Modeling bivariate long-range dependence with general phase*†

Stefanos Kechagias
SAS Institute

Vladas Pipiras
University of North Carolina

August 9, 2019

Abstract

Bivariate time series models are considered that are suitable for estimation, that have interpretable parameters and that can capture the general semi-parametric formulation of bivariate long-range dependence, including a general phase. The models also allow for short-range dependence and fractional cointegration. A simulation study to test the performances of a conditional maximum likelihood estimation method is carried out, under the proposed models. Finally, an application is presented to the U.S. inflation rates in goods and services where models not allowing for general phase suffer from misspecification.

1 Introduction

In this work, we are interested in modeling bivariate ($\mathbb{R}^2$–vector) stationary time series exhibiting long-range dependence (LRD, in short). In the univariate case, stationary long-range dependent (LRD) time series models have the autocovariance function decaying slowly like a power-law function at large lags, or the spectral density diverging like a power-law function at the zero frequency. The univariate LRD is understood well in theory and used widely in applications. See, for example, Park and Willinger (2000), Robinson (2003), Doukhan et al. (2003), Palma (2007), Giraitis et al. (2012), Beran et al. (2013), Pipiras and Taqqu (2017).

Bivariate and, more generally, multivariate (vector-valued) LRD time series models have also been considered by a number of researchers. But theoretical foundations for a general class of such models were laid only recently in Kechagias and Pipiras (2015). In particular, Kechagias and Pipiras (2015) stressed the importance of the so-called phase parameters. Turning to the bivariate case which is the focus of this work, the phase ϕ appears in the cross-spectrum of a bivariate LRD series around the zero frequency and controls the (a)symmetry of the series at large time lags $\pm\infty$. More specifically, the matrix-valued spectral density function[1] $f(\lambda)$ of a bivariate LRD series $\{X_n\}_{n\in\mathbb{Z}}=\{(X_{1,n}, X_{2,n})'\}_{n\in\mathbb{Z}}$ satisfies

$$f(\lambda) \sim \begin{pmatrix} \omega_{11}|\lambda|^{-2d_1} & \omega_{12}|\lambda|^{-(d_1+d_2)}e^{-i\mathrm{sign}(\lambda)\phi} \\ \omega_{12}|\lambda|^{-(d_1+d_2)}e^{i\mathrm{sign}(\lambda)\phi} & \omega_{22}|\lambda|^{-2d_2} \end{pmatrix}, \quad \text{as } \lambda \to 0, \tag{1.1}$$

*AMS subject classification. Primary: 62M10, 62M15. Secondary: 60G22, 42A16.

†Keywords and phrases: long-range dependence, bivariate time series, phase parameter, estimation, VARFIMA.

[1]The following convention is used here. The autocovariance function γ is defined as $\gamma(n) = \mathbb{E}X_nX_0'$ and the spectral density $f(\lambda)$ satisfies $\gamma(n) = \int_{-\pi}^{\pi} e^{in\lambda}f(\lambda)d\lambda$. The convention is different from Kechagias and Pipiras (2015), where $\mathbb{E}X_0X_n'$ is used as the autocovariance function, but is the same as in Brockwell and Davis (2009), Pipiras and Taqqu (2017).

where $\sim$ indicates the asymptotic equivalence, $\omega_{11} > 0$, $\omega_{22} > 0$, $\omega_{12} \in \mathbb{R}$ and $\phi \in (-\pi/2, \pi/2)$. The parameter $d_j \in (0, 1/2)$ is the LRD parameter of the component series $\{X_{j,n}\}_{n\in\mathbb{Z}}$, $j = 1, 2$. The case $d_j = 0$ is associated with short-range dependence or, for short, SRD (and $d_j < 0$ with the so-called anti-persistence), and it will also be included in the discussion below.

There are currently no parametric time series models of bivariate LRD with a general phase that can be used in estimation and applications.[2] To emphasize this perhaps surprising point, consider, for example, a commonly used bivariate LRD model known as VARFIMA$(0, D, 0)$ (Vector Autoregressive Fractionally Integrated Moving Average), defined as a natural extension of the univariate ARFIMA$(0, d, 0)$ model by fractionally integrating the component series of a bivariate white noise series, namely,

$$X_n = \begin{pmatrix} X_{1,n} \\ X_{2,n} \end{pmatrix} = \begin{pmatrix} (I-B)^{-d_1} & 0 \\ 0 & (I-B)^{-d_2} \end{pmatrix} \begin{pmatrix} \eta_{1,n} \\ \eta_{2,n} \end{pmatrix} = (I-B)^{-D}\eta_n = (I-B)^{-D}Q\epsilon_n, \quad (1.2)$$

where B is the backshift operator, $I = B^0$ is the identity operator, $D = \mathrm{diag}(d_1, d_2)$ and both $\{\eta_n\}_{n\in\mathbb{Z}} = \{(\eta_{1,n}, \eta_{2,n})'\}_{n\in\mathbb{Z}}$ and $\{\epsilon_n\}_{n\in\mathbb{Z}}$ are bivariate white noise series with zero means $\mathbb{E}\eta_n = 0$, $\mathbb{E}\epsilon_n = 0$, and covariances $\mathbb{E}\eta_n\eta_n' = \Sigma = QQ'$, $\mathbb{E}\epsilon_n\epsilon_n' = I_2$. Throughout the paper, the prime indicates the transpose and covariance matrices of error terms such as Σ are assumed to be positive definite. It turns out that the model (1.2) has necessarily a special phase

$$\phi = (d_1 - d_2)\pi/2. \quad (1.3)$$

That is, even if these models appear quite general, they are in fact very special from the phase perspective. The main goal of this work is to study a class of parametric models, that allows for a general phase, and to examine it through a simulation study and an application to real data, which suggest the need for models beyond the special phase (1.3). For the latter point, in particular, a phase different from (1.3) is suggested by a local Whittle analysis of Robinson (2008) (see also Baek et al. (2019)) and the fitted models of this work will match the local Whittle phase very closely.

To achieve the described goal, we follow Kechagias and Pipiras (2015) who constructed a two-sided VARFIMA$(0, D, 0)$ model with a general phase by taking

$$X_n = (I-B)^{-D}Q_+\epsilon_n + (I-B^{-1})^{-D}Q_-\epsilon_n, \quad (1.4)$$

where Q_+, Q_- are two real-valued invertible 2×2 matrices, and $\{\epsilon_n\}_{n\in\mathbb{Z}}$ is as in (1.2). The reason we refer to (1.4) as two-sided is the presence of B^{-1} in the second term of the right-hand side of (1.4), which translates into having the leads of the innovation process ϵ_n in a linear representation of X_n. Furthermore, as argued in Kechagias and Pipiras (2015), the two-sided fractional filters as in (1.4) are critical – one-sided filters as in (1.2) would necessarily give a special phase (1.3). The matrices Q_+, Q_- are responsible precisely for capturing bivariate (a)symmetry of the series at large lags $\pm\infty$, and thus lead to a general phase.

There are, however, several issues in using the model (1.4) for estimation. For example, it does not seem to be easy to determine or to use identifiability conditions for the model (1.4) in terms of D, Q_+ and Q_- – see Section 9.3 in Pipiras and Taqqu (2017) for parallel questions in the context of vector fractional Brownian motions. Another issue is that, in analogy with the univariate ARFIMA$(0, d, 0)$ model, we would like a parametric form of a bivariate LRD model to be consistent with the semi-parametric formulation (1.1). In particular, it is then reasonable for the sought model to have the same number of parameters as in (1.1), namely, 6 ($d_1, d_2, \omega_{11}, \omega_{12}, \omega_{22}$ and

[2]The work of Amblard and Coeurjolly (2011) is perhaps an exception but it uses increments of vector fractional Brownian motion which assume the scaling behavior across all frequencies.

ϕ). In this regard, note that even after an identification up to Q_+Q_+' and Q_-Q_-', the model (1.4) has 8 parameters (d_1, d_2 and the 6 different entries of Q_+Q_+' and Q_-Q_-'). In other words, there are multiple choices of Q_+Q_+' and Q_-Q_-' that lead to the same semi-parametric formulation (1.1). We thus seek a reduced parametrization for Q_+ and Q_- that leads to model with 6 parameters, and still yields the general semi-parametric form (1.1).

We will show here that the aforementioned goal can be achieved by taking Q_- as

$$Q_- = \begin{pmatrix} c & 0 \\ 0 & -c \end{pmatrix} Q_+ =: CQ_+, \tag{1.5}$$

for some real constant c. Note that model (1.4)–(1.5) indeed has 6 parameters (d_1, d_2, c and the three different entries of $\Sigma = Q_+Q_+'$). Under the relation (1.5) and letting $\{Z_n\}_{n\in\mathbb{Z}}$ be a zero mean bivariate white noise series with covariance matrix $\mathbb{E}Z_nZ_n' = Q_+Q_+' =: \Sigma$, we can rewrite (1.4) in the more succinct form

$$X_n = \Delta_c(B)^{-1}Z_n, \tag{1.6}$$

where the operator $\Delta_c(B)^{-1}$ is defined as

$$\Delta_c(B)^{-1} = (I-B)^{-D} + (I-B^{-1})^{-D}C. \tag{1.7}$$

We shall refer to (1.6) as a general phase VARFIMA$(0, D, 0)$ model or two-sided VARFIMA$(0, D, 0)$ model. Explicit formulas relating the parameters of the two-sided VARFIMA$(0, D, 0)$ model and the semi-parametric formulation (1.1) will be given below. As with the two-sided model (1.4), we emphasize again that the parameter c (or matrix C) plays a natural role of controlling the (a)symmetry. There might be other parametrizations that achieve this goal but not the one-sided representation (1.2).

The preceding description of the model (1.6), in fact, omits several important points that will be discussed in detail below. In particular, just touching upon these points, we shall distinguish between the ranges $|c| < 1$ and $|c| > 1$, leading to two competing models when $d_1, d_2 \in (0, 1/2)$. For the purposes of capturing the semi-parametric form (1.1), either of the two ranges is sufficient and can be used alone – this should not be too surprising since by symmetry, one could in principle use the filter (1.7) with C at $(I-B)^{-D}$ instead of at $(I-B^{-1})^{-D}$ and then the two filters could be related through the reciprocal relation $c \mapsto c^{-1}$ (see Section 2 for details). In addition to the symmetry, we shall use the two ranges for the following two reasons. The first reason is numerical: the numerical optimization over c should be aware of the competing models that may be selected since likelihoods have local extrema around their parameter values. The second reason is that the model (1.6) with $c \in \mathbb{R}$ will allow for one of the series to be SRD (with the corresponding $d = 0$) and for the bivariate model still to have a general phase. In the latter case, as argued below, the whole range of real c's is needed to have a general phase.

The model (1.6)–(1.7) will also be extended in the following two natural directions. In one direction, we shall include autoregressive and moving average parts, namely, consider a general phase VARFIMA(p, D, q) model

$$\Phi(B)X_n = \Delta_c(B)^{-1}\Theta(B)Z_n, \tag{1.8}$$

where $\Phi(B)$, $\Theta(B)$ are auto-regressive and moving average matrix polynomials of finite orders p and q satisfying the usual stability and invertibility conditions. In the other direction, we shall allow explicitly for a possibility of fractional cointegration. Fractional cointegration occurs when a linear combination of the two component series has a shorter memory (smaller d) than the individual component series. From a parametric standpoint, fractional cointegration is included in a model

by replacing X_n by AX_n where $A = [1 \ -\alpha; 0 \ 1]$ is a 2×2 matrix and its first row plays the role of a possible cointegrating vector reducing the memory.

In estimation of all our models, we shall use a possibly conditional Gaussian likelihood approach, where the conditional part refers to writing the likelihood for $\Phi(B)X_n$ (or for $\Phi(B)AX_n$ in the case of cointegration). The likelihood approach will naturally require calculating the autocovariance functions of some of the introduced models – explicit formulas for these will be given below. In the $c = 0$ case, Sowell (1986) was the first to calculate (numerically) the autocovariance function of the model (1.8) and perform exact likelihood estimation. Several estimation approaches (all in the case $c = 0$), have been suggested since then. Ravishanker and Ray (1997) used Bayesian methods to estimate parameters of VARFIMA models (see a recent corrigendum by Doppelt and O'Harra (2019) discussing errors of that work). For other approaches (all in the case $c = 0$), see Pai and Ravishanker (2009*a*, 2009*b*) who employed the EM and preconditioned conjugate gradient algorithms, as well as Dueker and Starz (1998), Martin and Wilkins (1999), Sela and Hurvich (2009), Sela (2010), Diongue (2010), Tsay (2010) and Doppelt and O'Harra (2018).

We should also emphasize that the analysis of this work is limited to the second-order properties of time series (that is, autocovariance, cross-correlation, cross-spectrum, and so on). Thus, although the models (1.4) and (1.8) are expressed through two-sided and hence non-causal linear representations, their non-causal nature is irrelevant to the extent that these models are used only as suitable parameterizations of bivariate long-range dependent models allowing for general phase through their second-order properties. Somewhat surprisingly perhaps, explicit one-sided representations for these two-sided models remain unknown; see Kechagias and Pipiras (2015).

The rest of the paper is structured as follows. General phase VARFIMA$(0, D, 0)$ and VARFIMA(p, D, q) series are presented in Sections 2 and 3, including the possibility of fractional cointegration. Estimation and other tasks are considered in Section 4. Section 5 contains a simulation study, and Section 6 contains an application to the U.S. inflation rates. Conclusions can be found in Section 7.

