



Conditional sum of squares estimation of k -factor GARMA models

Paul M. Beaumont¹ · Aaron D. Smallwood² 

Received: 24 December 2022 / Accepted: 15 September 2023
© Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

We analyze issues related to estimation and inference for the constrained sum of squares estimator (CSS) of the k -factor Gegenbauer autoregressive moving average (GARMA) model. We present theoretical results for the estimator and show that the parameters that determine the cycle lengths are asymptotically independent, converging at rate T , the sample size, for finite cycles. The remaining parameters lack independence and converge at the standard rate. Analogous with existing literature, some challenges exist for testing the hypothesis of non-cyclical long memory, since the associated parameter lies on the boundary of the parameter space. We present simulation results to explore small sample properties of the estimator, which 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 Gegenbauer processes · Asymptotic distributions · ARFIMA · Long memory · Equity trading volume · Bootstrap tests

Mathematics Subject Classification 91B84 · 62M10 · 62F12 · 62P20 · 65C05

Paul M. Beaumont and Aaron D. Smallwood have contributed equally to this work.

✉ Aaron D. Smallwood
smallwood@uta.edu

¹ Department of Economics, Florida State University, Tallahassee, FL 32306, USA

² Department of Economics, University of Texas Arlington, 701 S. West Street, mailbox: 19479, Arlington, TX 76019, USA

1 Introduction

The k -factor Gegenbauer autoregressive moving average (GARMA) model nests ARIMA, fractionally integrated ARMA (ARFIMA), seasonal ARFIMA, and single-factor GARMA models as special cases and may simultaneously include features of each. The k -factor GARMA(p, q) model is defined as

$$\phi(L) \prod_{i=1}^k (1 - 2\eta_i L + L^2)^{d_i} (x_t - \mu) = \theta(L) \varepsilon_t. \quad (1)$$

Here $|\eta_i| \leq 1$, d_i are memory parameters, and $\phi(L)$ and $\theta(L)$ are p and q order polynomials in the lag operator L such that $\phi(z) = 0$ and $\theta(z) = 0$ have all roots outside the unit circle and no common zeros. Further, $\{\varepsilon_t\}$ is a white noise disturbance sequence.

These long-memory models 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. Excellent recent reviews of the estimation methods for GARMA processes were provided by Dissanayake et al. (2018) and Hunt et al. (2022), who discuss difficulties with obtaining theoretical distribution results for estimators of the model parameters. Of particular note, there appears to be little existing distribution theory for maximum likelihood-based estimation methods in the time domain when $k > 1$. In this paper, we address this void by presenting a conditional sum of squares (CSS) estimator along with proposed joint asymptotic distributions for all parameters in the k -factor model. Simulation experiments generally validate the theoretical distributions. As an application, we model the trading volume of IBM equities, finding evidence of complex long-memory dynamics.

Long-memory models were popularized by Granger and Joyeux (1980), Hosking (1981) and Granger (1980, 1981) who introduced fractional differencing as a means of capturing complicated stochastic properties of data in the time and frequency domains. These models have proven especially useful by bridging the gap between infinite variance unit root processes and finite variance short memory processes. One shortcoming of the original fractionally differenced models, however, is that they are incapable of capturing long-memory processes with persistent cycles in the ACF. Gray et al. (1989), along with the correction in Gray et al. (1994), addressed this issue with the GARMA model, which was generalized by Woodward et al. (1998) to allow for multiple sources of cyclic long memory. The general 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(0,0) and GARMA(0,0) components, such that the ACF decays non-monotonically at a hyperbolic rate and is asymmetric about zero, such as shown in Fig. 1.

Due to its flexibility, the k -factor GARMA approach has proven very useful for modeling many physical, economic, and financial time series that exhibit complex long-memory features. For solar activity, Gray et al. (1989) and Chung (1996b) estimate a single-factor model for sunspots, while Maddanu and Proietti (2022)

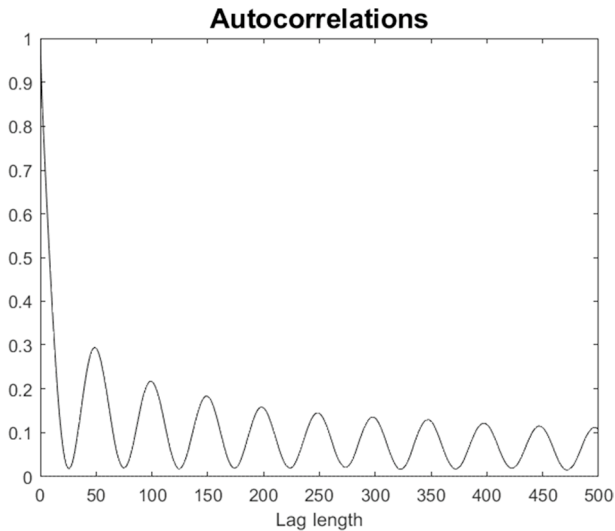


Fig. 1 ACF of a process with both an ARFIMA and a GARMA component

considered a model with $k = 4$, ultimately isolating a single long-memory cycle of about 11 years. Woodward et al. (1998) and Diongue and Ndong (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-Alana 2007; Asai et al. 2020), 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-Alana 2007; Beaumont and Smallwood 2022). The possibility of many 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 k -factor GARMA model, a unifying estimation approach does not appear to exist. Almost all studies assume the positions of the singularities are known (for example, Caporale and Gil-Alana (2011) and Arteche (2020)), or employ two-step procedures where the Gegenbauer frequencies are typically first estimated by inspection of the periodogram (for example, Hidalgo and Soulier (2004), Lu and Guegan (2011) and Asai et al. (2020), amongst others). Only a handful of studies have attempted to simultaneously estimate all model parameters, including memory parameters and the positions of the spectral poles, known as Gegenbauer frequencies. In this context, wavelet procedures have been used by Lu and Guegan (2011), Alomari et al. (2020), and Ayache et al. (2022) and offer a promising semi-parametric alternative to estimation of spectral poles. However, these methods have only generally been used to estimate models with $k = 1$. Specifically, Alomari et al. (2020) and Ayache et al. (2022) consider time series processes having spectra encompassing the 1-factor GARMA model as a special case. Alomari et al. (2020) do establish consistency

for the frequency parameter using the wavelet-based method of Bardet and Bertrand (2010) who introduced a nonparametric approach to spectral density estimation. The result was extended by Ayache et al. (2022) to establish asymptotic normality for the estimators. In the time domain, Dissanayake et al. (2016) provided distributional results using a state-space approach based on associated Gegenbauer polynomials and the Kalman filter to obtain likelihood-based estimates for the 1-factor GARMA(0,0) model with $|\eta| < 1$. Kouamé and Hili (2008, 2012) use minimum distance estimators and show consistency and asymptotic normality for estimators of differencing parameters, although specific knowledge of η_i is generally required.

A major difficulty in generalizing distribution theory for the full k -factor model lies in the fact that estimators of the parameters dictating the positions of the spectral poles appears to be non-standard, with rates of convergence that may differ relative to those of other parameters. Additionally, the relevant parameter space is closed, whereas successful attempts to establish distributional results for estimators in the time domain generally exclude the zero-frequency as an admissible value (see, Kouamé and Hili (2008) and Dissanayake et al. (2016), for example). Further, maximum likelihood-based estimators in the frequency domain typically use a discrete set of frequencies 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 in the models considered here, we are unaware of any study proposing a full set of distributional findings for any estimator. The strongest results appear to have been offered by Hidalgo (2005), who considers a semiparametric estimator of the memory parameter and position of the spectral pole for processes having spectra consistent with the GARMA process. Hidalgo (2005) rigorously establishes theoretical results for estimation of the underlying model parameters, even when the singularity occurs at the origin. 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. In the time domain, with a known spectral pole at the origin, Robinson (2006) establishes consistency and asymptotic normality for the CSS estimator of the parameters for a general model that includes stationary ARFIMA processes as a special case. As referenced above, for spectral poles that do not include the origin, partial results are available from Kouamé and Hili (2008) and Dissanayake et al. (2016).

With $k = 1$, promising results for the CSS method were proposed by Chung (1996a, 1996b), who attempted to establish complete distributional results for all parameters. The method of proof of Chung relied on the observation that, for the true parameter values, the expectation of the approximate likelihood function is zero. The results of Chung are seen as somewhat controversial, as there were no attempts to constrain the position of the unknown spectral pole. In fact, Chung (1996a) argues that there is a discontinuity in the distribution at the zero frequency. Perhaps more remarkably, with T denoting the sample size, Chung (1996a) argues that the associated estimate of the Gegenbauer frequency achieves a T^2 -rate of convergence when the spectral pole occurs at 0 or π , while it is otherwise $O_p(T^{-1})$. Most importantly, as initially pointed out by Giraitis et al. (2001), Chung (1996a) was unable to provide a rigorous initial proof establishing consistency. Additionally, Beaumont and

Smallwood (2022) provide extensive simulation evidence yielding some support for theoretical concerns when the position of the spectral pole occurs at the origin.

Although the results of Chung (1996a) may appear tenuous, the CSS estimator provides a feasible and relatively simple method to obtain joint estimation results for the GARMA model parameters. Additionally, the consistency proof established by Robinson (2006) for the CSS estimator likely extends to the k -factor GARMA process. Notwithstanding the concerns when the Gegenbauer frequency is 0, the simulation evidence of Beaumont and Smallwood (2022) otherwise generally supports the results of Chung (1996a). Beaumont and Smallwood (2022) also demonstrate that the CSS method generally obtains a smaller bias for estimation of the spectral pole relative to the Whittle counterpart. Diongue and Ndongo (2016) provide similar evidence, demonstrating that, compared to a Whittle-based estimator, 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.

Here, for the k -factor GARMA model, we study the CSS estimator described by Chung and Baillie (1993) for ARFIMA models and by Chung (1996a, 1996b) for single-factor GARMA models. All parameters are simultaneously estimated, including the ARMA components. Furthermore, we propose an asymptotic distribution for all parameters in the model, where, to our knowledge, only partial results are currently available. The results show that the estimates of each Gegenbauer frequency are asymptotically independent of all other model parameters. We provide simulation evidence to help validate the results. The simulation evidence, including additional results in Beaumont and Smallwood (2022), demonstrates that the theory can typically be reliably used to provide inference for the estimated parameters. To the extent that there are concerns with testing for models with a spectral pole at the origin, we provide a simple parametric bootstrap procedure based on our estimator.

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

2 k -Factor GARMA processes

The k -factor GARMA model, defined in Eq. (1), was originally discussed by Gray et al. (1989) and presented in greater detail by Woodward et al. (1998). More specifically, with $i = 1, \dots, k$, the d_i are memory parameters, and η_i dictate the periodic features of the process. Each Gegenbauer polynomial, $(1 - 2\eta_i L + L^2)^{d_i}$, has a pair of complex roots with modulus one and expands 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($p, 2d, q$) model (Granger and Joyeux 1980; Hosking 1981). Finally, we get an ARIMA model when $\eta = 1$ and $d = 0.5$, and an ARMA process when $d = 0$.

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

For stationary cases, the moving average representation is,

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

from which the spectral density function is obtained as

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

where $v_j = \cos^{-1}(\eta_j)$ are the Gegenbauer frequencies. The spectral density function is unbounded at v_j if $d_j > 0$ and vanishes there if $d_j < 0$. The autoregressive representation is most relevant for estimation of the CSS function considered here and is given as follows:

$$\frac{\phi(L)}{\theta(L)} \prod_{i=1}^k (1 - 2\eta_i L + L^2)^{d_i} (x_t - \mu) = \epsilon_t. \tag{4}$$

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

$$\gamma_j = 2 \int_0^\pi f(\omega) \cos(\omega j) d\omega, \tag{5}$$

where special attention must be given to the singularities in $f(\omega)$ as discussed by McElroy and Holan (2016). Convenient approximations for γ_j are only available for single frequency models. For example, when $\eta = 1$ and $d < 0.25$, the autocorrelations exhibit hyperbolic decay as demonstrated by Granger and Joyeux (1980) for fractional processes. For GARMA(0,0) models, Chung (1996a) shows that for large j , the autocorrelation function with $|\eta| < 1$ and $d < 0.5, d \neq 0$, can be approximated as $\rho_j \approx J \cos(jv)j^{2d-1}$, where the constant 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 Fig. 1, we illustrate a model that combines ARFIMA(0,0) and GARMA(0,0) models, which is of particular interest for economic and financial applications. This example used a model with parameters $(\eta_1, d_1) = (1, 0.15)$ and $(\eta_2, d_2) = (0.992, 0.25)$. Note that the first frequency corresponds to an unbounded spike at the origin of the spectrum. The second frequency corresponds to an unbounded spike at the frequency $v_2 = \cos^{-1}(0.992) = 0.1266$ radians, or

0.0201Hz, which is very close to the origin, with a cycle length of 50 periods. The ACF clearly demonstrates long cycles about the hyperbolic decay characteristic of fractional processes.

3 Estimation

As discussed above, several estimation procedures have been proposed for the k -factor model. In this section, we generalize the CSS estimator of Chung (1996a, 1996b) for single-factor GARMA models to models with $k > 1$.

3.1 The constrained sum of squares estimator

In this subsection, we define the CSS estimator we employ for the GARMA process and set preliminaries for the distribution theory proposed in the following subsection. In the case where a spectral pole exists at 0 or π , the CSS estimator of the k -factor GARMA model inherits the problems associated with time-domain estimation of μ for simple ARFIMA models as espoused by Cheung and Diebold (1994) and Chung (1996b). Therefore, in this section we impose that μ is known, leaving the issue of an unknown mean for future research.¹

To establish notation, let $\delta = (d_1, \dots, d_k)'$, $\tau = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$, and $\eta = (\eta_1, \dots, \eta_k)'$, where $\psi = (\delta', \tau', \eta)'$. We further have, $\delta \in \Psi_\delta$, $\tau \in \Psi_\tau$, and $\Psi_\eta = \prod_{i=1}^k [-1, 1]$, where Ψ_δ and Ψ_τ are compact subsets of \mathbb{R}^k and \mathbb{R}^{p+q} , respectively, and where $\Psi = \Psi_\delta \times \Psi_\tau \times \Psi_\eta$. The sum of squares function considered here is used to estimate the true, unknown values given by the associated vector denoted $\psi_0 = (\delta'_0, \tau'_0, \eta'_0)'$. If we assume that the initializing disturbances are zero, then the maximization of the CSS function is asymptotically equivalent to maximum likelihood estimation. The following additional assumptions are imposed for the distribution theory presented in the next subsection.

Assumption 1 $\{\varepsilon_t\}$ 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.

Assumption 2 δ_0 lies in the interior of the set $\prod_{i=1}^k [0, \bar{d}_i]$, where $\bar{d}_i = 0.25$ if $|\eta_{i,0}| = 1$, whereas $\bar{d}_i = 0.50$ if $|\eta_{i,0}| < 1$. Further, τ_0 is in the interior of Ψ_τ .

Assumption 3 The value of k is known, and $\eta_0 = (\eta_{1,0}, \eta_{2,0}, \dots, \eta_{k,0})'$ has no common elements, where $\eta_{i,0} \neq \eta_{j,0}, \forall i \neq j$.

¹ Below, we do consider estimation of μ using both the CSS estimator and sample mean and provide discussion where relevant.

The first assumption relaxes an unnecessarily strong normality condition, whereas, as illustrated below, estimation requires only the associated sum of squared errors. The second assumption is standard within the long-memory literature, specifically when developing consistency arguments (Robinson 2006), and the third condition is needed for identification. Below, we discuss methods that can be used to estimate the unknown value of k .

