Statistica Sinica Preprint No: SS-2016-0112R1	
Title	Inference of Bivariate Long-memory Aggregate Time Series
Manuscript ID	SS-2016-0112R1
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0112
Complete List of Authors	Henghsiu Tsai Heiko Rachinger and Kung-Sik Chan
Corresponding Author	Henghsiu Tsai
E-mail	htsai@stat.sinica.edu.tw

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

Inference of Bivariate Long-memory Aggregate Time Series

Henghsiu Tsai, Heiko Rachinger and Kung-Sik Chan
Academia Sinica, University of Vienna and University of Iowa

April 10, 2017

Abstract. With the increasing deployment of affordable and sophisticated sensors, multivariate time-series data are increasingly collected. These multivariate time series are often of long memory, the inference of which can be rather complex. We consider the problem of modeling long-memory bivariate time series that are aggregates from an underlying long-memory continuous-time process. We show that, with increasing aggregation, the resulting discrete-time process is approximately a linear transformation of two independent fractional Gaussian noises with the corresponding Hurst parameters equal to those of the underlying continuous-time processes. We use simulations to confirm the good approximation of the limiting model to aggregate data from a continuous-time process. The theoretical and numerical results justify modeling long-memory bivariate aggregate time series by this limiting model. The model parametrization does change drastically in the case of identical Hurst parameters. We derive the likelihood ratio test for testing the equality of the two Hurst parameters, within the framework of Whittle likelihood, and the corresponding maximum likelihood estimators. The limiting properties of the proposed test statistic and of the Whittle likelihood estimation are derived, and their finite sample properties are studied by simulation. The efficacy of the proposed approach is demonstrated with a 2-dimensional robotic positional error time

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

series, which shows that the proposed parsimonious model substantially outperforms a VAR(19) model.

Key words and phrases: Aggregation; asymptotic normality; fractional Gaussian noise; spectral maximum likelihood estimator; Whittle likelihood.

1 Introduction

Continuous-time long-memory models have found diverse applications in such fields as option pricing (Comte and Renault (1998)), volatility modeling (Casas and Gao (2008)), environmental study (Tsai and Chan (2005a)), and annual tree-ring measurements (Tsai and Chan (2005b)), among many others. For further developments of univariate continuous-time long-memory models, see, for example, Chambers (1996), Comte (1996), Comte and Renault (1996), and Brockwell and Marquardt (2005). To study interactions and comovements among a group of time series variables, one needs to consider multivariate time series models, and several long memory models have been proposed. Marquardt (2007) introduced a class of multivariate fractionally integrated continuous-time autoregressive moving-average (CARMA) processes and studied their probabilistic properties, but not their inference. Barndorff-Nielsen and Stelzer (2011) proposed the multivariate supOU (superpositions of Ornstein-Uhlenbeck-type) processes and proposed moment-based estimation methods for estimating the parameters; the finite- and large-sample properties of the estimator are unknown. Asai and McAleer (2013) proposed a fractionally integrated Wishart stochastic volatility model with a common long memory for multivariate stochastic volatility modeling. The case of more than one long memory parameter has, however, not been considered.

There are two ways to sample a continuous-time process over discrete epochs (Harvey (1990, p. 309)): Stock variables are instantaneous measurements of a continuous-time process, for example, as are instantaneous river-flows measured at a certain time in each

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

week. Flow variables, on the other hand, are aggregates which integrate the process over the sampling intervals, for example, annual tree ring growth. There is an extensive literature on temporal aggregation of univariate discrete-time long memory processes, see Beran et al. (2013, Section 2.2.1) and the references therein. For analogous results in the context of univariate continuous-time processes, see Chambers (1996, 1998), and Tsai and Chan (2005c).

Here, we are interested in modeling bivariate long-memory aggregate time series for at least two reasons. First, many data are aggregates, e.g., incidence rates of diseases, sales of products, industrial production, tree-ring widths, riverflows, and rainfall can only be obtained through aggregation over a certain time interval. Second, aggregation can simplify the model, e.g., we might be uninterested in the high-frequency variation that would require modeling the micro-noise structure. For a short-memory process, aggregate time-series data approach white noise with increasing aggregation due to the Central Limit Theorem. For a long-memory process, Tsai and Chan (2005c) show that the aggregates of a stationary univariate Continuous-time Auto-Regressive Fractionally Integrated Moving-Average (CARFIMA) process (Brockwell and Marquardt (2005) and Tsai and Chan (2005a and 2005b)) converges to a fractional Gaussian noise with increasing aggregation. The results of Tsai and Chan (2005c) can be readily extended to the case of stationary bivariate Continuous-time Auto-Regressive Fractionally Integrated (CARFI) processes driven by two independent standard fractional Brownian motions with two Hurst parameters: the approximation model of the aggregates is an instantaneous linear transformation of two independent fractional Gaussian noises that preserves the Hurst parameters. This approximation model provides a new framework for a simple yet general approach to analyzing bivariate aggregate long-memory time series. While the probabilistic properties of the limiting aggregate model are relatively well understood, its statistical inference can be complex in the parametric boundary where the two Hurst parameters agree, thereby requiring a careful study. We focus

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

on the bivariate case even though the limiting result holds for multivariate data, the inference of which will be pursued elsewhere.

The rest of the paper is organized as follows. The limiting behaviour of aggregates of bivariate continuous-time fractionally integrated AR processes are described in Section 2. The spectral maximum likelihood estimator (SMLE) and its large sample properties are discussed in Section 3. In Section 4, we derive the likelihood ratio test for testing the equality of two Hurst parameters within the framework of Whittle likelihood. In Section 5, we report some empirical performance of the fitting of the approximation model to aggregate data generated from a bivariate CARFI process,the SMLE, the test, and the effect of ignoring the equality of the two Hurst parameters on their estimation. We illustrate the use of the limiting aggregate model with application in Section 6. Section 7 concludes.

