

Fast Estimation of Time Series with Multiple Long-Range Persistencies

Tucker S. McElroy*

Scott H. Holan^{† ‡}

Abstract

Gegenbauer processes allow for flexible and convenient modeling of time series data with multiple spectral peaks, where the qualitative description of these peaks is via the concept of cyclical long-range dependence. The Gegenbauer class is extensive, including ARFIMA, seasonal ARFIMA, and GARMA processes as special cases. Model estimation is challenging for Gegenbauer processes when multiple zeroes and poles occur in the spectral density, because the autocovariance function is laborious to compute. The method of splitting – essentially computing autocovariances by convolving long memory and short memory dynamics – is only tractable when a single long memory pole exists. The main contribution of this paper is to propose an additive decomposition of the spectrum into a sum of spectra that each have a single singularity, so that an efficient splitting method can be applied to each term and then aggregated. This approach differs from McElroy and Holan (2012), which handles all poles in the spectral density at once through a careful analysis of truncation error. This paper's technique allows for fast estimation of time series with multiple long-range dependencies, which we illustrate numerically and through several case-studies.

Key Words: Gegenbauer, Long memory, Long-range dependence, Quasi-biennial oscillations, Seasonality, Spectral density

1. Introduction

This paper is concerned with the modeling of time series data with multiple persistent periodicities, where the persistency indicates a slow rate of decay in sample autocorrelations. Many time series of interest in macroeconomics, finance, and meteorology have persistency at nonzero frequencies, which are properly described as infinite (but integrable) peaks in the spectral density; see Ferrara and Guégan (2000), Bisaglia, Bordignon, and Lisi (2003), Soares and Souza (2006), Talamantes, Behseta, and Zender (2007), and Gil-Alana (2008). When two or more such peaks are present in the spectrum of the underlying process, the commonly used sets of models – ARFIMA (Hosking, 1981), seasonal ARFIMA (Porter-Hudak, 1990), Gegenbauer (Gray, Zhang, and Woodward, 1989), etc. – are challenging to compute. That is, the autocovariance function of such models is laborious to calculate, although in the case of a single long memory spectral peak there exists certain methods that facilitate computation. Recently, McElroy and Holan (2012) developed a general representation for processes with multiple spectral peaks, and showed how one can accurately determine the Wold coefficients of such processes from the cepstral representation (Bloomfield,

*Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100

[†]Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO, 65211-6100

[‡]Disclaimer: This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

1973). Further, McElroy and Holan (2012) produced error terms (expressed as hypergeometric functions) that accounted for truncation error in the calculation of autocovariances. The current paper presents a different algorithm that is both fast and accurate, and goes back to the older “splitting” method of Bertelli and Caporin (2002).

Briefly, in a nonparametric or semiparametric conception of the spectral density f , in which possibly several persistent phenomena exist, it is natural to write f as a product of functions that represent the poles in the spectrum, together with a term that corresponds to short memory dynamics, being both bounded and bounded away from zero. In our discussion, we may also be interested in anti-persistent behavior, or negative memory (McElroy and Politis, 2014), which corresponds to spectral troughs; i.e., a zero in f . The autocovariance sequence corresponding to f is equal to the convolution of the autocovariances of each term in the product; if there is only one pole/zero in the spectral density, then we must convolve the corresponding autocovariance function with that of the short memory portion, which can be safely truncated. This truncation of the short memory autocovariance function – presuming that the long memory autocovariance function is known analytically, or can be calculated to high accuracy – is called the splitting method, and is developed in Bertelli and Caporin (2002) and Hurvich (2002). This ingenious algorithm does not extend directly to having two or more poles/zeros, because then there are at least three functions being convolved, which means there are at least two infinite summations – and there seems to be no way to truncate these without a potentially tremendous loss of accuracy. The main insight of this paper is that a function f with $k \geq 1$ poles/zeros can be re-expressed as the sum of k distinct functions, each of which has a single pole/zero, together with its own distinct short memory spectrum. This is akin to a partial fraction decomposition, though instead of polynomials, we must work with somewhat more complicated frequency domain functions. The splitting method can be applied to each of the k distinct terms, and the corresponding autocovariances simply summed up.

This algorithm allows us to consider processes with a potentially large number of spectral poles/zeros. In fact, in one of our illustrations (Section 5), we have $k = 26$ seasonal poles! This is done without much worry about the potential loss of accuracy due to truncation, because we only have the approximation error due to the splitting method, which is typically negligible due to the rapid decay of autocovariances of short memory processes. Thus, complex time series data with multiple persistencies (where the frequency of persistency may be either unknown – in which case it is a parameter to be estimated – or known) can now be readily treated, whereas previously only special cases (e.g., seasonal ARFIMA, or the two-pole cases of McElroy and Holan (2012)) could realistically be handled. We demonstrate this algorithmic breakthrough on a broad array of series.

The resulting algorithms have been implemented for both a Gaussian likelihood and a Whittle likelihood, where the short memory portion is semi-parametrically described via the device of a cepstral model – this is just a Fourier expansion of the log spectrum. The aforementioned algorithms have been coded using the R programming language (R Development Core Team, 2014) and are available upon request from the first author. Adapting them to versions where the short memory portion is described via an ARMA process

is straightforward, and thus is not pursued here. Another application would be to consider several latent components, each of which are potentially cyclically persistent – in this case, the autocovariance of the observed process is given as the sum of the autocovariances associated with each latent process, as described in a more limited context in Holan and McElroy (2012).

The focus of this paper is primarily algorithmic and can be adapted across a wide-array of statistical models. Specifically, the algorithms herein should be useful in computing *any* of the cyclical long memory models currently being utilized to analyze time series with multiple persistencies. Section 2 provides the general semiparametric framework and shows how popular models can be embedded therein. Section 3 thoroughly develops the algorithm, whereas Section 4 presents numerical studies of speed and accuracy. We provide several illustrations in Section 5, utilizing our algorithm to fit data with multiple persistencies across a variety of scientific disciplines. Concluding discussion is provided in Section 6, with derivations in the Appendix.