Under the assumptions above, we can use the AR representation from (4) to define the sum of squares function. Specifically, define $\alpha_j(\psi)$ as the j th coefficient in the expansion of $\frac{\phi(L)}{\theta(L)} \prod_{i=1}^k (1 - 2\eta_i L + L^2)^{d_i}$. We define the truncated disturbances and sum of squares function, $s_T(\psi)$, as,

$$s_T(\psi) = \sum_{t=1}^T \varepsilon_t(\psi)^2 \tag{6}$$

where

$$\varepsilon_t(\psi) = \sum_{j=0}^{t-1} \alpha_j(\psi) x_{t-j}. \tag{7}$$

Under all above assumptions, the set of CSS estimates, $\hat{\psi} = (\hat{\delta}', \hat{\tau}', \hat{\eta}')'$, is then defined as follows:

$$\hat{\psi} = \arg \min_{\psi \in \Psi} s_T(\psi). \tag{8}$$

Conditions for consistency of the CSS estimators have been established by Robinson (2006). The following two assumptions establish consistency under the additional assumptions above and defining $\alpha(L; \psi) = \sum_{j=0}^{\infty} \alpha_j(\psi) L^j$.

Assumption 4 For the true parameter vector ψ_0 , we have $\psi_0 \in \Psi$, and for all $\psi \in \Psi \setminus \psi_0$, $\alpha(L; \psi) \neq \alpha(L; \psi_0)$.

Assumption 5 $\sum_{j=0}^{\infty} \sup_{\psi \in \Psi} |\alpha_j(\psi)| < \infty$.

The fourth assumption is also an identification condition, while the last assumption requires absolute summability of the coefficients in the autoregressive representation for x_t . Under the assumptions above, absolute summability is established if $d_i > 0$ for all $i \in \{1, \dots, k\}$, as provided in the following lemma whose proof is given in Appendix.

Lemma 1 Under Assumptions 1–4, the coefficients in the $AR(\infty)$ representation of x_t in Eq. (4) are absolutely summable provided $d_i > 0$ for all $i \in \{1, \dots, k\}$.

With $v_{i,0} = \cos^{-1}(\eta_{i,0})$, the elements of $I_{\psi_{\theta_0, \tau_0}}$ are defined as follows:

$$I_{d_{i,0}} = 2 \left[\frac{\pi^2}{3} - \pi v_{i,0} + v_{i,0}^2 \right], \quad i = 1, \dots, k \tag{11a}$$

$$I_{d_{i,0}d_{j,0}} = 2 \left[\frac{\pi^2}{3} - \pi v_{i,0} + \frac{v_{i,0}^2 + v_{j,0}^2}{2} \right], \quad v_{i,0} > v_{j,0}, \tag{11b}$$

$$I_{d_{i,0}\phi_{j,0}} = 2 \sum_{l=0}^{\infty} \phi_{l,0}^* \frac{\cos[(l+j)v_{i,0}]}{(l+j)}, \quad i = 1, \dots, k, \quad j = 1, \dots, p \tag{11c}$$

$$I_{d_{i,0}\theta_{m,0}} = 2 \sum_{l=0}^{\infty} \theta_{l,0}^* \frac{\cos[(l+m)v_{i,0}]}{(l+m)}, \quad i = 1, \dots, k, \quad m = 1, \dots, q \tag{11d}$$

where $\phi_{l,0}^*$ and $\theta_{l,0}^*$ denote the l th coefficients in the infinite order expansions of $\phi_0^{-1}(L)$ and $\theta_0^{-1}(L)$, respectively. The submatrices I_{ϕ_0} , I_{ϕ_0, θ_0} and I_{θ_0} consist of elements that are the same as the corresponding submatrices of the usual information matrix of an ARMA model.

To calculate the information matrix in Theorem 2, the coefficients of $\phi_{l,0}^*$ and $\theta_{l,0}^*$ are easily calculated recursively using the method of equating coefficients. Equipped with these values, it is straightforward to calculate the information matrix to obtain standard errors used in inference. In the application below, given the large number of potential permutations, with different values of k , p , and q , we use a straightforward computation that truncates relevant infinite sums with 10 million terms.

Throughout, we abstract from the case where μ is unknown, although a result is available if $|\eta_{i,0}| < 1$, $i = 1, \dots, k$. With $|\eta_{i,0}| < 1$ for all i , the CSS estimator of the true mean μ_0 , denoted $\hat{\mu}$, has the following distribution, where $f(0)$ denotes the spectral density function evaluated at frequency $\omega = 0$,

$$\sqrt{T}(\hat{\mu} - \mu_0) \rightsquigarrow N(0, 2\pi f(0)). \tag{12}$$

The distributions of $\hat{\mu}$ and the sample mean, \bar{x} , are equivalent. The proof is omitted as these results are a simple extension of Theorem 1 in Chung (1996b).

Theorem 3 is the central result and proposes the asymptotic distribution of $\hat{\eta}$ for all of our four cases.

Theorem 3 (Asymptotic distribution of $\hat{\eta}$) *Let $\hat{\eta}_1, \dots, \hat{\eta}_k$ be the estimators of $\eta_{1,0}, \dots, \eta_{k,0}$, based on Eq. (8) for a stationary and invertible k -factor GARMA model for a sample $\{x_t\}$, $t = 1, \dots, T$. Without loss of generality, order the elements of η_0 from smallest to largest. Then let $D_{\eta_{1,0}}$ denote a dummy variable that takes on the value 1 if $\eta_{1,0} = -1$ and 0 otherwise, and let $D_{\eta_{k,0}}$ denote a dummy variable that takes on the value 1 if $\eta_{k,0} = 1$ and 0 otherwise. Under Assumptions 1–5,*

$$T(\hat{\eta}_i - \eta_{i,0}) \rightsquigarrow \frac{\sin(v_{i,0}) \left[\int_0^1 W_{2i-1-D_{\eta_{1,0}}} dW_{2i-D_{\eta_{1,0}}} - \int_0^1 W_{2i-D_{\eta_{1,0}}} dW_{2i-1-D_{\eta_{1,0}}} \right]}{d_{i,0} \left[\int_0^1 W_{2i-1-D_{\eta_{1,0}}}^2(r) dr + \int_0^1 W_{2i-D_{\eta_{1,0}}}^2(r) dr \right]} \tag{13}$$

with $|\eta_{i,0}| < 1$, where $i = 1 + D_{\eta_{1,0}}, \dots, k - D_{\eta_{k,0}}$ and,

$$T^2(\hat{\eta}_1 + 1) \rightsquigarrow -\frac{1}{2d_{1,0}} \frac{\int_0^1 \left[\int_0^r W_1(s) ds \right] dW_1(r)}{\int_0^1 \left[\int_0^r W_1(s) ds \right]^2 dr}, \quad \text{if } \eta_{1,0} = -1, \tag{14}$$

$$T^2(\hat{\eta}_k - 1) \rightsquigarrow \frac{1}{2d_{k,0}} \frac{\int_0^1 \left[\int_0^r W_{2k-1-D_{\eta_{1,0}}}(s) ds \right] dW_{2k-1-D_{\eta_{1,0}}}(r)}{\int_0^1 \left[\int_0^r W_{2k-1-D_{\eta_{1,0}}}(s) ds \right]^2 dr}, \quad \text{if } \eta_{k,0} = 1, \tag{15}$$

where $W_1, W_2, \dots, W_{2k-D_{\eta_{1,0}}-D_{\eta_{k,0}}}$, are $2k - D_{\eta_{1,0}} - D_{\eta_{k,0}}$ independent Brownian motions.

The proof is given in ‘‘Appendix’’. An important result of this theorem relates to the asymptotic independence of the values in the vector $\hat{\eta}$. In addition, for each $\hat{\eta}_i$, $d_{i,0}$ and $v_{i,0}$ 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).

3.3 Estimation algorithm

These theorems provide practical information for designing an efficient algorithm. We know that the asymptotic distributions of the memory parameters are not independent of the ARMA parameters. Also, the asymptotic distribution of $\hat{\psi}_{\delta,\tau}$ and $\hat{\eta}$ are independent, but the elements of $\hat{\psi}_{\delta,\tau}$ are $O_p(T^{-1/2})$, whereas $\hat{\eta}_i$ is $O_p(T^{-1})$ if $|\eta_{i,0}| < 1$ and $O_p(T^{-2})$ if $|\eta_{i,0}| = 1$. These results suggest that the algorithm of Woodward et al. (1998), which estimates ARMA parameters independently of (η_i, d_i) , will produce inconsistent estimates. It might be preferable to use an extension of Chung’s method (Chung 1996a, b) by conducting a grid search over each element of η combined with a gradient method for $\psi_{\delta,\tau}$. However, Monte Carlo simulations indicate that the grid over each value of η_i must be very fine, since the objective function has many local minima. A k -dimensional line search for η coupled with a gradient-based search for $\psi_{\delta,\tau}$ would be computationally infeasible, unless the parameter space is bounded in some way or a very coarse grid is used.

The computational complexity of the CSS estimator for a k -factor GARMA model can be better appreciated if we consider the step of recursively computing

the residuals. The inverse of the i th Gegenbauer polynomial in the k -factor GARMA model can be expanded as (Gray et al. 1989)

$$(1 - 2\eta_i z + z^2)^{d_i} = \sum_{j=0}^{\infty} C_j^{(-d_i)}(\eta_i) z^j, \tag{16a}$$

where

$$C_j^{(-d_i)}(\eta_i) = \sum_{l=0}^{[j/2]} \frac{(-1)^l (2\eta_i)^{j-2l} \Gamma(-d_i - l + j)}{l! (j - 2l)! \Gamma(-d_i)}, \tag{16b}$$

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

$$C_j^{(-d_i)}(\eta_i) = 2\eta_i \left(\frac{-d_i - 1}{j} + 1 \right) C_{j-1}^{(-d_i)}(\eta_i) - \left(2 \frac{-d_i - 1}{j} + 1 \right) C_{j-2}^{(-d_i)}(\eta_i), \tag{17}$$

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

$$\frac{\phi(L)}{\theta(L)} \prod_{i=1}^k \left[\sum_{j=0}^{t-1} C_j^{(-d_i)}(\eta_i) L^j \right] (x_t - \mu) = \varepsilon_t. \tag{18}$$

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

To overcome computational issues, coupled with different rates of convergence of various model parameters, we use an extension of the iterative multi-step gradient-based algorithm developed by Ramachandran and Beaumont (2001). First, for a given k , we obtain a grid of starting values for each element of η . We use each set of starting values in this grid to estimate $\psi_{\delta,\tau}$. Conditional on the estimated value, $\hat{\psi}_{\delta,\tau}$, we then estimate the elements of η using an unconstrained gradient-based search.² Using the updated estimates of η , a new estimate of $\psi_{\delta,\tau}$ is obtained, which is then used to update the estimate of η . This procedure continues for all combinations of starting values for η_i . The final model results from the set of parameters that jointly 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 .

Our theoretical results assume that the number of spectral poles, k , is known, although this may be unlikely in many applications. It is beyond the scope of this paper to settle how k should be determined for all applications. However, we provide here some guidance based on the existing literature and also propose an additional

² The search occurs over all theoretically plausible values of η_i , only imposing a constraint to ensure $\eta_i \neq \eta_j, i \neq j$. All elements of η_i are estimated jointly, unless it is suspected that there exists a value $|\eta_i| = 1$, in which case this parameter is estimated separately at each iteration.

method that shows promise. Within the literature, k is most commonly selected through ocular inspection of the periodogram of the data to locate the dominant frequencies (Yajima 1991; Hidalgo and Soulier 2004; Arteché 2020). Although there is some theoretical support for this approach, the number of candidate frequencies could be low if the spectrum is dominated by behavior at the origin (Leschinski and Sibbertsen 2019). Hidalgo and Soulier (2004) introduce a procedure to determine the model order that sequentially identifies the largest periodogram frequency and then tests the significance of the persistence parameter at that frequency. If the parameter is found to be insignificant, the iterative procedure ends. Otherwise, the significant (η_i, d_i) -pair is added to the Gegenbauer filter, some neighborhood around that pole is excluded, and the sequential search continues. Leschinski and Sibbertsen (2019) propose a related iterative procedure that tests for significant poles in the spectrum after sequentially applying a Gegenbauer filter based on estimated memory parameters obtained using a Whittle method. The procedure terminates when the test-statistic for a singularity is insignificant.

We propose a relatively simple method that selects k based on the minimum value of the Bayesian Information Criterion (BIC) for integer values of $k \leq \bar{K}$, where \bar{K} is some sufficiently large upper bound.³ To test this procedure, we simulated various k -factor GARMA models, and present here results for a potentially interesting case with parameters $(\eta_{i,0}, d_{i,0})_{i=1}^2 = (0.5, 0.2)$ and $(-0.5, 0.4)$. The model also includes an AR(1) term with $\phi_0 = 0.8$. We simulate 1000 replications of the true model with sample sizes of $T = 100, 200, 300, 500, 1000,$ and 2000 . We select \bar{K} to be 4, which is large enough to explore the sensitivity of our results without placing undue burden on computational resources. In addition to recording the selected value of k based on the BIC, we also consider model selection based on the Hannan–Quinn (HQ) and Akaike (AIC) information criteria.

The results of these simulations are reported in Table 1, where the true value of k is 2. In the top panel, we report the proportion of times the AIC selects different values that range from 1 to 4. The correct value of k is selected a majority of times for all sample sizes. The success rate for choosing $k = 2$ increases slowly in sample sizes beginning with 52% for $T = 100$, and reaches nearly 70% when $T = 2000$. The second and third panels show the AIC only performs comparatively well when $T = 100$, whereas it is strictly dominated by the BIC and HQ criteria for larger samples. Specifically, the BIC and HQ criteria are extremely accurate when $T \geq 500$, with the BIC outperforming the HQ. The final panel of Table 1 shows the bias of the estimated parameters when k is set to 4. For large sample sizes, the bias induced by selecting $k > 2$ is quite small.

Additional simulations (available upon request) show that the consequences of choosing k too large are relatively minor unless T is small. We also observe that the estimation errors, particularly the RMSEs, associated with overestimating k are greater for the ARMA parameters than for η_i and d_i . Consequently, the more important the short-term dynamics are, the more critical it is to accurately estimate k . In

³ For our application below, we propose that \bar{K} be selected based on visual inspection of the periodogram of the differenced data to mitigate the impact of non-cyclic long memory.

Table 1 Proportion of times k is chosen by both sample size T and information criterion for a 2-factor GARMA(1,0) model

k/T	100	200	300	500	1000	2000
<i>AIC</i>						
1	0.1380	0.0220	0.0050	0.0000	0.0000	0.0000
2	0.5210	0.5570	0.5570	0.5600	0.6370	0.6980
3	0.2020	0.2690	0.2550	0.2460	0.2120	0.1780
4	0.1390	0.1520	0.1830	0.1940	0.1510	0.1240
<i>BIC</i>						
1	0.5740	0.2640	0.1250	0.0080	0.0000	0.0000
2	0.4120	0.7170	0.8660	0.9870	0.9990	1.0000
3	0.0120	0.0180	0.0090	0.0050	0.0010	0.0000
4	0.0020	0.0010	0.0000	0.0000	0.0000	0.0000
<i>HQ</i>						
1	0.3240	0.1090	0.0270	0.0020	0.0000	0.0000
2	0.5600	0.7950	0.8680	0.9140	0.9510	0.9620
3	0.0780	0.0810	0.0840	0.0690	0.0450	0.0350
4	0.0380	0.0150	0.0210	0.0150	0.0040	0.0030
<i>Bias of parameters when $k = 4$</i>						
ϕ_1	-0.2257	-0.1116	-0.0787	-0.0503	-0.0117	-0.0042
d_1	0.0062	0.0281	0.0285	0.0181	0.0026	0.0001
d_2	0.0002	0.0129	0.0056	0.0006	-0.0010	-0.0007
d_3	-0.1112	-0.0438	-0.0259	-0.0140	-0.0055	-0.0017
d_4	-0.0845	-0.0017	0.0082	0.0061	0.0020	0.0010
η_1	-0.0317	-0.0144	-0.0086	-0.0017	-0.0006	0.0006
η_2	-0.0668	-0.0096	0.0000	0.0002	0.0001	0.0001

The true model is a 2-factor GARMA(1,0) model with $\eta_{1,0} = 0.5$, $d_{1,0} = 0.2$, $\eta_{2,0} = -0.5$, $d_{2,0} = 0.4$, and $\phi_0 = 0.8$

many applications, until more definitive theoretical results for estimating k can be derived, we recommend that researchers use several methods to choose k and check the robustness of their estimation results.