2 Limiting behaviour of aggregates of bivariate continuous-time fractionally integrated AR processes

For simplicity, we consider an AR(1) process in this section; the limiting results remain the same for a general AR(p) process. A bivariate continuous-time autoregressive fractionally integrated process of order one (CARFI(1, H_1, H_2)) $\{Y_t\}$ is the solution of a first order stochastic differential equation with suitable initial condition and driven by two independent standard fractional Brownian motions with Hurst parameters H_1 and H_2. For definitions and discussions of fractional Brownian motion, see, e.g., Mandelbrot and Van Ness (1968). The (CARFI(1, H_1, H_2)) process is also termed fractional Ornstein-Uhlenbeck process (see e.g. Cheridito, Kawaguchi, and Maejima (2003)). Specifically, for $t \geq 0$,

$$dY_t = \Phi Y_t dt + \Sigma d\bar{B}_t^{(H_1,H_2)}, \tag{1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

where $\bar{B}_t^{(H_1,H_2)} = [B_t^{H_1}, B_t^{H_2}]'$, the superscript $'$ denotes transpose, $\{B_t^{H_1}, t \geq 0\}$ and $\{B_t^{H_2}, t \geq 0\}$ are independent stochastic processes, $\{B_t^{H_k}, t \geq 0\}$ is a standard fractional Brownian motion with Hurst parameter $0 < H_k < 1$, where $k = 1, 2$, Φ and Σ are 2×2 matrices. Similar to Equation (5) of Tsai and Chan (2005a) in the univariate case, the solution of (1) can be written as

$$Y_t = e^{\Phi t} Y_0 + \int_0^t e^{\Phi(t-u)} \Sigma d\bar{B}_t^{(H_1,H_2)}, \tag{2}$$

where $e^{\Phi t} = I_2 + \sum_{n=1}^{\infty} \{(\Phi t)^n (n!)^{-1}\}$, and I_2 is the 2×2 identity matrix. Under the condition that all eigenvalues of Φ have strictly negative real parts, $\{Y_t, t \geq 0\}$ is asymptotically stationary. The initial condition Y_0 is assumed to be normally distributed with zero mean and finite variance or, possibly, non-random. The estimation of the multivariate CARFI model, to our knowledge, has not been considered in the literature. To circumvent the identification and estimation problem of the highly-parameterized model, we simplify the model by aggregation.

Consider the case that the continuous-time process $\{Y_t\}$ defined by (1) is digitalized by aggregation over interval Δ,

$$Y_n^{\Delta} = \int_{(n-1)\Delta}^{n\Delta} Y_u du, \qquad n = 1, 2, ... \tag{3}$$

Due to the central limit effect, the short-memory structure of the CARFI(1, H) process can be expected to vanish with increasing aggregation, $\Delta \to \infty$. Indeed, it can be shown that, as $\Delta \to \infty$,

$$- \begin{bmatrix} \Delta^{-H_1} & 0 \\ 0 & \Delta^{-H_2} \end{bmatrix} \Sigma^{-1} \Phi Y_n^{\Delta} \xrightarrow{d} \begin{bmatrix} B_n^{H_1} - B_{n-1}^{H_1} \\ B_n^{H_2} - B_{n-1}^{H_2} \end{bmatrix}, \tag{4}$$

two independent fractional Gaussian noises with Hurst parameters H_i, $i = 1, 2$. See the online supplementary material for a proof of (4), and Beran (1994, p. 53) for further discussions of fractional Gaussian noise. Therefore, for large Δ, Y_n^{Δ} can be approximated by

$$Y_n = CX_n$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

for some non-singular 2×2 matrix C, where X_n is a vector of two independent fractional Gaussian noises with Hurst parameters H_i, $i = 1, 2$, and $C = -\Phi^{-1}\Sigma \text{diag}(\Delta^{H_1}, \Delta^{H_2})$, where $\text{diag}(\Delta^{H_1}, \Delta^{H_2})$ is a diagonal matrix with the $\Delta^{H_i}, i = 1, 2$, as its diagonal elements. This motivates us to consider the use of an instantaneous linear transformation of two independent fractional Gaussian noises to model bivariate long-memory aggregate time series.

The spectral density matrix of $\{Y_n\}$ is

$$f(\omega; H_1, H_2, A) \;=\; 2(1 - \cos\omega) A G(\omega; H_1, H_2) A', \tag{5}$$

where $A = C\text{diag}(e(H_1), e(H_2))$, $e(H) = \{\Gamma(2H + 1)\sin(\pi H)/(2\pi)\}^{1/2}$, $G(\omega; H_1, H_2) = \text{diag}(R_0(\omega; H_1), R_0(\omega; H_2))$, $R_0(\omega; H_i) = \sum_{v=-\infty}^{\infty} |\omega + 2v\pi|^{-2H_i-1}$, for $i = 1, 2$. Let f_{jk} be the (j, k)-th element of the spectral density matrix f, then $f_{jk}(\omega) = O(|\omega|^{1-2H_1})$, for all j and k. It can be verified that the auto-covariance function corresponding to the spectral density function defined by (5) is

$$\gamma(h) = AP(h; H_1, H_2)A',$$

where $P(h; H_1, H_2) = \text{diag}(P(h; H_1), P(h; H_2))$, and for each integer j,

$$P(h; H_j) = \frac{\pi\Gamma(2 - 2H_j)}{2H_j\Gamma(1.5 - H_j)\Gamma(H_j + 0.5)}\{|h + 1|^{2H_j} - 2|h|^{2H_j} + |h - 1|^{2H_j}\}. \tag{6}$$

