Conditional Sum of Squares Estimation of Multiple Frequency Long Memory Models☆

Paul M. Beaumont[a], Aaron D. Smallwood[b,*]

[a]*Department of Economics, Florida State University, Tallahassee, FL 32306, USA*
[b]*Department of Economics, University of Texas Arlington, 701 S. West Street. (mailbox: 19479), Arlington, TX 76019, USA*

Abstract

We review the multiple frequency Gegenbauer autoregressive moving average model, which is able to reproduce a wide range of autocorrelation functions. Extending the result of Chung (1996a), we propose the asymptotic distributions for a conditional sum of squares estimator of the model parameters. The parameters that determine the cycle lengths are asymptotically independent, converging at rate T for finite cycles. This result does not hold generally, most notably for the differencing parameters associated with the cycle lengths. Remaining parameters are typically not independent and converge at the standard rate of $T^{1/2}$. We present simulation results to explore small sample properties of the estimator, which strongly support most distributional results while also highlighting areas that merit additional exploration. We demonstrate the applicability of the theory and estimator with an application to IBM trading volume.

Keywords: k-factor Gegenbauer processes, Asymptotic distributions, ARFIMA, Conditional sum of squares
JEL Classification Codes: C22, C40, C58, G12

1. Introduction

Multiple frequency, or k-factor, Gegenbauer autoregressive moving average models (GARMA) include ARIMA, fractionally integrated ARMA (ARFIMA), seasonal ARFIMA, and single frequency GARMA models as special cases and may simultaneously include features of all of these methods. These methods are especially useful because they can capture complex but commonly observed patterns in the spectral density and autocorrelation functions (ACF) of a stochastic process using only a few parameters. In this

☆The authors thank Stefan Norrbin, Ian McKeague, and seminar participants at the University of Oklahoma. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that were used to generate simulation results in the paper. URL: http://www.tacc.utexas.edu.

*Corresponding author: Aaron D. Smallwood, Department of Economics, University of Texas Arlington, 701 S. West Street (mailbox: 19479), Arlington, TX 76019, USA

Email addresses: beaumont@fsu.edu (Paul M. Beaumont), smallwood@uta.edu (Aaron D. Smallwood)

paper, we present a conditional sum of squares (CSS) estimator along with proposed joint asymptotic distributions for all parameters in the multiple frequency model. Simulation experiments generally validate the theoretical distributions. As an application, we model the trading volume of IBM equities, which follows a very complex stochastic process.

Long memory models were popularized by Granger and Joyeux (1980) and Hosking (1981) who introduced fractional differencing as a means of capturing complicated stochastic properties of data in both the time and frequency domains. These models have proven especially useful in economics and finance by bridging the gap between infinite variance unit root processes and finite variance short memory processes. One shortcoming of fractionally differenced models, however, is that they are not capable of capturing long memory processes with persistent cycles in the autocorrelation function. Gray et al. (1989) and Woodward et al. (1998) addressed this issue with the k-factor Gegenbauer autoregressive moving average model (k-GARMA). This model is capable of generating many complex patterns in the ACF that have previously been very difficult to capture. One particularly interesting case is a process that contains both ARFIMA and GARMA components, such that the ACF decays non-monotonically at a hyperbolic rate and is asymmetric about zero such as shown in Figure 1.

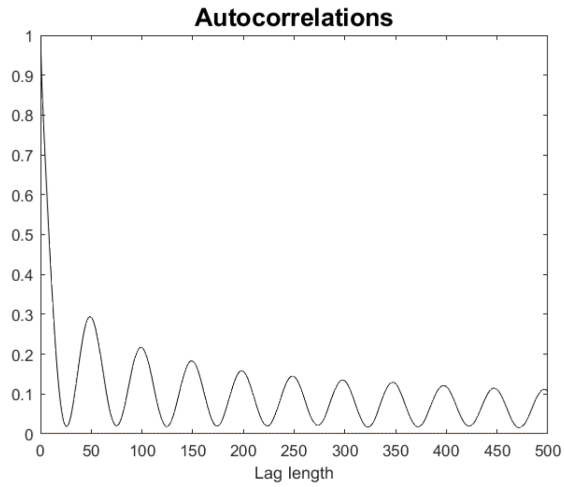

Figure 1: ACF of a process with both an ARFIMA and a GARMA component.

Due to its flexibility, the multiple frequency GARMA approach has proven very useful for modeling many physical, economic, and financial time series that exhibit complex long memory features. Chung (1996b) estimates a single factor model for sunspots and Woodward et al. (1998) and Diongue and Ndongo (2016) provide evidence supporting the existence of multiple sources of long memory in atmospheric CO2 and river flows. In economics and finance, these methods have been used to study interest rates (Ramachandran and Beaumont 2001; Gil-Alanaa 2007; Asai et al. 2018), exchange rates (Smallwood and Norrbin 2006), inflation (Caporale and Gil-Alana 2011; Peiris and Asai 2016), equity prices (Lu and Guegan 2011; Caporale and Gil-Alana 2014) and unemployment (Gil-Alanaa 2007) among many others. The possibility of multiple sources of long memory was illustrated recently by Leschinski and Sibbertsen (2019) who modeled California electricity load data using 14 independent long memory components.

Despite the increasing interest in the multi-factor GARMA model, a unifying estimation approach does not appear to exist. Only a handful of studies have attempted to simultaneously estimate all model parameters, including the relevant differencing parameters and the positions of the spectral poles, known as Gegenbauer frequencies. Almost all studies assume the positions of the singularities are known (see, for example, and Caporale and Gil-Alana (2011)) or they employ two-step procedures where the Gegenbauer frequencies are typically first estimated by inspection of the periodogram (see, for example, Lu and Guegan (2011) and Asai et al. (2018), amongst others). A major difficulty lies in the fact that estimation of the parameters dictating the positions of the spectral poles appears to be non-standard. Perhaps more importantly, the relevant parameter space is closed for these Gegenbauer frequencies, and there may exist a discontinuity in the distribution at the zero frequency (see, Chung (1996a)). Additionally, maximum likelihood based estimators in the frequency domain use a discrete set of parameters for the associated singularities. For these estimators, as argued by Giraitis et al. (2001), a full set of distributional results may not exist.

For inference for the models considered here, we are unaware of any study proposing a full set of distributional results for any estimator. For a single factor model, Giraitis et al. (2001) establish consistency for the Whittle estimator of the Gegenbauer frequency and provide normality results for the differencing parameter. Hidalgo and Soulier (2004) discuss the properties of the log-periodogram based semi-parametric estimates of the memory parameters for a multi-factor model where the Gegenbauer frequencies are obtained from inspection of the periodogram. For time domain maximum likelihood-based estimators, Chung (1996a) provided promising results for a constrained sum of squares (CSS) method for a single factor model based on the observation that, for the true parameter values, the expectation of the approximate likelihood function is zero. Regrettably, given the difficulties involved with the distributions of the estimates of the spectal poles, Chung (1996a) was unable to provide a rigorous initial proof establishing consistency, causing his results to be questioned by Giraitis et al. (2001).

Notwithstanding the concerns regarding the results of Chung (1996a), the CSS estimator provides a feasible and relatively simple method to obtain joint estimation results for the GARMA parameters. Recently, Beaumont and Smallwood (2019) provided a comprehensive simulation study that generally supports the results of Chung (1996a) excepting the case of parameter estimation of the Gegenbauer frequency when the true value is 0. Additionally, as the CSS estimator admits a continuous set of possible Gegenbauer frequencies, Beaumont and Smallwood (2019) show that the CSS method generally obtains a smaller bias for this parameter relative to the Whittle based counterpart. In comparison to an MCMC Whittle estimator, Diongue and Ndongo (2016) further demonstrate that the CSS method is relatively efficient in estimating differencing parameters for k-factor GARMA processes with infinite variance disturbances. Given these promising simulation results, it is worthwhile to consider the properties of the CSS estimator when applied to models with multiple Gegenbauer frequencies.

In this paper, we review the mulit-factor GARMA model and present the CSS estimator for all model parameters. Further, we extend the proofs of Chung (1996b) to obtain proposed distributional results, which show that the estimates of the Gegenbauer frequencies are asymptotically independent of each other and all other model parameters. Since we are unable to provide a rigorous consistency proof for the estimators of the sprectral poles, we provide simulation evidence to help validate the results. The simulation evidence, including additional results in Beaumont and Smallwood (2019), demonstrates that the theory can typically be reliably used to provide inference for the estimated parameters.

The rest of the paper is organized as follows. In the next section, we present the details of the multi-factor GARMA model. We present the CSS estimator and derive its properties in Section 3. In Section 4, we present Monte Carlo evidence for the finite sample precision of the iterative CSS estimation method that we propose. In Section 5, we show that the weekly trading volume of IBM stocks is best modelled with a five-frequency GARMA process. We summarize and draw conclusions in Section 6, and an appendix contains technical details.

2. Multiple Frequency Long Memory Processes

The multiple frequency GARMA model was originally discussed by Gray et al. (1989) and presented in greater detail by Woodward et al. (1998) who refer to the model as a k-factor GARMA model. The model is

$$\phi(L)\prod_{i=1}^{k}\left(1-2\eta_i L+L^2\right)^{\lambda_i}(x_t-\mu)=\theta(L)\,\varepsilon_t, \tag{1}$$

where $\phi(L)$ and $\theta(L)$ are polynomials in the lag operator L such that $\phi(z)=0$ and $\theta(z)=0$ have roots outside the unit circle, $\{\varepsilon_t\}$ is a white noise disturbance sequence, λ_i are differencing parameters, and the η_i are the parameters describing the periodic features of the process. The Gegenbauer polynomials $(1-2\eta_i L+L^2)^{\lambda_i}$ have a pair of complex roots with modulus one and expand to an infinite order polynomial in L. When $k=1$, we get the single frequency GARMA model (Hosking 1981; Gray et al. 1989), and when, in addition, $\eta=1$ the model further reduces to an ARFIMA(p,d,q) model (Granger and Joyeux 1980; Hosking 1981) where, in this context, $d=2\lambda$. Finally, we get an ARIMA model when $\eta=1$ and $\lambda=0.5$, and an ARMA process when $\lambda=0$.

Assuming that each η_i is distinct, the k-factor GARMA model is stationary if $\lambda_i<0.5$ whenever $|\eta_i|<1$, and $\lambda_i<0.25$ when $|\eta_i|=1$. The model is invertible if $\lambda_i>-0.5$ whenever $|\eta_i|<1$, and $\lambda_i>-0.25$ when $|\eta_i|=1$. Proofs for these results are available in Woodward et al. (1998).

For stationary cases, the moving average representation is,

$$(x_t-\mu)=\frac{\theta(L)}{\phi(L)}\prod_{i=1}^{k}\left(1-2\eta_i L+L^2\right)^{-\lambda_i}\varepsilon_t, \tag{2}$$

from which the spectral density function is obtained as

$$f(\omega)=\frac{\sigma^2}{2\pi}\left|\frac{\theta(e^{-i\omega})}{\phi(e^{-i\omega})}\right|^2\prod_{j=1}^{k}\{2\,|\cos(\omega)-\cos(\upsilon_j)|\}^{-2\lambda_j}, \omega\in[0,\pi] \tag{3}$$

where $\upsilon_j=\cos^{-1}(\eta_j)$ are the Gegenbauer frequencies. The spectral density function is unbounded at all υ_j if $\lambda_j > 0$ and vanishes there if $\lambda_j < 0$.

The autocovariances for a k-factor GARMA model can be computed as

$$\gamma_j = 2 \int_0^\pi f(\omega) \cos(\omega j)\, d\omega, \tag{4}$$

where special attention must be given to the singularities in $f(\omega)$ as recently discussed by McElroy and Holan (2016). Convenient expressions for γ_j are largely available for single frequency models only. For example, when $\eta = 1$ and $\lambda < 0.25$, the autocorrelations exhibit hyperbolic decay as demonstrated by Granger and Joyeux (1980) for fractional processes. For GARMA models, Chung (1996a) shows that for large j, the autocorrelation function with $|\eta| < 1$ and $\lambda < 0.5$, $\lambda \neq 0$, can be approximated as $\rho_j \approx J \cos(j\, \upsilon)\, j^{2\lambda - 1}$, where J does not depend upon j. This expression makes clear the hyperbolically damped sinusoidal pattern of the autocorrelation function of a stationary GARMA process with $|\eta| < 1$.