4 Finite sample performance

In this section, we report simulation results that examine the finite sample properties of the CSS estimator. We are interested in examining the bias in the parameter estimates and in comparing the finite sample standard errors of the estimates with their asymptotic counterparts. Ramachandran and Beaumont (2001) and Beaumont and Smallwood (2022) have done extensive simulations for the single-factor GARMA model, with the latter paying particular attention to the parametric region where η is close to one and d is close to one-half. Based on those results, we use sample sizes of 500, 1000, and 2000 and concentrate on two-factor models with parameter ranges that we believe are most relevant for economic and financial applications.

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

	η_1	η_2	d_1	d_2	μ	N
True	0.5	0	0.2	0.4	0	
Mean Bias	-0.0003	-0.0001	-0.0031	0.0068	0.0002	500
RMSE	0.0346	0.0082	0.0311	0.0397	0.0343	
MNSE	0.0078	0.0047	0.0309	0.0357	N/A	
Mean Bias	-0.00002	-0.00001	-0.0009	0.0018	-0.0004	2000
RMSE	0.0081	0.0022	0.0149	0.0182	0.0173	
MNSE	0.0020	0.0012	0.0153	0.0177	N/A	
True	0.5	-0.5	0.2	0.4	0	
Mean Bias	-0.0009	0.0001	0.0011	0.0061	0.0000	500
RMSE	0.0364	0.0080	0.0345	0.0380	0.0295	
MNSE	0.0079	0.0041	0.0352	0.0352	N/A	
Mean Bias	-0.0004	-0.0001	-0.0001	0.0014	0.0003	2000
RMSE	0.0082	0.0020	0.0177	0.0177	0.0145	
MNSE	0.0020	0.0011	0.0175	0.0175	N/A	
True	0	0.5	0.2	0.4	0	
Mean Bias	0.0015	-0.0001	-0.0030	0.0039	0.0015	500
RMSE	0.0435	0.0078	0.0344	0.0347	0.0408	
MNSE	0.0091	0.0040	0.0356	0.0309	N/A	
Mean Bias	0.0010	-0.0001	-0.0014	0.0004	0.0010	2000
RMSE	0.0079	0.0022	0.0172	0.0156	0.0196	
MNSE	0.0022	0.0011	0.0177	0.0153	N/A	
True	0	-0.5	0.2	0.4	0	
Mean Bias	0.0016	-0.0001	-0.0031	0.0048	0.0010	500
RMSE	0.0455	0.0079	0.0346	0.0346	0.0262	
MNSE	0.0091	0.0040	0.0356	0.0310	N/A	
Mean Bias	-0.0003	0.00001	-0.0014	0.0002	0.0007	2000
RMSE	0.0083	0.0019	0.0172	0.0162	0.0126	
MNSE	0.0022	0.0011	0.0177	0.0153	N/A	
True	-0.5	0.5	0.2	0.4	0	
Mean Bias	-0.0012	-0.0002	0.0015	0.0063	0.0006	500
RMSE	0.0373	0.0073	0.0335	0.0365	0.0360	
MNSE	0.0078	0.0040	0.0352	0.0351	N/A	
Mean Bias	0.0001	-0.00004	-0.0001	0.0004	0.0001	2000
RMSE	0.0084	0.0022	0.0176	0.0178	0.0188	
MNSE	0.0020	0.0011	0.0175	0.0175	N/A	
True	-0.5	0	0.2	0.4	0	
Mean Bias	-0.0020	0.00004	-0.0023	0.0060	-0.00002	500
RMSE	0.0371	0.0093	0.0321	0.0390	0.0279	
MNSE	0.0079	0.0047	0.0309	0.0357	N/A	

Table 2 (continued)

	η_1	η_2	d_1	d_2	μ	N
Mean Bias	0.0008	- 0.00002	- 0.0010	0.0008	0.0003	2000
RMSE	0.0074	0.0025	0.0155	0.0189	0.0138	
MNSE	0.0020	0.0012	0.0153	0.0177	N/A	

The means of the true asymptotic standard errors (MASE) for \hat{d}_1 with $T = 500$ and 2000 are $\{0.0306, 0.0349, 0.0353, 0.0353, 0.0348, 0.0305\}$ and $\{0.0153, 0.0174, 0.0176, 0.0176, 0.0174, 0.0153\}$. The corresponding values for \hat{d}_2 are $\{0.0353, 0.0349, 0.0306, 0.0306, 0.0349, 0.0353\}$ and $\{0.0176, 0.0174, 0.0153, 0.0153, 0.0174, 0.0174\}$, respectively

The initial simulation results are presented in Tables 2, 3, 4 and 5.⁴ Each column lists the parameters of the simulated model and each block in the tables gives the results from 1000 replications for each specific parameterization. For computational purposes, we use an iterative procedure to generate a large amount of data before discarding all but the last 500 or 2000 observations. Throughout, we report the true parameter values (True), the mean bias, the root mean squared error (RMSE), the mean of the numerical standard errors calculated from the estimated Hessian matrix in the last iteration (MNSE), and the mean of the true asymptotic standard errors (MASE) based on Theorem 2. We use the *estimated* values of all parameters to compute the true asymptotic standard errors for each of the 1000 replications and then average them to get the MASE. Since the mean bias is small, there will be inconsequential differences between the MASE computed this way and the true ASE computed using the *true* parameter values. Additionally, our MASE values will vary by sample size, since the standard errors are not multiplied by $T^{-1/2}$.

Table 2 presents the results for six different two frequency GARMA(0,0) models with values of $\eta_{i,0}$ set to $-\frac{1}{2}, 0, \frac{1}{2}$ and values of $d_{i,0}$ equal to 0.2 and 0.4. The estimation biases are all quite small, especially for the values of $\hat{\eta}_1$ and $\hat{\eta}_2$, which converge at a faster rate than the other parameters. Theoretically, $\hat{\eta}_i$ is $O_p(T^{-1})$ whereas remaining parameters have standard rate- \sqrt{T} convergence. It is therefore wholly consistent with the theoretical results to observe that the MNSE's of $\hat{\eta}_i$ in Table 2 are about 4 times larger for $T= 500$ relative to $T=2000$. In contrast, the MNSE's for \hat{d}_i are about 2 times larger for samples of 500 relative to samples of 2000.

Generally speaking, a larger value of $d_{i,0}$ mitigates the already small bias in $\hat{\eta}_i$, which appears to be marginally more sensitive to estimation outliers. This is likely due to the fact that an estimate of d_i near zero can lead to poor estimates of the corresponding η_i , because that Gegenbauer polynomial will have very little impact on the objective function no matter what the value of $\eta_{i,0}$ is. In all cases, μ is estimated with the sample mean, which is asymptotically equivalent to the CSS estimator of μ provided $|\eta_{i,0}| < 1, i = 1, \dots, k$. As noted above, the estimator

⁴ To conserve space, results with $T = 1000$ are omitted, but are available on request.

Table 3 Simulation results for 2-factor GARMA processes with $p, q \neq 0$

	η_1	η_2	d_1	d_2	ϕ	θ	μ	T
True	0.5	-0.5	0.2	0.4	0.8	N/A	0	
Mean Bias	-0.0005	0.0005	-0.0022	0.0026	-0.0075	-	0.0068	500
RMSE	0.0372	0.0079	0.0356	0.0432	0.0330	-	0.1441	
MNSE	0.0081	0.0041	0.0358	0.0399	0.0313	-	-	
MASE	-	-	0.0354	0.0397	0.0311	-	0.1414	
Mean Bias	-0.0004	0.000001	-0.0012	-0.0001	-0.0020	-	-0.0008	2000
RMSE	0.0084	0.0020	0.0175	0.0205	0.0162	-	0.0717	
MNSE	0.0020	0.0011	0.0178	0.0198	0.0154	-	-	
MASE	-	-	0.0177	0.0198	0.0154	-	0.0718	
True	0.5	-0.5	0.2	0.4	N/A	0.8	0	
Mean Bias	-0.0034	0.0001	0.0008	0.0071	-	-0.0097	-0.0010	500
RMSE	0.0346	0.0073	0.0391	0.0386	-	0.0350	0.0504	
MNSE	0.0078	0.0040	0.0400	0.0356	-	0.0314	-	
MASE	-	-	0.0397	0.0354	-	0.0313	0.0513	
Mean Bias	-0.0003	-0.000001	-0.0004	0.0022	-	-0.0027	-0.0002	2000
RMSE	0.0074	0.0021	0.0192	0.0180	-	0.0162	0.0260	
MNSE	0.0019	0.0011	0.0199	0.0178	-	0.0154	-	
MASE	-	-	0.0198	0.0177	-	0.0154	0.0259	
True	0.5	-0.5	0.2	0.4	-0.8	N/A	0	
Mean Bias	-0.0003	0.0002	-0.0028	0.0024	0.0047	-	-0.00003	500
RMSE	0.0341	0.0079	0.0397	0.0376	0.0329	-	0.0160	
MNSE	0.0079	0.0041	0.0399	0.0357	0.0310	-	-	
MASE	-	-	0.0396	0.0354	0.0309	-	0.0161	
Mean Bias	-0.0004	-0.000002	-0.0018	-0.0002	0.0011	-	-0.00003	2000
RMSE	0.0070	0.0019	0.0205	0.0180	0.0162	-	0.0082	
MNSE	0.0019	0.0011	0.0198	0.0178	0.0153	-	-	
MASE	-	-	0.0198	0.0177	0.0153	-	0.0080	
True	0.5	-0.5	0.2	0.4	N/A	-0.8	0	
Mean Bias	0.0008	-0.0001	-0.0014	0.0040	-	0.0111	0.0003	500
RMSE	0.0394	0.0078	0.0356	0.0440	-	0.0406	0.0066	
MNSE	0.0079	0.0040	0.0357	0.0400	-	0.0316	-	
MASE	-	-	0.0354	0.0397	-	0.0314	0.0061	
Mean Bias	-0.0001	0.000003	-0.0019	-0.0003	-	0.0022	0.0002	2000
RMSE	0.0075	0.0021	0.0185	0.0203	-	0.0165	0.0029	
MNSE	0.0020	0.0011	0.0178	0.0198	-	0.0154	-	
MASE	-	-	0.0177	0.0198	-	0.0154	0.0029	
True	0.5	-0.5	0.2	0.4	0.8	0.8	0	
Mean Bias	0.0012	-0.0001	0.0021	0.0024	-0.0049	-0.0276	0.0016	500
RMSE	0.0448	0.0067	0.0420	0.0437	0.0402	0.0565	0.2517	
MNSE	0.0078	0.0040	0.0428	0.0420	0.0334	0.0351	-	
MASE	-	-	0.0424	0.0418	0.0331	0.0349	0.2585	
Mean Bias	-0.0003	0.0001	-0.0012	-0.0003	-0.0013	-0.0063	0.0038	2000

Table 3 (continued)

	η_1	η_2	d_1	d_2	ϕ	θ	μ	T
RMSE	0.0070	0.0018	0.0206	0.0204	0.0166	0.0197	0.1314	
MNSE	0.0020	0.0011	0.0210	0.0209	0.0163	0.0165	–	
MASE	–	–	0.0209	0.0209	0.0163	0.0165	0.1296	

for the mean is $O_p(T^{-1/2})$, the same rate of convergence as the parameters in $\hat{\psi}_{\delta,\tau}$, so its bias is also quite small. The true asymptotic standard errors of the corresponding values of \hat{d}_1 and \hat{d}_2 are quite comparable to their numerical counterparts. Finally, in light of the results of Theorem 3, it is not surprising to see that the MNSE and RMSE for $\hat{\eta}_1$ and $\hat{\eta}_2$ are quite different, since the RMSE assumes convergence at the rate $T^{1/2}$.

To examine the influence of ARMA parameters, ϕ and θ , we choose a particular parameterization (second case from Table 2) and estimate various two-factor GARMA(p, q) models with p and q being either zero or one. The results are reported in Table 3 and are similar to those in Table 2. Again, for all of the cases considered in Table 3, the median and mean biases are quite small. Again, we see that the mean asymptotic standard errors are virtually identical to the RMSE and MNSE for $\hat{d}_i, \hat{\phi}$, and $\hat{\theta}$, particularly with $T = 2000$. These simulation results yield particularly strong evidence supporting the theoretical results in Theorem 2.

Table 4 examines the particularly interesting case where $\eta_{1,0}=1$ and $|\eta_{2,0}| < 1$, so that we get a combination ARFIMA and GARMA model. Compared to $\hat{\eta}_2, \hat{\eta}_1$ has very little bias and extremely small RMSE and MNSE, reflecting that this parameter may be $O_p(T^{-2})$ as reported in the theoretical results above. As expected, the MNSE for $\hat{\eta}_1$ is about $(2000/500)^2 = 16$ times larger when the sample size is 500 compared to when the sample size is 2000. The results for $\hat{\eta}_2$ when $|\eta_{2,0}| < 1$ are similar to those in Tables 2 and 3, as are the results for the \hat{d}_1 and \hat{d}_2 . When $\eta_{i,0} = 1$, however, the sample mean and CSS estimate of μ are no longer asymptotically equivalent. Thus, we use the CSS estimator for μ in these cases. The computational difficulties of time domain estimators for ARFIMA models when the mean is unknown have been well documented (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.

As noted above, the computational burden of the CSS estimator grows rapidly with the number of spectral poles due to the grid search over each η_i . Thus, if we could narrow the range of the grid search, we could improve the efficiency of the algorithm. With $i \neq j$, since $\hat{\eta}_i$ is independent of both $\hat{\eta}_j$ and $\hat{\psi}_{\delta,\tau}$, 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 5, we investigate this possibility. First, we estimate a 1-factor GARMA model and then filter the data with the resulting Gegenbauer polynomial before estimating the second frequency using a 1-factor model on these filtered data. This process should produce good starting values for η as long as the biases are not too large.