2 General phase VARFIMA$(0, D, 0)$ series

In this section, we consider the two-sided bivariate VARFIMA$(0, D, 0)$ model (1.6)–(1.7) and denote the corresponding series in this section as $X_n^{(c)}$. The following proposition and subsequent corollaries and discussion are key. The notation c_- and c_+ below should not be confused with the negative and positive parts of c.

Proposition 2.1 *Let $d_1, d_2 \in (0, 1/2)$ and Q_+ be an invertible 2×2 matrix with real-valued entries. For any $\phi \in (-\pi/2, \pi/2)$, there exist unique constants $c = c_-$ with $|c| < 1$ and $c = c_+$ with $|c| > 1$ such that the series $\{X_n^{(c)}\}_{n\in\mathbb{Z}}$ defined by (1.6)–(1.7) has the phase parameter ϕ in (1.1). Moreover, the constants $c = c_\pm$ have closed form given by*

$$c_\pm = c_\pm(\phi) = \frac{2(a_1 + a_2) \pm \sqrt{\Delta}}{2(a_1 - a_2 + \tan(\phi)(1 + a_1a_2))}, \tag{2.1}$$

where

$$a_1 = \tan\left(\frac{\pi d_1}{2}\right), \ a_2 = \tan\left(\frac{\pi d_2}{2}\right) \quad \text{and} \quad \Delta = 16a_1a_2 + 4(1 + a_1a_2)^2 \tan^2(\phi). \tag{2.2}$$

When $\phi = \arctan((a_2 - a_1)/(1 + a_1a_2))$, the expressions in (2.1) should be interpreted as $(a_1 - a_2)/(a_1 + a_2)$ for c_- and as ∞ for c_+, and make the functions $c_-(\phi)$ and $1/c_+(\phi)$ continuous at this point.

PROOF: By using Theorem 11.8.3 in Brockwell and Davis (2009), the VARFIMA$(0, D, 0)$ series in (1.6)–(1.7) has a spectral density matrix

$$f(\lambda) = \frac{1}{2\pi}\Delta_c(e^{-i\lambda})^{-1}\Sigma\Delta_c(e^{-i\lambda})^{-1^*}, \tag{2.3}$$

where the superscript $*$ denotes the complex conjugate operation. From (1.5) and by using the fact that $1 - e^{\pm i\lambda} \sim \mp i\lambda$, as $\lambda \to 0$, we have

$$f(\lambda) \sim \frac{1}{2\pi}\big((i\lambda)^{-D} + (-i\lambda)^{-D}C\big)\Sigma\big((-i\lambda)^{-D} + C(i\lambda)^{-D}\big), \quad \text{as} \quad \lambda \to 0. \tag{2.4}$$

Next, by denoting $\Sigma = (\sigma_{jk})_{j,k=1,2}$ and using the relation $\pm i = e^{\pm i\pi/2}$, we get that the (j, k) element of the spectral density $f(\lambda)$ satisfies

$$f_{jk}(\lambda) \sim g_{jk}\lambda^{-(d_j+d_k)}, \quad \text{as} \quad \lambda \to 0^+, \tag{2.5}$$

where the complex constant g_{jk} is given by

$$g_{jk} = \frac{\sigma_{jk}}{2\pi}(e^{-i\pi d_j/2} + (-1)^{j+1}ce^{i\pi d_j/2}) \cdot (e^{i\pi d_k/2} + (-1)^{k+1}ce^{-i\pi d_k/2}) \tag{2.6}$$

and $(-1)^{j+1}$, $(-1)^{k+1}$ in (2.6) account for the different signs next to c's in the diagonal matrix C in (1.5). Focusing on the $(1, 2)$ element, and by applying the polar-coordinate representation $z = \frac{z_1}{\cos(\phi)}e^{i\phi}$ of $z = z_1 + iz_2 \in \mathbb{C}$ with $\phi = \arctan(z_2/z_1)$ to the two multiplication terms below separately, we have

$$\begin{aligned} g_{12} &= \frac{\sigma_{12}}{2\pi}\left(\cos(\frac{\pi d_1}{2})(1+c) + i\sin(\frac{\pi d_1}{2})(c-1)\right) \cdot \left(\cos(\frac{\pi d_2}{2})(1-c) + i\sin(\frac{\pi d_2}{2})(1+c)\right) \\ &= \frac{\sigma_{12}}{2\pi}\frac{\cos(\frac{\pi d_1}{2})\cos(\frac{\pi d_2}{2})}{\cos(\phi_1)\cos(\phi_2)}(1-c^2)e^{-i\phi}, \end{aligned} \tag{2.7}$$

where

$$\phi = -(\phi_1 + \phi_2), \quad \phi_1 = \arctan\left(a_1\frac{c-1}{1+c}\right) \quad \text{and} \quad \phi_2 = \arctan\left(a_2\frac{1+c}{1-c}\right) \tag{2.8}$$

with a_1 and a_2 given in (2.2).

By using the arctangent addition formula $\arctan(u) + \arctan(v) = \arctan(\frac{u+v}{1-uv})$ for $uv < 1$ (in our case $uv = -a_1a_2 < 0$), we can rewrite ϕ as

$$\phi = -\arctan\left(\frac{a_1\frac{c-1}{1+c} + a_2\frac{1+c}{1-c}}{1 + a_1a_2}\right) =: h(c). \tag{2.9}$$

For all $d_1, d_2 \in (0, 1/2)$, the function $h : (-1, 1) \to (-\pi/2, \pi/2)$ is strictly decreasing (and therefore 1-1) and also satisfies

$$\lim_{c\downarrow -1} h(c) = \frac{\pi}{2}, \quad \lim_{c\uparrow 1} h(c) = -\frac{\pi}{2}.$$

Since h is continuous, it is also onto its range which completes the existence and uniqueness part of the proof for $c = c_- \in (-1, 1)$. For the statement about $c = c_+$, note that $h(1/c) = -h(c)$ so that the same conclusion also holds.

To obtain the formula (2.1), we invert the relation (2.9) to get the quadratic equation

$$(a_1 - a_2 + \tan(\phi)(1 + a_1a_2))c^2 - 2(a_1 + a_2)c + a_1 - a_2 - \tan(\phi)(1 + a_1a_2) = 0, \tag{2.10}$$

whose discriminant Δ is given by

$$\Delta = 16a_1a_2 + 4(1 + a_1a_2)^2 \tan^2(\phi)$$

and is always positive. The solutions of (2.10) are then given by

$$c_+ = \frac{2(a_1 + a_2) + \sqrt{\Delta}}{2(a_1 - a_2 + \tan(\phi)(1 + a_1a_2))}, \quad c_- = \frac{2(a_1 + a_2) - \sqrt{\Delta}}{2(a_1 - a_2 + \tan(\phi)(1 + a_1a_2))}.$$

It can be checked that $|c_+| > 1$ and $|c_-| < 1$. Finally, the statements about the continuity can be checked easily. $\square$

Proposition 2.1 and its proof have a number of important consequences presented next.

Corollary 2.1 *The spectral density of the time series $\{X_n^{(c)}\}_{n\in\mathbb{Z}}$ with $c = c_\pm(\phi)$ in Proposition 2.1 satisfies the asymptotic relation* (1.1) *with ϕ and*

$$\omega_{jj} = \frac{\sigma_{jj}}{2\pi}(1 + c^2 + (-1)^{j+1}2c\cos(\pi d_j)), \quad j = 1, 2, \tag{2.11}$$

$$\omega_{12} = \frac{\sigma_{12}}{2\pi}\frac{\cos(\frac{\pi d_1}{2})\cos(\frac{\pi d_2}{2})}{\cos(\phi_1)\cos(\phi_2)}(1 - c^2), \tag{2.12}$$

where $\Sigma = Q_+Q_+' = (\sigma_{jk})_{j,k=1,2}$ and ϕ_1, ϕ_2 are given in (2.8).

Proof: The relations (2.11)–(2.12) follow from (2.5)–(2.6) and (2.7)–(2.8). $\square$

Corollary 2.1 shows that the bivariate LRD model (1.6)–(1.7) when parametrized by $d_1, d_2, \Sigma = Q_+Q_+'$ and either $c = c_-$ or $c = c_+$, can be identified with the parameters $d_1, d_2, w_{11}, w_{22}, w_{12}$ and ϕ in the semi-parametric formulation (1.1). The following observations will contrast the models with c_- and c_+, and provide additional insights into the models with $|c| < 1$ and $|c| > 1$. The next results shows that the models (1.6)–(1.7) with c_- and c_+ are different, even if they both capture the same formulation (1.1). We continue using the notation of Proposition 2.1.

Corollary 2.2 *The series $\{X_n^{(c)}\}_{n\in\mathbb{Z}}$ with $c = c_-(\phi)$ and $\{X_n^{(c)}\}_{n\in\mathbb{Z}}$ with $c = c_+(\phi)$ satisfying (2.11)–(2.12) are different, in the sense that their individual component and cross spectral density functions are not the same.*

Proof: For example, by arguing as in the proof of Proposition 2.1, the spectral density of the first component of the series $X_n^{(c)}$ is given by

$$f_{11}(\lambda) = \frac{\sigma_{11}}{2\pi}\left|(1 - e^{-i\lambda})^{-d_1} + c(1 - e^{i\lambda})^{-d_1}\right|^2 = \frac{\sigma_{11}}{2\pi}|1 - e^{i\lambda}|^{2d_1}\left|(1 - e^{i\lambda})^{d_1}(1 - e^{-i\lambda})^{-d_1} + c\right|^2$$

$$= \frac{\sigma_{11}}{2\pi}|1 - e^{i\lambda}|^{2d_1}\left|e^{i(\lambda+\pi)d_1} + c\right|^2 = \frac{\sigma_{11}}{2\pi}\Big(2\sin\frac{\lambda}{2}\Big)^{-2d_1}\Big(1 + c^2 + 2c\cos((\lambda + \pi)d_1)\Big). \tag{2.13}$$

Let the superscripts $\pm$ indicate the quantities and parameters associated with the model having $c = c_\pm$. By (2.11),

$$\sigma_{11}^- = \sigma_{11}^+\frac{1 + c_+^2 + 2c_+\cos(\pi d_1)}{1 + c_-^2 + 2c_-\cos(\pi d_1)}$$

and hence, by (2.13),

$$f_{11}^-(\lambda) = f_{11}^+(\lambda)r_{11}(\lambda),$$

where

$$r_{11}(\lambda) = \frac{1 + c_-^2 + 2c_- \cos((\lambda + \pi)d_1)}{1 + c_+^2 + 2c_+ \cos((\lambda + \pi)d_1)} \frac{1 + c_+^2 + 2c_+ \cos(\pi d_1)}{1 + c_-^2 + 2c_- \cos(\pi d_1)}.$$

Though $r_{11}(\lambda) \sim 1$ as $\lambda \to 0$, $r_{11}(\lambda)$ is not equal to 1 across all λ, proving the statement for component spectral densities. A similar argument can be made for the cross spectral densities, which we omit for the shortness sake. $\square$

Inclusion of c_+ for the parameter c in the model (1.6)–(1.7) seems redundant in view of the discussion and the results above. We include c_+ for several reasons. One reason is that this inclusion naturally provides symmetry in the considered model in the following sense. Note that the model (1.6)–(1.7) can also be expressed as

$$X_n^{(c)} = \Delta_{c^{-1}}^+(B)^{-1} \widetilde{Z}_n, \tag{2.14}$$

where

$$\Delta_{\widetilde{c}}^+(B)^{-1} = (I - B)^{-D} \widetilde{C} + (I - B^{-1})^{-D}, \quad \widetilde{C} = \begin{pmatrix} \widetilde{c} & 0 \\ 0 & -\widetilde{c} \end{pmatrix} \tag{2.15}$$

and $\{\widetilde{Z}_n\}_{n \in \mathbb{Z}}$ is a zero mean bivariate white noise series with covariance matrix $\mathbb{E}\widetilde{Z}_n \widetilde{Z}_n' = C\Sigma C$. In other words, the model (1.6)–(1.7) with $c = c_+$ and $|c_+| > 1$ corresponds to the model (2.14)–(2.15) with $\widetilde{c} = 1/c_+$ and $|\widetilde{c}| < 1$, that is, the model symmetric to (1.6)–(1.7) where C with $|c| < 1$ is moved from $(I - B^{-1})^{-D}$ to $(I - B)^{-D}$. Though the models appear symmetric and can be used to capture the same semi-parametric form of bivariate LRD, Corollary 2.2 shows that they are not the same. Other reasons for including $c = c_+$ will appear below.