In Figure 1, we illustrated a model that combines ARFIMA and GARMA models, which is of particular interest for economic and financial applications. This example used a two frequency model with parameters $(\eta_1, \lambda_1) = (1, 0.15)$ and $(\eta_2, \lambda_2) = (0.992, 0.25)$. Note that the first frequency corresponds to an unbounded spike at the origin of the spectrum, and the second frequency corresponds to an unbounded spike at the frequency $\upsilon_2 = \cos^{-1}(0.992) = 0.1266$ radians, or 0.0201Hz, which is very close to the origin. The ACF clearly demonstrates long cycles about the hyperbolic decay characteristic of fractional processes.

3. Estimation

Several estimation procedures have been proposed for the k-factor model. A likelihood-based Whittle method was studied by Giraitis et al. (2001), while Hidalgo and Soulier (2004) advocate a semi-parametric method for λ_i after the Gegenbauer frequencies have been selected via maximization of the periodogram. In the frequency domain, wavelet procedures have been analyzed by Lu and Guegan (2011), while Dissanayake et al. (2018) use a state-space approach that uses associated Gegenbauer polynomials and the Kalman filter to obtain likelihood based estimates. Within this literature, there is no study that offers complete distributional results for estimation of both ν_i and associated differencing parameters.[1] Our approach is to estimate the k-factor GARMA model using a time domain parametric estimator that is asymptotically equivalent to maximum likelihood estimation. Specifically, we generalize the CSS estimator described by Chung and Baillie (1993) for fractional models and by Chung (1996a,b) for single frequency GARMA models. The procedure simultaneously estimates all parameters, including the ARMA components. Furthermore, we propose an analytic asymptotic distribution for all of the parameters of the model.

[1]For the single factor case, Giraitis et al. (2001) provide normality results for λ, but are only able to establish consistency for the Whittle estimator of ν.

Rewriting the MA representation of the model (2) in its AR form yields,

$$\frac{\phi_p(L)}{\theta_q(L)}\prod_{i=1}^{k}\left(1-2\eta_i L+L^2\right)^{\lambda_i}(x_t-\mu)=\varepsilon_t, \tag{5}$$

where p and q indicate the orders of the ARMA terms. If we assume that the initializing disturbances are zero, then the maximization of the CSS function is asymptotically equivalent to maximum likelihood estimation. Under the additional assumption that the disturbances, ε_t, are iid normal with variance σ^2, then the $p+q+2+2k$ parameters, $\Psi=\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q,\mu,\sigma^2,\eta_1,\lambda_1,\ldots,\eta_k,\lambda_k$, can be estimated by maximizing the CSS function

$$\mathcal{L}^*(\Psi)=-\frac{T}{2}\log(2\pi)-\frac{T}{2}\log(\sigma^2)-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2. \tag{6}$$

Note that the normality assumption is used here only to justify the construction of the CSS function, and is not necessary for the asymptotic theory proposed below. We require only that the $\{\varepsilon_t\}$ are martingale differences with respect to an increasing sequence of sigma-fields, F_t, such that, for some $\beta>0$, $\sup_t E(|\varepsilon_t|^{2+\beta}\|F_{t-1})<\infty$, almost surely, and $E(\varepsilon_t^2\|F_{t-1})=\sigma^2$, almost surely.[2]

3.1. Asymptotic Distributions

We extend the proofs of Chung (1996 a,b) to propose distributional theory for the CSS estimator in (6). Our proofs use the observation that the expectation of the score for the CSS function achieves a zero value for the true parameter set. We argue that an initial consistency proof for our full set of estimators may not be available. In particular, we argue that the distributional results for η_i are non-standard with a discontinuity occurring at $\eta_i=1$. In this specific case, it is not possible to constrain all parameters to lie in the interior of the parameter space, an assumption that would typically be employed in consistency proofs (see, for example, Andrews and Sun (2004)). Consequently, we use an extensive set of simulations to help validate results.

To extend Chung (1996a) and Chung (1996b), we consider four cases. The first case is for those models for which $|\eta_i|<1$ for all $i=1,\ldots,k$. The second case is for those models for which there exists a single $\eta_i=1$, and $|\eta_j|<1$ for all other frequencies. The third case is for those models for which there exists a value $\eta_i=-1$, and $|\eta_j|<1$ otherwise. The fourth case is for those models for which there exists two values η_i and η_j such that $\eta_i=1$ and $\eta_j=-1$, with $|\eta_m|<1$ otherwise. The first theorem establishes that the asymptotic information matrix for the k-factor GARMA model is block diagonal.

Theorem 1 (Independence of δ and η^*)**.** *Let $\delta=(\lambda_1,\ldots,\lambda_k,\phi',\theta')'$ and $\eta^*=(\eta_1,\ldots,\eta_k)'$ be the parameters associated with the CSS function for the k-factor GARMA model. The asymptotic distribution of δ is independent of η^*.*

The proof of this theorem is given in the Appendix. The essential idea is to establish the different rates of stochastic convergence for the elements of δ and η^*. Note that no

[2] As recently discussed by Peiris and Asai (2016), an additional advantage of the CSS estimator is that it is easily generalized to handle GARCH residuals and other types of non-normal distributional assumptions that often arise in financial and economic applications.

conditions are placed on the value of η_i relative to η_j, $i \neq j$, so this theorem holds for all four cases described above. Consequently, the asymptotic distribution of δ can be considered independently of η^*.

Theorem 2 yields the asymptotic distribution of the estimators of δ and μ.

Theorem 2 (Asymptotic distributions of δ and μ)**.** *Let $\hat{\delta}$ be the CSS estimator of δ for the stationary and invertible k-factor GARMA model. (If μ is unknown, we add the restriction that $\eta_i \neq 1$ for all $i = 1, \ldots, k$.) Then,*

$$\sqrt{T}(\hat{\delta} - \delta) \rightsquigarrow N(0, I_\delta^{-1}), \tag{7}$$

where $\rightsquigarrow$ denotes the weak convergence of the random vectors $\hat{\delta}$, and where

$$\underset{(k+p+q)\times(k+p+q)}{I_\delta} = \begin{bmatrix} I_{\lambda_1} & \cdots & I_{\lambda_1\lambda_k} & I_{\lambda_1,\phi} & I_{\lambda_1,\theta} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ I_{\lambda_1\lambda_k} & \cdots & I_{\lambda_k} & I_{\lambda_k,\phi} & I_{\lambda_k,\theta} \\ I_{\lambda_1,\phi} & \cdots & I_{\lambda_k,\phi} & I_\phi & I_{\phi,\theta} \\ I_{\lambda_1,\theta} & \cdots & I_{\lambda_k,\theta} & I_{\phi,\theta} & I_\theta \end{bmatrix}. \tag{8}$$

The elements of I_δ are defined as follows,

$$I_{\lambda_i} = 2\left[\frac{\pi^2}{3} - \pi\upsilon_i + \upsilon_i^2\right],\ i = 1, \ldots, k \tag{9a}$$

$$I_{\lambda_i\lambda_j} = 2\left[\frac{\pi^2}{3} - \pi\upsilon_i + \frac{\upsilon_i^2 + \upsilon_j^2}{2}\right],\ \upsilon_i > \upsilon_j, \tag{9b}$$

$$I_{\lambda_i\phi_j} = 2\sum_{l=0}^{\infty}\phi_l^*\frac{\cos[(l+j)\upsilon_i]}{(l+j)},\ i = 1, \ldots, k,\ j = 1, \ldots, p \tag{9c}$$

$$I_{\lambda_i\theta_m} = 2\sum_{l=0}^{\infty}\theta_l^*\frac{\cos[(l+m)\upsilon_i]}{(l+m)}, i = 1, \ldots, k,\ m = 1, \ldots, q \tag{9d}$$

where ϕ_l^* and θ_l^* denote the l^{th} coefficients in the infinite order expansions of $\phi^{-1}(L)$ and $\theta^{-1}(L)$, respectively. The submatrices I_ϕ, $I_{\phi,\theta}$ and I_θ consist of elements that are the same as the corresponding submatrices of the usual information matrix of an ARMA model. Finally, let $\eta_i < 1$, i=1,....,k. For the CSS estimator of the mean, $\hat{\mu}$, with $|\eta_i| < 1$ for all i, we have,

$$\sqrt{T}(\hat{\mu} - \mu) \rightsquigarrow N(0, 2\pi f(0)), \tag{10}$$

where $f(0)$ denotes the spectral density function evaluated at frequency $\omega = 0$. Further, the distribution of $\hat{\mu}$ is equivalent to the sample mean $\bar{x}$.

The proof of this theorem is given in the Appendix. Theorem 3 is the central result and proposes the asymptotic distribution of η^* for all of our four cases.

Theorem 3 (Aysymptotic distribution of η^*)**.** *Let $\hat{\eta}_1, \ldots, \hat{\eta}_k$ be the CSS estimators of $\eta_1, \ldots, \eta_k$, for a stationary and invertible k-factor GARMA model based on a sample $\{X_t\}$, $t = 1, \ldots, T$, with $\eta_i \neq \eta_j, i \neq j$. Without loss of generality, order the elements of*

η^* from smallest to largest. Then let I_{η_1} denote the indicator function, that takes on the value 1 if $\eta_1 = -1$ and 0 otherwise, and let I_{η_k} denote the indicator function that takes on the value 1 if $\eta_k = 1$ and 0 otherwise. If $\lambda_i \neq 0$, $i = 1, \ldots, k$, then,

$$T(\hat{\eta}_i - \eta_i) \rightsquigarrow \frac{\sin(\upsilon_i)}{\lambda_i} \frac{\left[\int_0^1 W_{2i-1-I_{\eta_1}}\, dW_{2i-I_{\eta_1}} - \int_0^1 W_{2i-I_{\eta_1}}\, dW_{2i-1-I_{\eta_1}}\right]}{\left[\int_0^1 W^2_{2i-1-I_{\eta_1}}(r)\, dr + \int_0^1 W^2_{2i-I_{\eta_1}}(r)\, dr\right]} \tag{11}$$

with $|\eta_i| < 1$, where $i = 1 + I_{\eta_1}, \ldots, k - I_{\eta_k}$ and,

$$T^2(\widehat{\eta}_1 + 1) \rightsquigarrow -\frac{1}{2\lambda_1} \frac{\int_0^1 \left[\int_0^r W_1(s) ds\right] dW_1(r)}{\int_0^1 \left[\int_0^r W_1(s)\, ds\right]^2 dr}, \text{ if } \widehat{\eta}_1 = -1 \tag{12}$$

$$T^2(\widehat{\eta}_k - 1) \rightsquigarrow \frac{1}{2\lambda_k} \frac{\int_0^1 \left[\int_0^r W_{2k-1-I_{\eta_1}}(s) ds\right] dW_{2k-1-I_{\eta_1}}(r)}{\int_0^1 \left[\int_0^r W_{2k-1-I_{\eta_1}}(s) ds\right]^2 dr}, \text{ if } \widehat{\eta}_k = 1, \tag{13}$$

where $W_1, W_2, \ldots., W_{2k-I_{\eta_1}-I_{\eta_k}}$, are $2k - I_{\eta_1} - I_{\eta_k}$ independent Brownian motions.

The proof is given in the Appendix. An important result of this theorem is the asymptotic independence of the values in the vector η^*. In addition, for each $\hat{\eta}_i$, the values of λ_i and υ_i enter the equation for the asymptotic distribution proportionally, so one only needs the values of the stochastic integrals depicted in Theorem 3 to calculate asymptotic confidence intervals. The values for these integrals are reported in Chung (1996a).