Table 4 Estimation of simulated ARFIMA/GARMA processes

	η_1	η_2	d_1	d_2	μ	T
True	1	0.75	0.2	0.3	0	
Mean Bias	-0.0014	-0.0005	-0.0096	0.0055	-0.0258	500
RMSE	0.0046	0.0120	0.0262	0.0331	0.8135	
MNSE	0.0003	0.0042	0.0221	0.0316	0.2512	
MASE	-	-	0.0218	0.0312	-	
Mean Bias	-0.0001	0.0001	-0.0034	0.0024	0.0011	2000
RMSE	0.0003	0.0027	0.0116	0.0157	0.6425	
MNSE	0.00002	0.0011	0.0107	0.0156	0.2338	
MASE	-	-	0.0107	0.0155	-	
True	1	0.5	0.2	0.3	0	
Mean Bias	-0.0014	0.0013	-0.0081	-0.0027	-0.0081	500
RMSE	0.0047	0.0149	0.0218	0.0325	0.6413	
MNSE	0.0004	0.0054	0.0188	0.0319	0.2015	
MASE	-	-	0.0186	0.0316	-	
Mean Bias	-0.0001	0.0003	-0.0026	0.0009	0.0016	500
RMSE	0.0003	0.0035	0.0098	0.0155	0.4913	
MNSE	0.00002	0.0014	0.0092	0.0158	0.1871	
MASE	-	-	0.0091	0.0158	-	
True	1	0	0.2	0.3	0	
Mean Bias	-0.0014	0.0006	-0.0068	-0.0015	-0.0018	500
RMSE	0.0040	0.0154	0.0209	0.0372	0.5151	
MNSE	0.0003	0.0063	0.0186	0.0363	0.1636	
MASE	-	-	0.0183	0.0360	-	
Mean Bias	-0.0001	0.0001	-0.0022	-0.0013	-0.0046	2000
RMSE	0.0002	0.0042	0.0095	0.0190	0.4076	
MNSE	0.00002	0.0016	0.0091	0.0181	0.1523	
MASE	-	-	0.0090	0.0180	-	
True	1	-0.5	0.2	0.3	0	
Mean Bias	-0.0012	-0.0003	-0.0089	-0.0051	-0.0035	500
RMSE	0.0037	0.0125	0.0262	0.0395	0.4530	
MNSE	0.0003	0.0055	0.0222	0.0375	0.1457	
MASE	-	-	0.0219	0.0373	-	
Mean Bias	-0.0001	0.0001	-0.0031	-0.0036	-0.0056	2000
RMSE	0.0002	0.0031	0.0115	0.0195	0.3595	
MNSE	0.00002	0.0014	0.0108	0.0186	0.1379	
MASE	-	-	0.0107	0.0185	-	

The first two models in Table 5 are cases from the previous simulations, and the third case represents a mixed ARFIMA/GARMA model in which the ARFIMA component is short memory ($d_{i,0} < 0$). The latter process, which is not covered by the theorems above, may result from differencing processes with a non-stationary

Table 5 Sequential estimation of 2-factor models with single-factor models

	η_1	η_2	d_1	d_2	ϕ	θ	μ	T
True	0.5	0.0	0.2	0.4	N/A	N/A	0	
Mean Bias	0.0018	0.0132	-0.0193	0.0487	-	-	0.0033	500
RMSE	0.0467	0.0704	0.0420	0.0795	-	-	0.0336	
MNSE	0.0084	0.0064	0.0302	0.0366	-	-	N/A	
True	0.5	0.0	0.2	0.4	N/A	N/A	0	
Mean Bias	0.0006	0.0053	-0.0095	0.0362	-	-	0.0010	2000
RMSE	0.0059	0.0121	0.0182	0.0417	-	-	0.0168	
MNSE	0.0020	0.0017	0.0152	0.0182	-	-	N/A	
True	0.5	-0.5	0.2	0.4	0.8	0.8	0	
Mean Bias	-0.0089	-0.0002	-0.0945	-0.1277	-0.0742	0.0525	-0.0038	500
RMSE	0.0876	0.0216	0.0979	0.1362	0.0838	0.0614	0.2580	
MNSE	0.0115	0.0042	0.0288	0.0325	0.0366	0.0240	N/A	
True	0.5	-0.5	0.2	0.4	0.8	0.8	0	
Mean Bias	-0.0006	0.0002	-0.0941	-0.1269	-0.0670	0.0658	-0.0031	2000
RMSE	0.0053	0.0021	0.0947	0.1284	0.0698	0.0671	0.1350	
MNSE	0.0026	0.0009	0.0143	0.0160	0.0179	0.0114	N/A	
True	1.0	0.75	-0.2	0.3	N/A	N/A	0	
Mean Bias	-0.0002	-0.0065	0.0608	-0.1277	-	-	0.0006	500
RMSE	0.0079	0.0287	0.0635	0.1306	-	-	0.0080	
MNSE	0.0001	0.0047	0.0186	0.0231	-	-	N/A	
True	1.0	0.75	-0.2	0.3	N/A	N/A	0	
Mean Bias	0.000001	-0.0008	0.0601	-0.1286	-	-	0.0002	2000
RMSE	0.00001	0.0035	0.0608	0.01294	-	-	0.0023	
MNSE	0.00001	0.0010	0.0093	0.0114	-	-	N/A	

ARFIMA component. For each of the cases considered in Table 5, the sample mean is used to estimate μ . We find that the method generally first selects the frequency with the largest corresponding value of $d_{i,0}$, thus capturing the most dominant feature of the ACF. The results in Table 5 indicate that the small sample biases in $\hat{\eta}_1$ and $\hat{\eta}_2$ are reasonable, suggesting that the method of choosing a tight grid around these point estimates might work, at least when k is small. The relatively large biases in the values of the vector $\hat{\psi}_{\delta,\tau}$, 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 multi-step gradient estimation algorithm, while largely validating the proposed distribution theory. Notably, the distribution of $\hat{\eta}_i$ appears independent of $\hat{\eta}_j$, $i \neq j$, and the distribution of these parameters is largely unaffected by the inclusion of ARMA dynamics. Additionally, the proposed distribution theory for \hat{d}_i is confirmed. Finally, as shown below, and in numerous other simulations that are available upon request, the estimator appears to achieve the proposed rates of convergence, even when we estimate multiple GARMA components.

For the single frequency case, Chung (1996a) uses a line grid search to estimate η , along with a gradient-based method for $\psi_{\delta,\tau}$. 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 near the true value. Based on the limited algorithm, Chung (1996a) provides support for the proposed theory and associated confidence bands for $\hat{\eta}$ for all cases except when $\eta_0 = 1$. Here, it would appear that the associated empirical test sizes for $\eta_0 = 1$ under the null are too large to be of practical use. Beaumont and Smallwood (2022) consider the consequences of using a two-dimensional grid search over both η and d without the use of boundary constraints for η , and show that the exact distributional results of Chung (1996a) are generally supported, with two exceptions. First, similar to Chung (1996a), Beaumont and Smallwood (2022) show that the theory under the hypothesis $\eta_0 = 1$ is problematic for testing purposes, with empirical sizes that are often much higher than their associated theoretical counterparts. Secondly, when $|\eta_{i,0}| < 1$, 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 $d_0 \rightarrow 0$. 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 experiments, including models where there exists a value of $\eta_{i,0} = 1$. For brevity, the full set of results are not reported here, but are available upon request. Here, we report results for four fairly complicated 2-factor parameterizations. Model 1 is a GARMA(0,0) model with $\{\eta_{1,0}, d_{1,0}\} = \{0.5, 0.4\}$, and $\{\eta_{2,0}, d_{2,0}\} = \{0, 0.2\}$. Given the distributional results above, this parameterization represents a case where the process is expected to be especially volatile.⁵ Model 2 is also a GARMA(0,0) model but with $\{\eta_{1,0}, d_{1,0}\} = \{0.98, 0.45\}$, and $\{\eta_{2,0}, d_{2,0}\} = \{-0.4, 0.3\}$. This parameterization approaches the region of the discontinuity in our theoretical distribution for $\hat{\eta}$ and is also a strongly persistent process with $d_{1,0}$ close to 0.50. Model 3 is the same as Model 1, except we add an AR(1) term with $\phi_0 = 0.80$. Finally, we consider a case with $\{\eta_{1,0}, d_{1,0}\} = \{1.00, 0.20\}$, and $\{\eta_{2,0}, d_{2,0}\} = \{-0.6, 0.45\}$. The theoretical results suggest that the estimates of η_1 and η_2 have different rates of convergence, and given the values of $d_{1,0}$ and $d_{2,0}$, the process is again close to the non-stationary border. This parameterization will allow us to explore how theoretical concerns regarding the CSS estimator when $|\eta_{i,0}| = 1$ impact results for $|\eta_{j,0}| < 1, i \neq j$.

First, we compare the theoretical and simulated distributions of $\hat{\eta}_i, i = 1, 2$. Figure 2 shows the empirical and theoretical normalized cumulative distribution functions (cdf) for $\hat{\eta}_1$ from Model 1 for sample sizes of 500 and 2000. For the empirical distributions, we plot $T(\hat{\eta}_1 - 0.50)$, where the elements of $\hat{\eta}$ are computed using the

⁵ Note that the scaling factor in Eq. (13) of Theorem 3 is $\frac{\sin(v_{i,0})}{d_{i,0}}$. As such, with $\sin(v_{2,0}) = 1$, the small value for $d_{2,0}$ is expected to generate estimated values of η_2 that are quite volatile.

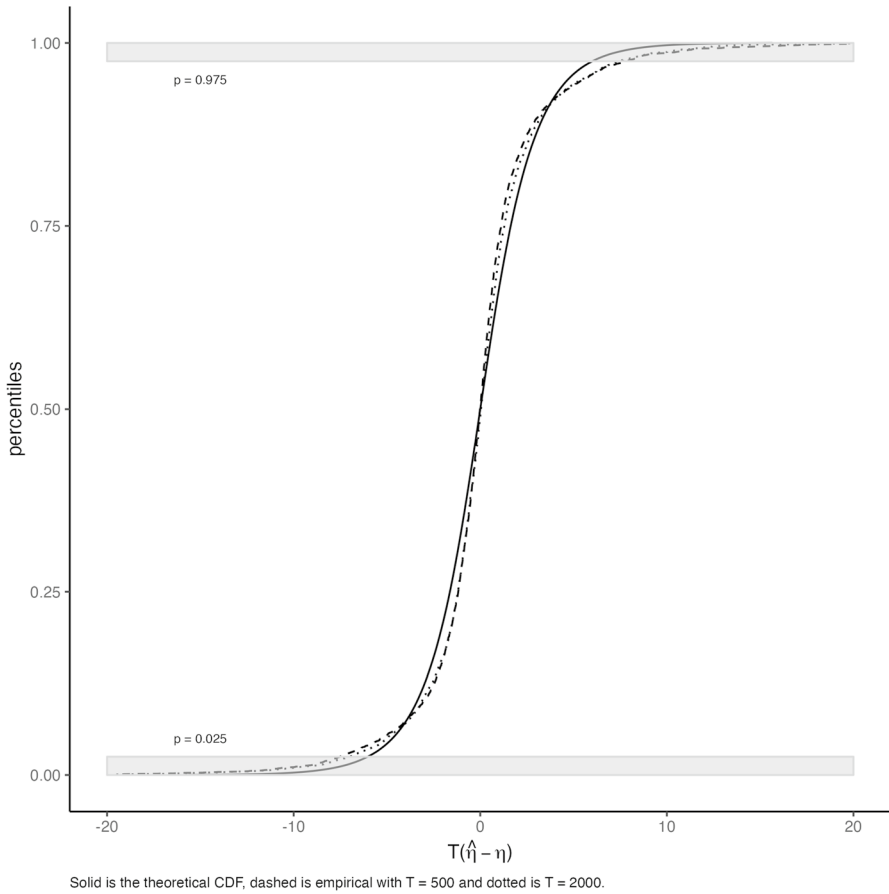


Fig. 2 Percentiles of theoretical/empirical CDFs of $\hat{\eta}_1$ in model 1 with $\{\eta_{1,0}, d_{1,0}\} = \{0.5, 0.4\}$

estimation algorithm described above, and the theoretical quantities have been calculated using Eq. (13) from Theorem 3. The vertical differences between the theoretical and empirical curves show the disagreements between the theoretically and empirically derived critical values for each percentile. The two 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 suggests that the proposed rate- T convergence 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 k -factor GARMA model, the calculation of confidence bands for $\hat{\eta}_i$ is likely the most important application of the theory. To get a sense of how applicable our proposed distribution theory and algorithm are, Table 6 provides the

estimated biases in calculating the upper and lower 68, 90, 95, and 99% confidence bands for the four models described above. As a reference point, the theoretical bands for each value of $\hat{\eta}_i$ with $T=500$ are provided in italic font. Below the theoretical bands, we show the bias associated with the empirical bands for $\hat{\eta}_1$, followed by those of $\hat{\eta}_2$, for each sample size.

For Model 1, and with relatively small samples of 500 observations, the 99% confidence bands are quite unreliable for $\eta_{2,0} = 0$. From the last two columns in the second block of Model 1 in Table 6, the theoretical confidence band for $\hat{\eta}_2$ with $\eta_{2,0} = 0$ when $T = 500$ is $[-0.0423, 0.0423]$. In contrast, among the 5050 simulations, 99% of the estimated values of $\hat{\eta}_2$ were within a range of $[-0.0910, 0.0683]$, thus producing a bias of the lower 99% band of -0.0487 (e.g., $-0.0910 + 0.0423$). 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 6 support the use of the proposed distribution theory in calculating these intervals. First, we note that the differences between the theoretical and estimated bands decrease sharply as T 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, 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 $\hat{\eta}_i$ are independent of ARMA components as implied by the proposed theory.

The simulations for the case with $\eta_{1,0} = 1$ merit additional discussion. First, we see that any potential concerns regarding estimation of η_1 likely do not impact estimation of η_2 . For example, with $\eta_{2,0} = -0.60$ and with $T = 500$, 99.5% of all values of $\hat{\eta}_2$ were less than -0.5793 , which is quite close to the theoretical upper 99% confidence band given by -0.5850 . Similar to other experiments, the biases in estimating theoretical percentiles decline with the sample size and become negligible for $T = 2000$. For estimates of η_1 , we see that the biases in calculating confidence bands are negligible, likely reflecting the proposed rate of convergence given by T^2 . Nonetheless, it is important to note that the T^2 factor also affects the test-statistic for the hypothesis $\eta_{1,0} = 1$. More specifically, using the distribution theory outlined in Theorem 3, we obtained the empirical sizes for the null hypothesis $H_0 : \eta_{1,0} = 1$ vs. the alternative $\eta_{1,0} < 1$ based on the test statistic $T^2(\hat{\eta}_1 - 1)$. The results show that substantial size distortion results. More specifically, the empirical sizes for $T = 500, 1000$, and 2000 observations were equal to 16.89, 18.48, and 18.75%, respectively, based on a 5% test size. This result matches the findings in Beaumont and Smallwood (2022), who show that the distribution theory under the null $\eta_{1,0} = 1$ can be unreliable. Computational methods likely offer resolution for researchers interested in determining if cycles are potentially infinite. In the next section, we briefly outline how to extend Beaumont and Smallwood (2022) to implement a simple parametric bootstrap in order to conduct tests of the hypothesis $|\eta_{i,0}| = 1$ in the multi-factor GARMA model.