Further insight on the differences between the models with c_- and c_+, and between the models with $|c| < 1$ and $|c| > 1$ can be given through the following observations. Since c_- and c_+ are the roots of the quadratic equation (2.10), note that

$$c_- c_+ = \frac{a_1 - a_2 - \tan(\phi)(1 + a_1 a_2)}{a_1 - a_2 + \tan(\phi)(1 + a_1 a_2)}.$$

For example, when $d_1 = d_2$, this relation reduces to

$$c_- = -\frac{1}{c_+}.$$

We stress again that $X_n^{(c_-)}$ and $X_n^{(c_+)}$ are two competing models for the same formulation of a bivariate LRD. In addition, the models $X_n^{(c)}$ and $X_n^{(1/c)}$ can also be considered closely related in the following sense. As in (2.13), the spectral density of the first component series of $X_n^{(c)}$ is

$$f_{11}(\lambda) = \frac{\sigma_{11}}{2\pi} \left| (1 - e^{-i\lambda})^{-d_1} + c(1 - e^{i\lambda})^{-d_1} \right|^2 = \frac{\sigma_{11} c^2}{2\pi} \left| (1 - e^{-i\lambda})^{-d_1} + c^{-1}(1 - e^{i\lambda})^{-d_1} \right|^2 \tag{2.16}$$

and thus also that of $X_n^{(1/c)}$ with the proper choice of σ_{11}. In other words, the component series of $X_n^{(c)}$ and $X_n^{(1/c)}$ have identical second-order properties after suitable normalization. The cross spectral density of $X_n^{(c)}$ and $X_n^{(1/c)}$ though are obviously different. In fact, as seen from the observations following (2.9) and, in particular, the fact that $h(c) = -h(1/c)$, the models $X_n^{(c)}$ and $X_n^{(1/c)}$ have the same phases which are of different signs. These observation have important practical consequences. If the cross spectral density is relatively weak (that is, $\sigma_{12}^2/(\sigma_{11}\sigma_{22})$ is small), it is natural to expect that the models $X_n^{(c)}$ and $X_n^{(1/c)}$ will be hard to distinguish. This is what we also

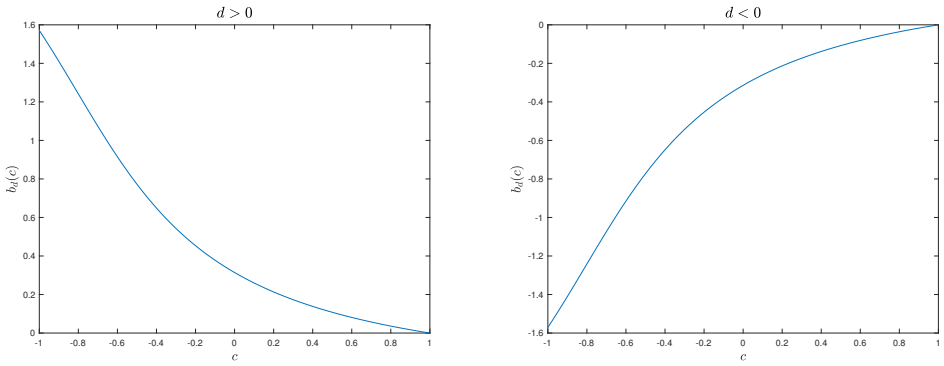

Figure 1: *The function* $b_d(c)$, $c \in (-1, 1)$, *in (2.17). Left:* $d > 0$. *Right:* $d < 0$.

see in estimation, where numerical optimization for c might converge to a local extremum in the neighborhood of the reciprocal of the chosen true value c.

The model (1.6)–(1.7) will be referred to as the *general phase VARFIMA*$(0, D, 0)$ *series* (*two-sided VARFIMA*$(0, D, 0)$ *series*).

Remark 2.1 Technical issues of Proposition 2.1 aside, there is a simple way to see why the proposed model will yield a general phase. Consider only the case $c \in (-1, 1)$. Note that a generic term $e^{-i\pi d/2} + ce^{i\pi d/2}$ entering (2.6) can be expressed in polar coordinates as

$$e^{-i\pi d/2} + ce^{i\pi d/2} = a_d(c)e^{-ib_d(c)}. \tag{2.17}$$

The generic shape of the function $b_d(c)$, $c \in (-1, 1)$, is given in Figure 1, left plot, for $d \in (0, 1/2)$, and right plot, for $d \in (-1/2, 0)$. When $d \in (0, 1/2)$, the range of $b_d(c)$, $c \in (-1, 1)$, is $(0, \pi/2)$ and when $d \in (-1/2, 0)$, it is $(-\pi/2, 0)$. When combined into the phase ϕ of (2.7), this obviously leads to the phase ϕ that covers the whole range $(-\pi/2, \pi/2)$. This discussion also shows that, for example, the choice $Q_- = cQ_+$, $c \in (-1, 1)$, would not lead to a general phase parameter for the resulting bivariate LRD models.

Remark 2.2 The case $c = 0$ corresponds to the phase $\phi = (d_1 - d_2)\pi/2$ (and in particular not necessarily $\phi = 0$.) If the two component series are interchanged (so that d_1 and d_2 are interchanged, and ϕ becomes $-\phi$), then the constant $c = c_\pm$ in (2.1) changes to $-c = -c_\pm$. We also note that the relation (2.1) does not involve the covariance matrix Σ of the innovation terms, but that (2.11) and (2.12) obviously do.

Remark 2.3 We also note the following important point regarding e.g. the boundaries $c_- = \pm 1$ of the range $c_- \in (-1, 1)$. As $c_- \to \pm 1$, the phase parameter $\phi \to \mp\pi/2$. In the specification $\omega_{12}e^{-i\phi}$ of the cross-spectrum constant, note that the cases $\phi = \pi/2$ and $\phi = -\pi/2$ are equivalent by changing the sign of ω_{12}, since $\omega_{12}e^{-i\pi/2} = (-\omega_{12})e^{-i(-\pi/2)}$. In the model (1.6)–(1.7), the sign of ω_{12} is the same as the sign of σ_{12}. From a practical perpective, this observation means that for the model (1.6)–(1.7) with c_- close to 1 (-1, resp.), it would be common to estimate c_- close to -1 (1, resp.) and σ_{12} with the opposite sign, since the respective models are not that different. This is also certainly what we observed in our simulations.

We conclude this section by arguing that the model (1.6)–(1.7) can also be used to capture general phase when one of the component series is SRD (with the corresponding $d = 0$) and

the other is LRD. (The component series can also be both SRD but this case is not particularly interesting since for two SRD series, their spectral density at zero is the sum of autocovariances at all lags and hence has zero phase $\phi = 0$.) We formulate this as another proposition analogous to Proposition 2.1. For the shortness sake, we suppose that $d_1 = 0$ and $d_2 \in (0, 1/2)$, but a similar result could be formulated when $d_2 = 0$ and $d_1 \in (0, 1/2)$.

Proposition 2.2 *Let $d_1 = 0, d_2 \in (0, 1/2)$ and Q_+ be an invertible 2×2 matrix with real-valued entries. For any $\phi \in (-\pi/2, \pi/2)$, there exists a unique constant $c \in \mathbb{R}$ such that the series $\{X_n^{(c)}\}_{n\in\mathbb{Z}}$ defined by (1.6)–(1.7) has the phase parameter ϕ in (1.1). Moreover,*

$$c = \frac{\tan\phi - a_2}{\tan\phi + a_2}, \tag{2.18}$$

where a_2 is defined in (2.2).

PROOF: As in (2.7)–(2.8) of the proof of Proposition 2.1, the relationship between a phase ϕ and a parameter c is

$$\phi = -\arctan\Big(a_2\frac{1+c}{1-c}\Big).$$

Solving for c leads to the relation (2.18). □

Note the key difference between Propositions 2.1 and 2.2: while there were two c's leading to the same phase in Proposition 2.1, there is only one c in Proposition 2.2. This is another reason for considering both c_- and c_+ in a general phase VARFIMA$(0, D, 0)$ model.

As in Corollary 2.1, the following relations among the rest of the parameters hold when $d_1 = 0$:

$$\omega_{11} = \frac{\sigma_{11}}{2\pi}(1+c)^2,\ \omega_{22} = \frac{\sigma_{22}}{2\pi}(1+c^2+(-1)^{j+1}2c\cos(\pi d_2)),\ \omega_{12} = \frac{\sigma_{12}}{2\pi}\frac{\cos(\frac{\pi d_2}{2})}{\cos(\phi_2)}(1-c^2).$$

These are the same as (2.11)–(2.12) by taking $d_1 = 0$.

Remark 2.4 There are a priori no issues with fitting the model (1.6)–(1.7) when a component series is anti-persistent (that is, with the corresponding $d < 0$). But when the two d's are of the opposite signs, the model (1.6)–(1.7) will not yield a general phase. The latter fact could be seen by employing informal arguments as in Remark 2.1. That is, note that the generic term (2.17) now enters (2.6) through two multiplicative terms with the same signs of the exponents of the exponentials. The resulting phase would then be the sum of $b_{d_j}(c)$ appearing in (2.17) and $b_{-d_k}(-c)$, and would cover only part of either $(0, \pi/2)$ or $(-\pi/2, 0)$ as seen from Figure 1.

3 General phase VARFIMA(p, D, q) series

In this section, we generalize the model (1.6)–(1.7) by introducing autoregressive (AR, for short) and moving average (MA, for short) components to capture potential short-range dependence effects. We also consider models with fractional cointegration. For the one-sided model (1.2), the ARMA extension has been achieved in a number of ways. Naturally, we focus on extensions that preserve the general phase and identifiability properties. We also consider the problem of computing (theoretically or numerically) the autocovariance functions of the introduced models, since these functions are used in estimation (see Section 4 below).

3.1 VARFIMA$(0, D, q)$ series

We begin with the case $p = 0$ (where there is no AR part). Define the *general phase VARFIMA$(0, D, q)$ series* (*two-sided VARFIMA$(0, D, q)$ series*) as

$$Y_n = \Delta_c(B)^{-1}\Theta(B)Z_n, \tag{3.1}$$

where $\Delta_c(B)^{-1}$ is the operator given by (1.7) and

$$\Theta(B) = I + \Theta_1 B + \ldots + \Theta_q B^q \tag{3.2}$$

is a matrix polynomial with 2×2 real-valued matrices $\Theta_s = (\theta_{jk,s})_{j,k=1,2}$, $s = 1, \ldots, q$. Here and throughout, $\{Z_n\}_{n\in\mathbb{Z}}$ is a white noise series with $\mathbb{E}Z_n Z_n' = \Sigma = (\sigma_{jk})_{j,k=1,2}$. Further conditions on Θ_s, $s = 1, \ldots, q$, will be discussed below.

In the next proposition, we compute the autocovariance function of the series in (3.1). Tsay (2010) calculated the autocovariance function of the one-sided analogue of (3.1) using the properties of the hypergeometric function. Our approach, which we find less cumbersome for the multivariate case, is similar to the one used for the two-sided VARFIMA$(0, D, 0)$ series in Proposition 5.1 of Kechagias and Pipiras (2015).

Proposition 3.1 *The (j, k) component $\gamma_{jk}(n)$ of the autocovariance matrix function $\gamma(n)$ of the bivariate two-sided VARFIMA$(0, D, q)$ series in (3.1) is given by*

$$\gamma_{jk}(n) = \frac{1}{2\pi}\sum_{u,v=1}^{2}\sum_{s,t=0}^{q}\theta_{ju,s}\theta_{kv,t}\sigma_{uv}\Big(a_{1,jk}\gamma^{(1)}_{st,jk}(n) + a_{2,j}\gamma^{(2)}_{st,jk}(n) + \gamma^{(3)}_{st,jk}(n) + a_{4,k}\gamma^{(4)}_{st,jk}(n)\Big), \tag{3.3}$$

where $\Theta_s = (\theta_{jk,s})_{j,k=1,2,s=1,\ldots,q}$, $\Sigma = (\sigma_{jk})_{j,k=1,2}$,

$$a_{1,jk} = c^2(-1)^{j+k}, \quad a_{2,j} = c(-1)^{j+1}, \quad a_{4,k} = c(-1)^{k+1}, \tag{3.4}$$

and

$$\begin{array}{rcccl} \gamma^{(1)}_{st,jk}(n) & = & \gamma^{(3)}_{st,kj}(n) & = & 2\Gamma(1-d_j-d_k)\sin(\pi d_k)\frac{\Gamma(n+t-s+d_k)}{\Gamma(n+t-s+1-d_j)}, \\ \gamma^{(4)}_{st,jk}(n) & = & \gamma^{(2)}_{ts,jk}(-n) & = & \begin{cases} 2\pi\frac{1}{\Gamma(d_j+d_k)}\frac{\Gamma(d_j+d_k+n+t-s)}{\Gamma(1+n+t-s)} & , \quad n \geq s-t, \\ 0 & , \quad n < s-t. \end{cases} \end{array} \tag{3.5}$$

Proof: By using Theorem 11.8.3 in Brockwell and Davis (2009), the VARFIMA$(0, D, q)$ series in (3.1) has a spectral density matrix

$$f(\lambda) = \frac{1}{2\pi}G(\lambda)\Sigma G(\lambda)^*, \tag{3.6}$$

where $G(\lambda) = \Delta_c(e^{-i\lambda})^{-1}\Theta(e^{-i\lambda})$. The (j, k) component of the spectral density is given by

$$f_{jk}(\lambda) = \frac{1}{2\pi}\sum_{u,v=1}^{2}\sum_{s,t=0}^{q}\theta_{ju,s}\theta_{kv,t}\sigma_{uv}e^{-i(s-t)\lambda}(f_{1,jk}(\lambda) + f_{2,jk}(\lambda) + f_{3,jk}(\lambda) + f_{4,jk}(\lambda)), \tag{3.7}$$

where

$$\begin{array}{cc} f_{1,jk}(\lambda) = a_{1,jk}(1-e^{i\lambda})^{-d_j}(1-e^{-i\lambda})^{-d_k}, & f_{2,jk}(\lambda) = a_{2,j}(1-e^{i\lambda})^{-(d_j+d_k)}, \\ f_{3,jk}(\lambda) = (1-e^{-i\lambda})^{-d_j}(1-e^{i\lambda})^{-d_k}, & f_{4,jk}(\lambda) = a_{4,k}(1-e^{-i\lambda})^{-(d_j+d_k)}. \end{array} \tag{3.8}$$