In (5), the computation of $f(\omega; H_1, H_2, A)$ requires the evaluation of infinite sums. Here, we adopt the method of Chambers (1996) to approximate $f(\omega; H_1, H_2, A)$ by

$$\tilde{f}(\omega; H_1, H_2, A) = 2(1 - \cos\omega) A \tilde{G}(\omega; H_1, H_2) A', \tag{7}$$

where $\tilde{G}(\omega; H_1, H_2) = \text{diag}(\tilde{R}_0(\omega; H_1), \tilde{R}_0(\omega; H_2))$ and, for each integer j, $\tilde{R}_0(\omega; H_j) = (4\pi H_j)^{-1}\{(2\pi M - \omega)^{-2H_j} + (2\pi M + \omega)^{-2H_j}\} + \sum_{k=-M}^{M} |\omega + 2k\pi|^{-2H_j-1}$ for some large integer M, with the approximation error of order $O(M^{-2H_j})$; see Chambers (1996).

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

3 Spectral maximum likelihood estimator and its large sample properties

In this section, we consider the application of the limiting aggregate model described in Section 2 for modeling bivariate aggregated long-memory data. Let $\{Y_t\}_{t=1}^N$ be a stationary bivariate time series with its spectral density matrix given by (5). For model identification, we assume $0 < H_2 \leq H_1 < 1$, and the first non-zero element in each column of the matrix A is positive. Because we are interested in applying model (5) for bivariate long memory time series analysis, we further assume $1/2 < H_1 < 1$.

The inference of the model depends on whether or not the parameter falls on the parametric boundary with identical Hurst parameters.

3.1 The case when $H_1 > H_2$

Let $I_Y(\omega) = J_Y(\omega)J_Y(\omega)^*/(2\pi N)$, where $J_Y(\omega) = \sum_{t=1}^N Y_t e^{it\omega}$, $J(\omega)^*$ denotes the conjugate transpose of $J(\omega)$. Let $\mathrm{tr}(A)$ and $\det A$ be the trace and the determinant of the matrix A, respectively, $\omega_j := 2\pi j/N \in (0, \pi)$ the Fourier frequencies, and T be the largest integer $\leq (N-1)/2$. Then the (negative) log-likelihood function of $\{Y_i\}$ can be approximated, up to a multiplicative constant, by the (negative) Whittle log-likelihood function (see Hosoya (1996))

$$\begin{aligned} -\tilde{l}(H_1, H_2, A) &= \sum_{i=1}^{T} \left[\log\det \tilde{f}(\omega_i; H_1, H_2, A) + \mathrm{tr}\{\tilde{f}(\omega_i; H_1, H_2, A)^{-1} I_Y(\omega_i)\} \right] \\ &= \sum_{i=1}^{T} \left\{ 2\log|\det A| + \log|2(1-\cos\omega_i)\tilde{G}(\omega_i; H_1, H_2)| \right\} \\ &\quad + \sum_{i=1}^{T} \mathrm{tr}\left[(A'^{-1}\{2(1-\cos\omega_i)\tilde{G}(\omega_i; H_1, H_2)\}^{-1} A^{-1} I_Y(\omega_i) \right]. \end{aligned} \tag{8}$$

The objective function (8) is minimized with respect to H_1, H_2, and A to get the spectral maximum likelihood estimators (SMLEs) $\widehat{H}_1$, $\widehat{H}_2$, and $\widehat{A}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

3.2 The case when $H_1 = H_2 = H$

If $H_1 = H_2 = H$, then the spectral density matrix given at (5) can be re-written as

$$f(\omega; H, B) \;=\; 2(1-\cos\omega)R_0(\omega; H)B, \tag{9}$$

where $B = [b_{ij}] = AA'$, and the auto-covariance function is $\gamma(h) = P(h; H)B$, where $P(h; H)$ is defined at (6). Under the condition that $H_1 = H_2$, we cannot identify the a_{ij}'s, we can only identify $b_{11} = a_{11}^2 + a_{12}^2$, $b_{12} = b_{21} = a_{11}a_{21} + a_{12}a_{22}$, and $b_{22} = a_{21}^2 + a_{22}^2$.

Taking the partial derivative of the log-likelihood function in (8) with respect to a_{ij} and equating the result to zero, we obtain

$$\begin{aligned} & 2T(A^{-1})' \\ = \;& (A^{-1})' \sum_{i=1}^{T} \{2(1-\cos\omega_i)\}^{-1} \tilde{G}(\omega_i; H_1, H_2)^{-1} A^{-1} \{I_Y(\omega_i) + I_Y(\omega_i)'^{-1})'. \end{aligned} \tag{10}$$

Under the condition that $H_1 = H_2 = H$, (10) becomes

$$2T(A^{-1})'^{-1})'^{-1} \sum_{i=1}^{T} \{2(1-\cos\omega_i)\tilde{R}_0(\omega_i; H)\}^{-1} \{I_Y(\omega_i) + I_Y(\omega_i)'^{-1})'. \tag{11}$$

As $B = AA'$, pre-multiplying both sides of (11) by B and post-multiplying by A', we get

$$B \;=\; \frac{1}{2T} \sum_{i=1}^{T} \left[\{2(1-\cos\omega_i)\tilde{R}_0(\omega_i; H)\}^{-1} \{I_N(\omega_i) + I_N(\omega_i)'\} \right]. \tag{12}$$

Substituting (12) into (8) yields the (profile) objective function

$$\begin{aligned} -\tilde{l}_0(H) \;&=\; 2T + \sum_{i=1}^{T} \log |2(1-\cos\omega_i)\tilde{R}_0(\omega_i; H)B| \\ &=\; 2T - 2T\log(2T) + 2\sum_{i=1}^{T} \log |(1-\cos\omega_i)\tilde{R}_0(\omega_i; H)| \\ &\quad + T \log \left| \sum_{i=1}^{T} \left[\{(1-\cos\omega_i)\tilde{R}_0(\omega_i; H)\}^{-1} \{I_N(\omega_i) + I_N(\omega_i)'\} \right] \right|. \end{aligned} \tag{13}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

The objective function (13) is minimized with respect to H to get the spectral maximum likelihood estimator (SMLE) $\widehat{H}$; the estimator of B is then calculated by (12).