Table 6 Empirical and theoretical confidence intervals of the η 's

OBS	68%L	68%U	90%L	90%U	95%L	95%U	99%L	99%U
<i>Model 1: $\eta_{1,0} = 0.50, d_{1,0} = 0.40; \eta_{2,0} = 0.00, d_{2,0} = 0.20$</i>								
η_1	<i>.4950</i>	<i>.5050</i>	<i>.4906</i>	<i>.5094</i>	<i>.4879</i>	<i>.5120</i>	<i>.4817</i>	<i>.5183</i>
500	0.0011	-0.0011	-0.0013	0.0017	-0.0030	0.0037	-0.0069	0.0110
1000	0.0004	-0.0007	-0.0007	0.0003	-0.0017	0.0013	-0.0054	0.0030
2000	0.0003	-0.0002	-0.0001	0.0003	-0.0005	0.0008	-0.0017	0.0016
η_2	<i>-0.0116</i>	<i>0.0116</i>	<i>-0.0217</i>	<i>0.0217</i>	<i>-0.0279</i>	<i>0.0279</i>	<i>-0.0423</i>	<i>0.0423</i>
500	-0.0044	0.0030	-0.0157	0.0111	-0.0224	0.0150	-0.0487	0.0260
1000	-0.0020	0.0015	-0.0069	0.0053	-0.0113	0.0077	-0.0203	0.0128
2000	-0.0010	0.0009	-0.0034	0.0030	-0.0049	0.0045	-0.0103	0.0076
<i>Model 2: $\eta_{1,0} = 0.98, d_{1,0} = 0.45; \eta_{2,0} = -0.40, d_{2,0} = 0.30$</i>								
η_1	<i>0.9790</i>	<i>0.9810</i>	<i>0.9781</i>	<i>0.9819</i>	<i>0.9775</i>	<i>0.9825</i>	<i>0.9763</i>	<i>0.9837</i>
500	0.0003	-0.0004	-1.7e-5	-0.0003	-0.0003	-0.0001	-0.0011	0.0004
1000	0.0001	-0.0001	-1.1e-6	-4.2e-5	-0.0001	0.0001	-0.0007	0.0003
2000	4.2e-5	-4.5e-5	-0.0001	8.6e-6	-0.0001	0.0001	-0.0005	0.0002
η_2	<i>-0.4071</i>	<i>-0.3929</i>	<i>-0.4133</i>	<i>-0.3867</i>	<i>-0.4170</i>	<i>-0.3830</i>	<i>-0.4258</i>	<i>-0.3742</i>
500	0.0004	0.0002	-0.0039	0.0059	-0.0072	0.0092	-0.0165	0.0175
1000	-0.0001	0.0002	-0.0025	0.0024	-0.0047	0.0042	-0.0079	0.0086
2000	0.0001	-4.5e-5	-0.0008	0.0010	-0.0019	0.0018	0.0048	0.0042
<i>Model 3: $\eta_{1,0} = 0.98, d_{1,0} = 0.45; \eta_{2,0} = -0.40, d_{2,0} = 0.30, \phi_0 = 0.80$</i>								
η_1	<i>0.9790</i>	<i>0.9810</i>	<i>0.9781</i>	<i>0.9819</i>	<i>0.9775</i>	<i>0.9825</i>	<i>0.9763</i>	<i>0.9837</i>
500	0.0002	-0.0004	-0.0003	-0.0002	-0.0009	0.0001	-0.0026	0.0008
1000	0.0001	-0.0001	-0.0001	-0.0001	-0.0002	0.0001	-0.0010	0.0004
2000	0.0001	-0.0001	-0.0001	1.5e-5	-0.0002	0.0001	-0.0004	0.0004
η_2	<i>-0.4071</i>	<i>-0.3929</i>	<i>-0.4133</i>	<i>-0.3867</i>	<i>-0.4170</i>	<i>-0.3830</i>	<i>-0.4258</i>	<i>-0.3742</i>
500	-0.0008	-0.0008	-0.0060	0.0045	-0.0104	0.0081	-0.0250	0.0158
1000	-0.0004	-0.0001	-0.0028	0.0020	-0.0049	0.0032	-0.0108	0.0075
2000	-0.0002	-0.0001	-0.0013	0.0010	-0.0023	0.0017	-0.0048	0.0038
<i>Model 4: $\eta_{1,0} = 1.00, d_{1,0} = 0.20; \eta_{2,0} = -0.60, d_{2,0} = 0.45$</i>								
η_1	<i>0.9997</i>	<i>1.0001</i>	<i>0.9994</i>	<i>1.0001</i>	<i>0.9990</i>	<i>1.0002</i>	<i>0.9980</i>	<i>1.0003</i>
500	-0.0001	-3.5e-5	-0.0005	-0.0001	-0.0009	-0.0001	-0.0019	-0.0002
1000	-3.4e-5	-1.4e-5	-0.0002	-3.1e-5	-0.0003	-4.3e-5	-0.0012	-0.0001
2000	-8.6e-6	-2.2e-6	-3.7e-5	-5.2e-6	-0.0001	-7.4e-6	-0.0001	-1.2e-5
η_2	<i>-0.6041</i>	<i>-0.5959</i>	<i>-0.6077</i>	<i>-0.5923</i>	<i>-0.6099</i>	<i>-0.5901</i>	<i>-0.6150</i>	<i>-0.5850</i>
500	0.0012	-0.0012	0.0004	0.0003	-0.0007	0.0018	-0.0034	0.0057
1000	0.0005	-0.0005	-1.6e-5	0.0002	-0.0007	0.0009	-0.0026	0.0018
2000	0.0001	-0.0003	-0.0002	0.0001	-0.0007	0.0005	-0.0019	0.0009

The true model is depicted in the header of each panel, with the sample size displayed in the first column. Quantities in italic font are theoretical confidence bands for $T = 500$ calculated using Theorem 3. Remaining values denote the difference between the percentiles from the simulations and the associated theoretical quantity for a given confidence band from Models 1-4

5 Application

Emerging research has demonstrated that cyclical long memory is an important characteristic of many financial time series.⁶ To demonstrate the applicability of the CSS estimator and the proposed theory, we consider the weekly trading volume of IBM equities from January 1, 1962, through March 28, 2022. Without loss of generality, the data have been rescaled by dividing by the maximum value for volume. The periodogram of the difference of the resulting series is depicted in Fig. 3.

From the visual inspection of Fig. 3, we identified as many as 9 frequencies as candidates for spectral poles, including the origin, which dominates the periodogram for the raw series. Based on the discussion above, we then used the BIC to select k and the number of autoregressive and moving average parameters. For each k , we considered all combinations of models with $p, q \leq 3$. Among the 144 estimated models, the BIC selected the 6-frequency GARMA(2,3) model, while the Hannan–Quinn marginally selected $k=8$ vs. $k = 6$ when considering $p = 2$ and $q = 3$. We therefore selected the 6-frequency GARMA(2,3) model whose estimation results appear in Table 7. Results for $k > 6$ produce similar findings that are available on request.

For the 6-factor model, one isolated frequency is at the origin and the other 5 estimated frequencies are depicted in Fig. 3 by the vertical dotted lines.⁷ Based on the simulation results as discussed above, we show 68% confidence bands under the assumption that $|\eta_{i,0}| < 1$. Additionally, for estimates of $\psi_{\delta,r}$, we present both numerical and asymptotic standard errors that are very similar and, thus, provide strong support for the proposed distribution theory.⁸

Because the estimated value of η_1 is only marginally less than 1, there is strong evidence of a spectral pole at the origin. As discussed above, however, the distribution theory building on Chung (1996a) is suspect when $\eta_{i,0}=1$. Consequently, we suggest that a bootstrap method may be a reliable alternative. Although the construction of a fully validated bootstrap test statistic is outside the scope of this paper, the existing literature provides guidance that we exploit here. First, note that under the null, $H_0 : \eta_{1,0} = 1$, the parameter η_1 lies on the boundary of the parameter space. In such cases, it has been established that bootstrap samples generated from unrestricted CSS estimation may yield invalid test statistics, failing to mimic the target distribution under the null (Andrews 2000; Cavaliere et al. 2017; Cavaliere and Rahbek 2021). A resolution to this problem is to use a restricted bootstrap, where

⁶ See Lu and Guegan (2011) and Caporale and Gil-Alana (2014) for applications to the Nikkei-based forward premia and price dividend ratios in the USA. Also, see Asai et al. (2020) who provide evidence of multiple sources of cyclical long memory in differenced interest rates for the USA and Australia at various maturities.

⁷ We selected $\omega = 0.4818$ as a potential candidate for a pole given the magnitude of the periodogram at this frequency relative to nearby ordinates. The remaining candidates were associated with the largest values of the periodogram of the differenced data.

⁸ As a robustness check, given potential non-stationarity when $\eta_{1,0} = 1$, we also estimated a model applied to the first difference of volume. The estimated value of d_1 is equal to -0.1923 , which implies a value of 0.3077 for the series in levels when $\eta_{1,0} = 1$. All other parameter estimates, which are available upon request, indicate no tangible disparities, including, most notably, the position of the spectral poles.

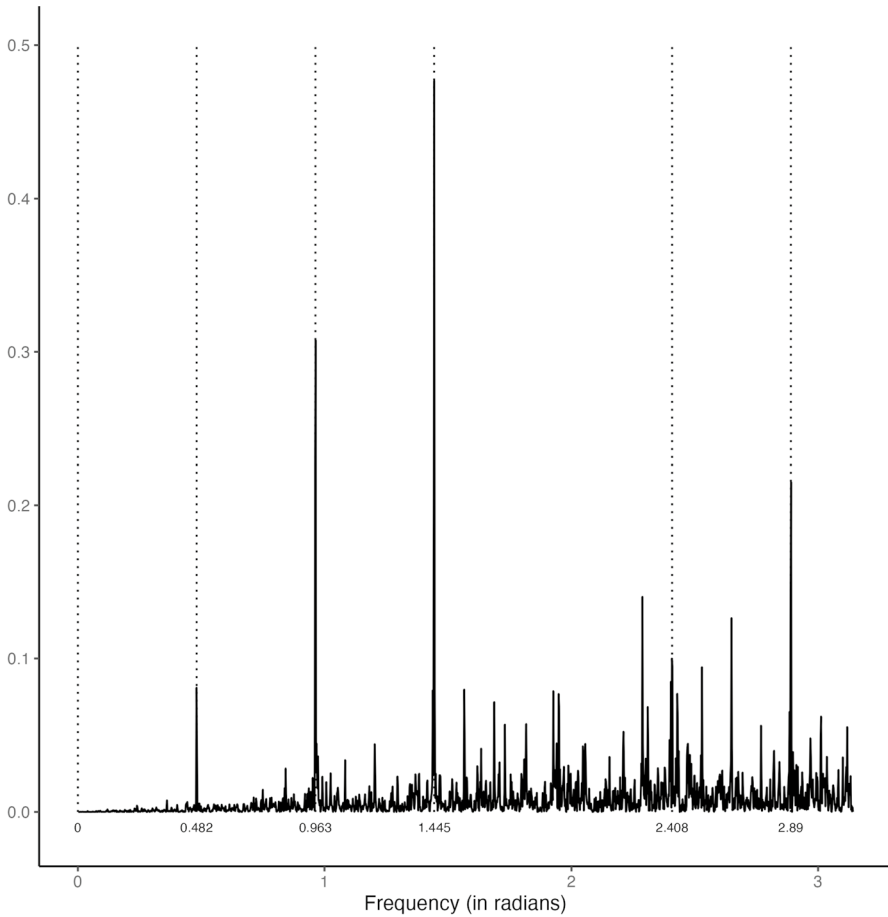


Fig. 3 Periodogram of the first difference of IBM trading volume

samples are formed from residuals and parameters estimated under the null (Cavaliere et al. 2017). Recently, for the single-frequency GARMA model, Beaumont and Smallwood (2022) propose a restricted bootstrap method to compute critical values and demonstrate that bootstrapped test statistics for the null $H_0 : \eta_0 = 1$ have correct nominal size, even under potential non-stationarity.

Following Beaumont and Smallwood (2022), we generate a test statistic for the null, $H_0 : \eta_{1,0} = 1$, through re-estimation of the selected 6-factor GARMA(2,3) model with $\eta_{1,0} = 1$ imposed. We sample with replacement from the estimated residuals to construct 1000 samples under the null hypothesis. We then estimate the unrestricted 6-factor model for each of the 1000 samples to obtain $T^2(\hat{\eta}_1^{(j)} - 1)$, for $j \in (1, 1000)$. The test-statistics are sorted to obtain bootstrapped critical values that are presented in Table 7 along with the critical values obtained using Theorem 3. As seen in the table, even with 3144 observations, the theoretical critical values appear to be far too small in absolute value when compared to the bootstrapped critical

Table 7 Estimation of 6-frequency GARMA(2,3) model for IBM volume

	d_1	d_2	d_3	d_4	d_5	d_6	
Estimates	0.3319	0.1871	0.2217	0.4973	0.1585	0.1312	
Asy. SE	[0.0206]	[0.0273]	[0.0234]	[0.0841]	[0.0228]	[0.0200]	
Num. SE	[0.0197]	[0.0244]	[0.0212]	[0.0838]	[0.0228]	[0.0194]	
	ϕ_1	ϕ_2	θ_1	θ_2	θ_3	μ	BIC
Estimates	0.0225	- 0.6747	- 0.5517	0.9689	- 0.3091	0.0212	- 9467.3
Asy. SE	[0.0361]	[0.0506]	[0.1036]	[0.0342]	[0.0906]	[0.1021]	
Num. SE	[0.0352]	[0.0507]	[0.0977]	[0.0332]	[0.0851]	[0.0237]	
Estimation results for η_i							
	η_1	η_2	η_3	η_4	η_5	η_6	
Estimates	1.0000	0.8863	0.5708	0.1258	- 0.7428	- 0.9684	
L-68% Band	0.9999997	0.8854	0.5694	0.1251	- 0.7442	- 0.9691	
U-68% Band	1.0000	0.8872	0.5721	0.1266	- 0.7411	- 0.9677	
Tests for $\eta_1 = 1$							
Statistic	Bootstrapped critical value			Chung-based critical value			
	90%	95%	99%	90%	95%	99%	
$T^2(\hat{\eta}_1 - 1)$	- 0.3799	- 54.49	- 121.96	- 335.52	- 30.29	- 48.94	- 111.94
ARIMA estimation results							
Estimates and statistics							
	ϕ_1	θ_1	c	BIC	DFG ^t	DFG ^c	
Variable	0.3251	- 0.9288	0.00002	- 9123.3	- 3.3835	- 1.7951	
ΔIBM_t	[0.0126]	[0.0052]	[0.0001]				

Confidence bands for $\hat{\eta}_i$ are constructed assuming $|\eta_{i,0}| < 1$. Values in brackets are standard errors based on the estimated model. Asymptotic standard errors (“Asy. SE”) are calculated using Theorem 2 assuming $|\eta_{i,0}| < 1$, for all i . Numerical standard errors (“Num. SE”) are based on the outer product of the estimated score. In the panel “Tests for $\eta_1 = 1$ ”, we report test results for the hypothesis $\eta_{1,0} = 1$ using both a parametric bootstrap procedure and theoretical results based on Chung (1996a). For ARIMA estimation results, DFG^t and DFG^c denote the DF-GLS test of Elliott et al. (1996) allowing for time trend and drift only, respectively. The 95% critical values for the test are given by - 2.890 and - 1.941, and lag lengths are selected by minimization of the BIC for up to 16 lags

values. In this example, the discrepancy does not alter the conclusion given an estimated value so close to unity.

It should be noted that more research is needed to determine the conditions under which the proposed bootstrap test is consistent. The main task would be an analysis of the distributional properties of the bootstrapped test statistic under the alternative. In general, as pointed out by Cavaliere and Rahbek (2021), this is a very difficult problem, and there is reason to believe the current environment presents unique challenges. In particular, the procedure above uses bootstrapped residuals obtained under the null. If the alternative hypothesis is true, the resulting disturbances are

expected to possess long memory of a potentially complicated form, since the correct filter, $(1 - 2\eta_{1,0}L + L^2)^{d_{1,0}}$, has not been applied to the data. Further, as our theory above shows, \hat{d}_i is not independent of \hat{d}_j , so that there are additional complications that arise under misspecification.⁹ The behavior of the residuals in this context will be important in future research exploring formal proofs for consistency.

