restricted to be a compact set then the estimates of all parameters are consistent, and with the exception of the Gegenbauer frequencies (which are then assumed to be known apriori), they are asymptotically Gaussian.

Dissanayake et al. (2016) derive a quasi-maximum likelihood (QML) estimation procedure (both AR and MA truncated versions) using a Kalman filter to evaluate the likelihood function for a $k = 1$ Gegenbauer process which includes estimation of the Gegenbauer frequency. Consistency and asymptotic normality are established only for the Gaussian white-noise case.

Dissanayake et al. (2014) and Dissanayake (2015) (Ch 4) further examined this estimation method for GARMA-GARCH models defined by (13) (without specific ARMA components). Whilst the exercise of putting a long memory process into a State Space representation results in a state vector with infinite length, it was shown by Chan and Palma (1998) Theorem 2.2 that for a long memory model, although the State Space representation is infinite-dimensional, the likelihood can be evaluated in a finite number of steps.

McElroy and Holan (2012) explored efficient methods for calculating the covariance matrix $\Sigma(v)$, and its determinant which allowed a relatively efficient calculation of the exact likelihood function. This was extended by McElroy and Holan (2016) who showed specifically how the technique could be applied for very efficient likelihood estimation specifically for the GEXP model (12), although they state the technique can be equally applied to other Gegenbauer based models like the GARMA model. McElroy and Holan (2016) illustrated this technique, as did Gray et al. (1989) before them, with the Mauna Loa data—382 monthly atmospheric CO_2 measurements collected at the summit of Mauna Loa in Hawaii beginning in March 1958 (Keeling et al. 1989). McElroy and Holan (2016) showed using 2-factor GEXP model with 1 short memory parameter (excepting the disturbance term variance) and for a process of length 500, a single maximum likelihood evaluation on their hardware took over 180 s, whilst their new technique took just 1.18 s.

3.2 Whittle and other frequency domain methods

Whilst the Whittle methods may technically be viewed as a type of approximate likelihood method, the utility and popularity of the method deserves separate treatment.

The Whittle likelihood can written as

$$\mathcal{L}_W(v) = \frac{1}{n} \left[\sum_{j=1}^n \frac{I(\omega_j)}{f(\omega_j)} + \log(f(\omega_j)) \right]. \quad (17)$$

This is derived directly from the Gaussian log-likelihood by a process of approximating the Fourier transforms of the autocovariance function by the eigenvectors of $\Sigma(v)$ —see Whittle (1953), Eqn 5.3.

Giraitis and Leipus (1995) obtained parameter estimates for (9) by minimizing (17) to obtain Whittle estimates for the process, and proved consistency. Giraitis and Taqqu (1999) showed for long memory models a $\sqrt{n}$ consistency for the Whittle based estimators under a Gaussian assumption and showed the parameters were asymptotically Gaussian under a range of conditions on the spectral density around the unbounded peak. Giraitis and Taqqu (1999) found that when the long memory parameter is outside the stationary and invertible region, or for non-Gaussian processes, the $\sqrt{n}$ consistency does not always hold and the limit distribution is not necessarily Gaussian.

For Gaussian Gegenbauer processes, Yajima (1996) examined (4) for $k = 1$ and found the maximum of the normalized periodogram—defined as $\arg \max_{\omega \in (0, \pi)} I(\omega)$ —is a strongly consistent estimator of the Gegenbauer frequency. Yajima (1996) conjectures the limiting distribution of this estimator is non-Gaussian but this is not proved. For the remainder of the parameters (excluding the process mean μ), Whittles' technique is shown to provide strong consistency of those estimates. Assuming the Gaussian process is stationary, the asymptotic distribution of the estimator is Gaussian, when the Gegenbauer frequency is estimated as above, subject to some regularity conditions on the spectral density.

Giraitis et al. (2001) established a strong, general result for a Whittle estimator of $k = 1$ Gegenbauer processes using a general definition which includes standard

Gegenbauer processes defined by (4) and also GEXP processes defined by (12), without the Gaussian assumption. The likelihood to be minimized is given by

$$\mathcal{L}_{GHR}(v) = \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \frac{I(\omega_j)}{k(\omega_j)}, \quad (18)$$

where $k(\omega) \equiv k(\omega; v)$ is given by $f(\omega) = \frac{\sigma_\epsilon^2}{2\pi} k(\omega)$. Giraitis et al. (2001) assume $\int_{-\pi}^{\pi} \log(k(\omega)) d\omega = 0$ and thus (18) is the Whittle likelihood (17).

Specifically, when the conditions in the Appendix are satisfied, they identify n -consistency for the Gegenbauer frequency, and $\sqrt{n}$ -consistency for the remainder of the parameters. Further for all parameters except the Gegenbauer frequency they established a central limit theorem.

Nordman and Lahiri (2006) established a frequency domain empirical likelihood function for a long memory process, and showed that this could be used to establish a Gaussian central limit theorem for Whittle estimates. However there has been no extension of this to Gegenbauer processes.

For the long memory FEXP model defined by (12), Moulines and Soulier (1999) looked at a broadband least-squares estimator. Hsu and Tsai (2009) extended this to a $k = 1$ Gegenbauer model as follows:

$$\begin{aligned} \log I(\omega_j) &= -2d \log |\cos \omega_j - \cos \lambda| + \sum_{r=0}^{m_e} \xi_r \cos(r\omega_j) - \gamma + \left(\log \left(\frac{I(\omega_j)}{f(\omega_j)} \right) + \gamma \right) \\ &= z_j' \psi + \eta_j, \end{aligned}$$

where $\eta_j = \log(I(\omega_j)/f(\omega_j)) + \gamma$, $\psi = (d, \xi_0 - \gamma, \xi_1 - \gamma, \dots, \xi_{m_e} - \gamma)'$, $z_j = (-2 \log |\cos \omega_j - \cos \lambda|, 1, \cos \omega_j, \cos(2\omega_j), \dots, \cos(m_e \omega_j))'$ and γ is the Euler–Mascheroni constant. This is in the form of a linear regression equation, and if λ is known apriori, then ψ can be estimated by

$$\hat{\psi} = (Z'Z)^{-1} Z'\zeta,$$

where $Z = (z_1, z_2, \dots, z_{\tilde{n}})$, and $\zeta = (\log I(\omega_1), \log I(\omega_2), \dots, \log I(\omega_{\tilde{n}}))$. For a $k = 1$ Gegenbauer process with sufficiently smooth short memory process (and some other technical conditions), Hsu and Tsai (2009) showed consistency and also that

$$\text{Var}(\hat{\psi}) = \begin{cases} \frac{\pi^2}{24} \frac{m_e}{n} + o\left(\frac{m_e}{n}\right) & \lambda = 0, \pi \\ \frac{\pi^2}{12} \frac{m_e}{n} + o\left(\frac{m_e}{n}\right) & 0 < \lambda < \pi. \end{cases}$$

Wavelets have also been used to calculate $\mathcal{L}_W$. Whitcher (2004) examines a $k = 1$ white noise process for (4) and derives a wavelet estimation method for a Whittle-like likelihood function. Boubaker (2015) extends the result of Whitcher to a full GARMA-GARCH model defined by (13). Neither paper examines the impact of estimating $\mathcal{L}_W$

in this way on the consistency or asymptotic distribution of the estimates; presumably the results above for a Whittle estimator are assumed to hold for these methods. The interested reader is also referred to Sect. 3.4 which documents some interesting work by Alomari et al. (2020) on semi-parametric estimation with wavelets in the context of functional time series.

Finally, Hunt et al. (2021) described a Whittle-like estimator, and whilst they did provide consistency and a central limit theorem in a variety of restricted situations, they do not provide full consistency results or CLT for all parameters where the underlying short memory process is an ARMA process. However they do indicate that the estimator is more resistant to highly non-Gaussian distributions, particularly where the distribution is highly skewed. Their estimator was defined by minimising

$$\begin{aligned}\mathcal{L}_{LW}(\nu) &= \frac{1}{n} \sum_{j=1}^n \log^2 \left(\frac{I(\omega_j)}{f(\omega_j)} \right) \\ &= \frac{1}{n} \sum_{j=1}^n [\log(I(\omega_j)) - \log(f(\omega_j))]^2.\end{aligned}\tag{19}$$

This estimator is technically not derived from a likelihood function but instead looks to find those values of the parameters ν which will make the spectral density most closely match the periodogram.