3.2. Estimation Algorithm

These theorems provide important practical information for designing an efficient estimator. We know that the asymptotic distributions of the $\hat{\lambda}_i$'s are not independent of ARMA parameters. Also, the asymptotic distribution of $\hat{\delta}$ and $\hat{\eta}^*$ are independent, but the elements of $\hat{\delta}$ are $O_p(T^{-1/2})$, whereas $\hat{\eta}_i$ is $O_p(T^{-1})$ if $|\eta_i| < 1$ and $O_p(T^{-2})$ if $|\eta_i| = 1$. These results suggest that the algorithm of Woodward et al. (1998), which estimates ARMA parameters independently of (η_i, λ_i), will produce inconsistent estimates. It would be more appropriate to use an extension of Chung's method (Chung 1996a,b) by conducting a grid search over η^* combined with a gradient method over δ. However, Monte Carlo simulations indicate that the grid over the η's must be very fine, since the likelihood function has many local minima. A k-dimensional line search for η_i coupled with a gradient based search for δ would be computationally infeasible, unless the parameter space is bounded in some way or a very coarse grid is used.

The computational complexity of an estimator for a k-factor GARMA model can be better appreciated when we consider the step of recursively computing the residuals for the CSS estimator. The i^{th} Gegenbauer polynomial in the k-factor GARMA model can be expanded as (Gray et al. 1989)

$$(1 - 2\eta_i z + z^2)^{-\lambda_i} = \sum_{j=0}^{\infty} C_j^{(\lambda_i)}(\eta_i)\, z^j, C_j^{(\lambda_i)}(\eta_i) = \sum_{l=0}^{[j/2]} \frac{(-1)^l (2\eta_i)^{j-2l}\, \Gamma(\lambda_i - l + j)}{l!\, (j-2l)!\, \Gamma(\lambda_i)}, \tag{14}$$

where $[j/2]$ is the integer part of $j/2$. As Chung (1996a) notes, the best way to calculate the coefficients $C_j^{(\lambda_i)}$ is via the following recursion,

$$C_j^{(\lambda_i)}(\eta_i) = 2\eta_i \left(\frac{\lambda_i - 1}{j} + 1\right) C_{j-1}^{(\lambda_i)}(\eta_i) - \left(2\frac{\lambda_i - 1}{j} + 1\right) C_{j-2}^{(\lambda_i)}(\eta_i), \tag{15}$$

where $C_0^{(\lambda_i)}(\eta_i) = 1$ and $C_1^{(\lambda_i)}(\eta_i) = 2\,\lambda_i\,\eta_i$. Under the assumption that $\varepsilon_0 = \varepsilon_{-1} = \ldots = 0$, the residuals can be calculated recursively from the expression

$$\phi(L)(x_t - \mu) = \prod_{i=1}^{k} \left[\sum_{j=0}^{t-1} C_j^{(\lambda_i)}(\eta_i)\, L^j\right] \theta(L)\,\varepsilon_t. \tag{16}$$

The combination of the k-dimensional product over the above sums create most of the computational burden.

To overcome computational issues, coupled with different rates of convergence of various model parameters, we employ an extension of the algorithm developed by Ramachandran and Beaumont (2001). First, through inspection of the periodogram and estimation of individual GARMA models, we choose k and obtain a grid of starting values for each element of η^*. We use each set of starting values in this grid to obtain estimates for the δ. Conditional on the estimated value, $\hat{\delta}$, we then estimate the elements of η^* using an unconstrained gradient based search.[3] Using the updated estimates of η^*, new estimates of δ are obtained, which are then used to update the estimates of η^*. This procedure continues for all combinations of starting values for η_i. The final model results from the set of parameters that produce the smallest sum of squared errors. Although computationally intensive, the use of this multi-step gradient based iterative algorithm provides large gains in computational time relative to the full k-dimensional line search for η_i, while also guaranteeing a continuous parameter space.

4. Finite Sample Performance

In this section, we report simulation results that examine the finite sample properties of the CSS estimation. We are interested in examining the bias in the parameter estimates and in comparing the finite sample standard errors of the parameter estimates with the asymptotic standard errors. Chung (1996a,b) and Ramachandran and Beaumont (2001) have done extensive simulations for the single frequency GARMA model, with the latter paying particular attention to the parametric region where η is close to one and λ is close to one-half. Based upon those results, we will use a sample size of 300, and we will concentrate on two frequency models with parameter ranges that we believe are most relevant for economic and financial applications. We pay particular attention to the mixed ARFIMA/GARMA case.

The simulation results are presented in Tables 1–4. The columns of each table list the parameters of the simulated model and each block in the tables gives the results for

[3] The search occurs over all theoretically plausible values of η_i, only imposing a constraint to insure $\eta_i \neq \eta_j$, $i \neq j$. All elements of η_i are estimated jointly, unless it suspected that there exists a value $|\eta_i| = 1$, in which case this parameter is estimated separately at each iteration.

a specific parameterization. Throughout, we report the true parameter values (TRUE), the mean and median biases (MEAN BIAS; MED. BIAS), along with the root mean squared error, mean of the numerical standard errors, and mean absolute deviations (RMSE, MNSE, MAD) based on 1000 replications. For computational purposes, we use an iterative procedure to generate a large amount of data before discarding all but the last 300 observations. On rare occasions, the generated data do not take on the properties of a multiple frequency GARMA model. For the 17000 generated series below, it was necessary to discard three.

Table 1 presents the results for six different two frequency GARMA(0,0) models with η values of $-\frac{1}{2}, 0, \frac{1}{2}$ and λ values of 0.2 and 0.4. The estimation biases are all quite small, especially for the η's that converge at a faster rate than the λ's. For λ_i, there is not much difference between the mean bias and the median bias or the RMSE and the MAD, indicating that the distribution of these parameters is quite robust. Generally speaking, a larger value of λ_i mitigates the already small bias in η_i, which appears to be marginally more sensitive to estimation outliers. This is likely due to the fact that an estimate of λ_i near zero can lead to wildly wrong estimates of the corresponding η_i, since that Gegenbauer polynomial will have very little impact on the likelihood function no matter what the value of η_i is. In these cases, the mean, μ, is estimated with the sample mean, which is again asymptotically equivalent to the CSS estimator of μ provided $\eta_i < 1,\ i = 1, ..., k$. As noted above, the estimator for the mean is $O_p(T^{-1/2})$, the same rate of convergence as the other parameters in δ, so its bias is also quite small.

To further validate the estimator, we compare the mean numerical standard errors calculated from the estimated Hessian matrix in the last iteration with the true asymptotic standard errors calculated with the aid of Theorem 2. In particular, for the six cases of Table 1, the true asymptotic standard errors of the corresponding values of λ_1 are 0.0394, 0.0450, 0.0394, 0.0455, 0.0450, and 0.0455, respectively. These values are quite comparable to the MNSE and RMSE of the corresponding numbers in Table 1. The true asymptotic standard errors for λ_2 are 0.0455, 0.0450, 0.0455, 0.0394, 0.0450, and 0.0394, which again are very close to their numerical counterparts. The true asymptotic standard errors for the values of μ are 0.0438, 0.0372, 0.0503, 0.0324, 0.0463, and 0.0351. Here, the RMSE is comparable to the true asymptotic standard error, although it is interesting to note that the RMSE slightly underestimates the standard deviation of the mean in small samples. Finally, in light of the results of Theorem 3, it is not surprising to see that the MNSE and RMSE for the η's are quite different, since the RMSE assumes convergence at the rate $T^{1/2}$.

To examine the influence of the ARMA parameters, ϕ and θ, we choose a particular parameterization (second case from Table 1) and estimate various two frequency GARMA(p, q) models with p and q being either zero or one. The results are reported in Table 2 and are similar to those in Table 1. Interestingly, the main consequence of the inclusion of ARMA parameters is a relatively wide distribution for the sample mean when a positive autoregressive parameter exists. Again, for all of the cases considered in Table 2, the median and mean biases are quite small, and the RMSE compares favorably with the MNSE.

Table 1: Simulations for the 2-factor GARMA(0,0) processes

	η_1	η_2	λ_1	λ_2	μ
True	0.5	0	0.2	0.4	0
Mean Bias	0.0017	-0.0011	-0.0033	0.0029	0.0006
Med. Bias	0.0003	-0.0004	-0.0035	0.0041	0.0004
RMSE	0.0473	0.0155	0.0507	0.0463	0.0356
MNSE	0.0080	0.0133	0.0461	0.0396	N/A
MAD	0.0296	0.0091	0.0329	0.0395	0.0289
True	0.5	-0.5	0.2	0.4	0
Mean Bias	0.0007	-0.0000	-0.0002	0.0061	-0.0008
Med. Bias	0.0004	-0.0000	0.0004	0.0076	-0.0004
RMSE	0.0428	0.0136	0.0448	0.0484	0.0314
MNSE	0.0134	0.0069	0.0454	0.0457	N/A
MAD	0.0267	0.0081	0.0363	0.0388	0.0253
True	0	0.5	0.2	0.4	0
Mean Bias	-0.0028	0.0001	-0.0016	0.0025	-0.0008
Me. Bias	-0.0014	-0.0002	-0.0037	0.0033	0.0007
RMSE	0.0495	0.0146	0.0457	0.0442	0.0422
MNSE	0.0154	0.0070	0.0459	0.0399	N/A
MAD	0.0317	0.0086	0.0365	0.0348	0.0340
True	0	-0.5	0.2	0.4	0
Mean Bias	0.0025	0.0003	-0.0009	0.0046	-0.0006
Med. Bias	0.0005	0.0002	-0.0032	0.0073	-0.0015
RMSE	0.0500	0.0135	0.0458	0.0448	0.0274
MNSE	0.0155	0.0067	0.0460	0.0399	N/A
MAD	0.0318	0.0080	0.0366	0.0360	0.0219
True	-0.5	0.5	0.2	0.4	0
Mean Bias	-0.0018	0.0004	-0.0011	0.0025	0.0002
Med. Bias	-0.0004	0.0001	-0.0007	0.0023	0.0008
RMSE	0.0446	0.0140	0.0442	0.0487	0.0383
MNSE	0.0134	0.0070	0.0454	0.0457	N/A
MAD	0.0274	0.0080	0.0352	0.0388	0.0308
True	-0.5	0	0.2	0.4	0
Mean Bias	-0.0010	0.0012	0.0004	0.0039	0.0002
Med. Bias	0.0001	0.0003	0.0019	0.0055	0.0003
RMSE	0.0447	0.0160	0.0381	0.0490	0.0285
MNSE	0.0131	0.0080	0.0397	0.0461	N/A
MAD	0.0283	0.0098	0.0299	0.0395	0.0228

Table 2: Simulation on the estimation of 2-factor GARMA processes with ARMA parameters

	η_1	η_2	λ_1	λ_2	ϕ	θ	μ
True	0.5	-0.5	0.2	0.4	0.8	N/A	0
Mean Bias	-0.0010	-0.0003	-0.0015	-0.0025	-0.0114	-	-0.0101
Med. Bias	0.0003	-0.0002	-0.0016	0.0003	-0.0099	-	-0.0107
RMSE	0.0444	0.0138	0.0456	0.0547	0.0429	-	0.1507
MNSE	0.0134	0.0070	0.0462	0.0519	0.0409	-	N/A
MAD	0.0275	0.0082	0.0363	0.0441	0.0335	-	0.1212
True	0.5	-0.5	0.2	0.4	N/A	0.8	0
Mean Bias	0.0014	0.0003	-0.0011	0.0054	-	-0.0099	-0.0007
Med. Bias	0.0005	-0.0001	-0.0011	0.0062	-	-0.0086	-0.0023
RMSE	0.0488	0.0125	0.0480	0.0463	-	0.0459	0.0567
MNSE	0.0137	0.0068	0.0518	0.0463	-	0.0407	N/A
MAD	0.0295	0.0077	0.0389	0.0368	-	0.0362	0.0454
True	0.5	-0.5	0.2	0.4	-0.8	N/A	0
Mean Bias	0.0020	0.0016	0.0013	0.0078	0.0042	-	0.0004
Med. Bias	0.0009	0.0001	0.0011	0.0109	-0.0003	-	-0.0000
RMSE	0.0461	0.0344	0.0516	0.0513	0.0431	-	0.0177

Table 3 examines the particularly interesting case where η_1=1 and $\eta_2 < 1$, so that we get a combination ARFIMA and GARMA model. Compared to η_2, the estimator for $\eta_1 = 1$ has very little bias and extremely small RMSE and MNSE, reflecting the fact that this parameter may be $O_p(T^{-2})$. The results for $|\eta_2| < 1$ are similar to those in Tables 1 and 2, as are the results for the λ's. When $\eta_i = 1$, however, the sample mean and CSS estimate of μ are no longer asymptotically equivalent. Thus, we use the CSS estimator for the mean in these cases. The computational difficulties of time domain estimators for ARFIMA models when the mean is unknown have been well documented (Adenstedt 1974; Yajima 1991; Chung and Baillie 1993; Cheung and Diebold 1994). In spite of these difficulties, the mean is fairly unbiased, albeit with a wide distribution. Again, the remaining parameters suffer from very little distortion.