2. A Semiparametric Description of Multiple Persistencies

The k -GEXP model was introduced in McElroy and Holan (2012), and is recapitulated here. Let $\{X_t\}$ be a mean zero covariance stationary time series. We will suppose that $\{X_t\}$ follows a k -factor Generalized Exponential model, or k -GEXP model, whose spectral density can be written compactly as

$$f(\lambda) = f_0^{-a}(\lambda) f_\pi^{-b}(\lambda) \prod_{j=1}^k f_{\omega_j}^{-c_j}(\lambda) g(\lambda), \quad (1)$$

where the functions of frequency f_ϑ are defined via

$$f_\vartheta(\lambda) = \begin{cases} |1 - e^{-i\lambda}|^2 & \text{if } \vartheta = 0 \\ |1 + e^{-i\lambda}|^2 & \text{if } \vartheta = \pi \\ |1 - e^{-i\vartheta} e^{-i\lambda}|^2 |1 - e^{i\vartheta} e^{-i\lambda}|^2 & \text{if } \vartheta \neq 0, \pi \end{cases}. \quad (2)$$

and the function g is bounded, representing the short memory portion of the spectrum. Also, the parameters a, b, c_1, \dots, c_k are each bounded in $(-1/2, 1/2)$ in order to guarantee stationarity. The frequencies ω_ℓ are distinct from one another, and not equal to zero or π . When a parameter $a, b,$ or c_j is positive, there is a corresponding pole in the spectral density at frequency zero, $\pi,$ or ω_j respectively – this is the case of long memory. On the other hand, negative parameters correspond to a zero in the spectrum, and correspond to intermediate memory (or negative memory, also called anti-persistence by some authors; see Beran (2010) and the references therein). Note that the total number of poles/zeros in (1) is at most $k + 2$, with a factor being omitted when the corresponding persistency parameter is zero.

The short memory spectrum g corresponds to an EXP(q) model (Bloomfield, 1973) so that

$$g(\lambda) = \sigma^2 \exp \left\{ \sum_{j=1}^q g_j \cos(\lambda j) \right\} = \exp \left\{ g_0 + \frac{1}{2} \sum_{0 < |j| \leq q} g_j e^{-i\lambda j} \right\}, \quad (3)$$

where $g_{-j} = g_j$. So the innovation variance σ^2 of the model is equal to $\exp(g_0)$. Here q can be viewed as a finite integer, with the case $q = \infty$ allowing for an arbitrary invertible short memory spectral density. Putting together the short memory and long memory portions yields the k -GEXP(q) model: k parameters govern the degree of persistence, with up to another k parameters governing the peak locations (when they are not known *a priori*) or periodicities, and q parameters governing the short memory dynamics. Typically a mean parameter would also be utilized, whereas the innovation variance can be estimated via utilizing a profile likelihood.

By shifting the various ω_ℓ , one can capture a variety of quasi-cyclical phenomena, such as seasonality, cyclicity, and long-term movements in the process. For a mean-zero Gaussian process, the spectrum f contains all of the model parameters of interest, so that one only needs compute the autocovariances, i.e., the inverse Fourier Transform (FT), of the spectrum in order to determine the likelihood function. Therefore both a classical and a Bayesian treatment of model estimation requires an efficient method of computing autocovariances for any given values of the associated model parameters.

One problem with the model as it stands, is that it is not identifiable when $k > 1$. This is because we can swap any two frequencies ω_1, ω_2 , along with their corresponding memory parameters, and obtain the exact same value of the spectral density. In order to guarantee identifiability in practice, one might put *a priori* restrictions on the frequencies ω_ℓ , such as insisting that each one belongs to a disjoint subset of $(0, \pi)$. In our own implementations, we ensure that the various ω_k are distinct via the following parametrization: for $\theta_1, \theta_2, \dots, \theta_k \in \mathbb{R}$, and $\theta_0 = 0$, set

$$\omega_{j+1} = \frac{\pi \exp(\theta_{j+1}) + \omega_j}{\exp(\theta_{j+1}) + 1}$$

for $j = 0, 1, \dots, k - 1$. This ensures that $\omega_{j+1} \in (\omega_j, \pi)$.

For many applications of the model, the frequencies are known ahead of time and therefore are not even parameters. For example, extremely persistent cyclical effects typically correspond to trends or seasonality, for which the corresponding frequencies are known. Seasonality is a term referring to a situation where the sampling frequency relative to annual time units is an integer. If there are p seasons, where p is integer, and the sampling frequency coincides with p , then seasonality might manifest in the data. If seasonality is present then it can often be described through peaks in the spectral density at frequencies $2\pi j/p$ for $1 \leq j \leq p$. Trend, on the other hand, is often described through polynomial functions or stochastic components, and tends to contribute spectral mass at the low frequencies.

However, there may be some occasions where the cyclical frequencies are not known ahead of time. For phenomena that are not tied to the terrestrial cycle (approximately 365 days), the sampling frequency may have no special meaning in terms of seasons. For astronomical data, the existence of long-range cycles with strong persistence is a known phenomenon, where the period can involve non-integer multiples of years. One of our case studies involves a time series of wind measurements, and has long-range persistencies – one of which does not correspond to a regular seasonal effect. Such phenomena might manifest as peaks in the spectral density, but the peak frequency is not known *a priori*, and must be estimated.

A secondary issue is weak identifiability problems due to possible confounding of the short memory g and the zero-pole (ZP) contributions to (1). Because values of the exponents close to zero make the spectrum approximately flat, excepting a kink, the same behavior can be approximated by the cepstral coefficients. In finite samples it is possible to have confusion between these elements, and therefore we have found it advisable to keep the cepstral order low (i.e., $q \leq 5$). In practice, this issue is not terribly important, because highly persistent peaks/troughs have a distinct spectral shape that is difficult to capture, or even approximate, with a short memory spectral density.