3.3 Bayesian methods

This section starts with a brief overview of the Bayesian methods for parameter estimation, relating to long memory processes generally, rather than specifically Gegenbauer processes.

Bayesian techniques have been used to estimate the parameters of ARFIMA and GARMA processes using both frequency domain methods and time domain methods. A key contribution made by almost all authors has been to find efficient ways to evaluate or approximate the full likelihood function, or use Bayesian techniques for some frequency domain approaches, particularly semi-parametric estimation approaches.

Specifically, and looking first at the frequency domain methods, a semi-parametric approach to estimating the long memory parameter was first explored by Petris (1997). Liseo et al. (2001) also developed a semi-parametric method to estimate the long memory parameter. Liseo et al. (2001) used a frequency domain method to divide the parameter space into regions where prior information is available and others where vague priors are adopted. Specifically, they restrict evaluation of the short memory parameters to those frequencies of the spectral density which are greater than some cut-off, and the long memory parameter being estimated from frequencies less than the cut-off—the cut-off being treated as a hyper parameter to be estimated as a part of the Bayesian estimation. Specifically, Liseo et al. (2001) specify a Bayes factor approach to estimate the hyper parameter. The likelihood is specifically estimated using a Whittle approximation.

For an ARFIMA process defined by (8), define a cut-off frequency as one of the Fourier frequencies, ω_m , for some $1 \leq m \leq \tilde{n}$. Then from (18) define

$$\mathcal{L}_l(v) = \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \frac{I(\omega_j)}{k(\omega_j)} = \frac{1}{\tilde{n}} \left\{ C \sum_{j=1}^m \frac{I(\omega_j)}{\omega_j^{2d}} + \sum_{j=m+1}^{\tilde{n}} \frac{I(\omega_j)}{f(\omega_j)} \right\}$$

as the approximate log-likelihood. As noted above, the cut-off m is then treated as a hyper-parameter.

Holan et al. (2009) examined the FEXP model (defined as (12) with $k = 1$ and $u = 1$) with Gaussian disturbances, and thus likelihood (15), and looked specifically at estimating the long memory parameter d . They estimate the autocovariance function by first estimating the autocovariances of the short memory portion of the model using the techniques of Pourahmadi (1983) and Hurvich (2002), and then the long memory components using a ‘splitting’ method documented by Hurvich (2002) and Bertelli and Caporin (2002), and from there constructing an estimate of Σ . The inverse and the determinant of this are then calculated, and the likelihood calculation is then used in an MCMC estimation method.

Makarava et al. (2011) looked at a Bayesian estimate of the long memory parameter and its variance in the context of linear-trend mixed models for continuous time processes.

Wavelet estimation via Bayesian methods has also been examined, notably by Ko and Vannucci (2006), who developed such a method for ARFIMA models.

Time-domain methods have also been used with Bayesian estimation. Pai and Ravishanker (1998) looked at univariate long memory ARFIMA models with a Gaussian distribution. In order to simplify and speed up the evaluation of the full likelihood, they propose a method (originally presented in Pai and Ravishanker (1995)) which involves treating the process values before those available as unobserved latent variables. This in turn allows them to calculate $\Sigma^{-1}(v)$ exactly, and use this to calculate the likelihood and then the posterior distribution of the ARFIMA parameters. Ravishanker and Ray (1997) and Doppelt and O’Hara (2019) extended this approach to multivariate long memory time series with Gaussian disturbances.

The above references all considered the ARFIMA/VARFIMA representation where the spectral density is defined by (6) with $k = 1$ and $\lambda = 0$. However there has also been some work specifically looking at estimating the parameters of Gegenbauer processes (6) with $\lambda \neq 0$.

Graves et al. (2015) looked at an approximate likelihood technique for ARFIMA models, based on a technique similar to Haslett and Raftery (1989) which truncated the infinite AR representation to a finite but large number. In their Appendix B, Graves et al. (2015) state that the MCMC methodology they develop can be easily extended to incorporate Gegenbauer models, provided the Gegenbauer frequency/frequencies are known a-priori.

Phillip et al. (2018) identified an efficient sampling scheme for estimating the parameters of (4) with disturbance term (14) when $p = 1$, $q = 0$. They used a truncated form of the Wold representation, similar to that used by Dissanayake et al. (2016) to estimate the likelihood.

Whilst this paper is looking specifically at methods to estimate the parameters of Gegenbauer processes, it would be remiss of us not to mention 2 further papers. Firstly, McElroy and Holan (2012) provide a method to efficiently calculate the autocovariances of a Gegenbauer process. This can be used be used equally easily for Bayesian estimates as frequentist maximum-likelihood estimates. Secondly Holan and McElroy (2012) adopted a structural approach to seasonal adjustment.

Durham et al. (2019) presented the SABL algorithm (sequentially adaptive Bayesian learning) applied to the estimation of ARFIMA models where the posterior density of the parameters is highly non-Gaussian, but where the number of parameters is relatively small. This algorithm differs from the more standard Hastings algorithms in that the updating of the Markov chain is done via a series of individual Bayesian estimation steps, rather than updating the Hamiltonian as in Metropolis-Hastings, Betancourt et al. (2014).

3.4 Semi-parametric methods

Semi-parametric methods refer generally to the estimation of the long memory parameter—the d parameter of either long memory ARFIMA or Gegenbauer processes (7) where $k = 1$. It is assumed that the unbounded peak can be located apriori, as indicated by the work of Yajima (1996), or from an understanding of the underlying process.

These methods became popular for two main reasons. First they allow a practitioner to test for the presence of long memory prior to model fitting; if no evidence is found then the practitioner can proceed to fit a more parsimonious model. Secondly these methods require fewer assumptions on the behavior of the underlying distribution in order to establish consistency and a central limit theorem.

The earliest of the semi-parametric methods for long memory ARFIMA models was Geweke and Porter-Hudak (1983), which although not in common use today, is reproduced here since it forms the basis of much subsequent work. Taking logs of the spectral density of an ARFIMA process, and with a small amount of manipulation, the following equation is obtained

$$\log(I(\omega_j)) = \log(f_Y(0)) - d \log(4 \sin^2(\omega_j/2)) + \log\left(\frac{f_Y(\omega_j)}{f_Y(0)}\right) + \log\left(\frac{I(\omega_j)}{f(\omega_j)}\right).$$

Geweke and Porter-Hudak pointed out this equation resembles a simple linear regression model with $\log(I(\omega_j))$ the response, $\log(4 \sin^2(\omega_j/2))$ the explanatory variables and $\log\left(\frac{I(\omega_j)}{f(\omega_j)}\right)$ representing the disturbance term. If the frequencies used are restricted to be close to 0, then $\log\left(\frac{f_Y(\omega_j)}{f_Y(0)}\right)$ will be close to 0. To formalize this, they introduce m as the number of frequencies in the neighborhood of 0 to include in the regression. The cutoff m must grow to infinity with n , but more slowly than n , satisfying $\lim_{n \rightarrow \infty} \frac{(\log n)^2}{m} = 0$. With this and some regularity conditions, d is estimated by

$$\hat{d}_{GPH} = -1/2 \frac{\sum_{j=1}^m (\tilde{v}_j - \bar{v}) \log I(\omega_j)}{\sum_{j=1}^m (\tilde{v}_j - \bar{v})^2},$$

where $\tilde{v}_j = \log |1 - e^{-i\omega_j}| = \log(4 \sin^2(\omega_j/2))$ and $\bar{v} = \frac{1}{m} \sum_{j=1}^m \tilde{v}_j$ so that if $d < 0$ then $\hat{d}_{GPH}$ behaves asymptotically as a $N\left(d, \frac{\pi^2}{6m}\right)$ variate. Geweke and Porter-Hudak (1983) suggested setting $m = \sqrt{n}$. Robinson (1994) however showed that the optimal value for m was proportional to $n^{4/5}$, although the exact optimal value depends also on the true value of d .

Robinson (1995b) explored the distributional properties of this estimator, and also the related estimator when $\tilde{v}_j = 2 \log |j|$ which is asymptotically the same, and showed that the consistency holds for $d > 0$ as well as $d < 0$, and further under some assumptions, the parameter estimate is asymptotically Gaussian.