Consequently, the (j,k) component of the autocovariance matrix satisfies $\gamma_{jk}(n) = \int_0^{2\pi} e^{in\lambda} f_{jk}(\lambda) d\lambda$, which in view of the relations (3.7)–(3.8) implies (3.3)–(3.4) with

$$\gamma^{(1)}_{st,jk}(n) = \gamma^{(3)}_{st,kj}(n) = \int_0^{2\pi} e^{i(n-s+t)\lambda}(1-e^{i\lambda})^{-d_j}(1-e^{-i\lambda})^{-d_k} d\lambda,$$

$$\gamma^{(2)}_{st,jk}(n) = \int_0^{2\pi} e^{i(n-s+t)\lambda}(1-e^{i\lambda})^{-x_{jk}} d\lambda, \quad \gamma^{(4)}_{st,jk}(n) = \int_0^{2\pi} e^{i(n-s+t)\lambda}(1-e^{-i\lambda})^{-x_{jk}} d\lambda,$$

where $x_{jk} = d_j + d_k$. The relations (3.5) follow from the evaluation of the integrals above as in the proof of Proposition 5.1 of Kechagias and Pipiras (2015). $\square$

We conclude this section with comments on parameter identification and another model specification. Since $\Theta(e^{-i\lambda}) \sim I_2 + \Theta_1 + \ldots + \Theta_q$ as $\lambda \to 0$, and since the relation (2.1) in Proposition 2.1 and the relation (2.18) in Proposition 2.2 do not involve Σ, the two-sided VARFIMA$(0,D,q)$ model has a general phase at the zero frequency (with the same relations (2.1) and (2.18) between the phase ϕ and the parameter c). This also means that the filter $\Delta_c(B)^{-1}$ can be identified at the zero frequency, assuming either $|c|<1$ or $|c|>1$ when $d_1, d_2 \in (0,1/2)$ (and without this assumption when one of the d's is zero). Under the latter assumption, the filter $\Delta_c(B)$ can be applied to the VARFIMA$(0,D,q)$ series to get a VARMA$(0,q)$ series without fractional integration. In particular, the parameters of Θ's are identifiable if and only if they are identifiable for the same VARMA$(0,q)$ model. For the latter, working with parameters associated with invertible models (e.g. Brockwell and Davis (2009)) is natural and we use them here, including in simulations. We expect the VARFIMA$(0,D,q)$ models to be different across $|c|<1$ and $|c|>1$ as well but currently do not have a proof of this conjecture.

Another way to define a VARFIMA$(0,D,q)$ series is to set $Y_n = \Theta(B)\Delta_c(B)^{-1}Z_n$. This is a different model from (3.1) since $\Theta(B)$ and $\Delta_c(B)^{-1}$ do not necessarily commute. Furthermore, by writing this model as $\Theta(B)^{-1}Y_n = \Delta_c(B)^{-1}Z_n$, it can be thought to allow for a form of fractional cointegration, in that a linear combination $\Theta(B)^{-1}Y_n$ of present and past observations of the two component series could reduce the memory. Fractional cointegration is modeled more directly in Section 3.3 below, and we shall work with a VARFIMA$(0,D,q)$ model defined by (3.1).

3.2 VARFIMA(p,D,q) series

We extend here the model (3.1) to a general phase fractionally integrated model containing both autoregressive and moving average components. As for the one-sided model (1.2), two possibilities can be considered for this extension. Let $\Phi(B) = I - \Phi_1 B - \ldots - \Phi_p B^p$ be the AR polynomial, where $\Phi_r = (\phi_{jk,r})_{j,k=1,2}$, $r=1,\ldots,p$, are 2×2 real-valued matrices. Further assumptions on Φ_r, $r=1,\ldots,p$, are discussed below. Following the terminology of Sela and Hurvich (2009), define the *general phase VARFIMA(p,D,q) series* (*two-sided VARFIMA(p,D,q) series*) $\{X_n\}_{n\in\mathbb{Z}}$ as

$$\Phi(B)X_n = \Delta_c(B)^{-1}\Theta(B)Z_n, \tag{3.9}$$

and the *general-phase FIVARMA(p,D,q) series*[3] (*two-sided FIVARMA(p,D,q) series*) as

$$\Phi(B)\Delta_c(B)X_n = \Theta(B)Z_n. \tag{3.10}$$

[3]The names VARFIMA and FIVARMA refer to the facts that the fractional integration (FI) operator $\Delta_c(B)^{-1}$ is applied to the MA part in (3.9), and after writing $X_n = \Delta_c(B)^{-1}\Phi(B)^{-1}\Theta(B)Z_n$, it is applied to the VARMA series in (3.10).

The one-sided FIVARMA(p, D, q) series (with $c = 0$ in (3.10)) have been more popular in the literature, with Lobato (1997), Sela and Hurvich (2009) and Tsay (2010) being notable exceptions. In particular, Sela and Hurvich (2009) investigated thoroughly the differences between the one-sided analogues of the models (3.9) and (3.10), focusing on models with no MA part.

We shall focus on and work with the two-sided VARFIMA(p, D, q) model (3.9). The following discussion explains this choice and also makes other important points. By using the reparametrizations of Dufour and Pelletier (2014), the FIVARMA model (3.10) can, in fact, take the VARFIMA model form (3.9) with diagonal $\Phi(B)$. Indeed, write first the model (3.10) as

$$\Delta_c(B)X_n = \Phi(B)^{-1}\Theta(B)Z_n.$$

Next, by using the relation $\Phi(B)^{-1} = |\Phi(B)|^{-1}\mathrm{adj}(\Phi(B))$, where $|\cdot|$ and $\mathrm{adj}(\cdot)$ denote the determinant and adjoint of a matrix, respectively, we can write

$$\Delta_c(B)|\Phi(B)|X_n = \mathrm{adj}(\Phi(B))\Theta(B)Z_n, \tag{3.11}$$

where the commutation of $\Delta_c(B)$ and $|\Phi(B)|$ is possible since $|\Phi(B)|$ is scalar-valued. Letting $\widetilde{\Phi}(B) = \mathrm{diag}(|\Phi(B)|)$ and $\widetilde{\Theta}(B) = \mathrm{adj}(\Phi(B))\Theta(B)$, the relation (3.11) yields

$$\Delta_c(B)\widetilde{\Phi}(B)X_n = \widetilde{\Theta}(B)Z_n. \tag{3.12}$$

Thus, a FIVARMA model with AR component of order p can indeed be written as a VARFIMA model with a diagonal AR part whose order will not exceed $2p$ (the maximum possible order of $|\Phi(B)|$).

When the VARFIMA(p, D, q) series (3.9) has a diagonal AR part, in which case the two models (3.9) and (3.10) are equivalent, the resulting model is identifiable in the following sense. As in the discussion for the VARFIMA$(0, D, q)$ model in the preceding section, such two-sided VARFIMA(p, D, q) model has a general phase, assuming either $|c| < 1$ or $|c| > 1$ for $d_1, d_2 \in (0, 1/2)$. Furthermore, its other parameters (than D and c) are similarly identifiable if and only if they are identifiable for the same VARMA(p, q) model. Conditions for the latter were studied recently in Dufour and Pelletier (2014). The latter essentially just require stable AR and invertible MA polynomials, which we also consider here, including in simulations.

We shall be using the VARFIMA(p, D, q) series with non-diagonal AR parts as well but the following caveats should be kept in mind. We currently do not have results on parameter identifiability for these models. Identifiability and constrained estimation are quite delicate even for general VARMA(p, q) models; see Dufour and Pelletier (2014), Roy et al. (2014). As with the other possible VARFIMA$(0, D, q)$ specification discussed at the end of Section 3.1, the model (3.9) with non-diagonal AR component also allows for a form of fractional cointegration, and thus it should be interpreted with caution. We generally prefer to model fractional cointegration explicitly as discussed in Section 3.3 below.

Finally, another reason for using the VARFIMA(p, D, q) series (3.9) is that the autocovariance function of its right-hand side can be computed explicitly as in Proposition 3.1; this will be useful in estimation in Section 4. Computing the autocovariance function of the series itself is complicated by the presence of the AR filter on the left-hand side of (3.9). Closed form formulas for the autocovariance function of the one-sided model (3.10) with $c = 0$ were provided by Sowell (1986), albeit his implementation is computationally inefficient as it requires multiple expensive evaluations of hypergeometric functions. The slow performance of Sowell's approach was also noted by Sela (2010), who proposed fast approximate algorithms for calculating the autocovariance functions of the one-sided models (3.9) and (3.10) with $c = 0$ when $p = 1$ and $q = 0$. Although not exact,

Sela's algorithms are fast with negligible approximation errors. In fact, it is straightforward to extend these algorithms to calculate the autocovariance function of a two-sided VARFIMA$(1, D, q)$ series. For models with AR components of higher orders, however, this extension seems to require restrictive assumptions on the AR coefficients and therefore we do not pursue this approach.

3.3 Fractionally cointegrated VARFIMA(p, D, q) series

For a bivariate LRD series $X_n = (X_{1,n}, X_{2,n})'$, fractional cointegration refers to the situation when there is a linear combination $X_{1,n} - \alpha X_{2,n}$ ($\alpha \neq 0$) having shorter memory than the individual component series. In fact, this is possible only when the component series have the same long-range dependence parameter. The possibility of fractional cointegration can be incorporated easily into a model as follows, similarly to Sela and Hurvich (2009), Robinson (2008). A *general phase VARFIMA(p, D, q) model with fractional cointegration* is defined as

$$\Phi(B)AX_n = \Delta_c(B)^{-1}\Theta(B)Z_n, \tag{3.13}$$

where

$$A = \begin{pmatrix} 1 & -\alpha \\ 0 & 1 \end{pmatrix} \tag{3.14}$$

and the rest of the model parameters are as in (3.9) but supposing $d_1 < d_2$. That is, the first row of the matrix A when applied to X_n, plays the role of a possible fractional cointegration and $d_1 < d_2$ is assumed to have shorter memory for the first component of AX_n after cointegration. The model (3.13)–(3.14) is fractionally cointegrated only when $\alpha \neq 0$. In practice, the model is fitted without the restriction $\alpha \neq 0$ and the results suggest fractional cointegration if the confidence interval for α does not include zero. We also note that some models of bivariate fractional cointegration would allow for the two component series to have different long-range dependence parameters (Johansen (2008), Johansen and Nielsen (2012)). In this case, cointegration would be "trivial" in that a cointegrating vector would select just the component series with the smaller long-range dependence parameter. The "trivial" cointegration effectively amounts to having $\alpha = 0$ in the model (3.13)–(3.14).

Note that the cointegrated model (3.13) has the same structure as the model (3.9), with $\Phi(B)$ replaced by $\Phi(B)A$. A number of subsequent developments, in particular, related to estimation and forecasting, can be adapted straightforwardly when $\Phi(B)$ is replaced by $\Phi(B)A$. For shortness sake, these will be written for $\Phi(B)$ and non-cointegrated models only.

4 Estimation and other tasks

In this section, we discuss estimation of the general phase VARFIMA(p, D, q) model (3.9) introduced in Section 3.2. (Cointegrated models can be treated similarly as noted in Section 3.3.) Knowledge of the explicit form of the autocovariance function (3.3), allows us to obtain parameter estimates via conditional maximum likelihood (CML) estimation. The major computational task in the CML computation is the Cholesky factorization of the model covariance matrix, which we found to be stable and efficient for small and moderate sample sizes.

4.1 Estimation

We start with some notation. Let $\{Y_n\}_{n=1,\ldots,N}$ be the two-sided VARFIMA$(0, D, q)$ series (3.1) and let $\Gamma(k) = \mathbb{E}Y_k Y_0'$ denote its autocovariance function. Let also $\Theta = (\mathrm{vec}(\Theta_1)', \ldots, \mathrm{vec}(\Theta_q)')'$ be

the vector containing the entries of the coefficient matrices of the MA polynomial $\Theta(B)$. Assuming that the bivariate white noise series $\{Z_n\}$ is Gaussian, and letting $\eta = (d_1, d_2, c, \sigma_{11}, \sigma_{12}, \sigma_{22}, \Theta')'$ be the $(6+4q)$–dimensional vector containing all the parameters of the model (3.1), the log-likelihood function of $\{Y_n\}_{n=1,\ldots,N}$ is given by

$$\ell(\eta|Y) = -\frac{1}{2}\log|\Omega(\eta)| - \frac{1}{2}Y'\Omega(\eta)^{-1}Y, \tag{4.1}$$

where $\Omega(\eta)$ is the $2N \times 2N$ covariance matrix of $\{Y_n\}_{n=1,\ldots,N}$. The matrix $\Omega(\eta)$ is populated by the 2×2 autocovariance matrices $\Gamma(k), k = -(N-1), \ldots, N-1$, in a Toeplitz fashion, that is, the (i,j)th 2×2 block matrix entry of $\Omega(\eta)$ is $\Gamma(j-i)$. Instead of $\Omega(\eta)$, one can also work with the covariance matrix $\widetilde{\Omega}(\eta) = \mathbb{E}(\mathbf{Y}\mathbf{Y}')$ of the stacked component series $\mathbf{Y} = \{Y_{1,1}, \ldots, Y_{1,N}, Y_{2,1}, \ldots, Y_{2,N}\}$. The latter is used in the SAS/IML numerical implementation of the likelihood.