For simplicity, if $H_1 > H_2$, let $\theta = (\theta_1, \cdots, \theta_6) = (H_1, H_2, a_{11}, a_{12}, a_{21}, a_{22})$, and if $H_1 = H_2 = H$, let $\theta = (\theta_1, \cdots, \theta_4) = (H, b_{11}, b_{12}, b_{22})$, and $\hat{\theta}$ be the spectral maximum likelihood estimator of θ. Proofs of our results are deferred to the Appendix with some further details provided in the Supplementary Material.

Theorem 1 *Assume the data $Y = \{Y_i\}_{i=1}^N$ are sampled from a stationary Gaussian long-memory process with the spectral density given by (5) or (9), and that the spectral maximum likelihood estimator $\hat{\theta} \in \Theta$, a compact parameter space, with the true parameter θ_0 in the interior of Θ. If the truncation parameter M increases with the sample size so that $M \to \infty$, and if $\sqrt{N}M^{-1} \to 0$ as $N \to \infty$, then $\sqrt{N}(\hat{\theta} - \theta_0)$ converges in distribution to a normal random vector with mean 0 and covariance matrix $\Gamma(\theta_0)^{-1}$, where the (i,j)-th element of $\Gamma(\theta)$ is given by*

$$\Gamma_{ij}(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} tr\left[f(\omega;\theta)^{-1} \frac{\partial f(\omega;\theta)}{\partial \theta_i} f(\omega;\theta)^{-1} \frac{\partial f(\omega;\theta)}{\partial \theta_j} \right] d\omega,$$

and $f(\omega;\theta)$ is the spectral density function defined by (5) or (9), respectively.

4 Testing the equality of the Hurst parameters

Given that in the boundary case of identical Hurst parameters, A is no longer identifiable and the model requires a different parametrization, it is pertinent to assess whether the boundary case occurs. Hence, we are interested in testing the hypotheses

$$\begin{cases} H_0 : H_1 = H_2 = H, \\ H_A : H_1 > H_2. \end{cases}$$

If $\widehat{H}_0$ is the SMLE of H under H_0, and $\widehat{H}_1$, $\widehat{H}_2$, and $\widehat{A}$ are the SMLEs of H_1, H_2, and A under H_A, then minus twice of the log-likelihood ratio is given by $\ell =$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

$2\{\tilde{l}(\widehat{H}_1, \widehat{H}_2, \widehat{A}) - \tilde{l}_0(\widehat{H}_0)\}$, where $\tilde{l}$ and $\tilde{l}_0$ are defined at (8) and (13), respectively, and one rejects H_0 if ℓ is too large. Under H_0, there are four parameters, whereas there are six parameters under H_A.

Theorem 2 *(a) If the conditions of Theorem 1 hold, under H_0, the asymptotic distribution of ℓ is an equal probabilistic mixture of an atom at zero and a χ^2 random variable with two degrees of freedom,and (b) under H_A, the test is consistent.*

Thus, for a test of size α we can simply use the 2α rather than the α significance point of a χ^2 distribution with two degrees of freedom. The non-standard asymptotic behavior of the log-likelihood-ratio statistic in Theorem 2 is due to the non-identifiability of the nuisance matrix A. Under the constraint that $H_1 = H_2$, the matrix A is not identifiable but rather it is the symmetric matrix $B = AA'$ that is identifiable. There is an extensive literature on hypothesis testing when some of the standard regularity conditions fail to hold; see, for example, Andrews (1998 and 2001) and references therein. However, our problem is different from the existing ones in the literature. In our setting, the nuisance parameters under the null are functions of the nuisance parameters under the alternative, with the number of identifiable nuisance parameters reduced from four under the alternative to three under the null. The proof technique relies on the Karush-Kuhn-Tucker (KKT) optimality condition for constrained optimization and exploits the manifold structure of the nuisance parameters; this is quite different than existing techniques used in the literature.

5 Simulation

In this section, we report some finite sample performance of the spectral maximum likelihood estimator and the test. For the performance of the proposed estimation method, we consider two data generating processes (DGP). In Subsection 5.2, the spectral density of the DGP takes the form (5) or (9). Given that aggregation is finite, fitting the

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

limiting model (5) to aggregate data may result in bias, even though the bias vanishes with increasing aggregation. So, it is pertinent to study the empirical performance of the estimator of the long-memory parameter by fitting the proposed limiting model to aggregate data generated by (1) and (3). We report some simulation results for clarification in Subsection 5.1. The size and the power of the test are studied in Subsections 5.3 and 5.4, respectively. To demonstrate the importance of the test, we report on a simulation study in Subsection 5.5 that investigated the efficiency loss of ignoring the equality of the Hurst parameters when they are indeed equal. All computations in this section were performed using Fortran code with IMSL subroutines with the minimization performed by the NCONF subroutine.