3. Derivation of the New Algorithm

A key challenge in the computation of the autocovariance function (acf) for multiple-pole long memory processes is the trade-off between computational efficiency and accuracy mentioned in McElroy and Holan (2012); truncation techniques attempt to balance the horns of this dilemma. A splitting method, such as that proposed in Bertelli and Caporin (2002) and Hurvich (2002), can be successful when there is only a single long memory pole in the spectrum; but with multiple poles in the spectrum, long memory autocovariances become convolved with other long memory autocovariances, implying that truncation of the convolution involves a substantial approximation error. The novel idea of this paper is to factor the multi-pole form of the spectrum into a sum of several components, each of which only contains a single long memory pole. Then each component can have its autocovariances calculated via splitting, and the results summed to produce the desired result.

Some notation that we require is given next. We use $z = e^{-i\lambda}$ where $\lambda \in [-\pi, \pi]$, and for any real-valued function $g(\lambda)$, its inverse FT is defined to be the (possibly complex) sequence $\gamma_h(g)$, defined via the formula

$$\gamma_h(g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) z^{-h} d\lambda.$$

Conversely, the FT of a sequence $\{a_h\}$ is the function $\sum_{h=-\infty}^{\infty} a_h z^h$. This establishes our conventions concerning 2π . In the case that g is even, its inverse FT sequence is symmetric. Furthermore, if g is the spectral density of a time series, its inverse FT is interpretable as the acf of the process.

The key result is a factorization that we describe below, which is akin to a partial fractions decomposition of the spectrum. Let $d^+ = \max\{d, 0\}$ and $d^- = \min\{d, 0\}$ where d can be any of the exponents a, b, c_j – if a particular term is not present in the spectrum, this is handled by setting the exponent to zero. So $d^+ + d^- = d$ throughout. Then, apart from the short memory portion g , we can split f into two factors f^- and f^+ corresponding to the terms having zeroes and the terms having poles, respectively. That is, $f = f^- f^+ g$ with

$$f^- = f_0^{-a^-} f_{\pi}^{-b^-} \prod_{j=1}^k f_{\omega_j}^{-c_j^-} \quad \text{and} \quad f^+ = f_0^{-a^+} f_{\pi}^{-b^+} \prod_{j=1}^k f_{\omega_j}^{-c_j^+}.$$

We wish to factorize f into a sum of ZP terms that each involve a single component of the f_{ω} terms, together

with a short memory portion. Toward that end, we define the ZP “version” of the spectrum via

$$f_{ZP} = f_0^{-a} 1_{\{a \neq 0\}} + f_\pi^{-b} 1_{\{b \neq 0\}} + \sum_{j=1}^k f_{\omega_j}^{-c_j} 1_{\{k > 0\}}. \quad (4)$$

With this notation, we can state the algebraic factorization result.

Proposition 1 *Given a k -GEXP spectrum of the form (1) and factors given by (2), there exists a short memory spectrum ν such that*

$$\nu = \left(f_0^{-a^-} f_\pi^{b^+} 1_{\{a \neq 0\}} + f_\pi^{-b^-} f_0^{a^+} 1_{\{b \neq 0\}} \right) \prod_{j=1}^k f_{\omega_j}^{c_j^+} + f_0^{a^+} f_\pi^{b^+} \sum_{j=1}^k \prod_{\ell \neq j} f_{\omega_\ell}^{c_\ell^+} f_{\omega_j}^{-c_j^-} 1_{\{k > 0\}}$$

$$\tau = f^- \cdot g / \nu$$

$$f = \tau \cdot f_{ZP}.$$

The function ν is bounded, and is also strictly positive if there are at least two ZP factors – in which case it corresponds to a short memory process. The function τ has only short memory and negative memory (it has no poles).

This factorization yields a decomposition of f into a sum of up to $k + 2$ terms, each of which involves τ divided by a factor f_ϑ^{-d} . Therefore, the inverse FT of each summand of f can be computed by convolving the inverse FT of numerator τ and denominator f_ϑ^{-d} via the splitting method. This is advantageous, because: (a) exact formulas for the acfs of the ZP factors f_ω^{-d} are known, and (b) the inverse FT for τ can be calculated to decent approximation (using the techniques of McElroy and Holan (2012)) since they have short/negative memory. The formulas needed are summarized in the following result.

Proposition 2 *The autocovariance sequence of the ZP factor τ / f_ϑ^d is given by the following formulas:*

$$\begin{aligned} \gamma_h \left(\tau / f_\vartheta^d \right) &= \sum_{j=-\infty}^{\infty} \gamma_j(\tau) \gamma_{h+j}(f_\vartheta^{-d}) \\ \gamma_h(f_\vartheta^{-d}) &= \begin{cases} \frac{\Gamma(h+d)\Gamma(1-2d)}{\Gamma(h-d+1)\Gamma(1-d)\Gamma(d)} & \text{if } \vartheta = 0 \\ (-1)^h \frac{\Gamma(h+d)\Gamma(1-2d)}{\Gamma(h-d+1)\Gamma(1-d)\Gamma(d)} & \text{if } \vartheta = \pi \\ \sqrt{\pi} \Gamma(1-2d) (2 \sin \vartheta)^{1/2-2d} \left\{ P_{h-1/2}^{2d-1/2}(\cos \vartheta) + (-1)^h P_{h-1/2}^{2d-1/2}(-\cos \vartheta) \right\} & \text{if } \vartheta \neq 0, \pi \end{cases} \\ P_{h-1/2}^{2d-1/2}(z) &= \left(\frac{2h-2}{h-2d} \right) z P_{h-3/2}^{2d-1/2}(z) - \left(\frac{2d+h-2}{h-2d} \right) P_{h-5/2}^{2d-1/2}(z). \end{aligned}$$

The functions $P^{2d-1/2}(z)$ are Legendre functions, and their recursive definition is initialized with

$$\begin{aligned} P_{-1/2}^{2d-1/2}(z) &= \left(\frac{1+z}{1-z} \right)^{d-1/4} \frac{F(1/2, 1/2, 3/2-2d; (1-z)/2)}{\Gamma(3/2-2d)} \\ P_{1/2}^{2d-1/2}(z) &= \left(\frac{1+z}{1-z} \right)^{d-1/4} \frac{F(-1/2, 3/2, 3/2-2d; (1-z)/2)}{\Gamma(3/2-2d)}. \end{aligned}$$