Using the fact that the series $\{Y_n\}_{n=1,\ldots,N}$ satisfies the relation

$$\Phi(B)X_n = Y_n, \tag{4.2}$$

where $\{X_n\}_{n=1,\ldots,N}$ is the two-sided VARFIMA(p, D, q) series (3.9), we can view $\{\Phi(B)X_n\}_{n=p+1,\ldots,N}$ as a two-sided VARFIMA$(0, D, q)$ series, whose log-likelihood function conditional on $X_1, \ldots, X_p$ and $\Phi = (\mathrm{vec}(\Phi_1)', \ldots, \mathrm{vec}(\Phi_p)')'$ is given by

$$\ell(\Phi, \eta; X_n|X_1, \ldots, X_p) \equiv \ell(\eta; \Phi(B)X_n), \quad n = p+1, \ldots, N. \tag{4.3}$$

The reason we do not absorb Φ in η, is to emphasize the different roles that these two parameters have in calculating the likelihood function in (4.3). More specifically, Φ is used to transform the available data $\{X_n\}_{n=1,\ldots,N,}$ to a two-sided VARFIMA$(0, D, q)$ series $\{Y_n\}_{n=1,\ldots,N,}$ while η is necessary to apply the for the covariance matrix.

The conditional likelihood estimators of Φ and η are then given by

$$(\widehat{\Phi}, \widehat{\eta}) = \underset{\Phi,\eta\in S}{\mathrm{argmax}}\ \ell(\Phi, \eta; X_n|X_1, \ldots, X_p), \tag{4.4}$$

where $S = \{\eta \in \mathbb{R}^{6+4q} : -\epsilon < d_1, d_2 < 0.5,\ |\Sigma| = \sigma_{11}\sigma_{22} - \sigma_{12}^2 > 0,\ (\sigma_{jj})_{j=1,2} \geq 0\}$ denotes the parameter space for η with some small $\epsilon > 0$, e.g. $\epsilon = 0.01$. Although there is no closed form solution for the estimates $\widehat{\Phi}$ and $\widehat{\eta}$, they can be computed numerically using the quasi-Newton algorithm of Broyden, Fletcher, Goldfarb, and Shanno (BFGS).

Theoretical properties of these estimators remain to be studied. Gaussian ML estimation for multivariate LRD series was studied theoretically by Hosoya (1996) but for models expressed through one-sided linear representations. As with many Gaussian likelihood methods, note that the model itself does not need to be Gaussian. The results should be extended to two-sided models, or alternatively, one-sided representations of the models considered here should be established. On the other hand, some insight might be gained (and certainly reflected in our simulations) from the theoretical analysis of local Whittle estimators in Robinson (2008), Baek et al. (2019). For example, these works suggest that the phase parameter might be the most difficult one to estimate, especially when cross correlation is "weak."

4.2 Forecasting

To compute forecasts and corresponding error, we use the multivariate Durbin-Levisnon (DL, for short) algorithm (see Brockwell and Davis (2009), p. 422). Given the parameter estimates, the DL algorithm yields the coefficients matrices $\Phi_{n,1}, \ldots, \Phi_{n,n}$ in the 1–step-ahead forecast (predictor)

$$\widehat{Y}_{n+1} := \widehat{Y}_{n+1|n} := \mathbb{E}(Y_{n+1}|Y_1, \ldots, Y_n) = \Phi_{n,1}Y_n + \ldots + \Phi_{n,n}Y_1, \tag{4.5}$$

as well as the associated forecast error matrix $V_n = \mathbb{E}(Y_{n+1} - \widehat{Y}_{n+1})(Y_{n+1} - \widehat{Y}_{n+1})'$. The h–step-ahead forecasts, $h \geq 1$, on the other hand, are given by

$$\widehat{Y}_{n+h|n} := \mathbb{E}(Y_{n+h}|Y_1, \ldots, Y_n) = F_{n,1}^h Y_n + \ldots + F_{n,n}^h Y_1, \tag{4.6}$$

where $F_{n,k}^h, k = 1, \ldots, n$, are 2×2 real-valued coefficient matrices, with the corresponding forecast error matrix

$$W_{n+h-1|n} = \mathbb{E}(Y_{n+h} - \widehat{Y}_{n+h|n})(Y_{n+h} - \widehat{Y}_{n+h|n})'. \tag{4.7}$$

The 1-step-ahead forecasts can be used recursively by repeated conditioning to obtain recursive expressions for the coefficient matrices $F_{h,k}^h$ and an expression for the error matrix $W_{n+h-1|n}$, as stated in the next result. The standard proof is omitted for shortness sake.

Proposition 4.1 *Let $h \geq 1$ and $\Phi_{n,k}$, $n \geq 1$, $k = 1, \ldots, n$, be as above. Then, the matrices $F_{n,k}^h$ in* (4.6) *satisfy the recursive relation*

$$F_{n,k}^h = \Phi_{n+h-1,h+k-1} + \sum_{j=1}^{h-1} \Phi_{n+h-1,j} F_{n,k}^{h-j}, \tag{4.8}$$

with $F_{n,k}^1 := \Phi_{n,k}$, $n \geq 1$, $k = 1, \ldots, n$. Moreover, the corresponding error matrices $W_{n+h-1|n}$ in relation (4.7)*, are given by*

$$W_{n+h-1|n} = \Gamma(0) - \sum_{j=1}^{n} F_{n,j}^h \Gamma(h + j - 1)', \tag{4.9}$$

where $\Gamma(n) = \mathbb{E} Y_n Y_0'$ is the autocovariance matrix function of $\{Y_n\}_{n \in \mathbb{Z}}$.

Remark 4.1 In the time series literature (e.g. Brockwell and Davis (2009)), it is more succinct and common to express the h–step-ahead forecasts by using the coefficient matrices appearing in the multivariate Innovations (IN, for short) algorithm. We use the DL algorithm since it is faster than the IN algorithm: the coefficient matrices $\Phi_{n,k}$ in (4.5) are computed in $O(n^2)$ number of steps, whereas the computational complexity for the analogous coefficients in the IN algorithm is $O(n^3)$. Note also that the log-likelihood function (4.1) can be rewritten as

$$(2\pi)^{-N} \Big(\prod_{j=0}^{N-1} V_j \Big)^{-1/2} \exp \Big\{ -\frac{1}{2} \sum_{j=0}^{N-1} (Y_{j+1} - \widehat{Y}_{j+1})' V_j^{-1} (Y_{j+1} - \widehat{Y}_{j+1}) \Big\} \tag{4.10}$$

(see relation (11.5.4) in Brockwell and Davis (2009)), and thus likelihood calculations can be carried out using the DL algorithm in a similar fashion with Tsay's (2010) approach. Despite the lower computational complexity DL offers through (4.10), compared to Cholesky decomposition required for (4.1), we found the latter to be faster for small and moderate sample sizes.

The DL algorithm used in the forecasting procedure and formulae above is based on the assumption that the autocovariance function of the time series $\{Y_n\}_{n \in \mathbb{Z}}$ can readily be computed, as for example for the two-sided VARFIMA$(0, D, q)$ series. We now turn our attention to the two-sided VARFIMA(p, D, q) series $\{X_n\}_{n \in \mathbb{Z}}$ defined through $\{Y_n\}_{n \in \mathbb{Z}}$ in (4.2). As we do not have an explicit form of the autocovariance function of $\{X_n\}_{n \in \mathbb{Z}}$, it is not immediately clear how to calculate the h–step-ahead forecasts

$$\widehat{X}_{n+h|n} = \mathbb{E}(X_{n+h}|X_1, \ldots, X_n)$$

and the corresponding error matrices

$$\widetilde{W}_{n+h-1|n} = \mathbb{E}(X_{n+h} - \widehat{X}_{n+h|n})(X_{n+h} - \widehat{X}_{n+h|n})',$$

$n \geq 1$, $h \geq 1$. In Proposition 4.2 below, we show that $\widehat{X}_{n+h|n}$ and $\widetilde{W}_{n+h-1|n}$ can be calculated approximately and recursively from $\widehat{Y}_{n+h|n}$ and $W_{n+h-1|n}$. For simplicity and since this order will be used in the simulations and the application below, we focus on the case $p = 1$. However, the proposition can be extended for larger values of p.

Proposition 4.2 *Let $F^h_{n,k}$, $n \geq 1$, $k = 1, \ldots, n$, be as in* (4.6). *Then, the h–step-ahead forecasts $\widehat{X}_{n+h|n} = \mathbb{E}(X_{n+h}|X_1, \ldots, X_n)$ satisfy*

$$\widehat{X}_{n+h|n} = \widehat{X}^{(a)}_{n+h|n} + R_{n+h|n}, \tag{4.11}$$

where

$$\widehat{X}^{(a)}_{n+h|n} = \Phi_1^h X_n + \sum_{s=0}^{h-1} \Phi_1^s \widehat{Y}_{n+h-s|n}, \tag{4.12}$$

$$R_{n+h|n} = \sum_{s=0}^{h-1} \Phi_1^s \Big(\mathbb{E}(Y_{n+h-s}|X_1, \ldots, X_n) - \mathbb{E}(Y_{n+h-s}|Y_1, \ldots, Y_n) \Big). \tag{4.13}$$

Moreover, the error matrices $\widetilde{W}^{(a)}_{n+h-1|n} = \mathbb{E}(X_{n+h} - \widehat{X}^{(a)}_{n+h|n})(X_{n+h} - \widehat{X}^{(a)}_{n+h|n})'$ can be computed by

$$\widetilde{W}^{(a)}_{n+h-1|n} = \sum_{s=0}^{h-1} \Phi_1^s W_{n+h-s-1|n} (\Phi_1^s)' + \sum_{\substack{s,t=0 \\ s \neq t}}^{h-1} \Phi_1^s A_{s,t}(n+h) (\Phi_1^t)', \tag{4.14}$$

where

$$A_{s,t}(n+h) = \Gamma(t-s) - \sum_{k=1}^{n} \Gamma(h-s+k-1)(F^{h-t}_{n,k})' \tag{4.15}$$

and $\Gamma(n) = \mathbb{E} Y_n Y_0'$ is the autocovariance matrix function of $\{Y_n\}_{n \in \mathbb{Z}}$.

Proof: By using the relation (4.2) recursively, we can write

$$X_{n+h} = \Phi_1^h X_n + \sum_{s=0}^{h-1} \Phi_1^s Y_{n+h-s}, \quad h = 1, 2, \ldots \tag{4.16}$$

which implies that

$$\widehat{X}_{n+h|n} = \Phi_1^h X_n + \sum_{s=0}^{h-1} \Phi_1^s \mathbb{E}(Y_{n+h-s}|X_1, \ldots, X_n). \tag{4.17}$$

Since $\mathbb{E}(Y_{n+h-s}|Y_1, \ldots, Y_N) = \widehat{Y}_{n+h-s|n}$, the relation (4.17) yields (4.12).

Next, we subtract (4.12) from (4.16) to get

$$X_{n+h} - \widehat{X}^{(a)}_{n+h|n} = \sum_{s=0}^{h-1} \Phi^s (Y_{n+h-s} - \widehat{Y}_{n+h-s|n}).$$

The h–step-ahead error matrix $\widetilde{W}^{(a)}_{n+h-1|n}$ is then given by

$$\begin{aligned}\widetilde{W}^{(a)}_{n+h-1|n} &= \mathbb{E}\Big(\sum_{s=0}^{h-1}\Phi_1^s(Y_{n+h-s}-\widehat{Y}_{n+h-s|n})\Big)\Big(\sum_{t=0}^{h-1}\Phi_1^t(Y_{n+h-t}-\widehat{Y}_{n+h-t|n})\Big)' \\ &= \sum_{s=0}^{h-1}\Phi_1^s W_{n+h-s-1|n}(\Phi_1^s)' + \sum_{\substack{s,t=0\\ s\neq t}}^{h-1}\Phi_1^s A_{s,t}(n+h)(\Phi_1^t)', \qquad (4.18)\end{aligned}$$

where $A_{s,t}(u) = \mathbb{E}(Y_{u-s}-\widehat{Y}_{u-s|n})(Y_{u-t}-\widehat{Y}_{u-t|n})'$. To show that $A_{s,t}(u)$ satisfies (4.15), note that for $s,t = 0,\ldots,u-n-1$, $s\neq t$, we have

$$\mathbb{E}\widehat{Y}_{u-s|n}Y'_{u-t} = \mathbb{E}(\mathbb{E}(\widehat{Y}_{u-s|n}Y'_{u-t}|Y_1,\ldots,Y_N)) = \mathbb{E}\widehat{Y}_{u-s|n}\mathbb{E}(Y_{u-t}|Y_1,\ldots,Y_N) = \mathbb{E}\widehat{Y}_{u-s|n}\widehat{Y}'_{u-t|n}.$$

Hence,

$$\begin{aligned}A_{s,t}(u) &= \mathbb{E}Y_{u-s}Y'_{u-t} - \mathbb{E}Y_{u-s}\widehat{Y}'_{u-t|n} - \mathbb{E}\widehat{Y}_{u-s|n}Y'_{u-t} + \mathbb{E}\widehat{Y}_{u-s|n}\widehat{Y}'_{u-t|n} \\ &= \Gamma(t-s) - \mathbb{E}Y_{u-s}\widehat{Y}'_{u-t|n} \\ &= \Gamma(t-s) - \mathbb{E}Y_{u-s}\Big(\sum_{k=1}^{n}F_{n,k}^{u-t-n}Y_{n-k+1}\Big)' \\ &= \Gamma(t-s) - \sum_{k=1}^{n}\Gamma(u-s-n+k-1)(F_{n,k}^{u-t-n})', \qquad (4.19)\end{aligned}$$

yielding the relations (4.14)–(4.15). □

Since $X_n - \Phi_1 X_{n-1} = Y_n$ for the VARFIMA$(1,D,q)$ series $\{X_n\}$ and the VARFIMA$(0,D,q)$ series $\{Y_n\}$, the approximation error $R_{n+h|n}$ in (4.13) becomes negligible for large n. For this reason, in the simulations and the application below, we shall use the approximate forecasts $\widehat{X}^{(a)}_{n+h|n}$ in (4.11) and their forecast error matrices $\widetilde{W}^{(a)}_{n+h-1|n}$ given by (4.14).