5.1 Fitting the limiting model to aggregate data

To assess the use of the limiting models established in Section 2 for estimating the Hurst parameters with aggregate data generated from a CARFI process, we conducted the following experiment. Consider the two data generating processes

Model 1:

$$\begin{bmatrix} dY_{1t} \\ dY_{2t} \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} dt + \begin{bmatrix} 2\Delta^{-0.7} & 0 \\ 0 & 2\Delta^{-0.6} \end{bmatrix} \begin{bmatrix} dB_t^{0.7} \\ dB_t^{0.6} \end{bmatrix},$$

Model 2:

$$\begin{bmatrix} dY_{1t} \\ dY_{2t} \end{bmatrix} = \begin{bmatrix} -2 & 0.5 \\ 0.5 & -1 \end{bmatrix} \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} dt + \begin{bmatrix} 2\Delta^{-0.7} & \Delta^{-0.6} \\ \Delta^{-0.7} & 2\Delta^{-0.6} \end{bmatrix} \begin{bmatrix} dB_t^{0.7} \\ dB_t^{0.6} \end{bmatrix},$$

where $\{B_t^{0.7}\}$ and $\{B_t^{0.6}\}$ are independent fractional Brownian motions with Hurst parameters 0.7 and 0.6, respectively. Model 1 is a vector of two independent univariate CARFI$(1, H_i)$, $i = 1, 2$, processes with $(H_1, H_2) = (0.70, 0.60)$. For positive integers m,

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

N, and Δ, let $T = N\Delta$. We used the method of Chambers (1995) to simulate observations $\{Y_{k/m}\}_{k=1}^{mT}$ from these models. Based on $\{Y_{k/m}\}_{k=1}^{mT}$, we computed $\{Y_n^{\Delta}\}$ (defined by (3)) by Euler approximation. Thus, for $n = 1, ..., N$, we approximated Y_n^{Δ} by

$$\frac{1}{m} \sum_{k=m(n-1)\Delta+1}^{mn\Delta} Y_{k/m}, \tag{14}$$

Then, we fit $\{Y_n^{\Delta}\}_{n=1}^{N}$ by (5), and found the estimates of H_1, H_2, and A by maximizing the Whittle likelihood (8). For the values of (m, Δ, N), we considered four cases: $(20, 10, 125)$, $(40, 10, 125)$, $(20, 20, 125)$, and $(20, 10, 250)$. The averages and standard errors of the estimates over 1,000 replications are summarized in Table 1.

From Table 1, it is clear that the biases of the estimates of the parameters are small, indicating the good approximation of the limiting aggregate model for the aggregate data generated from a CARFI process. The biases decrease, in general, with m, Δ, and N.

5.2 Spectral maximum likelihood estimation

In this subsection, we report some finite sample performance of the spectral maximum likelihood estimator for data simulated from stationary bivariate Gaussian processes with spectral density defined by (5) or (9). We considered two cases: (I) the two Hurst parameters are unequal so that A is identifiable, and (II) the two Hurst parameters are identical so only $B = AA'$ is identifiable. For case (I), we set

$$A = \begin{pmatrix} 2 & 1 \\ -3 & 1 \end{pmatrix}$$

with three pairs of increasingly separated Hurst parameters $(H_1 = 0.75, H_2 = 0.70)$, $(H_1 = 0.85, H_2 = 0.40)$ and $(H_1 = 0.95, H_2 = 0.10)$. When the Hurst parameters are close to each other, the A matrix is close to being non-identifiable. Thus, the elements of A are harder to estimate in this case. For case (II), data from three models were

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

Table 1: Averages and standard deviations (in parentheses) [asymptotic standard deviations] of 1,000 replications of the spectral maximum likelihood estimators of the parameters H_1, H_2, and $A = [a_{ij}]_{i,j=1}^{2,2}$ when the approximation model is applied to aggregates from the CARFI processes of Models 1 and 2.

(m, Δ, N)	Parameter	True value	Model 1	True value	Model 2
$(20, 10, 125)$	H_1	0.70	0.7196(0.0609)[0.0585]	0.70	0.7247(0.0580)[0.0585]
	H_2	0.60	0.6038(0.0608)[0.0573]	0.60	0.5984(0.0610)[0.0573]
	a_{11}	0.3999	0.3527(0.0856)[0.0262]	0.5713	0.5198(0.2336)[0.2005]
	a_{12}	0.0	0.0875(0.1166)[0.1590]	0.4667	0.3415(0.2461)[0.2292]
	a_{21}	0.0	-0.1823(0.2563)[0.3447]	0.6856	0.6493(0.4975)[0.4454]
	a_{22}	0.8168	0.6981(0.1703)[0.0541]	1.0501	0.8133(0.2993)[0.2813]
$(40, 10, 125)$	H_1	0.70	0.7226(0.0604)[0.0585]	0.70	0.7290(0.0596)[0.0585]
	H_2	0.60	0.6043(0.0625)[0.0573]	0.60	0.5963(0.0612) [0.0573]
	a_{11}	0.3999	0.3571(0.0838)[0.0262]	0.5713	0.5104(0.2360)[0.2005]
	a_{12}	0.0	0.0832(0.1154)[0.1590]	0.4667	0.3511(0.2414)[0.2292]
	a_{21}	0.0	-0.1725(0.2544)[0.3447]	0.6856	0.6260(0.4987) [0.4454]
	a_{22}	0.8168	0.6993(0.1618)[0.0541]	1.0501	0.8247(0.2908)[0.2813]
$(20, 20, 125)$	H_1	0.70	0.7148(0.0609)[0.0585]	0.70	0.7217(0.0601)[0.0585]
	H_2	0.60	0.5896(0.0628)[0.0573]	0.60	0.5805(0.0614) [0.0573]
	a_{11}	0.3999	0.3645(0.0768)[0.0262]	0.5713	0.5046(0.2289)[0.2005]
	a_{12}	0.0	0.0772(0.1094)[0.1590]	0.4667	0.3957(0.2307)[0.2292]
	a_{21}	0.0	-0.1618(0.2456)[0.3447]	0.6856	0.6020(0.4824)[0.4454]
	a_{22}	0.8168	0.7230(0.1511)[0.0541]	1.0501	0.8939(0.2829) [0.2813]
$(20, 10, 250)$	H_1	0.70	0.7170(0.0411)[0.0414]	0.70	0.7217(0.0393) [0.0414]
	H_2	0.60	0.6170(0.0419)[0.0405]	0.60	0.6122(0.0405) [0.0405]
	a_{11}	0.3999	0.3734(0.0620)[0.0185]	0.5713	0.5574(0.1768)[0.1418]
	a_{12}	0.0	0.0618(0.0981)[0.1124]	0.4667	0.3478(0.2067)[0.1621]
	a_{21}	0.0	-0.1261(0.2130)[0.2437]	0.6856	0.7056(0.3843) [0.3150]
	a_{22}	0.8168	0.7373(0.1228)[0.0382]	1.0501	0.8472(0.2587)[0.1989]