The first formula of Proposition 2, namely the convolution of the short memory τ autocovariances with the long/negative memory ZP autocovariances, is called a “splitting formula.” We can truncate the splitting formula at some threshold lag J without much cost to accuracy, because the $\{\gamma_j(\tau)\}$ sequence is absolutely summable, and hence decays exponentially. However, one caution is that the truncation of $\{\gamma_j(\tau)\}$ at J (i.e., setting $\gamma_j(\tau) = 0$ for $j > J$) can result in a sequence that is not positive definite, such that the approximation to the inverse FT

$$\gamma_h \left(\tau / f_{\vartheta}^d \right) \approx \gamma_0(\tau) \gamma_h(f_{\vartheta}^{-d}) + \sum_{j=1}^J \gamma_j(\tau) \left\{ \gamma_{h+j}(f_{\vartheta}^{-d}) + \gamma_{h-j}(f_{\vartheta}^{-d}) \right\}$$

need not be positive definite. This may seem to be a quibble, but we have found it to be quite important in practice, because a non positive definite sequence implies negative eigenvalues in the corresponding Toeplitz covariance matrix, obviating efforts at inversion. Our remedy is to construct a J -dependent process that has well-defined spectral density that closely approximates τ . One way to do this is to determine the Wold coefficients of τ (obtained via determining the cepstral coefficients, and utilizing the recursions described in Hurvich (2002)), and define $\tau^{(J)}$ to be spectral density of the MA(J) process corresponding to retaining only the first $J + 1$ Wold coefficients (the first Wold coefficient is always equal to one). Letting $\xi_k = 2\pi^{-1} \int_0^{\pi} \log \tau(\lambda) \cos(\lambda k) d\lambda$ be the k th cepstral coefficient, for $k \geq 1$, we have Pourahmadi’s recursions (Pourahmadi, 1984) involving the Wold coefficients ψ_j (with $\psi_0 = 1$):

$$\psi_j = \frac{1}{2j} \sum_{k=1}^j \psi_{j-k} k \xi_k. \quad (5)$$

Then $\tau^{(J)}(\lambda) = \left| \sum_{k=0}^J \psi_k e^{-i\lambda k} \right|^2$ by definition, and its autocovariance sequence is guaranteed to be positive definite; hence we utilize the approximation

$$\gamma_h \left(\tau / f_{\vartheta}^d \right) \approx \gamma_0(\tau^{(J)}) \gamma_h(f_{\vartheta}^{-d}) + \sum_{j=1}^J \gamma_j(\tau^{(J)}) \left\{ \gamma_{h+j}(f_{\vartheta}^{-d}) + \gamma_{h-j}(f_{\vartheta}^{-d}) \right\}. \quad (6)$$

There is no need to take J particularly large in practice, due to the rapid decay rate of the Wold coefficients of $\tau(z)$ – typically $J = 10$ is more than sufficient. We note that this approach to obtaining the J -dependent approximation is practicable, because the j th Wold coefficient only depends (recursively) upon the first j cepstral coefficients, as shown in Pourahmadi (1984) and Hurvich (2002). Therefore, one only needs the first J cepstral coefficients (we compute these using numerical integration, which is extremely accurate for short memory spectra). Using the recursions to obtain the first J Wold coefficients, one then computes the autocovariances of this MA(J) process.

Explicit steps of the entire algorithm are given as follows. Choose a truncation level J , and a total number of lags H for the acf sequence.

1. Read in the parameters
2. Compute $g(\lambda)$, $f^-(\lambda)$, $\nu(\lambda)$, and $\tau(\lambda)$ for a mesh of λ frequencies

3. The j th cepstral coefficient of τ is $2 \int_0^1 \log \tau(\pi x) \cos(\pi j x) dx$, which can be Riemann approximated at a given mesh
4. Obtain the Wold coefficients ψ_j for $1 \leq j \leq J$ of τ from the cepstral coefficients via (5), and compute the autocovariances of $\tau^{(J)}$ from these Wold coefficients
5. The acf of each long/negative memory pole is given by the direct formula in Proposition 2
6. Combine the acfs by convolution (6).

This algorithm might be used to do forecasting, taking the parameters in the first step to be known estimates; or this algorithm could be embedded in a routine to compute the Gaussian likelihood, in which case the parameters in Step 1 reflect the current iterates in a numerical optimization scheme. In the case of the Gaussian likelihood, the innovations algorithm (Brockwell and Davis, 1991) can be utilized, which has a loop over the sample size and is recursive in nature; similarly, a variant for the Whittle likelihood is available. The mean-zero scaled log Gaussian likelihood, based on a sample of size n from the data process – denoted by $X = [X_1, X_2, \dots, X_n]'$ – is

$$\mathcal{L}(\varphi; X) = X' \Sigma^{-1}(f_\varphi) X + \log \det \Sigma(f_\varphi), \quad (7)$$

where $\Sigma(f_\varphi)$ denotes the $n \times n$ Toeplitz covariance matrix corresponding to the spectral density f_φ , and φ abbreviates the full list of model parameters excluding the mean (in our applications, we also include a parameter for the mean). Here we utilize f_φ given by (1), and φ consists of all the memory parameters, together with the ZP locations (unless these are assumed to be known based on metadata) and the short memory cepstral coefficients g_j . This assumes some truncation order q for the EXP portion of the model. The innovations algorithm will evaluate both summands of (7), given only the acf sequence of f_φ . The Whittle likelihood is instead given by

$$\mathcal{W}(\varphi; X) = X' \Sigma(f_\varphi^{-1}) X + \log \sigma^2, \quad (8)$$

where recall that σ^2 equals the exponential of g_0 , the initial cepstral coefficient. Here we must compute the autocovariance of the inverse spectral density, but given the form of (1) this is quite easy to find: we just flip the sign on all the memory parameters d , as well as the cepstral coefficients g_j , and apply the algorithm above, followed by evaluation of (8). Together with a routine to conduct nonlinear optimization, either objective function can be utilized to fit these models to time series data (R code is available from the authors), resulting in Maximum Likelihood Estimates (MLEs) and Whittle Likelihood Estimates (WLEs) respectively. Mean effects can also easily be added if desired, either by direct optimization over specified regression effects or via a concentrated likelihood utilizing GLS formulas (cf. discussion in McElroy and Holan (2014)).