5 Simulation study

In this section, we perform a Monte Carlo simulation study to assess the performance of the CML estimation approach for the models (3.9) and (3.13)–(3.14) described in Section 3. We examine four different models with AR and MA components of orders $p,q = 0,1$, as well as a fifth fractionally cointegrated $(1,D,0)$ model. For the $(1,D,0)$ models, we shall consider a non-diagonal AR matrix. Though this should be treated with some caution in connection to identifiability as noted in Section 3.2, we sought to see what happens when the AR part is non-diagonal as well. The results for $(1,D,0)$ models with diagonal AR parts (not reported here) were qualitatively similar or better. For each model, we consider three sample sizes $N = 200, 400, 1000$. The Gaussian time series data are generated using the fast and exact synthesis algorithm of Helgason et al. (2011), while the number of replications is 5000.

To solve the maximization problem (4.4), we use the SAS/IML **nlpqn** function, which implements the BFGS quasi-Newton method, a popular iterative optimization algorithm. For our optimization scheme, we follow the approach found in Tsay (2010). A first step is to eliminate the nonlinear inequality constraint $|\Sigma| \geq 0$ in the parameter space S defined in Section 4.1, by

letting $\Sigma = U'U$, where $U = (U_{jk})_{j,k=1,2}$ is an upper triangular matrix. Then, the parameter vector θ can be written as $\theta = (d_1, d_2, c, U_{11}, U_{12}, U_{22}, \Theta')'$ while the parameter space becomes $S = \{\theta \in \mathbb{R}^{6+4q} : -\epsilon < d_1, d_2 < 0.5\}$, for some small positive ϵ (we considered $\epsilon = 0.01$).

Next, we describe our strategy on selecting initial parameter values (Φ^I, θ^I) for the BFGS method. Let

$$\Phi_1^0 = (\phi_{jk,1}^0)_{j,k=1,2}, \quad \theta^0 = (d_1^0, d_2^0, c^0, U_{11}^0, U_{12}^0, U_{22}^0, (\Theta_1^0)')', \tag{5.1}$$

where $\Theta_1^0 = (\theta_{jk,1}^0)_{j,k=1,2}$, be the true parameter values. We consider initial values

$$d_k^I = \frac{2d_k^0}{1+2d_k^0}, \quad c^I = \frac{2c^0}{1+|c^0|}, \quad U_{jk}^I = 1, \quad \theta_{jk,1}^I = \frac{e^{\theta_{jk,1}^0}-1}{e^{\theta_{jk,1}^0}+1}, \quad \phi_{jk,1}^I = \frac{e^{\phi_{jk,1}^0}-1}{e^{\phi_{jk,1}^0}+1}, \tag{5.2}$$

where $j, k = 1, 2$. Note that the transformations of the LRD parameters in (5.2) are essentially perturbations of the true parameter values that also retain the range of the parameter space S. For example, the value of d_k^I will be zero (or $1/2$) when d_k^0 is also zero (or $1/2$). The choice c^I is also a perturbed version of c that satisfies $|c^I| < 1$ when the true parameter c is located in the interval $(-1, 1)$. Moreover, even though the parameter space S does not include identifiability (including stability) constraints for the elements of the AR and MA polynomials as discussed in Section 3.2, we did not encounter any cases where the optimization algorithm considers such values.

In a practical scenario, when the true parameter values are unknown, initial estimates can be obtained as follows. One can follow the semiparametric approach of Robinson (2008) to obtain the local Whittle estimates of the parameters $d_1, d_2, \phi, \omega_{11}, \omega_{12}, \omega_{22}$, and subsequently use the relation (2.1) to get an estimate of c. For diagonal AR matrices, one can then apply the fractional filter to the data (see Jensen and Nielsen (2014) for a fast fractional differencing algorithm) and use techniques from the VARMA literature (e.g. least squares) to obtain estimates for short memory parameters in a quick manner. In addition to fast computation of initial estimates, this methodology allows one to identify possible weak cross-spectra in the form of small values for the ratio $\widehat{\omega}_{12}^2/(\widehat{\omega}_{11}\widehat{\omega}_{22})$. In the latter case, recalling the discussion under relation (2.16), the values c and $1/c$ lead to similar models and the optimization algorithm may converge to a local maxima around $1/c$ (with the values of U accordingly rescaled). We encountered this behavior in simulation schemes with weak cross-spectra in the form of bimodality for $\widehat{c}$ and the entries of $\widehat{U}$. A simple, yet effective strategy to deal with this issue is to refit the model a second time, using the initial value $c^I = 1/\widehat{c}$ where $\widehat{c}$ is the estimate from the first fit, and analogously transform the initial estimates for U. Then, for the two fitted models, the one with the larger likelihood value is selected.

Table 1 and Figure 2 present estimation results for the five models considered. To save space, we present the cointegrated $(1, D, 0)$ model in Figure 2 and the non-cointegrated $(1, D, 0)$ model in Table 1. For all models without MA components, we take (dropping the superscript 0 for simplicity) $d_1 = 0.2$, $d_2 = 0.4$, $c = 0.6$, $\Sigma_{11} = 3$, $\Sigma_{12} = 2$, $\Sigma_{22} = 3$, and wherever present, $\Phi_{11} = \phi_{11,1} = 0.5$, $\Phi_{12} = \phi_{12,1} = 0.2$, $\Phi_{21} = \phi_{21,1} = 0.4$, $\Phi_{22} = \phi_{22,1} = -0.8$, $\alpha = 0.5$.[4] In models with MA components we take the same d's and $c = 0.2$, $\Sigma_{11} = 3$, $\Sigma_{12} = 1.5$, $\Sigma_{22} = 3$, $\Theta_{11} = \theta_{11,1} = 0.7$, $\Theta_{12} = \theta_{12,1} = -0.1$, $\Theta_{21} = \theta_{21,1} = 0.2$, $\Theta_{22} = \theta_{22,1} = 0.4$. We also performed simulations for several other values of these parameters with similar results and therefore we omit them in favor of space economy.

Table 1 lists the median differences between the estimates and the corresponding true values, and the respective median absolute deviations. Figure 2, on the other hand, includes the boxplots of the estimates for the various parameters. The true parameter values are marked in dashed blue

[4]For this choice of d_1, d_2, c, the phase parameter is equal to $\phi = -1.15$. Taking $c = -0.1985$ with the same d's would yield zero phase.

lines. Moreover, the boxplots of $\widehat{U}_{11}$, $\widehat{U}_{12}$ and $\widehat{U}_{22}$ are centered at zero by subtracting the true parameter value, providing a uniform presentation scale. While the table concerns only the results for the sample sizes $N = 200$ and 400, the figure also includes the case of $N = 1000$.

(p,q)	$(0,0)$		$(0,1)$		$(1,0)$		$(1,1)$	
N	200	400	200	400	200	400	200	400
d_1	−0.007 0.036	−0.002 0.026	−0.009 0.043	−0.004 0.029	−0.019 0.062	−0.019 0.061	−0.029 0.097	−0.014 0.068
d_2	−0.008 0.036	−0.004 0.026	−0.011 0.040	−0.004 0.031	−0.005 0.028	−0.005 0.027	−0.005 0.038	−0.002 0.028
c	0.010 0.047	0.005 0.034	0.017 0.108	0.008 0.076	0.010 0.043	0.009 0.046	0.001 0.059	−0.003 0.043
Φ_{11}/Θ_{11}			0.010 0.048	0.004 0.033	0.017 0.059	0.016 0.060	0.016/0.032 0.111/0.080	0.005/0.022 0.081/0.057
Φ_{12}/Θ_{12}			0.007 0.058	0.003 0.041	−0.012 0.072	−0.015 0.074	−0.012 0.044	−0.010 0.029
Φ_{21}/Θ_{21}			0.004 0.051	0.002 0.037	−0.001 0.016	0.000 0.016	0.020 0.102	0.001 0.073
Φ_{22}/Θ_{22}			0.006 0.058	0.005 0.04	−0.002 0.017	−0.001 0.017	0.011/−0.001 0.032/0.039	0.008/−0.002 0.024/0.026
U_{11}	−0.017 0.075	−0.009 0.054	−0.038 0.154	−0.016 0.111	−0.019 0.065	−0.019 0.065	−0.028 0.086	−0.015 0.064
U_{12}	0.022 0.120	0.015 0.086	0.016 0.130	0.007 0.091	0.022 0.104	0.010 0.098	−0.001 0.140	−0.006 0.101
U_{22}	0.010 0.112	0.009 0.080	0.005 0.150	−0.001 0.107	0.012 0.087	0.005 0.091	−0.012 0.110	−0.008 0.083

Table 1: *Median differences between the estimates and the corresponding true values (top value in each cell) and median absolute deviations (bottom value in each cell) for the estimated parameters of VARFIMA series with* $(p,q) = (0,0), (0,1)$, *non-cointegrated* $(1,0)$ *and* $(1,1)$.

The results in Table 1 and Figure 2 indicate a satisfactory performance of the CML approach for all cases considered: the median differences are small overall and tend to decrease with the increasing sample size, and the decrease with the increasing sample size is also evident for the median deviations; moreover, many median deviations and box sizes are relatively small as well. The cases we show here have relatively strong cross-spectra and the optimization algorithm had no trouble finding the global maximum. We also considered two weak cross-spectra scenarios: a two-sided $(0, D, 0)$ model with parameters $d_1 = 0.2$ $d_2 = 0.4$, $c = 0.6$, $\Sigma_{11} = 3$, $\Sigma_{12} = 0.5$ $\Sigma_{22} = 3$ and a two-sided $(0, D, 1)$ model with the same d's, c, Σ and $\Theta_{11} = \theta_{11,1} = 0.7$, $\Theta_{12} = \theta_{12,1} = -0.1$, $\Theta_{21} = \theta_{21,1} = 0.2$, $\Theta_{22} = \theta_{22,1} = 0.4$. The models for these sets of parameters have small variance ratios $\gamma_{12}(0)^2/(\gamma_{11}(0)\gamma_{22}(0))$ (0.004 and 0.034 respectively) which similarly to $\omega_{12}^2/(\omega_{11}\omega_{22})$ can be viewed as measures of cross-dependence between the series (however without focusing on low frequencies as the spectral ratio does). In this case, the optimization algorithm produced bimodal estimates with small differences in likelihood values. After employing the "refit" strategy described above, however, we successfully estimated the true models and obtained similar results with those in Table 1 and Figure 2.