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

simulated,

$$H = 0.55, B = \begin{pmatrix} 26 & -13 \\ -13 & 13 \end{pmatrix}; H = 0.75, B = \begin{pmatrix} 13 & -11 \\ -11 & 17 \end{pmatrix};$$
$$H = 0.95, B = \begin{pmatrix} 13 & 5 \\ 5 & 17 \end{pmatrix}.$$

Thus, the simulated models were of increasing long-memory whereas the two component time series were negatively contemporaneously correlated in the first case, then very strong negatively contemporaneously correlated in the second case and finally mildly positively contemporaneously correlated. The sample sizes considered were $N = 512$ and $N = 1{,}024$. The value of M used in the computation of $\tilde{R}_0(\omega; H)$, defined below (7), was 100. We also used $M = 1{,}000$ in the simulations, but the results were essentially the same.

The averages and the standard errors of 1,000 replicates of the estimators for the case of $H_1 > H_2$ are reported in Table 2, whereas those for the case of $H_1 = H_2$ are summarized in Table 3. In Tables 2 and 3, we also report the asymptotic standard errors of the estimators. The true values of the parameters are given in the tables. Tables 2 and 3 show that the sample standard deviations of the estimates are generally close to the asymptotic standard errors for sample sizes 512 and 1,024, and the biases and the standard deviations decrease with increasing sample size. Also, the estimates of the elements in A have higher variances when the Hurst parameters are closer to each other. Quantile-quantile normal score plots (unreported) confirm that the estimates are close to normality. The proposed method works well even when A, respectively B, is almost singular. Corresponding simulation results are available upon request.

5.3 Empirical size of the test

To study the empirical size of the proposed likelihood ratio test for the equality of the two Hurst parameters, we simulated data from a stationary bivariate Gaussian process with

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

Table 2: Averages (standard deviations) [asymptotic standard deviations] of 1,000 simulations of the spectral maximum likelihood estimators of the parameters H_1, H_2, and $A = [a_{ij}]_{i,j=1}^{2,2}$.

parameter	true value	N=512	N=1,024
H_1	0.75	0.7571(0.0273)[0.0292]	0.7535(0.0203)[0.0206]
H_2	0.70	0.6923(0.0296)[0.0289]	0.6962(0.0209)[0.0204]
a_{11}	2.0	1.8481(0.4200)[0.4253]	1.8955(0.3259)[0.3007]
a_{12}	1.0	1.0162(0.5980)[0.8185]	1.0008(0.5295)[0.5787]
a_{21}	-3.0	-2.8768(0.3200)[0.4313]	-2.9054(0.2874)[0.3050]
a_{22}	1.0	0.8231(0.9753)[1.2271]	0.8840(0.8241)[0.8677]
H_1	0.85	0.8508(0.0295)[0.0295]	0.8508(0.0212)[0.0209]
H_2	0.40	0.3992(0.0278)[0.0262]	0.3994(0.0194)[0.0185]
a_{11}	2.0	1.9961(0.0795)[0.0742]	1.9979(0.0548)[0.0524]
a_{12}	1.0	0.9968(0.0885)[0.0832]	1.0007(0.0612)[0.0589]
a_{21}	-3.0	-3.0006(0.1061)[0.1031]	-2.9993(0.0729)[0.0729]
a_{22}	1.0	0.9928(0.1229)[0.1187]	0.9931(0.0869)[0.0840]
H_1	0.95	0.9513(0.0283)[0.0298]	0.9518(0.0214)[0.0211]
H_2	0.10	0.1036(0.0181)[0.0199]	0.1022(0.0123)[0.0141]
a_{11}	2.0	2.0044(0.0667)[0.0642]	2.0022(0.0467)[0.0454]
a_{12}	1.0	1.0069(0.0794)[0.0631]	1.0065(0.0533)[0.0446]
a_{21}	-3.0	-3.0067(0.0989)[0.0960]	-3.0034(0.0689)[0.0679]
a_{22}	1.0	1.0074(0.0846)[0.0731]	1.0041(0.0584)[0.0517]

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

Table 3: Averages (standard deviations) [asymptotic standard deviations] of 1,000 simulations of the spectral maximum likelihood estimators of the parameters H, and $B = [b_{ij}]_{i,j=1}^{22}$.

parameter	true value	N=512	N=1,024
H	0.55	0.5503(0.0197)[0.0197]	0.5500(0.0146)[0.0140]
b_{11}	26.0	25.96(1.7183)[1.6703]	25.96(1.1606)[1.1810]
b_{12}	-13.0	-13.00(0.9960)[1.0137]	-12.97(0.7030)[0.7168]
b_{22}	13.0	13.00(0.8102)[0.8351]	12.97(0.5664)[0.5905]
H	0.75	0.7512(0.0209)[0.0206]	0.7507(0.0153)[0.0146]
b_{11}	13.0	13.03(0.8166)[0.8253]	12.99(0.5685)[0.5836]
b_{12}	-11.0	-11.04(0.7993)[0.8264]	-10.99(0.5616)[0.5844]
b_{22}	17.0	17.04(1.0596)[1.0792]	16.99(0.7453)[0.7631]
H	0.95	0.9541(0.0219)[0.0211]	0.9528(0.0159)[0.0149]
b_{11}	13.0	13.10(0.8600)[0.8215]	13.07(0.5923)[0.5809]
b_{12}	5.0	5.015(0.7335)[0.6947]	5.037(0.4899)[0.4912]
b_{22}	17.0	17.15(1.1299)[1.0743]	17.08(0.7685)[0.7596]

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

Table 4: The empirical size of the likelihood ratio test for identical Hurst parameters. The results are based on 1,000 replications, with nominal sizes 0.01, 0.05, and 0.1.