4. Accuracy and Speed of the Algorithm

In order to assess the accuracy of the proposed algorithm in Section 3 we conduct a numerical study in which we compare estimates of the log determinant (ldet) of the autocovariance matrix (of dimension n) with those from other approaches. By comparing the log determinants, we provide an “aggregate” discrepancy measure. However, aside from a few special cases, the true value of the ldet is unknown and needs to be approximated. Under these scenarios, we compare the ldet from the proposed acf sequence and the asymptotic ldet of Proposition 1 of McElroy and Holan (2012), denoted as $\text{ldet}_{\text{prop}}$.

In the case of fractional Gaussian noise the ldet of the autocovariance sequence can be exactly calculated or approximated to any degree of accuracy using the splitting method. In this case, examining the ldet of the newly proposed approach, we find that all estimated ldets are within $1e-4$ of $\text{ldet}_{\text{prop}}$ and within $1e-12$ of the ldet obtained using splitting method – even for $d = .45$ (for both $n = 500$ and 1000). Similarly, for FEXP processes with unit innovation variance and various short memory specifications, all estimated ldets are within $1e-4$ of the asymptotic ldet ($\text{ldet}_{\text{prop}}$) and within $1e-12$ of the ldet obtained using splitting method – even for $d = .45$ (for both $n = 500$ and 1000).

For the 1-GEXP case, we consider sample sizes of $n = 500$ and 1000 with long memory parameter c_1 ranging from $.1$ to $.45$. Additionally, the seasonal frequency is set at $\omega = .56$, with the short memory portion $g = (g_0, g_1) = (0, .75)$ – which implies unit innovation variance. In particular, we compare the difference between the ldet of the new approach and the asymptotic ldet (i.e., $\text{ldet}_{\text{prop}}$) to the difference between the ldet from the truncation approach and the asymptotic ldet (Table 1). Note that, for the method of McElroy and Holan (2012), the truncation level J_{Trunc} was set between $5,000$ and $100,000$.

From Table 1, we see that both approaches are highly accurate and closely agree with the asymptotic approximation. Further, for the 1-pole case, the number of terms needed for accurate approximation increases as the degree of persistence increases. Even for the case of $c_1 = .25$, it happens that $25,000$ terms (Wold coefficients) are needed in order to obtain an accurate approximation (for both $n = 500$ and 1000). The number of terms quickly increases to $100,000$ for $c_1 = .35$. Consequently, computation is significantly more expensive for the truncation method relative to the newly proposed algorithm.

The 2-GEXP case presents additional complications, as there are now multiple seasonal poles in the spectral density. Here, we consider sample sizes of $n = 500$ and 1000 with long memory parameters (c_1, c_2) ranging from $(.1, .2)$ to $(.1, .45)$ and $(.45, .2)$ to $(.45, .4)$. Additionally, the seasonal frequency is set at $\omega = (.1, .56)$, with the short memory portion $g = (g_0, g_1) = (0, .75)$ (implying unit innovation variance). Again, we compare the difference between the ldet of the new approach and the asymptotic ldet to the difference between the ldet from the truncation approach and the asymptotic ldet (Table 2).

From Table 2, we see that both approaches are fairly accurate and are in agreement with the asymptotic approximation. However, in contrast to the 1-GEXP case, there is no definitive correspondence between the truncation needed for the approach of McElroy and Holan (2012) (MH_{2012}) and the degree of cyclical persistence; i.e., the truncation point in the MH_{2012} approach is a non-monotonic function of the level of cyclical

persistence. Also, although for moderate persistent cases the MH_{2012} method slightly outperforms the new approach relative to the asymptotic approximation, this performance highly depends on the truncation chosen, which would be unknown in practice. Further, for a “poor” choice of truncation the MH_{2012} method and asymptotic approach could differ significantly when there is strong persistence. Finally, it should be noted, that even in the “best” case scenarios it is not clear whether the MH approach outperforms the new approach, as there is no definitive “ground-truth.” Instead, our approximations are compared relative to the asymptotic approximation, which could exhibit error due to finite samples.

Aside from accuracy, the main advantages of the new algorithm are computational efficiency and the ability to easily accommodate spectra with more than two poles. The capacity to readily compute the autocovariance sequence in the multiple-pole setting is extremely advantageous and not feasible without the algorithm of Section 3. In terms of speed, we compare the computation time of one likelihood evaluation for several cases. In particular, we compute the likelihood with the autocovariance matrix calculated using the truncation approach of McElroy and Holan (2012) and using the new algorithm. All computations are carried out in the R programming language (R Development Core Team, 2014) using a MacBook Pro 2.7 GHz Intel Core i7 with 16 GB 1600 MHz DDR3 RAM.

The first example considers the 2-GEXP specified in Table 2 with $(c_1, c_2) = (.1, .45)$. For $n = 500$ ($J_{\text{Trunc}} = 100,000$), the computation time for the MH approach is 3.074 minutes, whereas the new algorithm (with mesh size 10001) is 1.180 seconds. Similarly, for $n = 1000$ ($J_{\text{Trunc}} = 50,000$), the MH approach takes 51.410 seconds, whereas the new algorithm took only 2.478 seconds (with mesh size 10001). For $(c_1, c_2) = (.45, .4)$, with $n = 500$ ($J_{\text{Trunc}} = 50,000$) the MH computation time was 50.167 seconds, whereas with the new approach the computation took 1.229 seconds. Finally for $(c_1, c_2) = (.45, .4)$, with $n = 1000$ ($J_{\text{Trunc}} = 25,000$) the MH computation time was 15.275 seconds, whereas with the new approach the computation took 2.548 seconds.

It is important to emphasize that all computations done here are completely in R (R Development Core Team, 2014). In order to make estimation feasible, McElroy and Holan (2012) made use of Fortran subroutines using the R package *inline*. Without the use of a lower level programming language such as Fortran or C++, maximum likelihood estimation would be impractical using the approach of McElroy and Holan (2012). In contrast, the new approach easily handles such computations without the need to “optimize” computer code.