To the extent that there is concern with the proposed bootstrap when the null is false, Cavaliere and Rahbek (2021) propose a hybrid approach to obtain bootstrapped samples using parameters estimated under the null, while using disturbances obtained from unrestricted estimation. This avoids the issue of sampling with long-memory residuals. Specifically, let ε_t^* denote the set of residuals obtained from the unconstrained model in Table 7. Then, resampling of ε_t^* with replacement is used along with parameter estimates with $\eta_1 = 1$ imposed to bootstrap samples consistent with the null under investigation. The remaining steps are the same as for the restricted bootstrap. As discussed extensively by Cavaliere and Rahbek (2021), the use of a hybrid bootstrap of this sort can be useful in instances where boundary conditions are met for a given parameter, but concerns also exist about the properties of residuals under the alternative. As a robustness check to the findings above, we conducted the hybrid bootstrap, and the results continue to yield a failure to reject the null $\eta_{1,0} = 1$ for any conventional test-size.

Finally, to put our findings into context relative to traditional time series methods, we provide estimation results associated with ARIMA models in the bottom panel of Table 7. First, unit root tests present somewhat contradictory results. Specifically, the DF-GLS test of Elliott et al. (1996) yields a rejection of the unit root null at the 5% level when a linear time trend is considered, where a failure to reject otherwise results. Further, coefficients on linear time trends are insignificant for ARIMA models estimated in levels, where the sum of autoregressive coefficients is quite close to one. We therefore proceed by estimating an ARIMA(1,1,1) model, which yielded the lowest BIC for all model combinations with p and q less than or equal to 3. The estimated moving average coefficient is large and negative, potentially contributing to the confusion rendered from standard unit root tests.

As evidenced by a much lower BIC value (-9467.3 vs. -9123.3), the estimated 6-factor GARMA model yields a superior in-sample fit relative to ARIMA methods. The GARMA estimation results further yield evidence against a unit root. Perhaps most importantly, the GARMA model can capture very diverse dynamics in the data and provides additional insights for researchers analyzing long-memory cycles in financial time series. In our example of IBM trading volume, the estimated values of the Gegenbauer frequencies, \hat{v}_i , range from 0.0003 to 2.8897, indicating cycle lengths of 13.05, 6.52, 4.35, 2.61 and 2.17 weeks. In addition, we detect an

⁹ Results from Cavaliere et al. (2022) imply that additional modifications might be needed for testing when nuisance parameters also lie on the boundary of the parameter space. For the k -factor GARMA model, $\eta_{k,0}$ could theoretically equal -1 , although it is not clear how empirically relevant this is. More importantly, our distributional theory indicates $\hat{\eta}_i$ is independent of $\hat{\eta}_j \forall i \neq j$, so it seems unlikely that the concern of Cavaliere et al. (2022) is relevant here. It does highlight, however, the difficulty with establishing bootstrap consistency in the current environment.

extremely long, potentially infinite, cycle associated with the value of $\hat{\eta}_1$ 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 may improve our understanding of 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 k -factor GARMA model generalizes existing long-memory models and has the particular advantage that the ACF can decay at a non-monotonic rate that is not necessarily symmetric about zero. In addition, the k -factor GARMA model can accommodate multiple poles in the spectral density function.

As noted by Hunt et al. (2022), providing a full set of distributional results for estimators of k -factor GARMA models has proven elusive. Building on the results in Chung (1996a, 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 $\hat{\eta}_i$ is independent of all other parameters, including $\hat{\eta}_j$, whenever $i \neq j$. It is important to note, however, that remaining parameters, notably memory 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 potential discontinuities in the distribution theory for $\hat{\eta}_i$. We attempt to overcome this shortcoming by conducting extensive simulations and drawing on the recent work of Beaumont and Smallwood (2022) 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. Further, the proposed theory can be used to accurately obtain confidence bands for $\hat{\eta}_i$. Finally, an application demonstrates the practical value of the k -factor GARMA model. The trading volume of IBM is shown to be well modeled by a six-factor GARMA model with a spectral singularity at the origin.

Given the early success of k -factor GARMA models, as discussed in 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 where specific interest lies in uncertainty regarding the periodicity of long-memory cycles. Nonetheless, challenges still remain. More work is likely needed to determine the appropriate number of spectral poles, although we are able to provide recommendations potentially complementing the recent breakthrough by Leschinski and Sibbertsen (2019). Perhaps more importantly, the proposed distribution theory for $\hat{\eta}_i$ directly follows Chung (1996a) in allowing the true value, $\eta_{i,0}$, to potentially lie on the boundary of the associated parameter space. Additionally, the theoretical results suggest that a discontinuity in the distribution of $\hat{\eta}_i$ occurs as $|\eta_{i,0}| \rightarrow 1$, where standard

rate- T convergence gives way to a T^2 rate when $|\eta_{i,0}| = 1$. While extensive simulation evidence provides strong support for the proposed theory when $|\eta_{i,0}| < 1$, the findings also suggest there are concerns when $\eta_{i,0} = 1$. A proposed bootstrap test offers one potential remedy for researchers interested in testing $\eta_{i,0} = 1$, although additional theory will be required to confirm consistency of the test.

Appendix: Lemma and Theorem Proofs

Proof of Lemma 1

Consider the AR representation of x_t in Eq. (4), where, without loss of generality, $\mu = 0$. Define the power series $a(L)$ and $b(L)$ as follows, recalling that all roots to $\phi(z) = 0$ and $\theta(z) = 0$ lie outside the unit circle:

$$a(L) = \frac{\phi(L)}{\theta(L)} = \sum_{n=0}^{\infty} a_n L^n$$

$$b(L) = \prod_{i=1}^k (1 - 2\eta_i L + L^2)^{d_i} = \sum_{n=0}^{\infty} b_n L^n$$

With $\alpha(L)x_t = \varepsilon_t$, we have $\alpha(L) = a(L)b(L)$. Here, the product represents a Cauchy product, where,

$$\alpha(L) = \left(\sum_{n=0}^{\infty} a_n L^n \right) \left(\sum_{n=0}^{\infty} b_n L^n \right) = \left(\sum_{n=0}^{\infty} c_n L^n \right), \tag{A.1}$$

and

$$c_n = \sum_{i=0}^n a_i b_{n-i}. \tag{A.2}$$

Clearly, $\sum_{n=0}^{\infty} |c_n| < \infty$ if both $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ are absolutely convergent (see page 80 of Rudin (1976) for a discussion). Additionally, $\sum_{n=0}^{\infty} a_n$ is known to be absolutely convergent (see, for example, Hassler (2018), Equation 3.11). Thus, if $\sum_{n=0}^{\infty} |b_n| < \infty$, then the k -factor GARMA process has an AR(∞) representation with absolutely summable coefficients. A useful form of the coefficients in the AR(∞) representation can be obtained by setting $d_i = -d_i$ in the MA(∞) representation of the k -factor GARMA(0,0) model provided on page 3255 of Kouamé and Hili (2008). We can start by assuming that $|\eta_i| < 1$, for all i . Let b_n denote the n th coefficient in the expansion of $\prod_{i=1}^k (1 - 2\eta_i L + L^2)^{d_i}$. Then, from Kouamé and Hili (2008), as $n \rightarrow \infty$,

$$b_n = 2 \sum_{i=1}^k D_i \frac{\Gamma(n - d_i)}{\Gamma(n + 1)\Gamma(-d_i)} \cos(v_i n + m_i) + o(n^{-2+\max(-d_1, -d_2, \dots, -d_k)}), \quad (\text{A.3})$$

where the final term is asymptotically negligible since $d_i > 0$. For the remaining terms, with $v_i = \cos^{-1}(\eta_i)$,

$$D_i = |2 \sin v_i|^{d_i} \prod_{j \neq i} |2(\cos(v_i) - \cos(v_j))|^{d_j}, \quad \text{and}$$

$$m_i = -v_i \sum_{j=1}^k d_j + \pi \sum_{j=1}^{i-1} d_j + \frac{\pi}{2} d_i.$$

It is important to note D_i and m_i are constants that do not depend on n . Ignoring the final term in A.3, define \tilde{b}_n , accordingly. We have,

$$\sum_{n=0}^{\infty} |\tilde{b}_n| \leq 2 \sum_{n=0}^{\infty} \left(\sum_{i=1}^k \left| \frac{\Gamma(n - d_i)}{\Gamma(n + 1)\Gamma(-d_i)} \cos(v_i n + m_i) \right| \right). \quad (\text{A.4})$$

If two series are absolutely convergent, then their sum is also absolutely convergent. Thus, we can consider each term in parentheses individually. Applying Stirling’s formula, we have,

$$\frac{\Gamma(n - d_i)}{\Gamma(n + 1)\Gamma(-d_i)} \sim \frac{n^{-d_i-1}}{\Gamma(-d_i)}. \quad (\text{A.5})$$

To prove absolute convergence, we apply this result and use the comparison test. We have, for individual i and fixed n

$$\left| \frac{D_i}{\Gamma(-d_i)} \frac{\cos(v_i n + m_i)}{n^{d_i+1}} \right| \leq \left| \frac{D_i}{\Gamma(-d_i)} \right| \frac{1}{n^{d_i+1}}. \quad (\text{A.6})$$

We further have, $\left| \frac{D_i}{\Gamma(-d_i)} \right| \sum_{n=0}^{\infty} n^{-d_i-1}$ is a convergent p-series, provided $d_i + 1 > 1$, or $d_i > 0$. This assumption is in place for all i , and therefore, by the comparison test, $\sum_{n=0}^{\infty} \tilde{b}_n$ is absolutely convergent.

Now, assume that there does actually exist a value of η_i , where $|\eta_i| = 1$. We then have that the i th term in parentheses from Eq. A.4 becomes:

$$\left| D_i \frac{\Gamma(n - 2d_i)}{\Gamma(n + 1)\Gamma(-2d_i)} \cos(v_i n) \right|, \quad (\text{A.7})$$

where $D_i = \prod_{j \neq i} |2(\cos(v_i) - \cos(v_j))|^{d_j}$. A proof that the affiliated sum is absolutely convergent is obvious in light of the above.

Proof of Theorem 1

From (6), with $\psi = (\psi'_{\delta,\tau}, \psi'_\eta)'$, we have $s_T(\psi) = s_T((\psi'_{\delta,\tau}, \psi'_\eta)')$, which is denoted below as s_T for simplicity. Consider the first-order Taylor series expansion of s_T of the invertible and stationary k -factor GARMA model for the process $\{x_t\}_{t=1}^T$, about the true parameter values $\psi_{\delta_0,\tau_0} = (d_{1,0}, \dots, d_{k,0}, \phi_{1,0}, \dots, \phi_{p,0}, \theta_{1,0}, \dots, \theta_{q,0})'$ and $\eta_0 = (\eta_{1,0}, \dots, \eta_{k,0})'$. We have,

$$\begin{aligned} & \left[\frac{1}{f_T} \frac{\partial s_T}{\partial \eta} \right] \Big|_{\psi=\psi_0} + \left[\frac{1}{T} \frac{\partial^2 s_T}{\partial \psi_{\delta,\tau} \partial \psi'_{\delta,\tau}} \quad \frac{1}{\sqrt{T}} \frac{1}{F_T} \odot \frac{\partial^2 s_T}{\partial \psi_{\delta,\tau} \partial \eta'} \right] \Big|_{\psi=\psi_0} \\ & \times \left[\sqrt{T} (\hat{\psi}_{\delta,\tau} - \psi_{\delta_0,\tau_0}) \right] = o_p(1) \end{aligned} \tag{A.8}$$

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 $|\eta_{j,0}| < 1$ and T^2 and $\frac{1}{T^2}$ when $|\eta_{j,0}| = 1$. $\frac{1}{f_T}$ denotes the matrix formed by stacking the vector, $\frac{1}{f_T}'$, on top of itself k times, and $\frac{1}{f_T f_T'} = \left(\frac{1}{f_T}\right) \left(\frac{1}{f_T}\right)'$.

We will show below that $\frac{1}{T} \frac{\partial^2 s_T}{\partial \psi_{\delta,\tau} \partial \psi'_{\delta,\tau}} \Big|_{\psi_{\delta,\tau}=\psi_{\delta_0,\tau_0}}$ and $\left(\frac{1}{f_T f_T'}\right) \odot \frac{\partial^2 s_T}{\partial \eta \partial \eta'} \Big|_{\eta=\eta_0}$ are $O_p(1)$,

while $\frac{1}{\sqrt{T}} \frac{1}{F_T} \odot \frac{\partial^2 s_T}{\partial \psi_{\delta,\tau} \partial \eta'} \Big|_{\psi_{\delta,\tau}=\psi_{\delta_0,\tau_0}, \eta=\eta_0}$ possesses elements that are all $o_p(1)$. We also

show that all remaining elements are bounded. To facilitate, let $I(\psi)$ denote the Fisher information matrix, which ultimately contains elements given by $E \frac{1}{T\sigma^2} \sum_{t=1}^T \frac{\partial \epsilon_t}{\partial \psi} \frac{\partial \epsilon_t}{\partial \psi'}$. $I(\psi)$ is partitioned as,

$$I(\psi) = \begin{bmatrix} I_{\psi_{\delta,\tau}} & I_{\psi_{\delta,\tau} \times \eta} \\ I_{\psi_{\delta,\tau} \times \eta} & I_{\eta} \end{bmatrix} \tag{A.9}$$

Given the assumptions as discussed above that all terms are $O_p(1)$ or $o_p(1)$, for large T , we have,

$$\begin{aligned} & \left[\sqrt{T} (\hat{\psi}_{\delta,\tau} - \psi_{\delta_0,\tau_0}) \right] = - \left[\begin{array}{cc} I_{\psi_{\delta_0,\tau_0}} & \frac{\sqrt{T}}{F_T} \odot I_{\psi_{\delta_0,\tau_0} \times \eta_0} \\ \frac{\sqrt{T}}{F_T} \odot I_{\psi_{\delta_0,\tau_0} \times \eta_0} & \frac{1}{\sigma^2} \frac{1}{f_T f_T'} \odot \sum_{t=1}^T \frac{\partial \epsilon_t}{\partial \eta} \frac{\partial \epsilon_t}{\partial \eta'} \Big|_{\eta=\eta_0} \end{array} \right]^{-1} \\ & \times \left[\begin{array}{c} \frac{1}{\sqrt{T} \sigma^2} \sum_{t=1}^T \epsilon_t \frac{\partial \epsilon_t}{\partial \psi_{\delta,\tau}} \\ \frac{1}{\sigma^2} \frac{1}{f_T} \odot \sum_{t=1}^T \epsilon_t \frac{\partial \epsilon_t}{\partial \eta} \end{array} \right] \Big|_{\psi=\psi_0} + o_p(1) \end{aligned} \tag{A.10}$$

For the individual information numbers in $I_{\psi_{\delta_0, \tau_0} \times \eta_0}$, consider $I_{d_{i,0}, \eta_{j,0}}$, $i = 1, \dots, k$, with $|\eta_{j,0}| < 1$. Using Gradshteyn and Ryzhik (1980) Eqs. 1.514 and 8.937.1, we find that the information matrix elements of $I_{d_{i,0}, \eta_{j,0}}$ are

$$I_{d_{i,0}, \eta_{j,0}} = E \left[\frac{4d_{j,0}}{\sigma^2 T} \sum_{t=1}^T \left(\sum_{l=1}^{\infty} \frac{\cos(lv_{i,0})}{l} \varepsilon_{t-l} \right) \left(\sum_{l=1}^{\infty} \frac{\sin(lv_{j,0})}{\sin(v_{j,0})} \varepsilon_{t-l} \right) \right]. \tag{A.11}$$

Under the assumptions governing ε_t , if $v_{j,0} > v_{i,0}$, and $v_{i,0} \neq v_{j,0}$, Gradshteyn and Ryzhik (1980) Equation 1.441.1 yields

$$\begin{aligned} I_{d_{i,0}, \eta_{j,0}} &= \frac{2d_{j,0}}{\sin(v_{j,0})} \left(\sum_{l=1}^{\infty} \frac{\sin[l(v_{i,0} + v_{j,0})] + \sin[l(v_{j,0} - v_{i,0})]}{l} \right) \\ &= \frac{2d_{j,0}(\pi - v_{j,0})}{\sin(v_{j,0})}. \end{aligned} \tag{A.12}$$

Thus, $I_{d_{i,0}, \eta_{j,0}} < \infty$. If $v_{j,0} < v_{i,0}$, then the infinite sums in (A.12) are equal to $\sum_{l=1}^{\infty} \frac{\sin[l(v_{j,0} - v_{i,0}) + 2\pi l]}{l}$. From Gradshteyn and Ryzhik (1980) Equation 1.444.1, we see that the infinite sum converges. The same is true if $v_{i,0} = v_{j,0}$. Chung (1996a, 1996b) has established that all remaining information numbers in $I_{\psi_{\delta_0, \tau_0} \times \eta_0}$ are finite. Thus, $\frac{\sqrt{T}}{F_T} \odot I_{\psi_{\delta_0, \tau_0} \times \eta_0} = o_p(1)$.