		Nominal level					Nominal level		
H	N	0.100	0.050	0.010	H	N	0.100	0.050	0.010
0.55	512	0.130	0.070	0.013	0.55	1024	0.126	0.057	0.010
0.65	512	0.132	0.063	0.012	0.65	1024	0.114	0.049	0.011
0.75	512	0.134	0.066	0.013	0.75	1024	0.124	0.053	0.010
0.85	512	0.134	0.069	0.014	0.85	1024	0.120	0.053	0.010
0.95	512	0.109	0.051	0.012	0.95	1024	0.107	0.050	0.008

its spectral density defined by (5), and $(a_{11}, a_{12}, a_{21}, a_{22}) = (1, 1, -1, 2)$. The common Hurst parameter H ranged from 0.55 to 0.95 with increment 0.1. Sample sizes were $N = 512$ and $N = 1,024$. From Table 4, we observe that the empirical sizes are generally close to their corresponding nominal sizes, with better agreement as sample size increases.

5.4 Empirical power of the likelihood ratio test

Next, we study the empirical power of the likelihood ratio test. Data were simulated from stationary bivariate Gaussian processes with the spectral density defined by (5), with various pairs of H_1 and H_2 and $(a_{11}, a_{12}, a_{21}, a_{22}) = (1, 1, -1, 2)$. The results are summarized in Table 5. Table 5 confirms that the power increases with the distance between H_1 and H_2, which corresponds to the deviation from the null hypothesis, $H_1 = H_2$. The dominating divergent component in the test is proportional to $H_1 - H_2$ as can be seen in the proof of Theorem 2(b). Further, for a constant difference between the memory parameters, the power decreases very slightly with H_1.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

Table 5: The empirical power of the likelihood ratio test for identical Hurst parameters. The results are based on $1,000$ replications and critical values of the nominal size 0.05.

H_1	0.55	0.65	0.75	0.80	0.85	0.85	0.90	0.95	0.95
H_2	0.50	0.60	0.70	0.70	0.70	0.80	0.70	0.10	0.90
$N = 512$	0.257	0.247	0.242	0.627	0.921	0.246	0.997	1.000	0.238
$N = 1{,}024$	0.441	0.419	0.404	0.903	0.996	0.403	0.999	1.000	0.379

5.5 The effect of ignoring the equality of the two Hurst parameters on their estimations

In this subsection, we report on a study on investigating the efficiency loss of ignoring the fact that $H_1 = H_2 = H$. We took the data to be generated from (9) with $\text{vec}A = [1, -3, 5, -2]'$. For each data set, the constrained and unconstrained estimators were computed with the constraint being that $H_1 = H_2 = H$. The root mean square errors (RMSE's) of the unconstrained estimators $\widehat{H}_i$, $i = 1, 2$, over the constrained estimator $\widehat{H}$ are summarized in Table 6 for the cases that the true values of H was $0.55, 0.65, 0.75, 0.85$, and 0.95. The results are summarized in Table 6. From the table, we see that the efficiency losses of ignoring the equality of the Hurst parameters are between 10% and 45%. These losses are substantial and illustrate the usefulness of the proposed methodology.

6 Application

We illustrate the proposed method with an equally-spaced trivariate time series of the positioning errors (in inches) in the"x", "y", and "z" direction, respectively, of an industrial robot named Stu. (Data were kindly supplied by Bill Fulkerson, Deere and Company, Technical Center, 3300 River Drive, Moline, IL, U.S.A.) Sample size were

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

Table 6: Ratio of the root mean square errors (RMSE's) of the estimators $\widehat{H}_i$, $i = 1, 2$, over $\widehat{H}$ when the true model is $H_1 = H_2 = H$.

	true value of H	0.55	0.65	0.75	0.85	0.95
ratio for $\widehat{H}_1$ over $\widehat{H}$	$N = 512$	1.368	1.360	1.443	1.426	1.106
	$N = 1{,}024$	1.332	1.331	1.381	1.352	1.213
ratio for $\widehat{H}_2$ over $\widehat{H}$	$N = 512$	1.365	1.390	1.385	1.387	1.235
	$N = 1{,}024$	1.313	1.338	1.323	1.332	1.226

$N = 2{,}703$.

The time series plots (Figure 1) and the sample ACFs (Figure 2) suggest that the time series of the positioning errors in the "x" and "y" directions have long memory, while that of the "z" direction appears to be nonstationary. For illustration, we restricted our analysis to the bivariate time series $\{(X_t, Y_t), t = 1, 2, \ldots, N\}$ consisting of the positioning errors in the "x" and "y" directions, respectively. As it is known that spurious cross correlations may occur because of the autocorrelation in each sequence, it is pivotal to study the correlation pattern after pre-whitening the time series, see, e.g., Cryer and Chan (2008, Section 11.3). Fig. 3 displays the sample cross-correlations between the pre-whitened X and Y when a cross-correlation is individually significant at 5% significance level if it exceeds the threshold $\pm 1.96/\sqrt{N} = \pm 0.0377$ in magnitude (the blue dashed lines in the figure). To avoid the problem of multiple testing, the threshold was adjusted by Bonferroni correction to be 0.0811 for achieving 5% family error rate. Consequently, Figure 3 shows that X and Y are contemporaneously correlated but not cross-correlated at any non-zero lag examined.