In the case of a Gegenbauer process defined by (4) where $k = 1$, a related estimate of d was suggested by Arteche (1998):

$$\hat{d}_A = -1/2 \frac{\sum_{j=\pm 1}^{\pm m} (\tilde{v}_j - \bar{v}) \log I(\lambda + \omega_j)}{\sum_{j=\pm 1}^{\pm m} (\tilde{v}_j - \bar{v})^2},$$

where $\tilde{v}_j = \log |j|$ and $\bar{v}$ defined as above, and where $\frac{1}{m} + \frac{m}{n} \rightarrow 0$ as $n \rightarrow \infty$. This estimator is shown to be asymptotically Gaussian under some conditions.

Modifications to these equations—originally suggested for long memory models by Robinson (1995b) are (i) trimming out the lower frequencies and (ii) pooling adjacent periodogram ordinates. If J ordinates are being pooled, the disturbance term is Gaussian, the spectral density is smooth in a neighborhood of the Gegenbauer frequency and other technical conditions, Arteche (1998) (Theorem 5) shows that $\hat{d}_A$ can be approximated by $N\left(d, \frac{J\psi'(J)}{\sqrt{8m}}\right)$ when n is large. The interested reader is referred to Arteche (1998) for further details.

Arteche (1998, 2000) and Arteche and Robinson (2000) also suggest a local-Whittle like estimator, based on the earlier work of Robinson (1995a), but restricted to the Fourier frequencies around the non-zero unbounded peak. The estimate is

$$\hat{d}_{LW} = \arg \min_{\Theta} R(d),$$

where Θ is a closed subset of $(-1/2, 1/2)$, and

$$R(d) = \log \tilde{C}(d) - \frac{2d}{m-l} \sum_{j=l+1}^m \log(\omega_j)$$

$$\tilde{C}(d) = \frac{1}{m-l} \sum_{j=l+1}^m \omega_j^{2d} I(\lambda + \omega_j).$$

Arteche and Robinson (2000) showed that $\hat{d}_{LW}$ can be approximated by $N\left(d, \frac{1}{4m}\right)$ when n is large and the disturbance terms are martingales which are 4-th order sta-

tionary and some other technical conditions including that m is an integer tending to infinity more slowly than n . Again the interested reader is referred to the original paper for specific details.

Narukawa (2016) looked at a local Whittle estimator for the GEXP model (12), assuming the Gegenbauer frequency was already known. In their notes it is suggested this frequency can be estimated using the method of Yajima (1996). Their estimate is given by

$$\hat{d}_N = \arg \min_d \frac{1}{\tilde{n}} \sum_{j=1, j \neq q_0}^{\tilde{n}} \left\{ \log(f(\omega_j)) + \frac{I(\lambda + \omega_j)}{f(\omega_j)} \right\},$$

where $q_0 = \arg \min_{j \in 1, \dots, \tilde{n}} |\lambda - \omega_j|$ so that ω_{q_0} is the closest Fourier frequency to λ , and if there are two such values, then the smaller of the two is chosen. However there are no notes provided on how to choose a suitable value for m_e , although the numerical studies of this paper illustrate the process by choosing values of m_e from 1 to 7. Narukawa (2016) discussed extensions to this for a general k -factor process. Narukawa (2016) was able to show that $\hat{d}_N$ is a consistent estimator of d under conditions including that the disturbance terms are iid, m_e varies with n and some other technical conditions. They also derived a central limit theorem for this estimator

$$\sqrt{\frac{n}{m_e}} (\hat{d} - d) \xrightarrow{d} \begin{cases} N(0, 1/4) & \lambda \in \{0, \pi\} \\ N(0, 1/2) & 0 < \lambda < \pi, \end{cases}$$

under further conditions including 4th order stationarity. See Narukawa (2016) for further details.

In a similar vein to Yajima (1996), Hidalgo and Soulier (2004) found that

$$\hat{\lambda}_{HS} = \frac{2\pi}{n} \arg \max_{l \leq j \leq \tilde{n}} I(\omega_j)$$

is also a strongly consistent estimator of the Gegenbauer frequency, with the primary condition being a finite eight-order moment, and other conditions. Having estimated the Gegenbauer frequency in this way, and if it is not found to be 0 or π , and if the underlying disturbance term is Gaussian, then a phase-shifted version of the Geweke and Porter-Hudak (1983) estimator is defined as

$$\hat{d} = -1/2 \frac{\sum_{j=1}^m (\tilde{v}_j - \bar{v}) \log I(\hat{\lambda}_{HS} + \omega_j)}{\sum_{j=1}^m (\tilde{v}_j - \bar{v})^2},$$

where $\tilde{v}_j = \log |1 - e^{-i\omega_j}|$ and $\bar{v} = \frac{1}{m} \sum_{j=1}^m \tilde{v}_j$, $m = [n^\delta]$ with $0 < \delta < \frac{2\beta}{2\beta+1}$, and $f_Y(\omega)$ satisfies a Hölder—or uniform Lipschitz condition of order β . The distribution of $\hat{d}$ can then be approximated by $N\left(d, \frac{\pi^2}{12m}\right)$ when n is large.

Hidalgo and Soulier (2004) outline a method for estimating the unbounded poles for a general k -factor Gegenbauer process and show consistency for these estimators.

The method is a sequential method, where

$$\hat{\lambda}_{s,HS} = \frac{2\pi}{n} \arg \max_{\Lambda_s} I(\omega_j), \quad s \geq 1,$$

where z_n is a non-decreasing sequence such that for any positive real numbers k and k' , $\log^k(n) \ll \frac{n}{z_n} \ll n^{k'}$, for instance $z_n = n e^{-\sqrt{\log(n)}}$, and $\Lambda_1 = \{1 \leq j \leq \tilde{n}\}$, $\Lambda_s = \left\{1 \leq j \leq \tilde{n}; |\omega_j - \hat{\lambda}_{s-1,HS}| \geq z_n/n; \dots; |\omega_j - \hat{\lambda}_{1,HS}| \geq z_n/n\right\}$ for $s > 1$.

Artiach and Arteche (2006) improve on the estimation methods of Yajima (1996) and Hidalgo and Soulier (2004) by suggesting an iterative approach to refine the estimate of the Gegenbauer frequency so that it need not fall on a Fourier frequency. As they point out, for small sample sizes, Fourier frequencies may not provide sufficient resolution to determine the Gegenbauer frequency. Their method involves starting with the Hidalgo and Soulier estimate, and then considering an interval between the frequency before and the frequency after the Fourier frequency, and dividing this into 20 sections, and looking for which of these continues to maximize the periodogram. This new interval is then recursively subdivided into a further 20 sections and the process repeated until the desired level of accuracy is obtained.

Bardet and Bertrand (2010) in an interesting paper looked at estimating the spectral density of a functional time series using wavelets and established consistency and a central limit theorem for each positive frequency. Using this key result, Alomari et al. (2020) established consistency of estimating the Gegenbauer frequency and exponent of a $k = 1$ Gegenbauer functional time series, under conditions involving sufficient smoothness of the short memory process, and some general assumptions on the smoothness of the wavelet function.

In a further result, Ayache et al. (2020) established a central limit theorem result for the estimators defined by Alomari et al. (2020).

3.5 Summary

For GARMA models defined by (4), a summary of the theoretical results for the various methods is presented in Table 1.

Quasi-maximum likelihood methods have consistency results and a Central Limit Theorem (CLT) from Dissanayake et al. (2016) only for a GARMA(0,1,0;u) (Gegenbauer white noise) case.

The semi-parametric methods give the strongest set of theoretical results; Yajima (1996), Hidalgo and Soulier (2004) and Artiach and Arteche (2006) give estimates, consistency and CLT results for locating the Gegenbauer frequencies, whilst the methods for estimating d given by Arteche and Robinson (2000) assume the Gegenbauer frequency is known a-priori, and further that the process can be transformed to a short memory process and the remaining parameters estimated from that. However, our simulations in Sect. 4 indicate that this process results in larger bias and root-mean-square error in the short memory parameters.