Table 3: Estimation of simulated ARFIMA/GARMA processes

	η_1	η_2	λ_1	λ_2	μ
True	1	0.75	0.2	0.3	0
Mean Bias	-0.0016	0.0026	-0.0127	0.0059	0.0417
Med. Bias	-0.0000	0.0010	-0.0121	0.0080	0.0612
RMSE	0.0062	0.0195	0.0345	0.0420	0.4345
MNSE	0.0005	0.0069	0.0288	0.0410	0.3679
MAD	0.0016	0.0109	0.0275	0.0338	0.3543
True	1	0.5	0.2	0.3	0
Mean Bias	-0.0023	0.0048	-0.0100	0.0041	0.0399
Med Bias	-0.0001	0.0014	-0.0085	0.0049	0.0539
RMSE	0.0085	0.0274	0.0283	0.0437	0.3486
MNSE	0.0006	0.0092	0.0245	0.0413	0.2812
MAD	0.0023	0.0150	0.0222	0.0348	0.2756
True	1	0	0.2	0.3	0
Mean Bias	-0.0040	0.0070	-0.0097	-0.0034	0.0023
Med. Bias	-0.0005	0.0019	-0.0096	-0.0016	0.0002
RMSE	0.0116	0.0393	0.0269	0.0498	0.2574
MNSE	0.0010	0.0108	0.0242	0.470	0.1859
MAD	0.0040	0.0176	0.0214	0.0387	0.2022
True	1	-0.5	0.2	0.3	0
Mean Bias	-0.0015	0.0041	-0.0079	-0.0032	-0.0074
Med. Bias	0.0000	0.0010	-0.0077	0.0002	-0.0203
RMSE	0.0085	0.0232	0.0316	0.0486	0.2797
MNSE	0.0004	0.0091	0.0287	0.0486	0.2365
MAD	0.0015	0.0132	0.0251	0.0386	0.2267

As noted above, the computational burden of the CSS estimator grows rapidly with the number of frequencies due to the grid search over each η_i. Thus, if we could narrow the range of the grid search, we could greatly improve the efficiency of the algorithm. With $i \neq j$, since η_i is independent of both η_j and the parameters in δ, it may be possible to first estimate each value of η_i sequentially to get good starting values. We could then re-estimate the entire model using fairly tight grids over each η_i. In Table 4, we investigate this possibility. First, we estimate a single frequency GARMA model and then filter the data with the resulting Gegenbauer polynomial before estimating the second frequency using a single frequency model on this filtered data. This process should produce good starting values for the η's as long as the biases are not too large.

Table 4: Estimation of simulated ARFIMA/GARMA processes with single frequency models

	η_1	η_2	λ_1	λ_2	ϕ	θ	μ
True	0.5	0.0	0.2	0.4	N/A	N/A	0
Mean Bias	0.0099	0.0265	-0.0274	0.0487	-	-	0.0006
Med. Bias	0.0023	0.0087	-0.0276	0.0535	-	-	0.0007
RMSE	0.0603	0.0486	0.0468	0.0656	-	-	0.0356
MNSE	0.0141	0.0473	0.0257	0.0103	-	-	N/A
MAD	0.0319	0.0279	0.0379	0.0566	-	-	0.0285
True	0.5	-0.5	0.2	0.4	0.8	0.8	0
Mean Bias	-0.0103	0.0027	-0.0975	-0.1278	-0.0796	0.0466	0.0045
Med. Bias	0.0159	0.0007	0.0971	-0.1256	-0.0752	0.0514	0.0009
RMSE	0.0787	0.0155	0.1010	0.1376	0.0955	0.0620	0.2663
MNSE	0.0202	0.0074	0.0374	0.0429	0.0481	0.0316	N/A
MAD	0.0406	0.0084	0.0975	0.1278	0.0814	0.0542	0.2112
True	1.0	0.75	-0.2	0.3	N/A	N/A	0
Mean Bias	-0.0019	-0.0151	0.0631	-0.1266	-	-	-0.0006
Med. Bias	0.0000	-0.0024	0.0611	-0.1254	-	-	-0.0008
RMSE	0.0336	0.0463	0.0689	0.1325	-	-	0.0130
MNSE	0.0007	0.0085	0.0243	0.0302	-	-	N/A
MAD	0.0019	0.0143	0.0631	0.1275	-	-	0.0105

The first two models in Table 4 are cases from the previous simulations, and the third case represents a mixed ARFIMA/GARMA model in which the ARFIMA component is short memory ($\lambda < 0$). The latter process may result from differencing a non-stationary ARFIMA process. For each of the cases considered in Table 4, the sample mean is used to estimate μ. We find that the single frequency estimator generally first selects the frequency with the largest corresponding value of λ, thus capturing the most dominate feature of the autocorrelation function. The results in Table 4 indicate that the small sample biases in η_1 and η_2 are quite reasonable, suggesting that the method of choosing a tight grid around these point estimates may work quite well. The relatively large biases in the values of the vector δ, however, confirm the results of Theorem 2 that a consistent estimator is obtained only through joint estimation of all parameters.

For a fixed sample size, these results strongly support the use of the estimation algorithm, while largely validating the proposed distribution theory. Notably, the distribution of η_i is independent of η_j, $i \neq j$, and the distribution of these parameters is largely unaffected by the inclusion of ARMA dynamics. Additionally, the proposed distribution theory for λ_i is confirmed. Finally, as shown below, and in numerous other simulations that are available upon request, the estimator achieves the proposed rates of convergence, even when we estimate multiple GARMA components. This implies that confidence bands for η_i will be calculated in precisely the same way they are for the single frequency case, as in Chung (1996a). As such, any concerns that may exist regarding the proposed distribution theory of Chung (1996a) will be inherited by the results here.

For the single frequency case, Chung (1996a) uses a line grid search to estimate η, along with a gradient based method for δ. This implies that the parameter space being searched over is a countable finite set that requires the use of boundary constraints, given that a fine grid would be needed to capture an estimate of η in the neighborhood of the true value. Based on the limited algorithm, Chung (1996a) provides strong support for the proposed theory and associated confidence bands for η for all cases except when $\eta = 1$. Here, it would appear that the associated empirical sizes of tests for $\eta = 1$ under the null are too large to be of practical use. Beaumont and Smallwood (2019) consider the consequences of using a two-dimensional grid search over both η and λ without the use of boundary constraints for η, and shows that the exact distributional results of Chung (1996a) are generally supported, with two exceptions. First, similar to Chung (1996a), Beaumont and Smallwood (2019) show the theory under the hypothesis $\eta = 1$ is inappropriate for testing purposes. Confidence bands tend to be much too wide, and empirical sizes are often much higher than their associated theoretical counterparts. Secondly, it is shown that with the use of the proposed algorithm, the resulting empirical distribution has slightly fatter tails and a more peaked density relative to the proposed theory. In terms of calculating confidence bands, the issue appears to be very minor and disappears as the sample size increases. Nonetheless, small biases in confidence bands can result, especially as λ approaches 0. We now consider more complete simulation evidence to analyze the extent to which these previous results carry over when $k > 1$.

For varying sample sizes, we considered a variety of parameterizations, including models where there exists a value of $\eta_i = 1$. For brevity, the full set of results are not reported here, but are available upon request. Here, we report only the results for three fairly complicated 2-frequency parameterizations. Model 1 is a GARMA(0,0) model with $\{\eta_1, \lambda_1\} = \{0.5, 0.4\}$, and $\{\eta_2, \lambda_2\} = \{0, 0.2\}$. Given the distributional results above, this parameterization represents a case where the process is expected to be especially volatile.[4] Model 2 is also a GARMA(0,0) model but with parameters $\{\eta_1, \lambda_1\} = \{0.98, 0.45\}$, and $\{\eta_2, \lambda_2\} = \{-0.4, 0.3\}$. This parameterization approaches the region of the discontinuity in our theoretical distribution for η^* and is also a strongly persistent process with λ_1 only marginally less than 0.50. Model 3 is the same as Model 1 except we add an AR(1) term with parameter $\phi = 0.80$.

First, we compare the theoretical and simulated distributions of η_i. Figure 2 shows the empirical and theoretical normalized cumulative distribution functions (cdf) for η_1 from Model 1 for sample sizes of 500 and 2000. For the empirical distributions we plot $T(\hat{\eta} - 0.50)$ where the $\hat{\eta}$'s are computed using the estimation algorithm described above, and the theoretical quantities have been calculated using equation (11) from Theorem 3. The horizontal differences between the theoretical and empirical curves show the disagreements between the theoretically and empirically derived critical values for each percentile. The two horizontal shaded regions show areas below the 0.025 and above the 0.975 percentiles, which would be relevant for the construction of a 95% confidence interval.

The first observation is that the empirical and theoretical distributions are in fairly close agreement, and this agreement is consistent as the sample size increases. This

[4] Note that the scaling factor in equation (11) of Theorem 3 is $\frac{\sin(\nu_i)}{\lambda_i}$ so that, given a small value for λ_2 coupled with $\nu_2 = \pi/2$, the estimated value of η_2 is expected to be quite volatile.

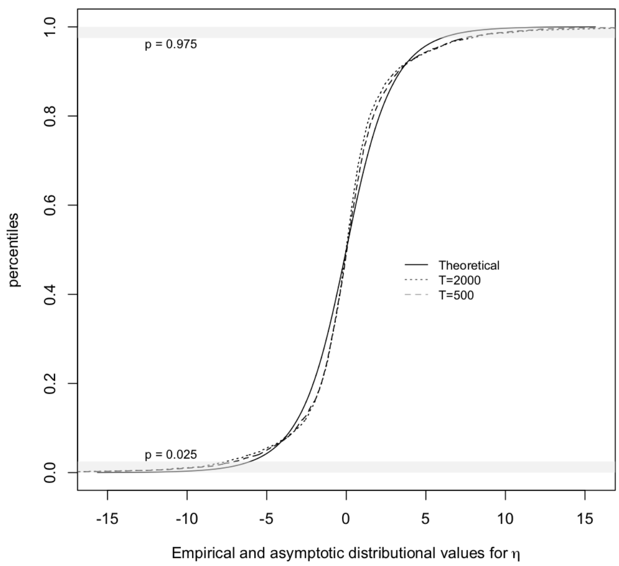

Figure 2: Percentiles of theoretical and empirical CDFs for T=500 and T=2000.

suggests that the proposed convergence rate of T in Theorem 3 is strongly supported. Second, there is some evidence that the empirical tails are larger than implied by the theory, so we will now explore the consequences of any such differences.

When estimating a multiple frequency GARMA model, the calculation of confidence bands for η_i is likely the most important application of the theory. To get a sense of how applicable our proposed distribution theory and algorithm are, Table 5 provides the estimated biases in calculating the upper and lower 68%, 90%, 95%, and 99% confidence bands for the three models described above. As a reference point, the theoretical bands for each value of η_i with a sample size of 500 are provided in bold font. Below the theoretical bands, we show the empirical bands for η_1 for each sample size, followed by the empirical bands for η_2.