If the remaining terms of all of the elements in (A.8) are $O_p(1)$ or $o_p(1)$, as shown below, then the second matrix in the top line of (A.8) is asymptotically block diagonal, and the distribution of $\sqrt{T}(\hat{\psi}_{\delta, \tau} - \psi_{\delta_0, \tau_0})$ can be considered independently of $f_T(\hat{\eta} - \eta_0)$ as claimed.

Proof of Theorem 2

From (A.10), the assumption that the remaining elements involving η in (A.10) are bounded, and the central limit theorem of Chan and Wei (1988),

$$\sqrt{T}(\hat{\psi}_{\delta, \tau} - \psi_{\delta_0, \tau_0}) = -I_{\psi_{\delta_0, \tau_0}}^{-1} \left[\frac{1}{\sqrt{T}\sigma^2} \sum_{t=1}^T \varepsilon_t \frac{\partial \varepsilon_t}{\partial \psi_{\delta, \tau}} \Big|_{\psi_{\delta, \tau} = \psi_{\delta_0, \tau_0}} \right] + o_p(1). \tag{A.13}$$

Given, A.13, we have

$$\sqrt{T}(\hat{\psi}_{\delta, \tau} - \psi_{\delta_0, \tau_0}) \rightarrow N(0, I_{\psi_{\delta_0, \tau_0}}^{-1}). \tag{A.14}$$

Information numbers for the diagonal terms of $I_{\psi_{\delta_0, \tau_0}}$ are given on page 251 in Chung (1996a). The off diagonal terms, $I_{d_{i,0}, d_{j,0}}$, which for large T and $i \neq j$ are,

$$I_{d_{i,0}, d_{j,0}} = E \frac{1}{T\sigma^2} \sum_{t=1}^T [\log(1 - 2\eta_{i,0}L + L^2)\varepsilon_t \log(1 - 2\eta_{j,0}L + L^2)\varepsilon_t]. \tag{A.15}$$

Using Gradshteyn and Ryzhik (1980) Eqs. 1.514 and 1.443.3 yields,

$$\begin{aligned}
 I_{d_{i,0}d_{j,0}} &= 2 \sum_{l=1}^{\infty} \frac{\cos(l(v_{i,0} + v_{j,0})) + \cos(l(v_{i,0} - v_{j,0}))}{l^2} \\
 &= 2 \left(\frac{\pi^2}{3} - \pi v_{i,0} + \frac{v_{i,0}^2 + v_{j,0}^2}{2} \right).
 \end{aligned}
 \tag{A.16}$$

For $i = 1, \dots, k$, $j = 1, \dots, p$, and $m = 1, \dots, q$, the proofs of the results for $I_{d_{i,0}\phi_{j,0}}$ and $I_{d_{i,0}\theta_{m,0}}$ follow 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 applied to any arbitrary values of η_i and η_j .

Lemma 2 Consider any stationary and invertible k -factor GARMA model defined in Eq. (1). Then, with $i \neq j$,

$$-\frac{1}{\sigma^2 T^\kappa} \sum_{t=1}^T \frac{\partial^2 \varepsilon_t}{\partial \eta_i \partial \eta_j} = o_p(1),$$

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

$$Z_{at} = -\frac{1}{2d_a} \frac{\partial \varepsilon_{t+1}}{\partial \eta_a} = \frac{\varepsilon_t}{(1 - 2\eta_a L + L^2)}, \quad a = 1, 2.
 \tag{A.17}$$

Applying Gradshteyn and Ryzhik (1980) Equation 8.937.1,

$$Z_{at} = \frac{1}{\sin(v_a)} \sum_{j=1}^t \sin[(t+1)v_a - jv_a] \varepsilon_j, \quad a = 1, 2
 \tag{A.18}$$

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

$$S_T(v_a, r) = \frac{\sqrt{2}}{\sqrt{T\sigma^2}} \sum_{j=1}^{[Tr]} \cos(jv_a) \varepsilon_j, \quad a = 1, 2
 \tag{A.19a}$$

$$T_T(v_a, r) = \frac{\sqrt{2}}{\sqrt{T}\sigma^2} \sum_{j=1}^{[Tr]} \sin(jv_a)\epsilon_j, \quad a = 1, 2 \tag{A.19b}$$

where $r \in [0, 1]$ and $[Tr]$ is the integer part. Finally, from the expressions in (A.19) and using $\omega_1 = v_1 + v_2, \omega_2 = v_1 - v_2, \omega_3 = v_2 - v_1$ along with a few rules of trigonometry, we get the following expression,

$$\begin{aligned} & \frac{4 \sin(v_1) \sin(v_2)}{\sigma^2} \frac{1}{T^2} \sum_{t=1}^{T-1} Z_{1t} Z_{2t} \\ &= \frac{1}{T} \sum_{t=1}^{T-1} (\cos[(t+1)\omega_2] - \cos[(t+1)\omega_1]) S_T(v_1, t/T) S_T(v_2, t/T) \\ & - \frac{1}{T} \sum_{t=1}^{T-1} (\sin[(t+1)\omega_1] + \sin[(t+1)\omega_3]) S_T(v_2, t/T) T_T(v_1, t/T) \tag{A.20} \\ & - \frac{1}{T} \sum_{t=1}^{T-1} (\sin[(t+1)\omega_1] + \sin[(t+1)\omega_2]) S_T(v_1, t/T) T_T(v_2, t/T) \\ & + \frac{1}{T} \sum_{t=1}^{T-1} (\cos[(t+1)\omega_1] + \cos[(t+1)\omega_2]) T_T(v_1, t/T) T_T(v_2, t/T). \end{aligned}$$

Consider the random elements

$$S_n^*(v_1) = \sum_{j=1}^n \cos(jv_1) \epsilon_j \text{ and } T_n^*(v_1) = \sum_{j=1}^n \sin(jv_1) \epsilon_j, \tag{A.21}$$

and similarly for $S_n^*(v_2)$ and $T_n^*(v_2)$. Let $\{X_n\} = \{S_n^*(v_1) S_n^*(v_2)\}$, and consider the first term in (A.20). It is clear from the definition of $S_T(v_1, \frac{t}{T})$ and $S_T(v_2, \frac{t}{T})$ that $\frac{1}{T} \sum_{t=1}^{T-1} \cos[(t+1)\omega_2] S_T(v_1, t/T) S_T(v_2, t/T) = o_p(1)$ if

$$\sup_{1 \leq j \leq T} \left| \sum_{n=1}^j e^{in\theta} X_n \right| = o_p(T^2). \tag{A.22}$$

First, observe that

$$E|S_n^*(v_1) S_n^*(v_2)| \leq \{ES_n^{*2}(v_1)\}^{1/2} \{ES_n^{*2}(v_2)\}^{1/2} \leq \sigma^2 n \tag{A.23}$$

so that $E|S_n^*(v_1) S_n^*(v_2)| = O(n)$. Now let $n \geq m$ and consider

$$|X_n - X_m| \leq |S_n^*(v_1)| |S_n^*(v_2) - S_m^*(v_2)| + |S_m^*(v_2)| |S_n^*(v_1) - S_m^*(v_1)|. \tag{A.24}$$

Noting that

$$E|S_n^*(v_1)|^2 = E(S_n^*(v_1))^2 = \sigma^2 \left(\sum_{j=1}^n \cos(jv_1) \right)^2 \leq \sigma^2 n \tag{A.25}$$

yields $E|S_n^*(v_1)|^2 = O(n)$. Given $m \leq n$, this also implies $E|S_m^*(v_2)|^2 \leq \sigma^2 n$.

Next consider the expression

$$E|S_n^*(v_2) - S_m^*(v_2)|^2 = \sigma^2 \left(\sum_{j=m+1}^n \cos^2(jv_2) \right) \leq \sigma^2(n - m). \tag{A.26}$$

Thus, $E|S_n^*(v_2) - S_m^*(v_2)|^2 = O(n - m)$. Similar reasoning implies that $E|S_n^*(v_1) - S_m^*(v_1)|^2 = O(n - m)$. If $v_1 \neq v_2$, by Theorem 2.1 in Chan and Wei (1988), we see that the first term in (A.20) is $o_p(1)$. By similar reasoning, the remaining terms in (A.20) are also seen to be $o_p(1)$. Thus, we have established that

$$\frac{4 \sin(v_1) \sin(v_2)}{\sigma^2} \frac{1}{T^2} \sum_{t=1}^{T-1} Z_{1t} Z_{2t} = o_p(1). \tag{A.27}$$

This expression is asymptotically equivalent to

$$-\frac{4d_1 d_2}{4 \sin(v_1) \sin(v_2)} \frac{4 \sin(v_1) \sin(v_2)}{\sigma^2} \frac{1}{T^2} \sum_{t=1}^{T-1} Z_{1t} Z_{2t}, \tag{A.28}$$

which is $o_p(1)$. This completes the proof of Case 1 in the lemma.

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

$$Z_{1t} = -\frac{1}{2d_1} \frac{\partial \varepsilon_{t+1}}{\partial \eta_1} = \frac{\varepsilon_t}{(1 - 2\eta_1 L + L^2)}, Z_{kt} = -\frac{1}{2d_k} \frac{\partial \varepsilon_{t+1}}{\partial \eta_k} = \frac{\varepsilon_t}{(1 - L)^2}. \tag{A.29}$$

Define the auxiliary process and its associated truncation.

$$Y_t = (1 - L)Z_{kt} = \sum_{j=1}^t \varepsilon_j. \tag{A.30}$$

This gives the following truncated series for Z_{kt} ,

$$Z_{kt} = \sum_{j=1}^t Y_t = \sum_{j=1}^t j \varepsilon_{t-j+1}. \tag{A.31}$$

For ease of exposition, define the random process

$$X_T(r) = \frac{1}{T} \frac{1}{\sqrt{T}\sigma} \sum_{j=1}^{[Tr]} Y_j, \tag{A.32}$$

and define $S_T(v_1, t/T)$ and $T_T(v_1, t/T)$ precisely as in (A.19). We then get,

$$\begin{aligned} \frac{\sqrt{2} \sin(v_1)}{\sigma^2} \frac{1}{T^3} \sum_{t=1}^{T-1} Z_{1t} Z_{kt} &= \frac{1}{T} \sum_{t=1}^{T-1} \sin[(t+1)v_1] S_T(v_1, t/T) X_T(t/T) \\ &- \frac{1}{T} \sum_{t=1}^{T-1} \cos[(t+1)v_1] T_T(v_1, t/T) X_T(t/T). \end{aligned} \tag{A.33}$$

Note that the expression

$$-\frac{4d_1 d_k}{\sqrt{2} \sin(v_1)} \frac{\sqrt{2} \sin(v_1)}{\sigma^2} \frac{1}{T^3} \sum_{t=1}^{T-1} Z_{1t} Z_{kt}, \tag{A.34}$$

is asymptotically equivalent to $-\frac{1}{\sigma^2} \frac{1}{T^3} \sum_{t=1}^T \frac{\partial \varepsilon_t}{\partial \eta_1} \frac{\partial \varepsilon_t}{\partial \eta_k}$. Define the processes

$$S_n^*(v_1) = \sum_{j=1}^n \cos(jv_1) \varepsilon_j, \quad T_n^*(v_1) = \sum_{j=1}^n \sin(jv_1) \varepsilon_j, \quad \text{and} \quad X_n^* = \sum_{j=1}^n Y_j, \tag{A.35}$$

to facilitate the analysis. It is easy to verify that

$$\frac{1}{T} \sum_{t=1}^{T-1} \sin[(t+1)v_1] S_T(v_1, t/T) X_T(t/T) = o_p(1) \tag{A.36}$$

if

$$\sum_{n=1}^{T-1} \sin[(n+1)v_1] S_n^*(v_1) X_n^* = o_p(n^3). \tag{A.37}$$

The same is true for the second term in (A.33). From (A.25) $ES_n^*(v_1)^2 \leq \sigma^2 n$. From Gradshteyn and Ryzhik (1980) Equation 0.121.2, we have

$$EX_n^{*2} = E \left[\sum_{j=1}^n j \varepsilon_{t-j+1} \right]^2 = \sigma^2 \sum_{j=1}^n j^2 = \sigma^2 \frac{2n^3 + 3n^2 + n}{6} \leq \sigma^2 n^3. \tag{A.38}$$

Given, $E|S_n^*(v_1)X_n^*| \leq \{ES_n^*(v_1)^2\}^{1/2} \{EX_n^{*2}\}^{1/2}$, we see that $E|S_n^*(v_1)X_n^*|$ is $O(n^2)$. Now let $n \geq m$ and consider

$$|S_n^*(v_1)X_n^* - S_m^*(v_1)X_m^*| \leq |S_n^*(v_1)| |X_n^* - X_m^*| + |X_m^*| |S_n^*(v_1) - S_m^*(v_1)|. \tag{A.39}$$

Clearly, $E|S_n^*(v_1)|^2 \leq \sigma^2 n$, and from (A.26), $E|S_n^*(v_1) - S_m^*(v_1)|^2 \leq \sigma^2(n - m)$. From (A.38) we have, $E|X_m^*|^2 \leq \sigma^3 m^3 \leq \sigma^2 n^3$. Finally, given Y_j from (A.30),

$$\begin{aligned}
 E|X_n^* - X_m^*|^2 &= E\left(\sum_{j=m+1}^n Y_j\right)^2 \\
 &= (n - m)^2 \sum_{j=1}^m \sigma^2 + \sigma^2 \sum_{j=1}^{n-m} j^2 \leq \sigma^2(n^3 - 2n^2m + n^2m) \\
 &= \sigma^2\{n^2(n - m)\}.
 \end{aligned}
 \tag{A.40}$$

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

$$\sup_{1 \leq j \leq n} \left| \sum_{t=1}^j e^{it\theta} S_t^* X_t^* \right| = o_p(n^3)
 \tag{A.41}$$

which implies that the first term in (A.33) is $o_p(1)$. Following the same reasoning, the second term in (A.33) 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,

$$Z_{1t} = -\frac{1}{2d_1} \frac{\partial \varepsilon_{t+1}}{\partial \eta_1} = \frac{\varepsilon_t}{(1 + L)^2} = \sum_{j=1}^t (-1)^{j+1} j \varepsilon_{t-j+1}.
 \tag{A.42}$$