Table 1 Summary of theoretical results available

Method	Consistency	CLT avail.
Grid Search	None	None
CSS	$k = 1$ Only	Yes
Autocorrelation distance	Yes	All except “u”
Quasi-ML	Gegenbauer white noise only	Gegenbauer white noise only
Whittle	Yes	All except “u”
Log-Whittle	Some	Some
Bayesian	Yes	Yes
Max periodogram $u = \cos(2\pi f)$	Yes	No
Local linear regression—d	Yes	Yes
Local Whittle—d	Yes	Yes

4 Monte-Carlo simulations

4.1 Simulations and results

To evaluate a range of the more popular methods for estimation of the parameters in (4), a short simulation study was undertaken looking specifically at a $k = 1$ GARMA(1, 0.22, 0, -0.8090) process. For this process the other parameters were $f_G = 0.4$, ($u \approx -0.8090$) $\sigma_\epsilon^2 = 1$, $\phi = 0.8$. The disturbance term was Gaussian. Since not all methods provide an estimate of the variance σ_ϵ^2 this parameter was not estimated.

1000 realizations were generated as detailed in Sect. 4.2. This process was repeated 3 times for processes of length 100, 500 and 1000.

The following algorithms were used to compare estimates: conditional sum-of-squares (CSS), Whittle method—Giraitis, Hidalgo & Robinson (Whittle), quasi-maximum likelihood with a MA truncation at 50 (QML), wavelets (Whitcher 2004) using the minimum bandwidth wavelet filter of length 8 due to Morris and Peravali (1999) (MB8), a Bayesian estimation technique, the log-Whittle estimator of Hunt et al. (2021) (log-Whittle), and 2 semi-parametric methods due to Arteche (1998, 2000) and Arteche and Robinson (2000)—a log-periodogram regression centered around the spectral peak (A–R) and a local-Whittle estimator (local Whittle).

These last two methods (A–R & local Whittle) rely firstly on using the maximum value of the periodogram to determine the Gegenbauer frequency and then both use the standard short-memory processes to determine the estimate of the autoregressive parameter.

The results—bias and mean-squared error to 4 decimal places—are presented in Tables 2, 3 and 4. In general the Gegenbauer frequency is determined very accurately, reflecting the faster convergence of this parameter for most methods (n -consistent whereas the other parameters are $\sqrt{n}$ consistent). Whilst the results are mixed, overall the Bayesian method appears to produce the best mix of small bias and MSE over the parameters, and over all three process lengths (where the parameter estimates for the Bayesian method are determined as the mean of the posterior distribution), although by

Table 2 Simulation results for processes of length 100

Method	d	f	ϕ	Avg secs	Median secs	Alg. Failure
A-R	- 0.1989 (0.0471)	- 0.0060 (0.0020)	- 0.3997 (0.2181)	0.00	0.00	0
Bayesian	0.0749 (0.0065)	0.0016 (0.0000)	- 0.0239 (0.0056)	8.19	7.67	0
CSS	0.0967 (0.0107)	- 0.0066 (0.0024)	- 0.0070 (0.0341)	0.16	0.16	0
Local Whittle	- 0.2142 (0.0478)	- 0.0060 (0.0020)	- 0.4123 (0.1957)	0.02	0.02	0
MB8	0.1086 (0.0128)	- 0.0100 (0.0007)	- 0.0015 (0.0121)	5.10	0.58	25
QML	0.0931 (0.0099)	- 0.0067 (0.0023)	- 0.0219 (0.0344)	2.44	2.43	0
Whittle	- 0.0256 (0.0047)	- 0.0063 (0.0022)	- 0.0803 (0.0418)	0.02	0.01	0
Log-Whittle	- 0.0090 (0.0053)	- 0.0054 (0.0021)	- 0.0499 (0.0339)	0.01	0.01	0

Format is "bias (mse)"

Table 3 Simulation results for processes of length 500

Method	d	f	ϕ	Avg secs	Median secs	Alg. failure
A-R	- 0.0437 (0.0051)	- 0.0017 (0.0006)	- 0.0471 (0.0082)	0.07	0.07	0
Bayesian	0.0357 (0.0017)	0.0000 (0.0000)	0.0042 (0.0008)	154.37	152.01	0
CSS	0.0366 (0.0018)	- 0.0019 (0.0007)	0.0059 (0.0061)	10.30	11.22	0
Local Whittle	- 0.0519 (0.0074)	- 0.0017 (0.0006)	- 0.0593 (0.0131)	0.10	0.10	0
MB8	0.0454 (0.0025)	- 0.0007 (0.0000)	- 0.0001 (0.0009)	0.52	0.50	5
QML	0.0518 (0.0032)	- 0.0022 (0.0007)	0.0065 (0.0061)	12.22	11.92	0
Whittle	- 0.0140 (0.0020)	- 0.0018 (0.0007)	- 0.0219 (0.0067)	0.09	0.09	0
Log-Whittle	- 0.0010 (0.0014)	- 0.0014 (0.0003)	0.0079 (0.0040)	0.09	0.09	0

Format is "bias (mse)"

Table 4 Simulation results for processes of length 1000

Method	d	f	ϕ	Avg secs	Median secs	Alg. failure
A-R	- 0.0270 (0.0029)	- 0.0024 (0.0008)	- 0.0247 (0.0037)	0.55	0.54	0
Bayesian	0.0177 (0.0005)	- 0.0001 (0.0000)	0.0019 (0.0005)	897.42	789.99	1
CSS	0.0177 (0.0007)	- 0.0025 (0.0009)	- 0.0007 (0.0069)	81.61	86.54	0
Local Whittle	- 0.0286 (0.0039)	- 0.0024 (0.0008)	- 0.0307 (0.0079)	0.59	0.59	0
MB8	0.0190 (0.0006)	0.0008 (0.0000)	- 0.0063 (0.0005)	0.50	0.49	29
QML	0.0313 (0.0014)	- 0.0029 (0.0010)	0.0004 (0.0081)	23.89	23.27	0
Whittle	- 0.0084 (0.0014)	- 0.0022 (0.0009)	- 0.0162 (0.0080)	0.57	0.56	0
Log-Whittle	- 0.0255 (0.0015)	- 0.0119 (0.0003)	- 0.0151 (0.0013)	0.57	0.56	0

Format is "bias (mse)"

far with the longest computation time. In particular, the MSE for the shorter processes was substantially smaller than other methods for the $n = 100$ process length.

The wavelet (MB8) method also performed quite well, however the source code to run this needed to be modified to assist with convergence, and to estimate the short memory component, and even so there were a number of cases where the algorithm failed to converge. Some further details on this are given in Sect. 4.2 and in the source code itself.

The CSS, Whittle and log-Whittle methods all performed quite well, with the log-Whittle better at the shorter process lengths.

The semi-parametric methods (A–R and local-Whittle)—when used this way to estimate all the parameters of the process, performed worst of all, particularly on smaller process lengths.

It is also interesting to view the average time taken to produce each estimate. In general, for a length of $n = 500$, the time domain methods—CSS and QML—took 10–12 s and the Bayesian method took 2½ minutes per simulated process. The frequency domain methods ranged from 0.07 to 0.10 of a second. The wavelet MB8 method recorded some exceptionally long convergence times for some processes (although convergence was achieved in those cases) for the $n = 100$ case and this explains the much longer time for that method - on average than at the longer process lengths. It can be seen that the median time is much shorter.

As noted in Sect. 4.2, the wavelet method MB8 did have difficulty with numerically integrating the spectral density to calculate the bandwidth variances. The original source code was modified to assist in this quadrature. However the algorithm still did not converge to a solution at all for a limited number of cases.

The Bayesian estimation also exhibited some issues. The Rhat statistic (Gelman and Rubin 1992) is used to identify these issues—an estimation was deemed to have failed when the Rhat statistic was greater than 1.1. This happened only for a single run—in $n = 1000$ - run 128. However the $n = 100$ also had 132 runs where the number of effective samples was less than 400 (Vehtari et al. 2019, Section 2). In practice this is usually relatively straight forward to rectify by increasing the number of iterations per chain. The $n = 1000$ sample which exhibited high Rhat statistics—run 128—also had too few effective samples, although most likely this was a result of the four Markov chains not fully mixing.

4.2 Algorithm and software details

To generate realizations of Gegenbauer processes, the algorithm originally documented by Ford (1991) and later also by Woodward et al. (1998) was used. In summary, the algorithm generates a theoretical autocorrelation function, by numerically integrating the theoretical spectrum and then uses the method of Hosking (1981) to generate realizations from a process with that autocorrelation structure.