For Model 1, and with relatively small sample sizes of 500 observations, the 99% confidence bands are quite unreliable for $\eta_2 = 0$. The theoretical confidence band for $\eta_2 = 0$ when $T = 500$ is $[-0.0423, 0.0423]$, whereas, amongst the 5050 simulations, 99% of the estimated values of η_2 were within a range from $-0.0423 - 0.0487 = -0.091$ to $0.0423{+}0.026 = 0.068$. In general, with small sample sizes, there are small but potentially non-negligible biases when using the 99% confidence bands. Otherwise, the results in Table 5 support the use of the proposed distribution theory in calculating these intervals. First, we note that the differences between the estimated and theoretical bands decrease sharply as the sample size increases, and become negligible in most cases when $T = 2000$. Throughout, 68% and 90% bands are surprisingly accurate, such that multiple confidence bands could be presented for researchers wishing to take a conservative approach. Finally,

Table 5: Theoretical and empirical confidence intervals of the η's

	68%L	68%U	90%L	90%U	95%L	95%U	99%L	99%U
Model 1: $\eta_1 = 0.50$, $\lambda_1 = 0.40$; $\eta_2 = 0.00$, $\lambda_2 = 0.20$								
$\eta_1 : 500$	**0.4950**	**0.5050**	**0.4906**	**0.5094**	**0.4879**	**0.5120**	**0.4817**	**0.5183**
$\hat{\eta}_1 : 500$	-0.0011	0.0011	0.0013	-0.0017	0.0030	-0.0037	0.0069	-0.0110
$\hat{\eta}_1 : 1000$	-0.0004	0.0007	0.0007	-0.0003	0.0017	-0.0013	0.0054	-0.0030
$\hat{\eta}_1 : 2000$	-0.0003	0.0002	0.0001	-0.0003	0.0005	-0.0008	0.0017	-0.0016
$\eta_2 : 500$	**-0.0116**	**0.0116**	**-0.0217**	**0.0217**	**-0.0279**	**0.0279**	**-0.0423**	**0.0423**
$\hat{\eta}_2 : 500$	0.0044	-0.0030	0.0157	-0.0111	0.0224	-0.0150	0.0487	-0.0260
$\hat{\eta}_2 : 1000$	0.0020	-0.0015	0.0069	-0.0053	0.0113	-0.0077	0.0203	-0.0128
$\hat{\eta}_2 : 2000$	0.0010	-0.0009	0.0034	-0.0030	0.0049	-0.0045	0.0103	-0.0076
Model 2: $\eta_1 = 0.98$, $\lambda_1 = 0.45$; $\eta_2 = -0.40$, $\lambda_2 = 0.30$								
$\eta_1 : 500$	**0.9790**	**0.9810**	**0.9781**	**0.9819**	**0.9775**	**0.9825**	**0.9763**	**0.9837**
$\hat{\eta}_1 : 500$	-0.0003	0.0004	0.0000	0.0003	0.0003	0.0001	0.0011	-0.0004
$\hat{\eta}_1 : 1000$	-0.0001	0.0001	0.0000	0.0000	0.0001	-0.0001	0.0007	-0.0003
$\hat{\eta}_1 : 2000$	-0.0000	0.0000	0.0001	-0.0000	0.0001	-0.0001	0.0005	-0.0002
$\eta_2 : 500$	**-0.4071**	**-0.3929**	**-0.4133**	**-0.3867**	**-0.4170**	**-0.3830**	**-0.4258**	**-0.3742**
$\hat{\eta}_2 : 500$	-0.0004	-0.0002	0.0039	-0.0059	0.0072	-0.0092	0.0165	-0.0175
$\hat{\eta}_2 : 1000$	0.0001	-0.0002	0.0025	-0.0024	0.0047	-0.0042	0.0079	-0.0086
$\hat{\eta}_2 : 2000$	-0.0001	0.0000	0.0008	-0.0010	0.0019	-0.0018	-0.0048	-0.0042
Model 3: $\eta_1 = 0.98$, $\lambda_1 = 0.45$; $\eta_2 = -0.40$, $\lambda_2 = 0.30$, $\phi = 0.80$								
$\eta_1 : 500$	**0.9790**	**0.9810**	**0.9781**	**0.9819**	**0.9775**	**0.9825**	**0.9763**	**0.9837**
$\hat{\eta}_1 : 500$	-0.0002	0.0004	0.0003	0.0002	0.0009	-0.0001	0.0026	-0.0008
$\hat{\eta}_1 : 1000$	-0.0001	0.0001	0.0001	0.0001	0.0002	-0.0001	0.0010	-0.0004
$\hat{\eta}_1 : 2000$	-0.0001	0.0001	0.0001	-0.0000	0.0002	-0.0001	0.0004	-0.0004
$\eta_2 : 500$	**-0.4071**	**-0.3929**	**-0.4133**	**-0.3867**	**-0.4170**	**-0.3830**	**-0.4258**	**-0.3742**
$\hat{\eta}_2 : 500$	0.0008	0.0008	0.0060	-0.0045	0.0104	-0.0081	0.0250	-0.0158
$\hat{\eta}_2 : 1000$	0.0004	0.0001	0.0028	-0.0020	0.0049	-0.0032	0.0108	-0.0075
$\hat{\eta}_2 : 2000$	0.0002	0.0001	0.0013	-0.0010	0.0023	-0.0017	0.0048	-0.0038

Notes: Quantities in bold font are theoretical confidence bands for $T = 500$ calculated using Theorem 3. Remaining values denote the difference between the theoretical quantity for a given confidence band and the associated percentiles from simulations from Model 1, Model 2, and Model 3.

we observe that there are no qualitative differences between the estimated bands from the GARMA(0,0) and GARMA(1,0) models, represented as Model 2 and Model 3, suggesting that the values of η_i are independent of ARMA components as implied by the proposed theory.

We also ran numerous simulations with $\eta_1 = 1$, with results available upon request, that match the findings in Chung (1996a) and Beaumont and Smallwood (2019). The results show that the distribution theory under the null $\eta_1 = 1$ are quite unreliable. In these cases, however, Beaumont and Smallwood (2019) argue that more reliable results can likely be obtained by using the confidence bands and testing procedures under the alternative $\eta < 1$, a suggestion we echo here.

5. Application

Emerging research has demonstrated that cyclical long memory is an important characteristic of many financial time series.[5] To demonstrate the applicability of the CSS estimator and the proposed theory, we consider the weekly trading volume of IBM equities measured in thousands from January 1, 1962 through July 1, 2019. Sequential estimation of single frequency models suggests the strong possibility of at least 3 sources of long memory, and inspection of the periodogram of the differenced series, which is depicted in Figure 3, suggests up to two more long memory frequencies.

We consider all combinations of k-factor GARMA(p,q) models with $k \leq 5$ and $p, q \leq 2$. Ultimately, a 5-factor GARMA(2,2) model was selected on the basis of the Akaike information criteria, and the theory outlined above, where all parameters are found to be statistically significant. Results are presented in Table 6. Based on the simulation results as discussed above, we show confidence bands for the 68% and 95% quantities under the assumption that $|\eta_i| < 1$.[6] Unambiguously, the results indicate that there potentially exists a singularity at the origin, as all confidence bands contain the value 1 for η_1.[7]

Given that the estimated values of the associated Gegenbauer frequencies, ν_i, range from 0.0009 to 2.8898, we detect cycle lengths of 13.05, 6.52, 4.35, and 2.17 weeks, in addition to the extremely long, potentially infinite cycle associated with a value of η_1 that cannot be distinguished from unity. To our knowledge, we are the first to document the potential for multiple sources of long memory in equity trading volumes, a finding that could be important in better understanding stock market behavior.

6. Conclusions

In this paper we review the properties of a model that captures very diverse patterns in the autocorrelation functions of data. The multiple frequency, or k-factor, GARMA

[5] See, Lu and Guegan (2011) and Caporale and Gil-Alana (2014) for recent applications to the Nikkei-based forward premia and price dividend ratios associated with the S&P index. Also, see Asai et al. (2018) who provide evidence of multiple sources of cyclical long memory in differenced interest rates for the US and Australia at various maturities.

[6] Recall that simulation results show that when $\eta_i = 1$, the confidence bands under the alternative are more conservative and potentially more reliable than those under the null.

[7] Estimation results applied to the first difference of volume yield similar results. The estimated value of λ_1 is equal to -0.2114, which implies a value of 0.2886 for the series in levels. All other parameter estimates, which are available upon request, indicate no tangible disparities, including, most notably, the position of the spectral poles.

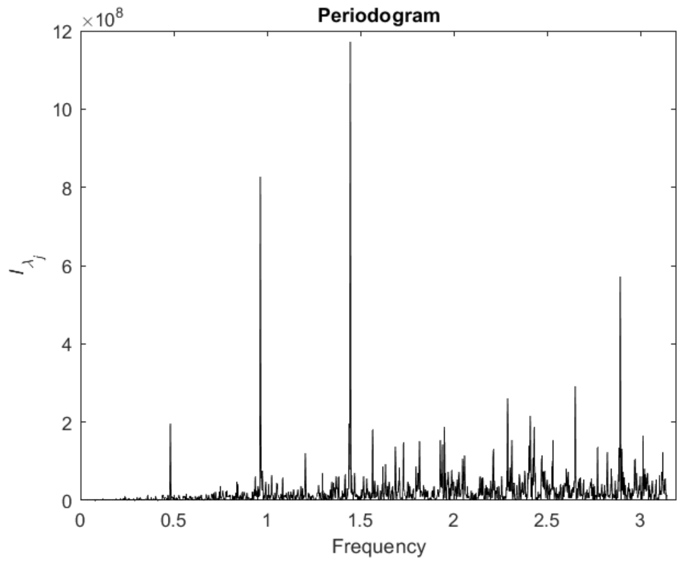

Figure 3: Periodogram of the first difference of IBM trading volume.

model generalizes existing long memory models and has the particular advantage that the autocorrelations can decay at a non-monotonic rate that is not necessarily symmetric about zero. In addition, the k-factor GARMA model can accommodate multiple singularities in the spectral density function.

Providing a full set of distributional results for estimators of k-factor GARMA models has proven elusive. Building on the results in Chung (1996a) and Chung (1996b), we study a conditional sum of squares estimator and propose its asymptotic properties. The key feature of our results is that, for all possible values, the asymptotic distribution of η_i is independent of all other parameters, including η_j whenever $i \neq j$. It is important to note, however, that remaining parameters, notably differencing parameters, are not asymptotically independent of each other, and therefore methods that sequentially estimate these values will likely suffer from severe bias. Finally, the model parameters are shown to converge at differing rates. This greatly complicates attempts to establish rigorous initial consistency proofs, especially given discontinuities in the distribution theory for η_i. We attempt to overcome this shortcoming by conducting extensive simulations and drawing on the recent work of Beaumont and Smallwood (2019) to show that the estimator performs in precisely the way our theory predicts in nearly all cases.

The simulation results show that the estimator performs well and that the finite sample standard errors are close to the asymptotic calculations. In addition, the simulation results suggest that the computational complexity associated with a k-dimensional grid search can be greatly reduced via repeated estimation of a single frequency GARMA model to obtain starting values. Further, the proposed theory can be used to accurately obtain confidence bands for estimated values of η_i. Finally, an application demonstrates the practical value of the k-factor GARMA model. The trading volume of IBM is shown

Table 6: Estimation of 5-frequency GARMA(2,2) model for IBM volume

	η_1	η_2	η_3	η_4	η_5
Estimated values	1.0000	0.8863	0.5707	0.1276	-0.9685
Lower 68% Bands	0.999999	0.8853	0.5696	0.1241	-0.9695
Upper 68% Bands	1.0000	0.8872	0.5719	0.1275	-0.9675
Lower 95% Bands	0.999997	0.8840	0.5680	0.1217	-0.9708
Upper 95% Bands	1.0000	0.8885	0.5735	0.1298	-0.9661

	λ_1	λ_2	λ_3	λ_4	λ_5
Estimated values	0.3151	0.1926	0.2779	0.2246	0.0986
Asymp. Std Errors	[0.0176]	[0.0298]	[0.0300]	[0.0318]	[0.0284]
Num. Std Errors	[0.0165]	[0.0269]	[0.0282]	[0.0282]	[0.0252]

	ϕ_1	ϕ_2	θ_1	θ_2	μ
Estimated values	-0.5855	-0.3808	-0.2198	0.1847	8409.4
Asymp. Std Errors	[0.0893]	[0.0537]	[0.1442]	[0.0862]	[7462.5]
Num. Std Errors	[0.0866]	[0.0499]	[0.1343]	[0.0809]	[4431.7]

Notes: Confidence bands are constructed assuming $\eta_i < 1$. Values in brackets are asymptotic standard errors based on the estimated model, with $\hat{\eta_i} < 1$, for all i using Theorem 2. Numerical standard errors are based on the outer product of the estimated score.

to be well modeled by a five-frequency GARMA model with a spectral singularity at the origin.