Finally, we also comment on the model selection task concerning the one- and two-sided models when using BIC and AIC. Figure 3, the left plot, presents the proportion of times that these informa-

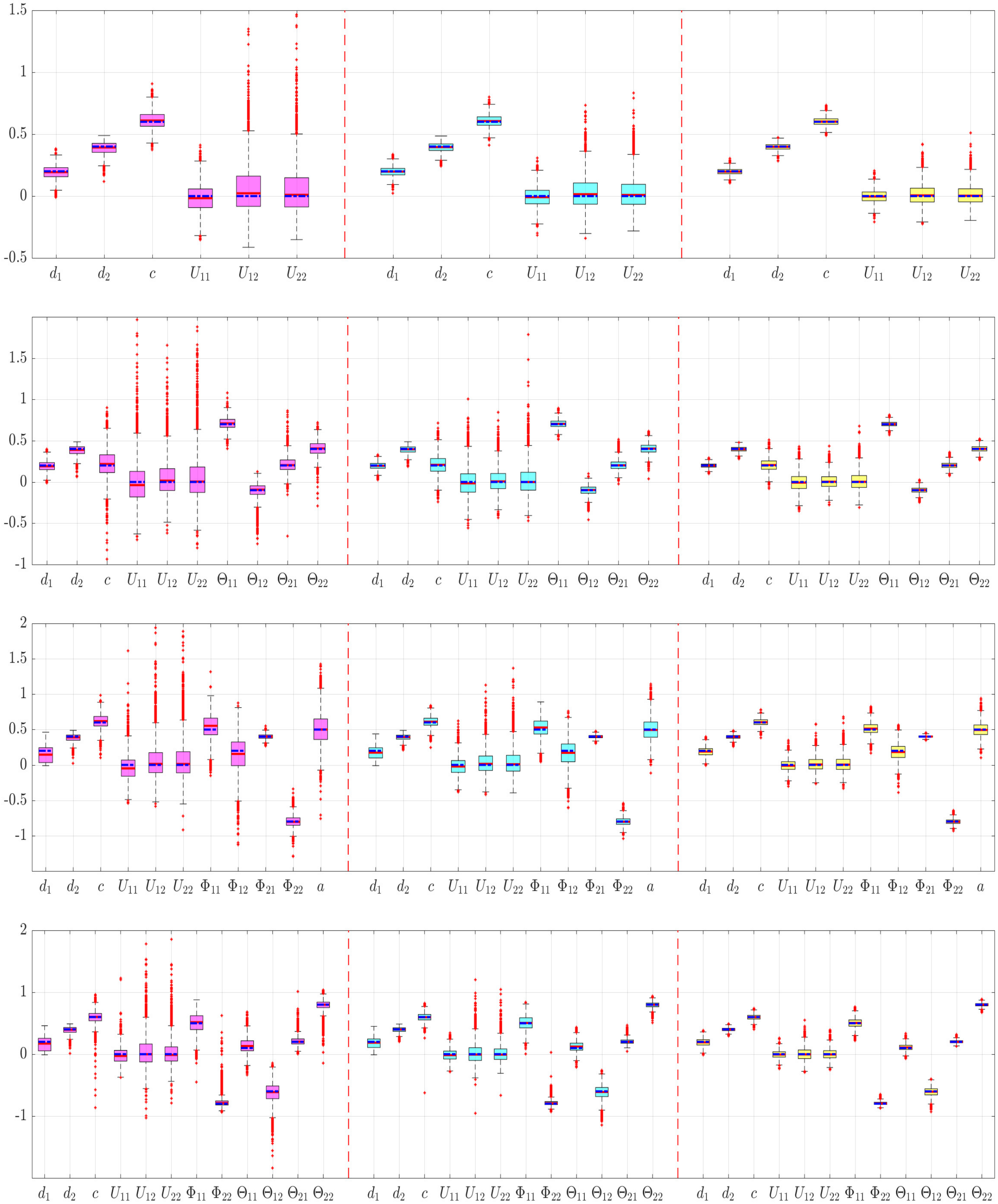

Figure 2: *The red solid lines are the medians and the blue dashed lines are the true parameter values (except for U_{11}, U_{12}, U_{22} which are centered at 0). Top to bottom:* $(p, q) = (0, 0)$, $(0, 1)$, *cointegrated* $(1, 0)$ *and* $(1, 1)$. *Left to right:* $N = 200, 400, 1000$.

tion criteria select the one-sided VARFIMA$(0, D, 0)$ model over the two-sided VARFIMA$(0, D, 0)$, when in fact the latter model is true, for the same parameter values as in Table 1 in the case $p = q = 0$. The right plot of the figure presents the analogous plot for a different set of values of the parameters d_1, d_2 and c. The performance of the model selection criteria is satisfactory overall.

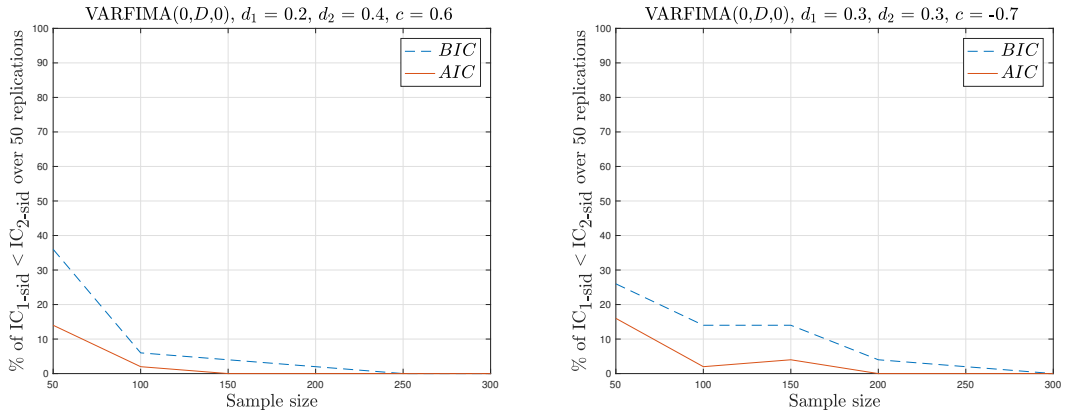

Figure 3: *The proportion of times that the considered information criteria select the one-sided VARFIMA$(0, D, 0)$ model over the two-sided one, when the latter is true.*

6 Application

In this section, we apply the estimation techniques discussed in Section 4.1 to analyze inflation rates in the U.S. under the two-sided, possibly cointegrated, VARFIMA(p, D, q) models discussed in Section 3. Models with both $|c| < 1$ and $|c| > 1$ will be fitted. Long-range dependence in inflation rates has been reported and studied in a number of works (see, for example, Baillie et al. (1996), Doornik and Ooms (2004), Sela and Hurvich (2009), Baillie and Morana (2012) and references therein), and might also be expected due to aggregation effects (Granger (1980)). More specifically, Sela and Hurvich (2009) tested the fit of several long- and short-range dependent models on the annualized monthly inflation rates for goods and services in the U.S. during the period of February 1956–January 2008 ($N = 624$ months) and selected a one-sided VARFIMA model as the best choice. Besides their long memory features, however, the time series of inflation rates often exhibit asymmetric behavior, and therefore call for multivariate LRD models that allow for a general phase.

Following the notation of Sela (2010),[5] we denote the Consumer Price Indices series for commodities as $\{CPI_n^c\}_{n=0,\ldots,N}$ and the corresponding series for services as $\{CPI_n^s\}_{n=0,\ldots,N}$. Then, we define the annualized monthly inflation rates for goods and services as

$$g_n = 1200\frac{CPI_n^c - CPI_{n-1}^c}{CPI_{n-1}^c} \quad \text{and} \quad s_n = 1200\frac{CPI_t^s - CPI_{n-1}^s}{CPI_{n-1}^s},$$

respectively. The two series $\{g_n\}_{n=1,\ldots,N}$, $\{s_n\}_{n=1,\ldots,N}$ are depicted in Figure 4.[6]

The two plots in Figure 5 provide some motivation for why a general phase model is needed for this dataset. More specifically, the left plot in Figure 5 depicts the sample cross-correlation function

[5] See also the accompanying R code.

[6] The consumer price indices (raw) data are available online from the Bureau of Labor Statistics.

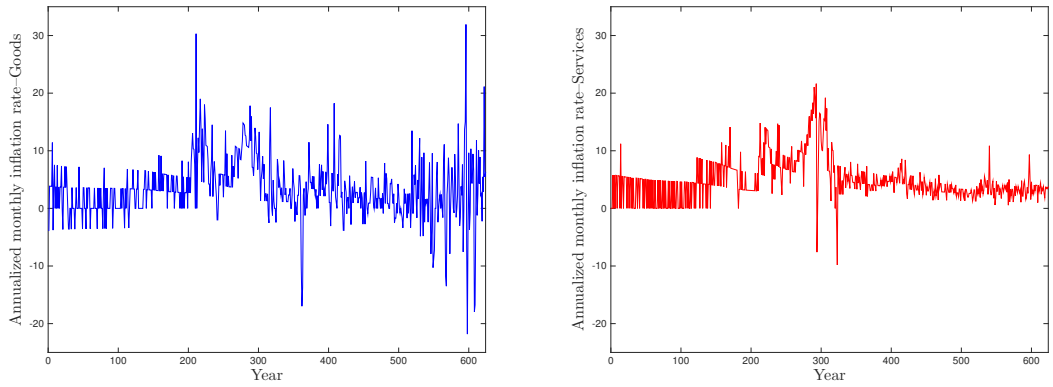

Figure 4: *Annualized monthly inflation rates for goods (left) and services (right) from February 1956 to January 2008.*

$\widehat{\rho}_{12}(h)$ of the two series for all lags such that $|h| < 25$. Observe that for negative lags the sample cross-correlation function decays faster than for positive lags suggesting time-non-reversibility of the series and hence non-zero phase.

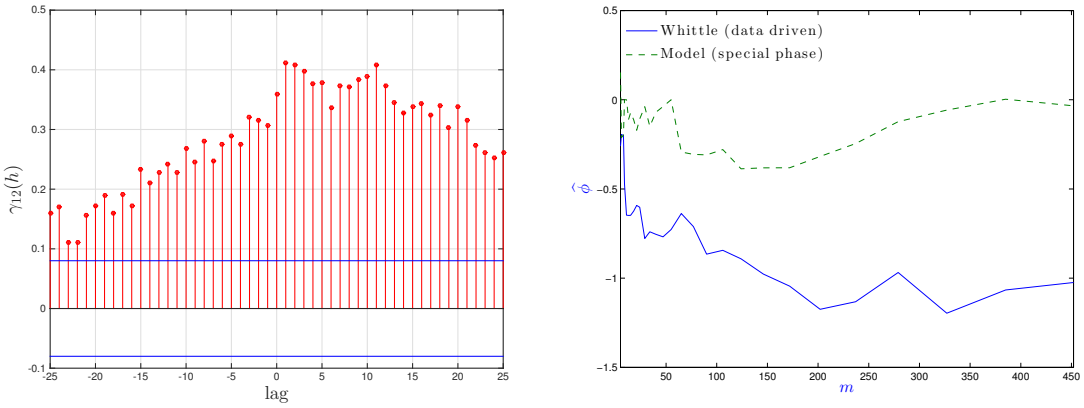

Figure 5: *Left plot: Sample cross-correlation* $\widehat{\rho}_{12}(h)$ *of the series* $\{g_n\}_{n=1,\ldots,N}$ *and* $\{s_n\}_{n=1,\ldots,N}$ *depicted in Figure 4 for* $|h| \leq 25$. *Right plot: Local Whittle phase estimates, one corresponding to the one-sided VARFIMA (dashed line) and one estimated directly from the data (solid line). Both estimates are plotted as functions of a tuning parameter* $m = N^{0.25+0.0125k}$, $k = 1, \ldots, 51$, *where* $N = 624$ *is the sample size of the series.*

Further evidence for general phase can be obtained from the local Whittle estimation of Robinson (2008) which can be used to estimate the phase and the LRD parameters directly from the data. The estimation is semiparametric in the sense that it only requires specification of the spectral density at low frequencies. The right plot in Figure 5 is a local Whittle plot, depicting two local Whittle estimates of the phase parameter ϕ as functions of m – a tuning parameter representing the number of lower frequencies used in the estimation. The dashed line corresponds to the special phase estimate $\widehat{\phi} = (\widehat{d}_1 - \widehat{d}_2)\pi/2$ of the one-sided VARFIMA model based on the local Whittle estimates of the two d's. On the other hand, the solid line shows the phase parameter estimated directly from the data. The two lines being visibly different suggest that the special phase parameter

and the associated VARFIMA model are not appropriate. A more detailed local Whittle analysis of the dataset can be found in Baek et al. (2019): it includes confidence intervals in local Whittle plots and also the results under fractional cointegration.

In the analysis of Sela and Hurvich (2009), the one-sided VARFIMA$(1, D, 0)$ model was selected as the best choice (based on AIC), amongst vector autoregressive models of both low and high orders and also amongst one-sided VARFIMA$(p, D, 0)$ and FIVARMA$(p, D, 0)$ models with $p \leq 1$. The estimated VARFIMA$(1, D, 0)$ model, in particular, was

$$\begin{array}{rcl} g_n &=& 0.3027 g_{n-1} + 0.4245 s_{n-1} + \epsilon_{1,n}, \\ s_n &=& -0.0237 g_{n-1} - 0.3085 s_{n-1} + \epsilon_{2,n}, \end{array} \tag{6.1}$$

where

$$\begin{pmatrix} \epsilon_{1,n} \\ (I-B)^{0.4835}\epsilon_{2,n} \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} 20.23 & 0.46 \\ 0.46 & 7.08 \end{pmatrix}\right). \tag{6.2}$$

We should note here that fitting the 1-sided model using the SAS/ETS VARMAX procedure produced estimates similar to those of Sela's algorithm (implemented in R), for all parameters except d_1 which Sela estimates to be zero while for this model we estimated it to be 0.191. More specifically, the estimated one-sided VARFIMA$(1, D, 0)$ model is

$$\begin{array}{rcl} g_n &=& 0.132\,(0.064)\, g_{n-1} + 0.076\,(0.080)\, s_{n-1} + \epsilon_{1,n}, \\ s_n &=& 0.056\,(0.023)\, g_{n-1} - 0.308\,(0.044)\, s_{n-1} + \epsilon_{2,n}, \end{array} \tag{6.3}$$

where

$$\begin{pmatrix} (I-B)^{0.191\,(0.052)}\epsilon_{1,n} \\ (I-B)^{0.475\,(0.062)}\epsilon_{2,n} \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} 20.21 & 0.75 \\ 0.75 & 7.02 \end{pmatrix}\right) \tag{6.4}$$

and the underlying U in $\Sigma = U'U$ has $U_{11} = 4.605\,(0.131)$, $U_{12} = 0.164\,(0.109)$ and $U_{22} = 2.645\,(0.075)$, with standard errors of the estimates added in the parentheses throughout.