Although Whitcher (2001) provided an alternative, wavelet based method for generating realizations, that paper found no overall improvement from using the wavelet algorithm, although its' computation time was reduced. Also the techniques of McElroy and Holan (2012, 2016) allow for the very accurate and fast evaluation of the

autocovariance function for a given finite sample. However, for the purposes of this paper the additional time spent in generating the realizations was not significant—the main criterion is the time taken to estimate the parameters—so it was decided to continue to use the “Hosking” method as modified by Ford.

A ‘run-in’ period of 500 observations was used for processes of length 100 and 500, and a run-in period of 1000 observations was used for processes of length 1000. Once the autocovariance function of the process had been estimated, the function in the R package *waveslim* (Whitcher 2015) called “hosking.sim” was used to generate the specific realisation from the autocovariance function.

Numerical integration was required for this algorithm. Numerical integration was done via adaptive Gauss-Kronrod quadrature, as supplied by the R *pracma* package (Borchers 2018), or in some cases alternatively by a fine-grained custom trapezoidal quadrature.

In order to ensure the realisation was sufficiently close to the theoretical definition, a check was made to ensure there was a peak evident at the Gegenbauer frequency (away from the autoregressive peak near the origin); if this was not evident, the realization was thrown away and the process re-generated. This was done since the frequency domain estimation techniques do require a sufficiently distinguishable peak to allow the identification and estimation of the long memory process.

For maximum likelihood and quasi-maximum likelihood estimation, a Kalman filter approach was used to calculate the likelihood, and the R package “FKF” was used for this purpose (Luethi et al. 2018). For wavelet estimation methods, the *waveslim* package (Whitcher 2015) modified as detailed below was used.

For semi-parametric algorithms, the Gegenbauer frequency was determined by the method of Yajima (1996) and Hidalgo and Soulier (2004)—choosing the Fourier frequency which maximized the periodogram. After estimating the long memory parameter, those processes were then filtered back to a short memory process using (10); specifically if $X = (X_1, X_2, \dots, X_n)^T$ and $Y = (Y_1, Y_2, \dots, Y_n)^T$ and defining

$$T = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ C_1^{(d)}(u) & 1 & 0 & \cdots & 0 \\ C_2^{(d)}(u) & C_1^{(d)}(u) & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \ddots & \vdots \\ C_n^{(d)}(u) & C_{n-1}^{(d)}(u) & \cdots & C_1^{(d)}(u) & 1 \end{pmatrix}$$

where $C_n^{(d)}(u)$ represent the Gegenbauer polynomials defined by (5), then $X = T^{-1}Y$ and since T is lower-triangular with determinant 1, the matrix inversion is comparatively straight-forward and numerically stable.

For semi-parametric algorithms, once the short-memory vector X is obtained, the standard R function “*arima*” was used to estimate the short-memory parameters.

The wavelet based methods were based on Whitcher (2004) using the code written by the author from the R package “*waveslim*”—specifically the routine “*spp.mle*” which can be used to determine the Gegenbauer frequency and the long memory

parameter d . To support the simulations, the source code for this was modified to include an AR(1) parameter; also some stability and accuracy improvements were made, particularly around the numerical integration of the spectral density to estimate the bandpass variances.

Other estimation algorithms were coded directly. The authors gratefully acknowledge that the R code for the quasi-maximum likelihood implementation was supplied by Hao Wu from the University of Sydney.

A key challenge to be overcome was associated with the autoregressive component of the spectral density. An example appears in Chung (1996b) (Fig. 1). In this figure it can clearly be seen that, although the peak at the Gegenbauer frequency is apparent, in a finite sample, the peak near the origin (or in the case of a negative autoregressive parameter—near π) can, and in shorter series often does, have a higher value than the peak at the Gegenbauer frequency. Note that the height of the peak near the origin or near π is related to how close the autoregressive parameter is to ± 1 —in other words a unit-root—the closer it is the higher the peak, and the greater difficulty the algorithms have in determining the correct location of the Gegenbauer frequency.

Even when this is not a problem, for frequency domain estimation methods there is a tendency for the algorithm to converge to the incorrect peak for the Gegenbauer frequency. To avoid this, the algorithms were specifically coded to look for a peak well away from the origin. (It is assumed a practitioner will already have determined that a Gegenbauer process rather than a long memory process is an appropriate model; perhaps by an examination of the autocorrelation function to identify the characteristic damped sine curve as illustrated by Hosking (1982), Fig. 2). The specific approach taken is to examine frequencies which exclude the bottom and top 10% of the Fourier frequencies.

However, as shown in Sect. 4.1, the wavelet based functions were still unable to converge to the correct Gegenbauer frequency in a few cases.

For Bayesian estimation, the “stan” package (Carpenter et al. 2017) was used, via the “rstan” interface and this estimated the parameters via a time-domain technique. Relatively diffuse priors were used for this— $\phi \sim N(0, 1/2^2)$, $f \sim \log N(0, 0.2^2)$, and $d \sim N(0, 1/4^2)$. Parameters were specifically restricted to $d, f \in (0, 1/2)$, and $\phi \in (-1, 1)$. These priors and parameter limits assume that the model has been correctly identified as a stationary and invertible GARMA model. The posterior distribution was estimated using a NUTS MCMC method (Hoffman and Gelman 2014), with 4 Markov chains, 500 warmup samples and 1500 post-warmup samples, and in contrast to other methods, the chains used multi-threading due to the additional time required for single threaded operation. Even so, as is noted in Tables 2, 3 and 4, the time taken for the Bayesian estimation was clearly the longest.

The “Rhat” statistic and the number of effective samples were monitored to ensure the convergence of the algorithm (Vehtari et al. 2019). Any variation in the Rhat statistic was to be reported as an ‘algorithm failure’. As noted above, this happened only once.

5 Sunspot data

Box and Jenkins (1976), noted that the analysis of sunspot data has had a long history in time series literature, including Schuster (1906) and Yule (1927).

Recently the Royal Observatory of Belgium has undertaken a considerable project to improve and standardise the sunspot data, and these data have been chosen to illustrate the above techniques (Clette et al. 2014).

It is emphasized here that the intention is not to provide an exhaustive analysis of modeling of this process but more to illustrate the techniques and how they may be used.

5.1 Background

The original sunspot data was collated and some observed directly by Rudolf Wolf (and followed by his successor Alfred Wolfer), and labeled the (international) sunspot number (SN). Much of the information reproduced in this section is due to Clette et al. (2014).

Many authors have illustrated their techniques using the Wolfer sunspot time series, although this is available in a number of forms and lengths. Most authors use the annual series from 1749 to 1924, and credit the data to either Anderson (1971) (A.3) or directly Waldmeier (1961). Anderson in turn credits the data to Waldmeier, who as director of the Zurich Observatory continued to collate the Wolfer sunspot numbers after Wolfer. When Waldmeier retired in 1980, the process was taken over by the Royal Observatory of Belgium, Brussels, and this led to the foundation of the SIDC “Sunspot Index Data Center”.

Over time alternative series with different ways of counting sunspots, and applying “corrections” were introduced. This included the group number (GN) from Hoyt and Schatten (1998a, b). A key problem arose since these alternative series metrics did not agree in some respects.

Starting in 2011 a major project was undertaken by the “Sunspot Index and Long-term Solar Observations”—SILSO—run from the Royal Observatory of Belgium to collate, validate and integrate the many sources of information. A key requirement was to try and keep as much of the history of the process as possible, since changes in solar activity can happen over considerable periods of time.

An example of this was the Maunder minimum which occurred very early in the series—between 1645 and 1715—when sunspot activity was greatly reduced. Measurements of this time allowed analysis to show how natural solar forcing impacted past climate changes on scales of 50–1000 years (Shapiro et al. 2011).

5.2 Model fitting

In this section the annual data is analysed. The official series is now kept with a specific version number—this analysis uses version 2.0.

The first step is a model identification step where an attempt is made to identify the best form of the model, before proceeding to estimate the parameters of the model.

Fig. 2 SILSO Sunspot Number—plot and periodogram

The plot and periodogram of the process are presented in Fig. 2. It appears there is a single, very large spike in the periodogram at frequency 0.0906, corresponding to a period of 11 years and 13 days.

As found by Gray et al. (1989) and Chung (1996b), a $k = 1$ GARMA process appears to be the most appropriate fit. Using the method of Arteche (1998, 2000) and Arteche and Robinson (2000) with $l = 0$, the value of d is estimated as 0.3689. Using these values, as detailed in Sect. 4.2, the original process was transformed to a short memory process.