Given the early success of multiple frequency GARMA models as discussed in the introduction, our proposed estimator should find a number of important applications in a myriad of fields. Further, the proposed distribution theory will likely be useful in a number of contexts. For example, in instances where full model selection is not needed or desired, our distribution theory and simulation results suggest that single factor models can be used to robustly select Gegenbauer frequencies, thus complementing the use of methods based on maximization of the periodogram. Several challenges still remain. First, we are unable to provide a rigorous initial consistency proof for our asymptotic results, which we believe may prove quite elusive for the full model. However, it may be possible to consider a subset of Gegenbauer frequencies in this context. Perhaps related, there appears to be some concern with the use of the proposed theory when $\eta_i = 1$. Finally, computational difficulties may arise as the number of frequencies, k, becomes large. Although our estimation procedures can limit this problem by refining the grid for each η_i, recent research by Leschinski and Sibbertsen (2019) has identified cases where k could be as large as 14. It remains to be seen if the resulting complexity would limit the practical implementation of the CSS estimator for models that have a high number of spectral singularities.

Appendix

Proof of Theorem 1:

Consider the first order Taylor series expansion of the CSS estimators of the invertible and stationary k-factor GARMA model of the process $\{x_t\}_{t=1}^{T}$ about the true parameter values $\delta = (\lambda_1, \ldots, \lambda_k, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ and $\eta^* = (\eta_1, \ldots, \eta_k)$:

$$\begin{bmatrix} \frac{1}{\sqrt{T}} \frac{\partial \mathcal{L}^*}{\partial \delta} \\ \frac{1}{f_T} \odot \frac{\partial \mathcal{L}^*}{\partial \eta^*} \end{bmatrix} + \begin{bmatrix} \frac{1}{T} \frac{\partial^2 \mathcal{L}^*}{\partial \delta\, \partial \delta'} & \frac{1}{\sqrt{T}} \frac{1}{F_T} \odot \frac{\partial^2 \mathcal{L}^*}{\partial \delta\, \partial \eta^{*\prime}} \\ \frac{1}{\sqrt{T}} \frac{1}{F_T} \odot \frac{\partial^2 \mathcal{L}^*}{\partial \delta\, \partial \eta^{*\prime}} & \left(\frac{1}{f_T}\right) \left(\frac{1}{f_T}\right)' \odot \frac{\partial^2 \mathcal{L}^*}{\partial \eta^*\, \partial \eta^{*\prime}} \end{bmatrix} \times \begin{bmatrix} \sqrt{T}\left(\hat{\delta} - \delta\right) \\ f_T \odot \left(\hat{\eta^*} - \eta^*\right) \end{bmatrix} = o_p\left(1\right) \quad \text{(A.1)}$$

where $\odot$ denotes element by element multiplication, f_T and $\frac{1}{f_T}$ denote k$\times$1 vectors whose j^{th} elements are T and $\frac{1}{T}$ when $|\eta_j| < 1$ and T^2 and $\frac{1}{T^2}$ when $|\eta_j| = 1$. $\frac{1}{F_T}$ denotes the matrix formed by stacking the vector, $\frac{1}{f_T}'$, on top of itself k times.

We will show that $\frac{1}{T} \frac{\partial^2 \mathcal{L}^*}{\partial \delta\, \partial \delta'}$ and $\left(\frac{1}{f_T}\right) \left(\frac{1}{f_T}\right)' \odot \frac{\partial^2 \mathcal{L}^*}{\partial \eta^*\, \partial \eta^{*\prime}}$, are $\mathrm{O}_p(1)$, while $\frac{1}{\sqrt{T}} \frac{1}{F_T} \odot \frac{\partial^2 \mathcal{L}^*}{\partial \delta\, \partial \eta^{*\prime}}$ possesses elements that are all $o_p(1)$. Below, we show that the remaining elements are bounded and find their proposed distribution. For large T, we get

$$\begin{bmatrix} \frac{1}{\sqrt{T}\,\sigma^2} \sum_{t=1}^{T} \varepsilon_t \frac{\partial \varepsilon_t}{\partial \delta} \\ \frac{1}{\sigma^2} \frac{1}{f_T} \odot \sum_{t=1}^{T} \varepsilon_t \frac{\partial \varepsilon_t}{\partial \eta^*} \end{bmatrix} + \begin{bmatrix} I_\delta & \sqrt{T} \frac{1}{F_T} \odot I_{\delta \eta^*} \\ \sqrt{T} \frac{1}{F_T} \odot I_{\delta \eta^*} & \frac{1}{\sigma^2} \left(\frac{1}{f_T}\right) \left(\frac{1}{f_T}\right)' \odot \sum_{t=1}^{T} \frac{\partial^2 \varepsilon_t}{\partial \eta^*\, \partial \eta^{*\prime}} \end{bmatrix} \times \begin{bmatrix} \sqrt{T}\left(\hat{\delta} - \delta\right) \\ f_T \odot \left(\hat{\eta^*} - \eta^*\right) \end{bmatrix} = o_p\left(1\right). \quad \text{(A.2)}$$

The cases for $I_{\phi_i,\eta_j}, I_{\theta_i,\eta_j}$, and I_{λ_i,η_j} when $|\eta_j| = 1$ follow from Chung (1996a,b). Using Gradshteyn and Ryzhik (1980) equations 1.514 and 8.937.1, we find that the information matrix elements of I_{λ_i,η_j} are

$$-E\left[\frac{1}{T} \frac{\partial^2 \mathcal{L}^*}{\partial \lambda_i\, \partial \eta_j}\right] = E\left[\frac{4\lambda_j}{\sigma^2 T} \sum_{t=1}^{T} \left(\sum_{l=1}^{\infty} \frac{\cos(l\upsilon_i)}{l} \varepsilon_{t-l}\right) \left(\sum_{l=1}^{\infty} \frac{\sin(l\upsilon_j)}{\sin(\upsilon_j)} \varepsilon_{t-l}\right)\right]. \quad \text{(A.3)}$$

Under the assumptions governing ε_t, if $\upsilon_j > \upsilon_i$, and $\upsilon_i \neq \upsilon_j$, Gradshteyn and Ryzhik (1980) equation 1.441.1 yields

$$I_{\lambda_i\, \eta_j} = \frac{2\lambda_j}{\sin(\upsilon_j)} \left(\sum_{l=1}^{\infty} \frac{\sin[l(\upsilon_i + \upsilon_j)]}{l} + \sum_{l=1}^{\infty} \frac{\sin[l(\upsilon_j - \upsilon_i)]}{l}\right) = \frac{2\lambda_j(\pi - \upsilon_j)}{\sin(\upsilon_j)}. \quad \text{(A.4)}$$

Thus, $I_{\lambda_i\, \eta_j} < \infty$. If $\upsilon_j < \upsilon_i$, then the infinite sums in (A.4) are equal to $\sum_{l=1}^{\infty} \frac{\sin[l(\upsilon_j - \upsilon_i) + 2\pi l]}{l}$. From Gradshteyn and Ryzhik (1980) equation 1.444.1, we see that the infinite sum converges. The same is true if $\upsilon_i = \upsilon_j$.

If the remaining terms of all of the elements in (A.2) are $O_p(1)$ as shown below, then the matrix in (A.2) is asymptotically block diagonal, and the distribution of $\sqrt{T}(\hat{\delta} - \delta)$ can be considered independently of $f_T\ (\hat{\eta} - \eta)$ as claimed.

Proof of Theorem 2:.

From (A.2), the assumption that the remaining elements involving η^* in (A.2) are bounded, and the central limit theorem of Chan and Wei (1988),

$$\sqrt{T}(\hat{\delta}-\delta) = -I_\delta^{-1}\left[\frac{1}{\sqrt{T}\sigma^2}\sum_{t=1}^{T}\varepsilon_t\frac{\partial\varepsilon_t}{\partial\delta}\right] + o_p(1) \rightsquigarrow N(0, I_\delta^{-1}). \tag{A.5}$$

Information numbers for the diagonal terms of I_δ are given in Chung (1996a) (page 251). The off diagonal terms, $I_{\lambda_i\,\lambda_j}$, which for large T and $i \neq j$ are,

$$-E\left(T^{-1}\frac{\partial^2\mathcal{L}^*}{\partial\lambda_i\partial\lambda_j}\right) = E\frac{1}{T\sigma^2}\sum_{t=1}^{T}\left[\log(1-2\eta_i L + L^2)\varepsilon_t \, \log(1-2\eta_j L + L^2)\varepsilon_t\right]. \tag{A.6}$$

Using Gradshteyn and Ryzhik (1980) equations 1.514 and 1.443.3 yields,

$$I_{\lambda_i\lambda_j} = 2\sum_{l=1}^{\infty}\frac{\cos(l(\upsilon_i+\upsilon_j)) + \cos(l(\upsilon_i-\upsilon_j))}{l^2} = 2\left(\frac{\pi^2}{3} - \pi\upsilon_i + \frac{\upsilon_i^2+\upsilon_j^2}{2}\right), \tag{A.7}$$

For the CSS estimator for $\hat{\mu}$, the information number is:

$$I_\mu = -E\left(T^{-1}\frac{\partial^2\mathcal{L}^*}{\partial d_\mu^2}\right) = \frac{1}{\sigma^2}\left|\frac{\phi(1)}{\theta(1)}\right|^2\prod_{i=1}^{k}(2-2\eta_i)^{2\lambda_i} = \frac{1}{2\pi}f(0)^{-1}, \tag{A.8}$$

where $f(0)$ denotes the spectral density function evaluated at $\omega = 0$. Now consider the variance for $\bar{x}$. We have:

$$var\left[\sqrt{T}(\bar{x}-\mu)\right] = \frac{\sigma^2}{T}\left[T\left\{\left|\frac{\theta(1)}{\phi(1)}\right|^2\prod_{i=1}^{k}(2-2\eta_i)^{-\lambda_i}\right\}\right] = 2\pi f(0). \tag{A.9}$$

By the central limit theorem of Chan and Wei, we also have $\sqrt{T}(\bar{x}-\mu) \rightsquigarrow N(0, 2\pi f(0))$. The proof of the results for $I_{\lambda_1\phi_j}, I_{\lambda_1\theta_m}, \ldots, I_{\lambda_k\phi_j}$, and $I_{\lambda_k\theta_m}$ follows directly from Chung (1996b) in the single frequency case.

Proof of theorem 3:

Before proving Theorem 3 we state and prove the following useful lemma.

Lemma 1. *Let $\hat{\eta}_1, \ldots, \hat{\eta}_k$ be the CSS estimators for $\eta^* = (\eta_1, \ldots, \eta_k)$ in a stationary and invertible k-factor GARMA model. Then, with $i \neq j$,*

$$\frac{1}{T^\alpha}\frac{\partial^2\mathcal{L}^*}{\partial\eta_i\partial\eta_j} = o_p(1),$$

where $\alpha = 2$ if $|\eta_i|, |\eta_j| < 1, \{i, j \in [1,k] : i \neq j\}$ (case 1), $\alpha = 3$ if $\eta_i = \pm 1$ and $|\eta_j| < 1$ (cases 2 and 3), and $\alpha = 4$ if $\eta_i = -1$ and $\eta_j = 1$ (case 4).