5. Case Studies

To illustrate the application of the algorithm introduced in Section 3, we examine four different datasets. The four examples demonstrate the effectiveness of our approach under various model specifications (i.e., varying number of long range persistencies). As such, no formal attempt has been made toward selecting the “optimal” candidate model (i.e., via model selection).

5.1 Quasi-biennial Oscillations

We consider a series consisting of 11,000 daily observations beginning November 19, 1964, corresponding to the east/west wind component from Truk Island (see Cressie and Wikle, 2011, Chapter 3), which are down-sampled at a rate of 50 days in order to produce suitable separation between the annual frequency and the secondary spectral peak (corresponding to a period of roughly 2.3 years). This latter effect is referred to as a quasi-biennial oscillation (QBO), and is further described in Andrews, Holton, and Leovy (1987) and Wikle et al. (1995). We then fitted a 2-GEXP(5) model with two seasonal poles to the differenced series. The seasonal poles were located at frequencies $\omega \in (0, \pi)$, with the frequency ordinates and peak strengths estimated as parameters. The cepstral parameter MLEs were $-2.644, -.567, .311, .145, -.180$. The peak locations were estimated to be .3608 and .8586, which correspond to periods of 2.385 and 1.003 years respectively, i.e., the QBO and annual effect. The latter peak was quite strong, with memory parameter estimated to be .478, whereas the QBO peak was similar, with memory parameter .472. The mean estimate was -4.589 .

Repeating the exercise with the Whittle likelihood, the cepstral parameter WLEs were $-2.516, -.526, .266, .076$, and $-.188$. The peak locations were .3597 and .8584, corresponding to periods of 2.393 and 1.003 years each. The memory parameters were .472 and .478 respectively, which very closely match the Gaussian likelihood results. Then mean estimate was -4.596 .

5.2 Industrial Production

The monthly series of Industrial Production (Federal Reserve Board) is available from 1919 onwards. Having first adjusted the data for a trading day effect, Easter holiday effects, and additive outliers, we then studied the years 1949.1 through 2007.12 in the log-scale. Nonstationary trend effects in the data required a single differencing; although a seasonal difference could be applied as well, we decided to attempt an estimation of the seasonal poles through a GEXP model instead. The data has terrestrial periodicities, being sampled at monthly intervals, and hence we expect the spectral peaks to have the form $2\pi j/12$. Therefore, the differenced log series was fitted by a 7-GEXP(3) model where the trend/seasonal poles were fixed at $\pi j/6$ for $0 \leq j \leq 6$. We might expect $a, b \neq 0$ in the model, and $c_j \neq 0$ for $1 \leq j \leq 5$, which is indeed borne out by our estimates. The choice of the order of the EXP portion was only meant for illustration and could be formally chosen through model selection methods if desired.

The cepstral MLEs were 2.447, 1.045, .410. The memory parameter MLEs were .046, .136, .470, .490, .489, .475, and .464 respectively, indicating the trend memory is fairly weak (differencing having removed most of the structure). Moreover, the first seasonal peak is somewhat weak, but the memory at peaks two through six all exceed .45, yielding pronounced effects. The estimated mean was .0027.

The cepstral WLEs were similar: 2.910, 1.249, .513. The memory parameters were $-.146, .056, .477, .491, .491, .482$, and .474; it is interesting that the frequency zero memory parameter is negative in this case, which indicates that for the raw data the point estimate for the order of trend integration is .854 – in

the nonstationary range for memory parameters. The point estimate for the first seasonal peak is also hardly distinguishable from short memory, and the results for the other peaks resemble those from the Gaussian likelihood. The mean was .0026.

5.3 Mauna Loa

The Mauna Loa time series has now been studied in several papers that treat seasonal long memory. The dataset consists of 382 monthly atmospheric CO₂ measurements collected at the summit of Mauna Loa in Hawaii, beginning in March 1958. While twin-peak models have been considered before, such as the 2-factor GARMA of Woodward, Cheng, and Gray (1998) and the 2-GEXP(4) of McElroy and Holan (2012), the periodogram indicates that a model accommodating additional peaks is warranted. Because the data is terrestrial and monthly, it is plausible that the spectral peak frequencies take the form $\pi j/6$ for $1 \leq j \leq 6$, as with the previous example. Therefore we again consider the 7-GEXP(3) model, and we compute the MLEs and WLEs for the once-differenced data (which was adjusted for missing values in a pre-processing stage). The cepstral MLEs were $-3.340, .524, -.393$, with a mean of .101. The memory parameters were .331, .498, .488, .371, .314, .035, $-.051$ respectively, which indicate a strong trend effect, as well as strong seasonality in the first through fourth seasonal frequencies. The point estimate for the fifth seasonal frequency differs little from short memory, and a mildly negative memory effect is present at the sixth frequency.

The WLEs for the cepstrum were $-1.352, 1.557, -.460$, with a mean parameter of .100. The memory parameters were $-.160, .498, .497, .388, .414, .029, -.124$. The point estimate for the trend memory is now negative, which differs substantially from the results from the Gaussian likelihood, although the other memory parameters are broadly similar.

5.4 Unemployment Insurance Claims

The weekly time series of unemployment insurance claims from the Bureau of Labor Statistics (1987 through 2007, for a total of 1040 observations) is clearly seasonal, and is a fairly important series for understanding movements in the employment sector of the economy. Like the industrial production example, the spectral peak frequencies will be taken as known, where $p = 52$ is the number of seasons. Therefore, we assume that spectral peaks are located at frequencies $\pi j/26$ for $1 \leq j \leq 26$, in addition to peaks at frequency 0 and π . Hence we fit a 27-GEXP(5) model via the Gaussian and Whittle likelihoods. There are 32 parameters in the model, so that the ratio of sample size to parameters is over thirty. The calculation of autocovariances for such a process is now feasible, which we illustrate through providing model-fitting results.