The autocorrelation function and partial autocorrelation function of this short memory process are shown in Fig. 3. They appear to indicate that an autoregressive process with parameters at lags (years) 1 and 9 might be an appropriate fit. For a point of comparison—as in Gray et al. (1989)—a process with a single autoregressive parameter was also fit.

Having identified an initial model, the next step is to estimate the parameters. As identified in Sect. 4, above, a Bayesian method appears most appropriate, particularly since there is earlier work identifying the likely values of the parameters using the data up to 1924—the priors specified were $\mu \sim N(44.78, 20^2)$, $d \sim N(0.42, 0.1^2)$, $f \sim N(0.0883, 0.04^2)$, and $\sigma_\epsilon^2 \sim \log N(20.0, 20.0^2)$. For the AR(1) model the prior $\phi_1 \sim N(0.49, 0.1^2)$ was used, and for the AR(1,9) model, relatively diffuse priors were selected— $\phi_1, \phi_9 \sim N(0.0, 1.0^2)$. The ‘Stan’ program was used to perform this analysis (Stan Development Team 2019). The results are shown in Table 5, and for comparative purposes the values estimated on the original series by both Gray et al. (1989) and Chung (1996b) are reproduced and labeled as the “1924 data”, with the analysis done here denoted as “SILSO data”.

Fig. 3 SILSO Sunspot Number—short memory component

Table 5 Model parameters for SILSO and 1924 sunspot numbers processes

GARMA model	f	d	ϕ_1	ϕ_9	μ	σ_ϵ^2
AR(1) (SILSO data)	0.0932	0.42	0.52		77.4	614.7
Gray et al. (1989) AR(1) (1924 data)	0.0883	0.42	0.49		44.8	NR
Chung (1996b) AR(1) (1924 data)	0.0891	0.42	0.49		46.7	232.2
AR(1,9) (SILSO data)	0.0928	0.35	0.57	0.23	67.5	582.4

“NR” indicates the value was not reported

It should be noted that the mean and variance of the 1924 data is different from the SILSO data; however the Gegenbauer and autoregressive parameters are very similar.

Comparing the Bayesian estimates for the AR(1) and AR(1,9) models, the smaller residual variance is given by the AR(1,9) model. For this model, only 0.025% of the samples for the parameter d were over 0.5 indicating that the Gegenbauer component is very likely stationary.

The autocorrelation function (ACF) of the residuals for the AR(1,9) model are presented in Fig. 4 and show some small evidence of autocorrelation at lags 3 and 11. A Box–Ljung test at these lags however indicated these were not significant.

The Gegenbauer frequency calculated for this model actually corresponds to a period of 10 years and 9 months and 11 days—a bit shorter than previously calculated by other authors. However, it is interesting to note that the maximum of the periodogram is at $f = 0.0906250$ which corresponds to 11 years and 13 days.

test lag	test statistic	df	p-value
3	6.2255	3	0.1011
11	16.595	11	0.1204

Fig. 4 ACF for residuals of Sunspot process from model GARMA(1,0.41,0;0.84) & Box–Ljung statistics

6 Conclusion

This paper has documented a variety of variations of Gegenbauer models which have been developed and further documented the various estimation techniques used for a GARMA process.

Section 4 documented some Monte Carlo simulations using the main estimation methods. This established that a Bayesian methodology appeared the most accurate at estimating the parameters of a GARMA model, although other methods such as the Whittle, CSS and log-Whittle methods also performed very well. An example of the use of Bayesian estimation in practice was provided in Sect. 5.

It is noted that there is still work to be done on the theoretical results behind these estimators however (refer to Sect. 3.5). In particular the theoretical asymptotic distribution and standard errors have not been derived for a number of estimators. There is no way to determine the exact model unless some knowledge from the series itself can be used (for instance monthly observations might be expected to have a frequency close to 1/12).

Recent work by Beaumont and Smallwood (2020) shows that some methods of inference on parameter values can have remarkably less power than others, and the choice of method appears crucial.

With long memory models generally, techniques like bootstrap that may be used in other areas of statistics are typically not feasible.

In general the semi-parametric methods which estimate only a single parameter at a time showed the highest bias and MSE. Whilst these techniques are very useful for model identification, the authors would caution against using them for final estimation or forecasting.

The final observation to be made is that the time taken to evaluate the time-domain based methods is in general considerably longer than for frequency domain methods.

Acknowledgements The authors would like to thank the 2 anonymous referees and the editor who suggested substantial improvements to this paper.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data Availability The sunspot data used in Sect. 5 was obtained from <https://www.bis.sidc.be/silso/datafiles>, specifically the series labeled “Yearly mean total sunspot number [1700 - now]”. The specific version used

includes data up to 2017. Data for the Monte Carlo simulations referred to in Sect. 4 is included in the github repository used for the code—please refer to the next section.

Declarations

Conflict of interest The authors have no conflicts of interest or competing interests.

Code availability The code and simulated data and full results for the Monte Carlo simulations in Sect. 4 may be obtained from github located at: https://github.com/rph50/survey_paper All other code may be obtained from the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix: Regularity conditions for Giraitis et al. (2001)

GHR A.1

$\{Y_t\}$ has spectral density defined by (6), and $f_X(\omega)$ is an even function bounded away from zero with parameters $\theta \in \Theta$, $\phi \in \Phi$, with Θ, Φ compact sets, and the 1st and 2nd order derivatives w.r.t. θ, ϕ are continuous.

GHR A.2

Parameterize $f(\omega) = \frac{\sigma_\epsilon^2}{2\pi} k(\omega; \theta, \phi, d, \lambda)$ $\omega \in (-\pi, \pi]$ where $\theta \in \mathbb{R}^q$, $\phi \in \mathbb{R}^p$, $d \in \mathbb{R}$, $\lambda \in [0, \pi]$. Assume $k(\omega; \theta, \phi, d, \lambda) > 0$ $\omega \in (-\pi, \pi]$, and $\int_{-\pi}^{\pi} k(\omega; \theta, \phi, d, \lambda) d\omega = 0$.

GHR A.3

$$\inf_{(\theta, \phi, d, \lambda) \in \mathbb{R}^{p+q+1} \times [0, \pi]} (2\pi)^{-1} \int_{-\pi}^{\pi} \frac{k(\omega; \theta_0, \phi_0, d_0, \lambda_0)}{k(\omega; \theta, \phi, d, \lambda)} d\omega = 1$$

and the set $\{\omega : k(\omega; \theta, \phi, d, \lambda) \neq k(\omega; \theta_0, \phi_0, d_0, \lambda_0), \theta \neq \theta_0, \phi \neq \phi_0, d \neq d_0, \lambda \neq \lambda_0\}$ has positive Lebesgue measure, and the matrix

$$\begin{aligned} \Omega &= (e_{i,j})_{i,j=1,\dots,(r+1)} \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \nabla_{\tau,d} \log k(\omega; \theta_0, \phi_0, d_0, \lambda_0) \nabla_{\tau,d}^T \log k(\omega; \theta_0, \phi_0, d_0, \lambda_0) d\omega \end{aligned}$$

is positive definite.

GHR A.4

θ_0, ϕ_0 are interior points to Θ, Φ and $\lambda \in [0, \pi]$ and $d \in (0, 1/2)$ unless $\lambda \in 0, \pi$ when $d \in (0, 1/4)$.

GHR A.5

$Y_t = \sum_{j=0}^{\infty} v_j \epsilon_{t-j}$ with $\sum_{j=0}^{\infty} v_j^2 < \infty$, and $v_0 = 1$ and ϵ_j is ergodic and $E[\epsilon_t | F_{t-1}] = 0$, $E[\epsilon_t^2 | F_{t-1}] = \sigma_\epsilon^2$ a.s. and $E[\epsilon_t^i | F_{t-1}] = \mu_i$ a.s. $i=3,4$, such that μ_3, μ_4 are non-stochastic and F_t is the σ -algebra generated by $\epsilon_s, s \leq t$ also for some $\delta > 0 \max_t E[\epsilon_t]^{4+\delta} < \infty$.

GHR A.6

Uniformly in $\omega \in [0, \pi] \setminus \lambda$, the function $v(\omega) = \sum_{j=0}^{\infty} v_j e^{ij\omega}$ has the property $|(d/d\omega)v(\omega)| = O\left(\frac{|v(\omega)|}{|\omega - \lambda|}\right)$.