Proof of the Lemma:

Case 1: $|\eta_i|, |\eta_j| < 1, \{i, j \in [1,k] : i \neq j\}$. Without loss of generality, and for ease of notation, rearrange the terms in η^* such that $\eta_i = \eta_1, \eta_j = \eta_2$. Let,

$$Z_{at} = -\frac{1}{2\lambda_a}\frac{\partial \varepsilon_{t+1}}{\partial \eta_a} = \frac{\varepsilon_t}{(1 - 2\eta_a L + L^2)}, \ a = 1, 2. \tag{A.10}$$

Applying Gradshteyn and Ryzhik (1980) equation 8.937.1,

$$Z_{at} = \frac{1}{\sin(\upsilon_a)} \sum_{j=1}^{t} \sin[(t+1)\upsilon_a - j\upsilon_a]\varepsilon_j, \ a = 1, 2 \tag{A.11}$$

which follows if $\varepsilon_0 = \varepsilon_{-1} = ... = 0$. Now, define the random elements

$$S_T(\upsilon_a, r) = \frac{\sqrt{2}}{\sqrt{T\sigma^2}} \sum_{j=1}^{[Tr]} \cos(j\upsilon_a)\varepsilon_j, \ a = 1, 2 \tag{A.12a}$$

$$T_T(\upsilon_a, r) = \frac{\sqrt{2}}{\sqrt{T\sigma^2}} \sum_{j=1}^{[Tr]} \sin(j\upsilon_a)\varepsilon_j, \ a = 1, 2 \tag{A.12b}$$

where $r \in [0, 1]$ and $[Tr]$ is the integer part. Finally, from the expressions in (A.12) and using $\omega_1 = \upsilon_1 + \upsilon_2$, $\omega_2 = \upsilon_1 - \upsilon_2$, $\omega_3 = \upsilon_2 - \upsilon_1$ along with a few rules of trigonometry, we get the following expression,

$$\begin{aligned}
&\frac{4\sin(\upsilon_1)\sin(\upsilon_2)}{\sigma^2}\frac{1}{T^2}\sum_{t=1}^{T-1} Z_{1t} Z_{2t} \\
&= \frac{1}{T}\sum_{t=1}^{T-1}(\cos[(t+1)\,\omega_2] - \cos[(t+1)\omega_1])S_T(\upsilon_1, t/T)S_T(\upsilon_2, t/T) \\
&- \frac{1}{T}\sum_{t=1}^{T-1}(\sin[(t+1)\omega_1] + \sin[(t+1)\omega_3])S_T(\upsilon_2, t/T)\,T_T(\upsilon_1, t/T) \\
&- \frac{1}{T}\sum_{t=1}^{T-1}(\sin[(t+1)\omega_1] + \sin[(t+1)\omega_2])S_T(\upsilon_1, t/T)T_T(\upsilon_2, t/T) \\
&+ \frac{1}{T}\sum_{t=1}^{T-1}(\cos[(t+1)\omega_1] + \cos[(t+1)\omega_2])T_T(\upsilon_1, t/T)T_T(\upsilon_2, t/T).
\end{aligned} \tag{A.13}$$

Consider the random elements

$$S_n^*(\upsilon_1) = \sum_{j=1}^{n} \cos(j\upsilon_1)\,\varepsilon_j \quad \text{and} \quad T_n^*(\upsilon_1) = \sum_{j=1}^{n} \sin(j\upsilon_1)\,\varepsilon_j, \tag{A.14}$$

and similarly for $S_n^*(\upsilon_2)$ and $T_n^*(\upsilon_2)$. Let $\{X_n\} = \{S_n^*(\upsilon_1)\,S_n^*(\upsilon_2)\}$, and consider the first term in (A.13). It is clear from the definition of $S_T(\upsilon_1, \frac{t}{T})$ and $S_T(\upsilon_2, \frac{t}{T})$ that

$\frac{1}{T}\sum_{t=1}^{T-1}\cos[(t+1)\omega_2]S_T(\upsilon_1,t/T)S_T(\upsilon_2,t/T)=o_p(1)$ if

$$\underset{1\le j\le T}{sup}\left|\sum_{n=1}^{j}e^{in\theta}X_n\right|=o_p(T^2). \tag{A.15}$$

First, observe that

$$E|S_n^*(\upsilon_1)S_n^*(\upsilon_2)|\le\{ES_n^{*2}(\upsilon_1)\}^{1/2}\{ES_n^{*2}(\upsilon_2)\}^{1/2}\le\sigma^2 n \tag{A.16}$$

so that $E|S_n^*(\upsilon_1)S_n^*(\upsilon_2)|=O(n)$. Now let $n\ge m$ and consider

$$|X_n-X_m|\le|S_n^*(\upsilon_1)||S_n^*(\upsilon_2)-S_m^*(\upsilon_2)|+|S_m^*(\upsilon_2)||S_n^*(\upsilon_1)-S_m^*(\upsilon_1)|. \tag{A.17}$$

Noting that

$$E|S_n^*(\upsilon_1)|^2=E(S_n^*(\upsilon_1))^2=\sigma^2\left(\sum_{j=1}^{n}\cos(j\upsilon_1)\right)^2\le\sigma^2 n \tag{A.18}$$

yields $E|S_n^*(\upsilon_1)|^2=O(n)$. Given $m\le n$, this also implies $E|S_m^*(\upsilon_2)|^2\le\sigma^2 n$.

Next consider the expression

$$E|S_n^*(\upsilon_2)-S_m^*(\upsilon_2)|^2=\sigma^2\left(\sum_{j=m+1}^{n}\cos^2(j\upsilon_2)\right)\le\sigma^2(n-m). \tag{A.19}$$

Thus, $E|S_n^*(\upsilon_2)-S_m^*(\upsilon_2)|^2=O(n-m)$. Similar reasoning implies that $E|S_n^*(\upsilon_1)-S_m^*(\upsilon_1)|^2=O(n-m)$. If $\upsilon_1\ne\upsilon_2$, by Theorem 2.1 in Chan and Wei (1988), we see that the first term in (A.13) is $o_p(1)$. By similar reasoning, the remaining terms in (A.13) are also seen to be $o_p(1)$. Thus, we have established that

$$\frac{4\sin(\upsilon_1)\sin(\upsilon_2)}{\sigma^2}\frac{1}{T^2}\sum_{t=1}^{T-1}Z_{1t}Z_{2t}=o_p(1). \tag{A.20}$$

This expression is asymptotically equivalent to

$$-\frac{4\lambda_1\lambda_2}{4\sin(\upsilon_1)\sin(\upsilon_2)}\frac{4\sin(\upsilon_1)\sin(\upsilon_2)}{\sigma^2}\frac{1}{T^2}\sum_{t=1}^{T-1}Z_{1t}Z_{2t}, \tag{A.21}$$

which is $o_p(1)$. So this completes the proof of Case 1 in the Lemma.

Case 2: Without loss of generality, let $\eta_k=1$, $|\eta_j|<1$, and $j\ne k$. Rearrange the polynomials in η^* such that $\eta_j=\eta_1$, and define the following elements,

$$Z_{1t}=-\frac{1}{2\lambda_1}\frac{\partial\varepsilon_{t+1}}{\partial\eta_1}=\frac{\varepsilon_t}{(1-2\eta_1L+L^2)}\ ,\ Z_{2t}=-\frac{1}{2\lambda_k}\frac{\partial\varepsilon_{t+1}}{\partial\eta_k}=\frac{\varepsilon_t}{(1-L)^2}. \tag{A.22}$$

Define the auxiliary process and its associated truncation.

$$Y_t = (1-L)Z_{2t} = \sum_{j=1}^{t} \varepsilon_j. \tag{A.23}$$

This gives the following truncated series forZ_{2t},

$$Z_{2t} = \sum_{j=1}^{t} Y_t = \sum_{j=1}^{t} j\,\varepsilon_{t-j+1}. \tag{A.24}$$

For ease of exposition, define the random process

$$X_T(r) = \frac{1}{T}\frac{1}{\sqrt{T}\sigma}\sum_{j=1}^{[Tr]} Y_j, \tag{A.25}$$

and define $S_T(\upsilon_1, t/T)$ and $T_T(\upsilon_1, t/T)$ precisely as in (A.12). We then get,

$$\begin{aligned}\frac{\sqrt{2}\sin(\upsilon_1)}{\sigma^2}\frac{1}{T^3}\sum_{t=1}^{T-1} Z_{1t}Z_{2t} = \frac{1}{T}\sum_{t=1}^{T-1} &\sin[(t+1)\upsilon_1]\, S_T(\upsilon_1, t/T)\, X_T(t/T) \\ &- \frac{1}{T}\sum_{t=1}^{T-1}\cos[(t+1)\upsilon_1]\, T_T(\upsilon_1, t/T)\, X_T(t/T).\end{aligned} \tag{A.26}$$

Note that the expression

$$-\frac{4\lambda_1\lambda_k}{\sqrt{2}\sin(\upsilon_1)}\frac{\sqrt{2}\sin(\upsilon_1)}{\sigma^2}\frac{1}{T^3}\sum_{t=1}^{T-1} Z_{1t}Z_{2t}, \tag{A.27}$$

is asymptotically equivalent to $\frac{1}{T^3}\frac{\partial^2\mathcal{L}^*}{\partial\eta_1\partial\eta_k}$. Define the processes

$$S_n^*(\upsilon_1) = \sum_{j=1}^{n}\cos(j\upsilon_1)\varepsilon_j,\ T_n^*(\upsilon_1) = \sum_{j=1}^{n}\sin(j\upsilon_1)\varepsilon_j, \text{ and } X_n^* = \sum_{j=1}^{n} Y_j, \tag{A.28}$$

to facilitate the analysis. It is easy to verify that

$$\frac{1}{T}\sum_{t=1}^{T-1}\sin[(t+1)\upsilon_1]\, S_T(\upsilon_1, t/T)\, X_T(t/T) = o_p(1) \tag{A.29}$$

if

$$\sum_{n=1}^{T-1}\sin[(n+1)\upsilon_1]\, S_n^*(\upsilon_1)X_n^* = o_p(n^3). \tag{A.30}$$

The same is true for the second term in (A.26). From (A.18) $ES_n^*(\upsilon_1)^2 \le \sigma^2 n$. From

Gradshteyn and Ryzhik (1980) equation 0.121.2, we have

$$EX_n^{*2} = E\left[\sum_{j=1}^{n} j\,\varepsilon_{t-j+1}\right]^2 = \sigma^2\sum_{j=1}^{n} j^2 = \sigma^2\frac{2n^3+3n^2+n}{6} \leq \sigma^2 n^3. \tag{A.31}$$

Given, $E|S_n^*(\upsilon_1)X_n^*| \leq \{ES_n^*(\upsilon_1)^2\}^{1/2}\{EX_n^{*2}\}^{1/2}$, we see that $E|S_n^*(\upsilon_1)X_n^*|$ is $O(n^2)$. Now let $n \geq m$ and consider

$$|S_n^*(\upsilon_1)X_n^* - S_m^*(\upsilon_1)X_m^*| \leq |S_n^*(\upsilon_1)||X_n^* - X_m^*| + |X_m^*||S_n^*(\upsilon_1) - S_m^*(\upsilon_1)|. \tag{A.32}$$

Clearly, $E|S_n^*(\upsilon_1)|^2 \leq \sigma^2 n$, and from (A.19), $E|S_n^*(\upsilon_1) - S_m^*(\upsilon_1)|^2 \leq \sigma^2(n-m)$. From (A.31) we have, $E|X_m^*|^2 \leq \sigma^3 m^3 \leq \sigma^2 n^3$. Finally, given Y_j from (A.23),

$$E|X_n^* - X_m^*|^2 = E\left(\sum_{j=m+1}^{n} Y_j\right)^2 \tag{A.33}$$

$$= (n-m)^2\sum_{j=1}^{m}\sigma^2 + \sigma^2\sum_{j=1}^{n-m} j^2 \leq \sigma^2(n^3 - 2n^2m + n^2m) = \sigma^2\{n^2(n-m)\}. \tag{A.34}$$

Thus, from Theorem 2.1 in Chan and Wei (1988),

$$\underset{1\leq j\leq n}{sup}\left|\sum_{t=1}^{j} e^{it\theta}S_t^*X_t^*\right| = o_p(n^3) \tag{A.35}$$

which implies that the first term in (A.26) is $o_p(1)$. Following the same reasoning, the second term in (A.26) is also $o_p(1)$ and this proves Case 2 of the Lemma.