The MLEs for the cepstrum were $-1.672, -1.385, -.282, .369, -.040$, with a mean of $-.0007$. The memory MLEs were $-.004, .452, .484, .431, .470, .381, .397, .202, .423, .293, .249, .346, .481, .171, .240, .210, .413, .220, .472, .431, .317, .258, .454, .325, .470, .321, \text{ and } .357$. Excepting the trend frequency, all

of these are positive, and some are quite large, indicating substantial persistence. It is interesting that one of the weakest peaks is the thirteenth (.171), which corresponds to quarterly phenomena, whereas the first seasonal peak (.452), which corresponds to once-a-year phenomena, is one of the strongest.

The cepstral WLEs were $-1.559, -1.284, -.231, .388, -.040$, with a mean of $-.0009$. The memory WLEs were $.003, .443, .486, .437, .475, .389, .373, .208, .405, .298, .251, .337, .486, .168, .238, .204, .419, .210, .471, .428, .308, .254, .457, .319, .475, .315, \text{ and } .351$. These point estimates are remarkably consistent with the MLEs, in most cases only differing in the third decimal place.

6. Discussion

This paper presents a new algorithm, that allows the computation of the autocovariance function of time series models with multiple persistencies. This is accomplished by rewriting the model spectral density as a sum of single pole factors, and using the fast splitting method on each summand. Our numerical work shows the accuracy of this approach, and the gains in computation time over previous methods are substantial. More importantly, the encoding for multiple peak processes is straightforward, unlike in the algorithm of McElroy and Holan (2012).

Our focus is not on modeling, although four case studies are provided in order to show that likelihood evaluation is feasible. We have used the k -GEXP model to demonstrate the new algorithm, but k -GARMA variants could also be devised along the same lines, with similar gains to speed and accuracy. Previous analysis of time series with multiple persistencies has been hampered by computation time; this work enables practitioners to more readily utilize multiple-pole models.

Acknowledgments

The authors thank Christopher Wikle for suggesting and providing the dataset used in Section 5.1 and for useful discussion. This research was partially supported by the U.S. National Science Foundation (NSF) and the U.S. Census Bureau under NSF grant SES-1132031, funded through the NSF-Census Research Network (NCRN) program.

Appendix

Proof of Proposition 1. Because $f = f^- f^+ g$, it suffices to prove that $f_{ZP} = f^+ \cdot \nu$. Using $d = d^+ + d^-$ for all values of d , and multiplying each term of ν by f^+ , we obtain f_{ZP} at once. The assertions about ν follow because the function evidently has no poles, and being a sum of non-negative terms can only have a zero if this zero is common to all the summands; but this is impossible due to the distinct locations of the various ϑ frequencies. (In the case where there is only one ZP factor, ν would indeed have a zero.) \square

Proof of Proposition 2. The first formula is straight convolution, while the second formula for the ZP factor follows from the ARFIMA(0,d,0) results described in Brockwell and Davis (1991) in the cases that $\vartheta = 0, \pi$. For the third case, we refer to the GARMA(0,d,0) results of Chung (1996). \square

REFERENCES

- Andrews, D.G., Holton, J.R., and Leovy, C.B. (1987), *Middle Atmospheric Dynamics*, Academic Press.
- Beran, J. (2010), “Long-range Dependence,” *WIREs Comput. Stat.*, 2, 26–35.
- Bertelli, S. and Caporin, M. (2002), “A Note on Calculating Autocovariances of Long-Memory Processes,” *J. Time Ser. Anal.* 23, 503–508.
- Bisaglia, L., Bordinon, S. and Lisi, F. (2003), “k-Factor GARMA Models for Intraday Volatility Forecasting,” *Appl. Econom. Lett.*, 10, 251–254.
- Bloomfield, P. (1973), “An Exponential Model for the Spectrum of a Scalar Time Series,” *Biometrika*, 60, 217–226.
- Cressie, N. and Wikle, C.K. (2011), *Statistics for Spatio-Temporal Data*, Hoboken, NJ: Wiley.
- Brockwell, P. and Davis, R. (1991), *Time Series: Theory and Methods*, New York: Springer.
- Chung, C. (1996), “Estimating a Generalized Long Memory Process,” *Journal of Econometrics*, 73, 237–259.
- Ferrara, L. and Guégan, D. (2000), “Forecasting Financial Time Series with Generalized Long Memory Processes,” in *Advances in Quantitative Asset Management*, ed. C. L. Dunis, Kluwer Academic Publishers, pp. 319–342.
- Gil-Alana, L. A. (2008), “Cyclical Long-Range Dependence and the Warming Effect in a Long Temperature Time Series,” *Internat. J. Climatology*, 28, 1435–1443.
- Gray, H., Zhang, N. and Woodward, W. (1989), “On Generalized Fractional Processes,” *J. Time Ser. Anal.*, 10, 233–257.
- Holan, S. and McElroy, T. (2012), “On the Seasonal Adjustment of Long Memory Time Series,” in *Economic Time Series: Modeling and Seasonality*, eds. W. Bell, S. Holan, and T. McElroy, New York: Chapman and Hall.
- Hosking, J. (1981), “Fractional Differencing,” *Biometrika*, 68, 165–176.
- Hurvich, C. (2002), “Multistep Forecasting of Long Memory Series Using Fractional Exponential Models,” *Internat. J. Forecasting*, 18, 167–179.
- McElroy, T. and Holan, S. (2012), “On the Estimation of Autocovariances for Generalized Gegenbauer Processes,” *Statistica Sinica*, 22, 1661–1687.
- McElroy, T. and Holan, S. (2014), “Asymptotic Theory of Cepstral Random Fields,” *Annals of Statistics*, 42, 64–86.
- McElroy, T. and Politis, D. (2014), “Spectral Density and Spectral Distribution Inference for Long Memory Time Series via Fixed-b Asymptotics,” *Journal of Econometrics*, 182, 211–225.
- Porter-Hudak, S. (1990), “An Application of the Seasonal Fractionally Differenced Model to the Monetary Aggregates,” *J. Amer. Statist. Assoc.*, 85, 338–344.
- Pourahmadi, M. (1984), “Taylor Expansion of $\exp(\sum_{k=0}^{\infty} a_k z^k)$ and Some Applications,” *Amer. Math. Monthly*, 91, 303–307.
- R Development Core Team (2014), R: A Language and Environment for Statistical Computing.
- Soares, L. and Souza, L. (2006), “Forecasting Electricity Demand Using Generalized Long Memory,” *Internat. J. Forecasting*, 22, 17–28.
- Talamantes, J., Behseta, S. and Zender, C. (2007), “Statistical Modeling of Valley Fever Data in Kern County, California,” *Internat. J. Biometeorology*, 51, 307–313.
- Wikle, C. K., Sherman, P.J., and Chen, T-C. (1995), “Identifying Periodic Components in Atmospheric Data Using a Family of Minimum Variance Spectral Estimators,” *Journal of Climate*, 8, 2352–2363.
- Woodward, W., Cheng, Q. and Gray, H. (1998), “A k-Factor GARMA Long Memory Model,” *J. Time Ser. Anal.*, 19, 485–504.