References

- Adenstedt R (1974) On large-sample estimation for the mean of a stationary random sequence. *Ann Stat* 2(6):1095–1107
- Alomari H, Ayache A, Fradon M, Olenko A (2020) Estimation of cyclic long-memory parameters. *Scand J Stat* 47:104–133. <https://doi.org/10.1111/sjos.12404>
- Anderson T (1971) The statistical analysis of time series. Wiley, New York
- Arteche J (1998) Seasonal and cyclical long-memory in time series. PhD thesis, London School of Economics
- Arteche J (2000) Gaussian semiparametric estimation in seasonal/cyclical long memory time series. *Kybernetika* 36(3):279–310
- Arteche J (2007) The analysis of seasonal long memory: the case of Spanish inflation. *Oxford Bull Econ Stat* 69(6):749–772
- Arteche J, Robinson P (2000) Semiparametric inference in seasonal and cyclical long memory processes. *J Time Ser Anal* 21(1):1–25
- Artiach M, Arteche J (2006) Estimation of frequency in SCLM models. *Proceedings in computational statistics 2006*. Physica-Verlag, New York, pp 1147–1155
- Ayache A, Fradon M, Nanayakkara R, Olenko A (2020) Asymptotic normality of simultaneous estimators of cyclic long-memory processes. <https://arxiv.org/abs/2011.06229v1>
- Baillie R, Bollerslev T, Mikkelsen H (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J Econom* 74:3–30
- Bardet J, Bertrand P (2010) A non-parametric estimator of the spectral density of a continuous-time Gaussian process observed at random times. *Scand J Stat* 37:458–476. <https://doi.org/10.1111/j.1467-9469.2009.00684.x>
- Baumont P, Smallwood A (2019) Conditional sum of squares estimation of multiple frequency long memory models. Working paper. <https://doi.org/10.13140/RG.2.2.12795.87845>
- Baumont P, Smallwood A (2020) Inference for estimators of generalized long memory processes. Working paper. <https://www.researchgate.net/publication/345018058>
- Beran J (1993) Fitting long-memory models by generalized linear regression. *Biometrika* 80(4):817–822
- Beran J, Feng Y, Ghosh S, Kulik R (2013) Long memory processes. Wiley, New York
- Bertelli S, Caporin M (2002) A note on calculating autocovariances of long-memory processes. *J Time Ser Anal* 23(5):503–508

- Betancourt M, Byrne S, Livingstone S, Girolami M (2014) The geometric foundations of Hamiltonian Monte Carlo. [arXiv:1410.5110](https://arxiv.org/abs/1410.5110)
- Bloomfield P (1973) An exponential model for the spectrum of a scalar time series. *Biometrika* 60(2):217–226
- Bollerslev T, Mikkelsen H (1996) Modeling and pricing long memory in stock market volatility. *J Econom* 73:151–184
- Borchers H (2018) *pracma*: practical numerical math functions. <https://CRAN.R-project.org/package=pracma>. R package version 2.1.8
- Bordignon S, Caporin M, Lisi F (2007) Generalized long-memory GARCH models for intra-daily volatility. *Comput Stat Data Anal* 51:5900–5912
- Bos C, Koopman S, Ooms M (2014) Long memory with stochastic variance model: a recursive analysis for US inflation. *Comput Stat Data Anal* 76:144–157
- Boubaker H (2015) Wavelet estimation of Gegenbauer processes: simulation and estimation application. *Comput Econ* 46:551–574
- Box G, Jenkins G (1976) *Time series analysis: forecasting and control*. Holden-Day, San Francisco
- Brockwell P, Davis R (1991) *Time series: theory and methods*, 2nd edn. Springer, New York
- Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017) Stan: A probabilistic programming language. *J Stat Softw* 76(1):1–32
- Chan N, Palma W (1998) State space modelling of long-memory processes. *Ann Stat* 26(2):719–740
- Chung C (1996a) Estimating a generalized long memory process. *J Econom* 73:237–259
- Chung C (1996b) A generalized fractionally integrated autoregressive moving-average process. *J Time Ser Anal* 17(2):111–140
- Clette F, Svalgaard L, Vaquero J, Cliver E (2014) Revisiting the sunspot number—a 400 year perspective on the solar cycle. *Space Sci Rev* 186(1–4):35–103
- Diongue A, Guégan D (2008) The k-factor Gegenbauer asymmetric power GARCH approach for modelling electricity spot price dynamics. Documents de travail du Centre d’Economie de la Sorbonne
- Dissanayake G (2015) Advancement of fractionally differenced Gegenbauer processes with long memory. PhD thesis, School of Mathematics and Statistics, University of Sydney
- Dissanayake G, Peiris S (2011) Generalized fractional processes with conditional heteroscedasticity. *Sri Lankan Journal of Applied Statistics* 12:1–12
- Dissanayake G, Peiris S, Proietti T (2014) Estimation of generalized fractionally differenced processes with conditionally heteroskedastic errors. In: Ruiz I, Garcia G (eds) *International work conference on time series analysis*, Proceedings ITISE, vol 2. Granada, Copicentro Granada S L., pp 871–890
- Dissanayake G, Peiris S, Proietti T (2016) State space modelling of Gegenbauer processes with long memory. *Comput Stat Data Anal* 100:115–130
- Doppelt R, O’Hara K (2019) Posterior sampling in two classes of multivariate fractionally integrated models. *Aust N Z J Stat* 39(3):295–311
- Durham G, Geweke J, Porter-Hudak S, Sowell F (2019) Bayesian inference for ARFIMA models. *J Time Ser Anal* 40:388–410
- Espejo R, Leonenko N, Ruiz-Medina M (2014) Gegenbauer random fields. *Random Oper Stoch Equ* 22(1):1–16
- Espejo R, Leonenko N, Olenko A, Ruiz-Medina M (2015) On a class of minimum contrast estimators for Gegenbauer random fields. *TEST* 24:657–680
- Ford C (1991) The Gegenbauer and Gegenbauer AR & MA long memory time series models. PhD thesis, Southern Methodist University
- Gelman A, Rubin D (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7(4):457–472
- Geweke J, Porter-Hudak S (1983) The estimation and application of long memory time series models. *J Time Ser Anal* 4(4):221–238
- Giraitis L, Leipus R (1995) A generalized fractionally differencing approach in long-memory modelling. *Lith Math J* 35(1):53–65
- Giraitis L, Taqqu M (1999) Whittle estimator for finite-variance non-Gaussian time series with long memory. *Ann Stat* 27(1):178–203
- Giraitis L, Hidalgo J, Robinson P (2001) Gaussian estimation of parametric spectral density with unknown pole. *Ann Stat* 29(4):987–1023
- Graves T, Gramacy R, Franzke C, Watkins N (2015) Efficient Bayesian inference for natural time series using ARFIMA processes. *Nonlinear Process Geophys* 22:679–700