Case 3: Without loss of generality, let $\eta_1 = -1$ and $|\eta_j| < 1$, $j \neq 1$. Rearrange the polynomials in η^* such that $\eta_2 = \eta_j$. Then,

$$Z_{1t} = -\frac{1}{2\lambda_1}\frac{\partial\varepsilon_{t+1}}{\partial\eta_1} = \frac{\varepsilon_t}{(1+L)^2} = \sum_{j=1}^{t}(-1)^{j+1}j\,\varepsilon_{t-j+1}. \tag{A.36}$$

Define the process

$$X_T(t/T) = \begin{cases} \frac{1}{T}\frac{1}{\sqrt{T}\sigma}\sum_{j=1}^{t}(-1)^{j+1}j\,\varepsilon_{t-j+1} \text{ if } t \text{ is odd} \\ \frac{1}{T}\frac{1}{\sqrt{T}\sigma}\sum_{j=1}^{t}(-1)^{j}j\,\varepsilon_{t-j+1} \text{ if } t \text{ is even} \end{cases}. \tag{A.37}$$

Let $\omega_1 = (\upsilon_2 + \pi)$ and $\omega_2 = (\upsilon_2 - \pi)$. Noting that $T^{3/2}\sigma\, X_T(t/T)\, cos[(t+1)\pi] = Z_{1t}$, and defining Z_{2t} as in (A.10), we get

$$
\begin{aligned}
&\frac{2\sqrt{2}\sin(\upsilon_2)}{\sigma^2}\frac{1}{T^3}\sum_{t=1}^{T-1} Z_{1t}Z_{2t} \\
&= \frac{1}{T}\sum_{t=1}^{T-1}(\sin[(t+1)\omega_1] + \sin[(t+1)\omega_2])S_T(\upsilon_2, t/T)X_T(t/T) \\
&- \frac{1}{T}\sum_{t=1}^{T-1}(\cos[(t+1)\omega_1] + \sin[(t+1)\omega_2])S_T(\upsilon_2, t/T)\,X_T(t/T) \qquad \text{(A.38)}
\end{aligned}
$$

Construct the variable $S_n^*(\upsilon_2)$ as above and the auxiliary variable X_n^* as

$$
X_n^* = \begin{cases} \sum\limits_{j=1}^{n}(-1)^{j+1}j\varepsilon_{n-j+1} \text{ if } n \text{ is odd} \\ \sum\limits_{j=1}^{n}(-1)^{j}j\varepsilon_{n-j+1} \text{ if } n \text{ is even.} \end{cases} . \qquad \text{(A.39)}
$$

Using these definitions we get

$$
\frac{1}{T}\sum_{t=1}^{T-1}\sin[(t+1)\omega_1]\,S_T(\upsilon_2, t/T)\,X_T(t/T) = o_p(1) \qquad \text{(A.40)}
$$

if $\sum_{n=1}^{T-1}\sin[(n+1)\omega_1]\,S_n^*(\upsilon_2)\,X_n^* = o_p(n^3)$. Again, $\{ES_n^*(\upsilon_2)^2\}^{1/2} \leq \sigma\sqrt{n}$ and $E|S_n^*(\upsilon_2)X_n^*| \leq \{ES_n^*(\upsilon_2)^2\}^{1/2}\{EX_n^{*2}\}^{1/2}$. Now, if n is odd, we have

$$
E\left(X_n^*\right)^2 = E\left(\sum_{j=1}^{n}(-1)^{j+1}j\,\varepsilon_{t-j+1}\right)^2 = \sigma^2\sum_{j=1}^{n}j^2 \leq \sigma^2 n^3, \qquad \text{(A.41)}
$$

and precisely the same reasoning holds if n is even. This implies that $E|S_n^*(\upsilon_2)X_n^*|$ is $O(n^2)$. We know that

$$
|S_n^*(\upsilon_2)X_n^* - S_m^*(\upsilon_2)X_m^*| \leq |S_n^*(\upsilon_2)||X_n^* - X_m^*| + |X_m^*||S_n^*(\upsilon_2) - S_m^*(\upsilon_2)| \qquad \text{(A.42)}
$$

where the bounds on $|S_n^*(\upsilon_2)|$ and $|S_n^*(\upsilon_2) - S_m^*(\upsilon_2)|$ were established in (A.18) and (A.19), respectively, and the bound on $|X_m^*|$ was established in the discussion above (A.33). Now, choosing $n \geq m$ for n odd and m even, gives

$$
E|X_n^* - X_m^*|^2 = E[(n-m)\sum_{j=1}^{m}(-1)^{j+1}\varepsilon_j + \sum_{j=1}^{n-m}(-1)^{j+1}j\varepsilon_{n-j+1}]^2 \leq \sigma^2[n^2(n-m)]. \qquad \text{(A.43)}
$$

The result holds for any permutations of n and m. By Theorem 2.1 of Chan and Wei (1988), the first term in (A.38) is $o_p(1)$ and, by exactly the same reasoning, the remaining terms are also $o_p(1)$. This completes the proof of Case 3.

Case 4: Without loss of generality, let $\eta_1 = -1, \eta_k = 1$, with $|\eta_j| < 1$, for $j \neq 1, k$.

Define the following elements:

$$Z_{1t} = -\frac{1}{2\lambda_1}\frac{\partial \varepsilon_{t+1}}{\partial \eta_1} = \frac{\varepsilon_t}{(1+L)^2} = \sum_{j=1}^{t}(-1)^{j+1}j\ \varepsilon_{t-j+1} \tag{A.44}$$

$$Z_{kt} = -\frac{1}{2\lambda_k}\frac{\partial \varepsilon_{t+1}}{\partial \eta_k} = \frac{\varepsilon_t}{(1-L)^2} = \sum_{j=1}^{t} j\ \varepsilon_{t-j+1} \tag{A.45}$$

$$X^*_{1t} = \begin{cases} \sum_{j=1}^{t}(-1)^{j+1} j\varepsilon_{t-j+1} \text{ if } t \text{ is odd} \\ \sum_{j=1}^{t}(-1)^{j} j\varepsilon_{t-j+1} \text{ if } t \text{ is even} \end{cases}. \tag{A.46}$$

Then,

$$\frac{4\lambda_1\lambda_k}{T^4}\sum_{t=1}^{T-1} Z_{1t}\, Z_{kt} = \frac{4\lambda_1\lambda_k}{T^4}\sum_{t=1}^{T-1}\cos[(t+1)\pi]\, X^*_{1t}\, X^*_{kt} \tag{A.47}$$

where X^*_{kt} is defined similarly to X^*_n in (A.28). This allows us to apply Theorem 2.1 in Chan and Wei (1988) to show that the last expression is $o_p(1)$ if

$$\underset{1\le j\le n}{sup}\left|\sum_{t=1}^{j} e^{it\theta} X^*_{1t}X^*_{kt}\right| = o_p(n^4). \tag{A.48}$$

Now let X^*_{1n} and X^*_{kn} be defined equivalently to X^*_{1t} and X^*_{kt} with the sequence of partial sums running to n rather than t. From the definition of X^*_{1n} we have

$$EX^{*2}_{1n} = \sigma^2\sum_{j=1}^{n} j^2 \le \sigma^2 n^3. \tag{A.49}$$

From (A.31)

$$E|X^*_{1n}X^*_{kn}| \le \{EX^{*2}_{1n}\}^{1/2}\{EX^{*2}_{kn}\}^{1/2} \le \sigma^2\{n^{3/2}\}\{n^{3/2}\}. \tag{A.50}$$

Choose n and m as integers greater than 0 with $n \ge m$. Then,

$$|X^*_{1n}X^*_{kn} - X^*_{1m}X^*_{km}| \le |X^*_{1n}||X^*_{kn} - X^*_{km}| + |X^*_{km}||X^*_{1n} - X^*_{1m}|. \tag{A.51}$$

From (A.31) and (A.49), we know that $E|X^*_{km}|^2$ and $E|X^*_{1n}|^2$ are both $O(n^3)$, while from (A.34) and (A.43), $E|X^*_{kn} - X^*_{km}|^2$ and $E|X^*_{1n} - X^*_{1m}|^2$ are $O(n^2(n-m))$. By Theorem 2.1 in Chan and Wei (1988), the sequence in (A.48) is $o_p(n^4)$ and thus the sequence in (A.47) is $o_p(1)$. This completes the proof of the Lemma.

Theorem 3 follows from the lemma, Theorem 2.2 in Chan and Wei (1988), and Theorem 1 in Chung (1996a). Note that for the jth element of η^*, we get

$$T^a(\hat{\eta}_j - \eta_j) = -[\frac{1}{T^{2a}}\sum_{t=1}^{T}(\frac{\partial \varepsilon_t}{\partial \eta_j})^2]^{-1}[\frac{1}{T^a}\sum_{t=1}^{T}\varepsilon_t\frac{\partial \varepsilon_t}{\partial \eta_j}] + o_p(1), \tag{A.52}$$

where $a = 1$ if $|\hat{\eta}_j| < 1$, and $a = 2$ if $|\hat{\eta}_j| = 1$. The theorem is complete as this is precisely the k-factor version of equation (A.5) in Chung (1996a).

References

References

Adenstedt, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *The Annals of Statistics*, 2(6):1095–1107.

Andrews, D. W. and Sun, Y. (2004). Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica*, 72(2):569–614.

Asai, M., Peiris, S., McAleer, M., and Allen, D. (2018). Cointegrated dynamics for a generalized long memory process. Technical Report EI2018-32, Econometric Institute Research Papers.

Beaumont, P. M. and Smallwood, A. D. (2019). Inference for likelihood-based estimators of generalized long-memory processes. *Working Paper, University of Texas-Arlington*.

Caporale, G. M. and Gil-Alana, L. (2014). Long-run and cyclical dynamics in the US stock market. *Journal of Forecasting*, 33(2):147–161.

Caporale, G. M. and Gil-Alana, L. A. (2011). Multi-factor Gegenbauer processes and European inflation rates. *Journal of Economic Integration*, pages 386–409.

Chan, N. H. and Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *The Annals of Statistics*, 16(1):367–401.

Cheung, Y.-W. and Diebold, F. X. (1994). On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean. *Journal of Econometrics*, 62(2):301 – 316.

Chung, C.-F. (1996a). Estimating a generalized long memory process. *Journal of Econometrics*, 73(1):237 – 259.

Chung, C.-F. (1996b). A generalized fractionally integrated autoregressive moving average process. *Journal of Time Series Analysis*, 17:111–140.

Chung, C.-F. and Baillie, R. T. (1993). Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA models. *Empirical Economics*, 18(4):791–806.

Diongue, A. K. and Ndongo, M. (2016). The k-factor GARMA process with infinite variance innovations. *Communications in Statistics-Simulation and Computation*, 45(2):420–437.

Dissanayake, G. S., Peiris, M. S., and Proietti, T. (2018). Fractionally differenced Gegenbauer processes with long memory: A review. *Statistical Science*, 33(3):413–426.

Gil-Alanaa, L. A. (2007). Testing the existence of multiple cycles in financial and economic time series. *Annals of Economics & Statistics*, 8(1):1–20.

Giraitis, L., Hidalgo, J., and Robinson, P. M. (2001). Gaussian estimation of parametric spectral density with unknown pole. *The Annals of Statistics*, 29(4):987–1023.

Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Tables of integrals, series, and products*. Academic Press, New York, NY, 4th edition.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–24.

Gray, H. L., Zhang, N., and Woodward, W. A. (1989). On generalized fractional processes. *Journal of Time Series Analysis*, 10:233–257.

Hidalgo, J. and Soulier, P. (2004). Estimation of the location and exponent of the spectral singularity of a long memory process. *Journal of Time Series Analysis*, 25(1):55–81.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68:165–76.

Leschinski, C. and Sibbertsen, P. (2019). Model order selection in periodic long memory models. *Econometrics and Statistics*, 9:78 – 94.

Lu, Z. and Guegan, D. (2011). Estimation of time-varying long memory parameter using wavelet method. *Communications in Statistics—Simulation and Computation*, 40(4):596–613.

McElroy, T. S. and Holan, S. H. (2016). Computation of the autocovariances for time series with multiple long-range persistencies. *Computational Statistics & Data Analysis*, 101:44–56.

Peiris, M. and Asai, M. (2016). Generalized fractional processes with long memory and time dependent volatility revisited. *Econometrics*, 4(4):37.

Ramachandran, R. and Beaumont, P. (2001). Robust estimation of GARMA model parameters with an application to cointegration among interest rates of industrialized countries. *Computational Economics*, 17(2/3):179–201.

Smallwood, A. D. and Norrbin, S. C. (2006). Generalized long memory processes, failure of cointegration tests and exchange rate dynamics. *Journal of Applied Econometrics*, 21(4):409–417.

Woodward, W. A., Cheng, Q. C., and Gray, H. L. (1998). A k-factor GARMA long-memory model. *Journal of Time Series Analysis*, 19(485-504).

Yajima, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *The Annals of Statistics*, 19(1):158–177.