Define the process

$$X_T(t/T) = \begin{cases} \frac{1}{T} \frac{1}{\sqrt{T}\sigma} \sum_{j=1}^t (-1)^{j+1} j \varepsilon_{t-j+1} & \text{if } t \text{ is odd} \\ \frac{1}{T} \frac{1}{\sqrt{T}\sigma} \sum_{j=1}^t (-1)^j j \varepsilon_{t-j+1} & \text{if } t \text{ is even} \end{cases}.
 \tag{A.43}$$

Let $\omega_1 = (v_2 + \pi)$ and $\omega_2 = (v_2 - \pi)$. Noting that $T^{3/2}\sigma X_T(t/T) \cos[(t + 1)\pi] = Z_{1t}$, and defining Z_{2t} as in (A.17), we get

$$\begin{aligned}
 &\frac{2\sqrt{2} \sin(v_2)}{\sigma^2} \frac{1}{T^3} \sum_{t=1}^{T-1} Z_{1t} Z_{2t} \\
 &= \frac{1}{T} \sum_{t=1}^{T-1} (\sin[(t + 1)\omega_1] + \sin[(t + 1)\omega_2]) S_T(v_2, t/T) X_T(t/T) \\
 &\quad - \frac{1}{T} \sum_{t=1}^{T-1} (\cos[(t + 1)\omega_1] + \sin[(t + 1)\omega_2]) S_T(v_2, t/T) X_T(t/T)
 \end{aligned}
 \tag{A.44}$$

Construct the variable $S_n^*(v_2)$ as above and the auxiliary variable X_n^* as

$$X_n^* = \begin{cases} \sum_{j=1}^n (-1)^{j+1} j \varepsilon_{n-j+1} & \text{if } n \text{ is odd} \\ \sum_{j=1}^n (-1)^j j \varepsilon_{n-j+1} & \text{if } n \text{ is even.} \end{cases} \tag{A.45}$$

Using these definitions, we get

$$\frac{1}{T} \sum_{t=1}^{T-1} \sin[(t+1)\omega_1] S_T(v_2, t/T) X_T(t/T) = o_p(1) \tag{A.46}$$

if $\sum_{n=1}^{T-1} \sin[(n+1)\omega_1] S_n^*(v_2) X_n^* = o_p(n^3)$. Again, $\{ES_n^*(v_2)^2\}^{1/2} \leq \sigma\sqrt{n}$ and $E|S_n^*(v_2)X_n^*| \leq \{ES_n^*(v_2)^2\}^{1/2} \{EX_n^{*2}\}^{1/2}$. Now, if n is odd, we have

$$E(X_n^*)^2 = E\left(\sum_{j=1}^n (-1)^{j+1} j \varepsilon_{n-j+1}\right)^2 = \sigma^2 \sum_{j=1}^n j^2 \leq \sigma^2 n^3, \tag{A.47}$$

and precisely the same reasoning holds if n is even. This implies that $E|S_n^*(v_2)X_n^*|$ is $O(n^2)$. We know that

$$|S_n^*(v_2)X_n^* - S_m^*(v_2)X_m^*| \leq |S_n^*(v_2)||X_n^* - X_m^*| + |X_m^*||S_n^*(v_2) - S_m^*(v_2)|, \tag{A.48}$$

where the bounds on $|S_n^*(v_2)|$ and $|S_n^*(v_2) - S_m^*(v_2)|$ were established in (A.25) and (A.26), respectively, and the bound on $|X_m^*|$ was established in the discussion above (A.40). Now, choosing $n \geq m$ for n odd and m even, gives

$$\begin{aligned} E|X_n^* - X_m^*|^2 &= E[(n-m) \sum_{j=1}^m (-1)^{j+1} \varepsilon_j \\ &+ \sum_{j=1}^{n-m} (-1)^{j+1} j \varepsilon_{n-j+1}]^2 \leq \sigma^2 [n^2(n-m)]. \end{aligned} \tag{A.49}$$

The result holds for any values of n and m . By Theorem 2.1 of Chan and Wei (1988), the first term in (A.44) 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 $j \neq 1, j \neq k$. Define the following elements:

$$Z_{1t} = -\frac{1}{2d_1} \frac{\partial \varepsilon_{t+1}}{\partial \eta_1} = \frac{\varepsilon_t}{(1+L)^2} = \sum_{j=1}^t (-1)^{j+1} j \varepsilon_{t-j+1} \tag{A.50}$$

$$Z_{kt} = -\frac{1}{2d_k} \frac{\partial \varepsilon_{t+1}}{\partial \eta_k} = \frac{\varepsilon_t}{(1-L)^2} = \sum_{j=1}^t j \varepsilon_{t-j+1} \tag{A.51}$$

$$X_{1t}^* = \begin{cases} \sum_{j=1}^t (-1)^{j+1} j \varepsilon_{t-j+1} & \text{if } t \text{ is odd} \\ \sum_{j=1}^t (-1)^j j \varepsilon_{t-j+1} & \text{if } t \text{ is even} \end{cases}. \tag{A.52}$$

Then,

$$\frac{4d_1 d_k}{T^4} \sum_{i=1}^{T-1} Z_{1i} Z_{kt} = \frac{4d_1 d_k}{T^4} \sum_{i=1}^{T-1} \cos[(t+1)\pi] X_{1i}^* X_{kt}^* \tag{A.53}$$

where X_{kt}^* is defined similarly to X_n^* in (A.35). This allows us to apply Theorem 2.1 in Chan and Wei (1988) to show that the last expression is $o_p(1)$ if

$$\sup_{1 \leq j \leq n} \left| \sum_{i=1}^j e^{it\theta} X_{1i}^* X_{kt}^* \right| = o_p(n^4). \tag{A.54}$$

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

$$EX_{1n}^{*2} = \sigma^2 \sum_{j=1}^n j^2 \leq \sigma^2 n^3. \tag{A.55}$$

From (A.38)

$$E|X_{1n}^* X_{kn}^*| \leq \{EX_{1n}^{*2}\}^{1/2} \{EX_{kn}^{*2}\}^{1/2} \leq \sigma^2 \{n^{3/2}\} \{n^{3/2}\}. \tag{A.56}$$

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

$$|X_{1n}^* X_{kn}^* - X_{1m}^* X_{km}^*| \leq |X_{1n}^*| |X_{kn}^* - X_{km}^*| + |X_{km}^*| |X_{1n}^* - X_{1m}^*|. \tag{A.57}$$

From (A.38) and (A.55), we know that $E|X_{km}^*|^2$ and $E|X_{1n}^*|^2$ are both $O(n^3)$, while from (A.40) and (A.49), $E|X_{kn}^* - X_{km}^*|^2$ 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.54) is $o_p(n^4)$ and thus the sequence in (A.53) is $o_p(1)$. This completes the proof of the lemma.

From the lemma, replace η_i and η_j with $\eta_{i,0}$ and $\eta_{j,0}$. Theorem 3 then follows from the lemma, Theorem 2.2 in Chan and Wei (1988), and Theorem 1 in Chung (1996a). Specifically, we have

$$T^b(\hat{\eta}_j - \eta_{j,0}) = - \left[\frac{\frac{1}{T^b} \sum_{t=1}^T \varepsilon_t \frac{\partial \varepsilon_t}{\partial \eta_j}}{\frac{1}{T^{2b}} \sum_{t=1}^T \left(\frac{\partial \varepsilon_t}{\partial \eta_j}\right)^2} \right] \Bigg|_{\eta_j = \eta_{j,0}} + o_p(1), \tag{A.58}$$

where $b = 1$ if $|\eta_{j,0}| < 1$, and $b = 2$ if $|\eta_{j,0}| = 1$. The theorem is complete as this is precisely the k -factor version of Equation (A.5) in Chung (1996a).

Funding Not applicable

Availability of data and materials Available upon request

Code availability Available upon request.

Declarations

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

Ethics approval Not applicable

Consent to participate Not applicable

Consent for publication Not applicable

References

- Alomari, H.M., Ayache, A., Fradon, M., Olenko, A.: Estimation of cyclic long-memory parameters. *Scand. J. Stat.* **47**(1), 104–133 (2020)
- Andrews, D.W.: Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 399–405 (2000)
- Andrews, D.W., Sun, Y.: Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica* **72**(2), 569–614 (2004)
- Arteche, J.: Exact local Whittle estimation in long memory time series with multiple poles. *Econom. Theo.* **36**(6), 1064–1098 (2020)
- Asai, M., Peiris, S., McAleer, M., Allen, D.: Cointegrated dynamics for a generalized long memory process. *J. Time Ser. Econom.* **12**(1), 1–18 (2020)
- Ayache, A., Fradon, M., Nanayakkara, R., Olenko, A.: Asymptotic normality of simultaneous estimators of cyclic long-memory processes. *Electron. J. Stat.* **16**(1), 84–115 (2022)
- Bardet, J.-M., Bertrand, P.R.: A non-parametric estimator of the spectral density of a continuous-time Gaussian process observed at random times. *Scand. J. Stat.* **37**(3), 458–476 (2010)
- Beaumont, P.M., Smallwood, A.D.: Inference for estimators of generalized long memory processes. In: *Communications in Statistics - Simulation and Computation*, Forthcoming (2022)
- Caporale, G.M., Gil-Alana, L.: Long-run and cyclical dynamics in the US stock market. *J. Forecast.* **33**(2), 147–161 (2014)
- Caporale, G.M., Gil-Alana, L.A.: Multi-factor Gegenbauer processes and European inflation rates. *J. Econom. Integr.* 386–409 (2011)
- Cavaliere, G., Nielsen, H.B., Rahbek, A.: On the consistency of bootstrap testing for a parameter on the boundary of the parameter space. *J. Time Ser. Anal.* **38**(4), 513–534 (2017)
- Cavaliere, G., Nielsen, H.B., Pedersen, R.S., Rahbek, A.: Bootstrap inference on the boundary of the parameter space, with application to conditional volatility models. *J. Econom.* **227**(1), 241–263 (2022)
- Cavaliere, G., Rahbek, A.: A primer on bootstrap testing of hypotheses in time series models: With an application to double autoregressive models. *Econom. Theo.* **37**(1), 1–48 (2021)
- Chan, N.H., Wei, C.Z.: Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Stat.* **16**(1), 367–401 (1988)
- Cheung, Y.-W., Diebold, F.X.: On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean. *J. Econom.* **62**(2), 301–316 (1994)
- Chung, C.-F.: Estimating a generalized long memory process. *J. Econom.* **73**(1), 237–259 (1996)

- Chung, C.-F.: A generalized fractionally integrated autoregressive moving average process. *J. Time Ser. Anal.* **17**, 111–140 (1996)
- Chung, C.-F., Baillie, R.T.: Small sample bias in conditional sum-of squares estimators of fractionally integrated ARMA models. *Empir. Econ.* **18**(4), 791–806 (1993)
- Diongue, A.K., Ndongo, M.: The k-factor GARMA process with infinite variance innovations. *Commun. Stat. Simul. Comput.* **45**(2), 420–437 (2016)
- Dissanayake, G., Peiris, M.S., Proietti, T.: State space modeling of Gegenbauer processes with long memory. *Comput. Stat. Data Anal.* **100**, 115–130 (2016)
- Dissanayake, G., Peiris, M.S., Proietti, T.: Fractionally differenced Gegenbauer processes with long memory: A review. *Stat. Sci.* **33**(3), 413–426 (2018)
- Elliott, G., Rothenberg, T.J., Stock, H.: Efficient tests for an autoregressive unit root. *Econometrica* **64**(4), 813–836 (1996)
- Gil-Alana, L.A.: Testing the existence of multiple cycles in financial and economic time series. *Ann. Econom. Fin.* **8**(1), 1–20 (2007)
- Giraitis, L., Hidalgo, J., Robinson, P.M.: Gaussian estimation of parametric spectral density with unknown pole. *Ann. Stat.* **29**(4), 987–1023 (2001)
- Gradshteyn, I.S., Ryzhik, I.M.: *Tables of Integrals, Series, and Products*, 4th edn. Academic Press, New York, NY (1980)
- Granger, C.W.J.: Long memory relationships and the aggregation of dynamic models. *J. Econom.* **14**(2), 227–238 (1980)
- Granger, C.W.J.: Some properties of time series data and their use in econometric model specification. *J. Econom.* **16**(1), 121–130 (1981)
- Granger, C.W.J., Joyeux, R.: An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.* **1**(1), 15–24 (1980)
- Gray, H.L., Zhang, N., Woodward, W.A.: On generalized fractional processes. *J. Time Ser. Anal.* **10**, 233–257 (1989)
- Gray, H.L., Zhang, N.-F., Woodward, W.A.: On generalized fractional processes: A correction. *J. Time Ser. Anal.* **15**(5), 561–562 (1994)
- Hassler, U.: *Time Series Analysis with Long Memory in View*. John Wiley & Sons, Hoboken (2018)
- Hidalgo, J.: Semiparametric estimation for stationary processes whose spectra have an unknown pole. *Ann. Stat.* **33**(4), 1843–1889 (2005)
- Hidalgo, J., Soulier, P.: Estimation of the location and exponent of the spectral singularity of a long memory process. *J. Time Ser. Anal.* **25**(1), 55–81 (2004)
- Hosking, J.R.M.: Fractional differencing. *Biometrika* **68**, 165–76 (1981)
- Hunt, R., Peiris, S., Weber, N.: Estimation methods for stationary Gegenbauer processes. *Stat. Pap.* **63**, 1707–1741 (2022)
- Kouamé, E.F., Hili, O.: Minimum distance estimation of k-factors GARMA processes. *Stat. Prob. Lett.* **78**(18), 3254–3261 (2008)
- Kouamé, E.F., Hili, O.: A new time domain estimation of k-factors GARMA processes. *Comptes Rendus Mathématique* **350**(19), 925–928 (2012)
- Leschinski, C., Sibbertsen, P.: Model order selection in periodic long memory models. *Econom. Stat.* **9**, 78–94 (2019)
- Lu, Z., Guegan, D.: Estimation of time-varying long memory parameter using wavelet method. *Commun. Stat. Simul. Comput.* **40**(4), 596–613 (2011)
- Maddanu, F., Proietti, T.: Modelling persistent cycles in solar activity. *Sol. Phys.* **297**(1), 1–22 (2022)
- McElroy, T.S., Holan, S.H.: Computation of the autocovariances for time series with multiple long-range persistencies. *Comput. Stat. Data Anal.* **101**, 44–56 (2016)
- Peiris, M., Asai, M.: Generalized fractional processes with long memory and time dependent volatility revisited. *Econometrics* **4**(3), 37 (2016)
- Ramachandran, R., Beaumont, P.: Robust estimation of GARMA model parameters with an application to cointegration among interest rates of industrialized countries. *Comput. Econom.* **17**(2/3), 179–201 (2001)
- Robinson, P.M.: Conditional-sum-of-squares estimation of models for stationary time series with long memory. *Lect. Notes-Monogr. Ser.* **52**, 130–137 (2006)
- Rudin, W., et al.: *Principles of Mathematical Analysis*, vol. 3. McGraw-Hill, New York (1976)
- Smallwood, A.D., Norrbin, S.C.: Generalized long memory processes, failure of cointegration tests and exchange rate dynamics. *J. Appl. Econom.* **21**(4), 409–417 (2006)

Woodward, W.A., Cheng, Q.C., Gray, H.L.: A k -factor GARMA longmemory model. *J. Time Ser. Anal.* **19**, 485–504 (1998)

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

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.