| log Σ : 1-GEXP: $g = (0, .75)$; $\omega = .56$ | | | | | | |
|--|-----------------|----------------------------|--|------------------|----------------------------|--|
| | ldet: $n = 500$ | | | ldet: $n = 1000$ | | |
| c_1 | Prop | New | MH ₂₀₁₂ | Prop | New | MH ₂₀₁₂ |
| .1 | 0.4299917 | .4299917 (1.696498e-5) | .429932 (1.568073e-5) $J_{\text{Trunc}} = 5,000$ | 0.44379 | .4437882 (9.213278e-6) | .443785 (5.240159e-6) $J_{\text{Trunc}} = 5,000$ |
| .25 | 1.582358 | 1.582393 (3.501754e-5) | 1.582377 (1.854933e-5) $J_{\text{Trunc}} = 25,000$ | 1.669002 | 1.66902 (1.775161e-5) | 1.668987 (-1.433973e-5) $J_{\text{Trunc}} = 25,000$ |
| .35 | 3.058414 | 3.058385 (-2.889742e-5) | 3.058331 (-8.284450e-5) $J_{\text{Trunc}} = 100,000$ | 3.228235 | 3.228223 (-1.230519e-5) | 3.228115 (-1.201864e-4) $J_{\text{Trunc}} = 100,000$ |
| .45 | 5.973976 | 5.973767 (-2.088053e-4) | 5.972568 (-1.407956e-3) $J_{\text{Trunc}} = 100,000$ | 6.254700 | 6.2546 (-9.965885e-5) | 6.252199 (-2.501926e-3) $J_{\text{Trunc}} = 100,000$ |

Table 1: Log determinant of the autocovariance sequence for a 1-GEXP model obtained using asymptotic approximation of Proposition 1 of McElroy and Holan (2012) (Prop), the algorithm of Section 3 (New), and the truncation approach of McElroy and Holan(2012) (MH₂₀₁₂). In the case of the MH₂₀₁₂ approach we provide the truncation used, J_{Trunc} . The number in parenthesis denotes the difference between the estimated ldet and ldet_{Prop}. In the case of MH₂₀₁₂, J_{Trunc} was chosen to reflect the smallest difference between ldet_{Prop} and the ldet from MH₂₀₁₂.

| log Σ : 2-GEXP: $g = (0, .75)$; $\omega = (.1, .56)$ | | | | | | |
|--|-----------------|----------------------------|--|------------------|----------------------------|--|
| | ldet: $n = 500$ | | | ldet: $n = 1000$ | | |
| (c_1, c_2) | Prop | New | MH ₂₀₁₂ | Prop | New | MH ₂₀₁₂ |
| (.1, .2) | 1.528072 | 1.497591 (-3.048125e-2) | 1.527665 (-4.072432e-4) $J_{\text{Trunc}} = 100,000$ | 1.597387 | 1.570535 (-2.685184e-2) | 1.596123 (-1.263973e-3) $J_{\text{Trunc}} = 50,000$ |
| (.1, .3) | 2.714834 | 2.687255 (-2.757927e-2) | 2.714331 (-5.028845e-4) $J_{\text{Trunc}} = 100,000$ | 2.853464 | 2.830055 (-2.340929e-2) | 2.852651 (-8.126799e-4) $J_{\text{Trunc}} = 100,000$ |
| (.1, .45) | 6.538299 | 6.515872 (-2.242651e-2) | 6.535380 (-2.918457e-4) $J_{\text{Trunc}} = 100,000$ | 6.832886 | 6.811977 (-2.090914e-2) | 6.834062 (1.175967e-3) $J_{\text{Trunc}} = 50,000$ |
| (.45, .2) | 8.134616 | 8.093399 (-4.121709e-2) | 8.13965 (5.0335e-3) $J_{\text{Trunc}} = 50,000$ | 8.470792 | 8.437261 (-3.353121e-2) | 8.486669 (1.587655e-2) $J_{\text{Trunc}} = 50,000$ |
| (.45, .3) | 9.492050 | 9.457665 (-3.438547e-2) | 9.494429 (2.378529e-3) $J_{\text{Trunc}} = 25,000$ | 9.897542 | 9.874612 (-2.292996e-2) | 9.856134 (-4.140792e-2) $J_{\text{Trunc}} = 50,000$ |
| (.45, .4) | 11.682510 | 11.66138 (-2.113188e-2) | 11.273010 (-0.409498) $J_{\text{Trunc}} = 50,000$ | 12.185040 | 12.173 (-1.200251e-2) | 11.807010 (-0.378036) $J_{\text{Trunc}} = 25,000$ |

Table 2: Log determinant of the autocovariance sequence for a 2-GEXP model obtained using asymptotic approximation of Proposition 1 of McElroy and Holan (2012) (Prop), the algorithm of Section 3 (New), and the truncation approach of McElroy and Holan(2012) (MH₂₀₁₂). In the case of the MH₂₀₁₂ approach we provide the truncation used, J_{Trunc} . The number in parenthesis denotes the difference between the estimated ldet and ldet_{Prop}. In the case of MH₂₀₁₂, J_{Trunc} was chosen to reflect the smallest difference between ldet_{Prop} and the ldet from MH₂₀₁₂.