- Gray H, Zhang N, Woodward W (1989) On generalized fractional processes. *J Time Ser Anal* 10(3):233–257
- Gray H, Zhang N, Woodward W (1994) On generalized fractional processes—a correction. *J Time Ser Anal* 15(5):561–562
- Hannan E (1973) The asymptotic theory of linear time-series models. *J Appl Probab* 10(1):130–145
- Haslett J, Raftery A (1989) Space-time modelling with long-memory dependence: assessing Ireland's wind power resource. *J R Stat Soc C* 38:1–50
- Hidalgo J (2005) Semiparametric estimation for stationary processes whose spectra have an unknown pole. *Ann Stat* 33(4):1843–1889
- Hidalgo J, Soulier P (2004) Estimation of the location and exponent of the spectral density of a long memory process. *J Time Ser Anal* 25(1):55–81
- Hoffman D, Gelman A (2014) The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res* 15:1593–1623
- Holan S, McElroy T (2012) Bayesian seasonal adjustment of long memory time series, chapter 17. Chapman and Hall/CRC, pp 403–429
- Holan S, McElroy T, Chakraborty S (2009) A Bayesian approach to estimating the long memory parameter. *Bayesian Anal* 4(1):159–190
- Holan S, Lund R, Davis G (2010) The ARMA alphabet soup: a tour of ARMA model variants. *Stat Surv* 4:232–274
- Hosking J (1981) Fractional differencing. *Biometrika* 68(1):165–176
- Hosking J (1982) Some models of persistence in time series. In: Anderson OD (ed) *Time series analysis: theory and practice, Proceedings of the international conference held at Valencia, June 1981*, vol 1, pp 641–653
- Hoyt D, Schatten K (1998a) Group sunspot numbers: a new solar activity reconstruction. *Sol Phys* 179:189–219
- Hoyt D, Schatten K (1998b) Group sunspot numbers: a new solar activity reconstruction. *Sol Phys* 181:491–512
- Hsu N, Tsai H (2009) Semiparametric estimation for seasonal long-memory time series using generalized exponential models. *J Stat Plan Inference* 139:1992–2009
- Hunt R, Peiris M, Weber N (2021) A general frequency domain estimation method for Gegenbauer processes. *J Time Ser Econ* 13(2):119–144. <https://doi.org/10.1515/jtse-2019-003>
- Hurst H (1951) Long-term storage capacity of reservoirs. *Trans Am Soc Civ Eng* 116:770
- Hurst H (1956) The problem of long-term storage in reservoirs. *Int Assoc Sci Hydrol* 1(3):13–27
- Hurvich C (2002) Multistep forecasting of long memory series using fractional exponential models. *Int J Forecast* 18(2):167–179
- Keeling C, Bacastow R, Carter A, Piper S, Whorf T, Heimann M, Mook W, Roeloffzen H (1989) A three-dimensional model of atmospheric CO₂ transport based on observed winds: 1. analysis of observational data. Aspects of climate variability in the Pacific and the Western Americas 55:165–236
- Ko K, Vannucci M (2006) Bayesian wavelet analysis of autoregressive fractionally integrated moving-average processes. *J Stat Plan Inference* 136(10):3415–3434
- Kouamé F, Hili O (2008) Minimum distance estimation of k-factor GARMA processes. *Stat Probab Lett* 78(18):3254–3261
- Li D, Robinson P, Shang H (2020) Long-range dependent curve time series. *J Am Stat Assoc* 115(530):957–971. <https://doi.org/10.1080/01621459.2019.1604362>
- Ling S, Li W (1997) On fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity. *J Am Stat Assoc* 92(439):1184–1194
- Liseo B, Marinucci D, Petrella L (2001) Bayesian semiparametric inference on long-range dependence. *Biometrika* 88(4):1089–1104
- Lu Z, Guegan D (2011) Estimation of time-varying long memory parameter using wavelet method. *Commun Stat Simul Comput* 40(4):596–613
- Luethi D, Erb P, Oztiger S (2018) FKF: Fast Kalman Filter. <https://CRAN.R-project.org/package=FKF>. R package version 0.1.5
- Lustig A, Charlot P, Marimoutou V (2017) The memory of ENSO revisited by a 2-factor Gegenbauer process. *Int J Climatol* 37:2295–2303
- Makarava N, Benmehdi S, Holschneider M (2011) Bayesian estimation of self-similarity exponent. *Phys Rev E* 84(2):021109
- Mann H, Wald A (1943) On the statistical treatment of linear stochastic difference equations. *Econometrica* 11(3–4):173–220

- McElroy T, Holan S (2012) On the computation of autocovariances for generalized Gegenbauer processes. *Stat Sin* 22:1661–1687
- McElroy T, Holan S (2016) Computation of the autocovariances for time series with multiple long-range persistencies. *Comput Stat Data Anal* 101:44–56
- Morris J, Peravali R (1999) Minimum-bandwidth discrete-time wavelets. *Signal Process* 76:181–193
- Moulines E, Soulier P (1999) Broadband log-periodogram regression of time series with long-range dependence. *Ann Stat* 27(4):1415–1439, 08
- Narukawa M (2016) Semiparametric Whittle estimation of a cyclical long-memory time series based on generalised exponential models. *J Nonparametric Stat* 28(2):272–295
- Nordman D, Lahiri S (2006) A frequency domain empirical likelihood for short- and long-range dependence. *Ann Stat* 34(6):3019–3050
- Pai J, Ravishanker N (1995) Exact likelihood function forms for an ARFIMA process. In: Bock H, Polasek W (eds) *Proceedings of the 19th annual conference of the GfKI, data analysis and information systems*, vol 19. Springer, Heidelberg, pp 323–331
- Pai J, Ravishanker N (1998) Bayesian analysis of autoregressive fractionally integrated moving average processes. *J Time Ser Anal* 19(1):99–112
- Palma W (2006) *Long-memory time series theory and methods*. Wiley, Hoboken
- Palma W, Chan N (2005) Efficient estimation of seasonal long-range dependent processes. *J Time Ser Anal* 26:863–892
- Petris G (1997) *Bayesian spectral analysis of long memory time series*. PhD thesis, Duke University
- Phillip A, Chan J, Peiris S (2018) Bayesian estimation of Gegenbauer long memory processes with stochastic volatility: methods and applications. In: *Studies in nonlinear dynamics & econometrics*, vol 22. De Gruyter, pp 1–29
- Pourahmadi M (1983) Exact factorization of the spectral density and its application to forecasting and time series analysis. *Commun Stat Theory Methods* 12:2085–2094
- Rainville E (1960) *Special functions*. Macmillan, New York
- Ravishanker N, Ray B (1997) Bayesian analysis of vector ARFIMA processes. *Aust N Z J Stat* 39(3):295–311
- Robinson P (1994) Rates of convergence and optimal spectral bandwidth for long range dependence. *Probab Theory Relat Fields* 99:443–473
- Robinson P (1995a) Gaussian semiparametric estimation of long range dependence. *Ann Stat* 23(5):1630–1661
- Robinson P (1995b) Log-periodogram regression of time series with long range dependence. *Ann Stat* 23(3):1048–1072
- Robinson P (2006) Conditional-sum-of-squares estimation of models for stationary time series with long memory. *IMS lecture notes monograph series, Time series and related topics* 52:130–137. <https://doi.org/10.1214/074921706000000996>
- Rypdal K, Rypdal M, Fredriksen H (2015) Spatio-temporal long-range persistence in earth's temperature field: analysis of stochastic-diffusive energy balance models. *J Clim* 28:8379–8395
- Schuster A (1906) On the periodicities of sunspots. *Philos Trans R Soc A* 206:69–100
- Shapiro AI, Schmutz W, Rozanov E, Schoell M, Haberleiter M, Shapiro AV, Nyeki S (2011) A new approach to the long-term reconstruction of the solar irradiance leads to large historical solar forcing. *Astronomy Astrophys* 529:A67
- Smallwood A, Beaumont P (2004) Multiple frequency long memory models. Technical report, University of Oklahoma Working Papers
- Stan Development Team (2019) *Stan modeling language users guide and reference manual*. <http://mc-stan.org>
- Sutcliffe J, Hurst S, Awadallah A, Brown E, Hamed K (2016) Harold Edwin Hurst: the Nile and Egypt, past and future. *Hydrol Sci J* 61(9):1557–1570
- Szegő G (1959) *Orthogonal polynomials*. AMS, New York
- Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P (2019) Rank-normalization, folding, and localization: an improved $\hat{R}$ for assessing convergence of MCMC. <https://arxiv.org/abs/1903.08008>
- Waldmeier M (1961) The sunspot activity in the years 1610–1960. *Zurich Schulthess, Zurich*
- Whitcher B (2001) Simulating Gaussian stationary processes with unbounded spectra. *J Comput Graph Stat* 10(1):112–134
- Whitcher B (2004) Wavelet-based estimation for seasonal long-memory processes. *Technometrics* 46(2):225–238

- Whitcher B (2015) waveslim: basic wavelet routines for one-, two- and three-dimensional signal processing. URL <https://CRAN.R-project.org/package=waveslim>. R package version 1.7.5
- Whittle P (1953) The analysis of multiple stationary time series. *J Roy Stat Soc B* 15(1):125–139
- Woodward W, Cheng Q, Gray H (1998) A k-factor GARMA long-memory model. *J Time Ser Anal* 19(4):485–504
- Woodward W, Gray H, Elliott A (2017) *Applied time series analysis with R*, 2nd edn. CRC Press, Taylor and Francis Group, Boca Raton
- Wu H, Peiris S (2018) An introduction to vector Gegenbauer processes with long memory. *Stat* 7:e197
- Yajima Y (1996) Estimation of the frequency of unbounded spectral densities. In: *Proceedings of the business and economic statistical section 4-7*. American Statistical Association
- Yule U (1927) On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philos Trans R Soc A* 226:267–298

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.