The parameter estimates in (6.1)–(6.2) reveal an interesting feature, noted by Sela (2010). In particular, while the lagged services inflation has a significant influence on goods inflation, the lagged goods inflation seems to have a small effect on services inflation. This behavior is potentially related to the so-called gap between the prices in services and the prices in goods which was studied by Peach et al. (2004). More specifically, the term gap refers to the tendency of prices in services to increase faster than prices in goods. But note that this effect is not present in the estimated model (6.3)–(6.4) and, in fact, is reversed.

For comparison, we present next two estimated two-sided VARFIMA$(1, D, 0)$ models with and without fractional cointegration. Moreover, we focus on diagonal AR components, in part to avoid confusion with fractional cointegration as noted in Section 3.2. The estimated two-sided non-cointegrated VARFIMA$(1, D, 0)$ model is

$$\begin{array}{rcl} g_n &=& 0.106\,(0.053)\, g_{n-1} + \epsilon_{1,n}, \\ s_n &=& -0.439\,(0.080)\, s_{n-1} + \epsilon_{2,n}, \end{array} \tag{6.5}$$

where

$$\begin{pmatrix} ((I-B)^{-0.231\,(0.043)} + 0.344\,(0.158)(I-B^{-1})^{-0.231})^{-1}\epsilon_{1,n} \\ ((I-B)^{-0.439\,(0.044)} - 0.344(I-B^{-1})^{-0.439})^{-1}\epsilon_{2,n} \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} 12.141 & 0.948 \\ 0.948 & 11.680 \end{pmatrix}\right) \tag{6.6}$$

and the corresponding U in $\Sigma = U'U$ has $U_{11} = 3.484\,(0.408)$, $U_{12} = 0.272\,(0.154)$ and $U_{22} = 3.407\,(0.471)$. On the other hand after expanding relation (3.13), we obtain the estimated two-sided VARFIMA$(1, D, 0)$ model with fractional cointegration as

$$\begin{array}{rcl} g_n &=& 0.146\,(0.057)\,g_{n-1} + 1.11\,(0.131)\,s_n - 0.1621\,s_{n-1} + \epsilon_{1,n}, \\ s_n &=& -0.166\,(0.043)\,s_{n-1} + \epsilon_{2,n}, \end{array} \tag{6.7}$$

where

$$\begin{pmatrix} ((I-B)^{-0.021\,(0.045)} - 2.604\,(1.219)(I-B^{-1})^{-0.021})^{-1}\epsilon_{1,n} \\ ((I-B)^{-0.483\,(0.022)} + 2.604(I-B^{-1})^{-0.483})^{-1}\epsilon_{2,n} \end{pmatrix} \sim \mathcal{N}\Bigg(0, \begin{pmatrix} 10.65 & 1.415 \\ 1.415 & 0.645 \end{pmatrix}\Bigg) \tag{6.8}$$

and the corresponding U in $\Sigma = U'U$ has $U_{11} = 3.263\,(2.504)$, $U_{12} = 0.433\,(0.169)$ and $U_{22} = 0.675\,(0.241)$. Note, that in the first equation of (6.7), the coefficients of the terms g_{n-1}, s_n and s_{n-1} correspond to $\widehat{\phi}_{11}$, $\widehat{a}$ and $\widehat{a}\widehat{\phi}_{11}$ respectively.

We have a number of interesting observations related to the preceding fits. In terms of AIC, listed from smallest to largest, the estimated models are (6.7)–(6.8), (6.3)–(6.4) and (6.5)–(6.6). In terms of BIC, the models are (6.7)–(6.8), (6.5)–(6.6) and (6.3)–(6.4). We also fitted $(0, D, 0)$ models (both one- and two-sided) allowing for fractional cointegration. Both selection criteria preferred two-sided models over their one-sided analogues with the sole exception of the BIC in the non-cointegrated $(1, D, 0)$ case. Additionally, all cointegrated models were selected over non-cointegrated ones.

Finally, we mention that the degrees of dependence as well as the asymmetry behavior implied by both fitted models (6.5)–(6.6) and (6.7)–(6.8), agree with the ones obtained by the semiparametric approach of Baek et. al (2019). More specifically, a visual inspection of the local Whittle plots in Figure 7 of Baek et. al (2019) yields approximate estimates $\widehat{d}_{1,nc} \approx 0.23$, $\widehat{d}_{2,nc} \approx 0.42$, $\widehat{\phi}_{nc} \approx -0.78$ and $\widehat{d}_{1,c} \approx 0.03$, $\widehat{d}_{2,c} \approx 0.45$, $\widehat{\phi}_c \approx 0.5$ for the non-cointegrated (nc) and cointegrated (c) cases respectively. Plugging in the maximum likelihood estimates of d_1, d_2 and c from (6.7)–(6.8) into the relation (2.9) yields the phase estimates $\phi_{nc} = -0.85$, $\phi_c = 0.33$.

7 Conclusions

In this work, we studied modeling approaches for bivariate stationary series exhibiting long-range dependence that allow for general phase. The study was motivated by the fact that commonly considered bivariate long-range dependent models could capture only a very special phase, and by an inflation time series data that called for models beyond those with the special phase.

Several open questions related to our models were already raised in this work, including identifiability conditions (Sections 3.1 and 3.2), theoretical properties of the estimators (Section 4.1), and the one-sided representations of the considered models (Section 1). A natural extension of this work is to consider multivariate (that is, higher-order) long-range dependent series with general phase. In fact, the authors have made an incursion into this extension in a conference paper in Baek et al. (2017). But the models suggested in Baek et al. (2017) cannot be interpreted as nicely as the models studied here, for example, they do not model phases through dedicated parameters. Among all these lingering issues, discovering the forms of one-sided representations is probably the most fundamental, as this might also suggest modeling approaches that go beyond those based on covariance structures as in this work.

Acknowledgments

Vladas Pipiras was supported in part by NSA grant H98230-13-1-0220 and NSF grant DMS-1712966. The authors are thankful to Rick Wicklin (SAS Institute) for helping us optimize the efficiency of the SAS code and to Richard Davis (Columbia University) for his comments on an earlier version of this paper. The authors would also like to thank the three anonymous Referees and the Associate Editor for many insightful comments and useful suggestions that helped improving the paper considerably.

Data Availability Statement

The data that support the findings of this study are openly available through the US Bureau of Labor Statistics at https://download.bls.gov/pub/time.series/cu/cu.data.20.USCommoditiesServicesSpecial, reference number as the series CUSR0000SAC and CUSR0000SAS.

References

Amblard, P.-O. & Coeurjolly, J.-F. (2011), 'Identification of the multivariate fractional Brownian motion', *Signal Processing, IEEE Transactions on* **59**(11), 5152–5168.

Baek, C., Kechagias, S. & Pipiras, V. (2017), Semiparametric, parametric, and possibly sparse models for multivariate long-range dependence, *in* 'Wavelets and Sparsity XVII', Vol. 10394, International Society for Optics and Photonics, p. 103941S.

Baek, C., Kechagias, S. & Pipiras, V. (2019), Asymptotics of bivariate local Whittle estimators with applications to fractal connectivity, Preprint. Available at http://pipiras.web.unc.edu/papers/.

Baillie, R. T. & Morana, C. (2012), 'Adaptive ARFIMA models with applications to inflation', *Economic Modelling* **29**(6), 2451–2459.

Baillie, R. T., Chung, C.-F. & Tieslau, M. A. (1996), 'Analysing inflation by the fractionally integrated ARFIMA-GARCH model', *Journal of Applied Econometrics* **11**(1), 23–40.

Beran, J., Feng, Y., Ghosh, S. & Kulik, R. (2013), *Long-Memory Processes*, Springer, Heidelberg.

Brockwell, P. J. & Davis, R. A. (2009), *Time Series: Theory and Methods*, Springer Series in Statistics, Springer, New York. Reprint of the second (1991) edition.

Diongue, A. K. (2010), 'A multivariate generalized long memory model', *Comptes Rendus Mathematique* **348**(5), 327–330.

Doornik, J. A. & Ooms, M. (2004), 'Inference and forecasting for ARFIMA models with an application to US and UK inflation', *Studies in Nonlinear Dynamics & Econometrics*.

Doppelt, R. & O'Hara, K. (2019), 'Posterior sampling in two classes of multivariate fractionally integrated models: corrigendum to ravishanker, n. and bk ray (1997) australian journal of statistics 39 (3), 295–311', *Australian & New Zealand Journal of Statistics* **61**(1), 85–87.

Doppelt, R., O'Hara, K. et al. (2018), Bayesian estimation of fractionally integrated vector autoregressions and an application to identified technology shocks, *in* '2018 Meeting Papers', number 1212, Society for Economic Dynamics.

Doukhan, P., Oppenheim, G. & Taqqu, M. S. (2003), *Theory and Applications of Long-Range Dependence*, Birkhäuser Boston Inc., Boston, MA.

Dueker, M. & Startz, R. (1998), 'Maximum-likelihood estimation of fractional cointegration with an application to US and Canadian bond rates', *Review of Economics and Statistics* **80**(3), 420–426.

Dufour, J.-M. & Pelletier, D. (2014), 'Practical methods for modeling weak VARMA processes: identification, estimation and specification with a macroeconomic application', *Preprint*.

Giraitis, L., Koul, H. L. & Surgailis, D. (2012), *Large Sample Inference for Long Memory Processes*, Imperial College Press, London.

Granger, C. W. J. (1980), 'Long memory relationships and the aggregation of dynamic models', *Journal of Econometrics* **14**(2), 227–238.

Helgason, H., Pipiras, V. & Abry, P. (2011), 'Fast and exact synthesis of stationary multivariate Gaussian time series using circulant embedding', *Signal Processing* **91**(5), 1123–1133.

Hosoya, Y. (1996), 'The quasi-likelihood approach to statistical inference on multiple time-series with long-range dependence', *Journal of Econometrics* **73**, 217–236.

Jensen, A. N. & Nielsen, M. Ø. (2014), 'A fast fractional difference algorithm', *Journal of Time Series Analysis* **35**(5), 428–436.

Johansen, S. (2008), 'A representation theory for a class of vector autoregressive models for fractional processes', *Econometric Theory* **24**(3), 651–676.

Johansen, S. & Nielsen, M. Ø. (2012), 'Likelihood inference for a fractionally cointegrated vector autoregressive model', *Econometrica* **80**(6), 2667–2732.

Kechagias, S. & Pipiras, V. (2015), 'Definitions and representations of multivariate long-range dependent time series', *Journal of Time Series Analysis* **36**(1), 1–25.

Lobato, I. N. (1997), 'Consistency of the averaged cross-periodogram in long memory series', *Journal of Time Series Analysis* **18**(2), 137–155.

Martin, V. L. & Wilkins, N. P. (1999), 'Indirect estimation of ARFIMA and VARFIMA models', *Journal of Econometrics* **93**(1), 149–175.

Pai, J. & Ravishanker, N. (2009*a*), 'Maximum likelihood estimation in vector long memory processes via EM algorithm', *Computational Statistics & Data Analysis* **53**(12), 4133–4142.

Pai, J. & Ravishanker, N. (2009*b*), 'A multivariate preconditioned conjugate gradient approach for maximum likelihood estimation in vector long memory processes', *Statistics & Probability Letters* **79**(9), 1282–1289.

Palma, W. (2007), *Long-Memory Time Series*, John Wiley & Sons, Inc., Hoboken, New Jersey, USA.

Park, K. & Willinger, W. (2000), *Self-Similar Network Traffic and Performance Evaluation*, Wiley Online Library.

Peach, R., Rich, R. & Antoniades, A. (2004), 'The historical and recent behaviour of goods and services inflation', *Economic Policy Review* pp. 19–31.

Pipiras, V. & Taqqu, M. S. (2017), *Long-Range Dependence and Self-Similarity*, Cambridge University Press, Cambridge.

Ravishanker, N. & Ray, B. K. (1997), 'Bayesian analysis of vector ARFIMA processes', *Australian Journal of Statistics* **39**(3), 295–311.

Robinson, P. M. (2003), *Time Series with Long Memory*, Advanced Texts in Econometrics, Oxford University Press, Oxford.

Robinson, P. M. (2008), 'Multiple local Whittle estimation in stationary systems', *The Annals of Statistics* **36**(5), 2508–2530.

Roy, A., McElroy, T. S. & Linton, P. (2014), Estimation of causal invertible VARMA models, Preprint.

Sela, R. J. (2010), Three essays in econometrics: multivariate long memory time series and applying regression trees to longitudinal data, PhD thesis, New York University.

Sela, R. J. & Hurvich, C. M. (2009), 'Computationally efficient methods for two multivariate fractionally integrated models', *Journal of Time Series Analysis* **30**(6), 631–651.

Sowell, F. (1986), 'Fractionally integrated vector time series', PhD thesis, Duke University.

Tsay, W. J. (2010), 'Maximum likelihood estimation of stationary multivariate ARFIMA processes', *Journal of Statistical Computation and Simulation* **80**(7), 729–745.

Stefanos Kechagias
SAS Institute
100 SAS Campus Drive
Cary, NC 27513, USA
Stefanos.Kechagias@sas.com

Vladas Pipiras
Dept. of Statistics and Operations Research
UNC at Chapel Hill
CB#3260, Hanes Hall
Chapel Hill, NC 27599, USA
pipiras@email.unc.edu