In order to check the forecasting performance of the bivariate limiting aggregate long memory model, we saved the last 100 data point from fitting the model with which we compared the long-run forecast for these 100 observations: we computed the k-step

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

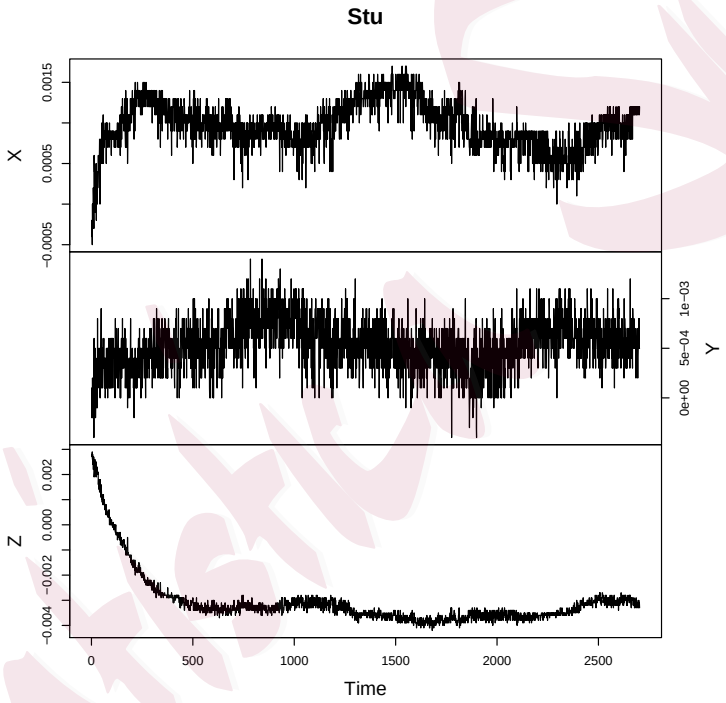

Figure 1: Time series plots of the positioning errors in the three directions

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

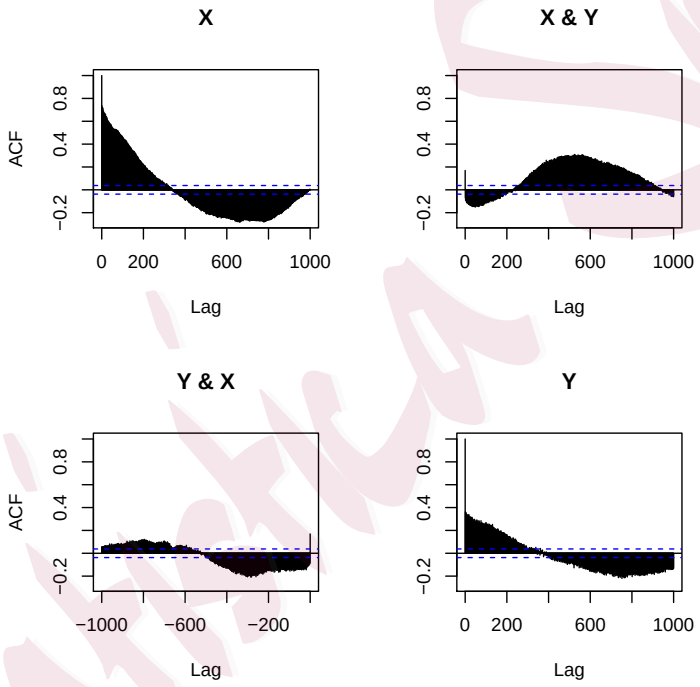

Figure 2: Sample ACFs of X and Y, and cross-correlations of X vs Y

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

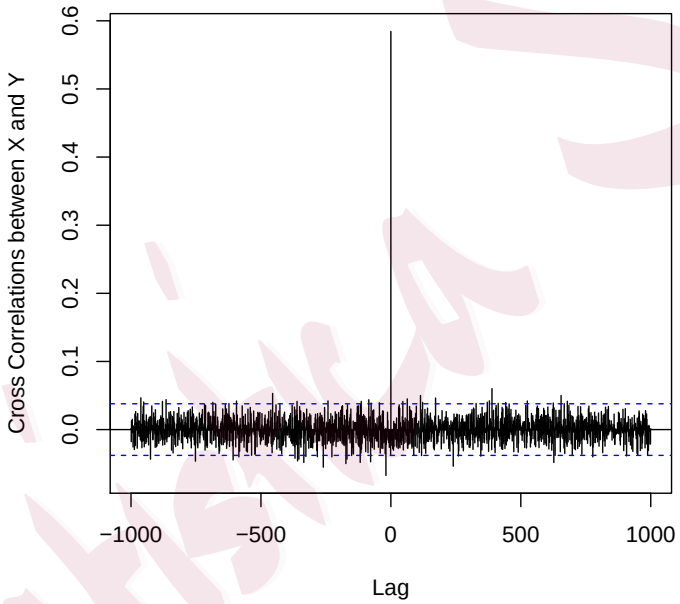

Figure 3: Cross-correlations between X & Y, after pre-whitening; cut-off for 5% family error rate is 0.0811.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

ahead predictor to forecast the k-th saved observation, for $k = 1, 2, \ldots, 100$. We also compared the forecasting performance of our model with a VAR(19) model, with the VAR order determined by AIC, both in terms of the 95% prediction ellipsoids and the point predictors. It can be seen from the plots in Figure 4 that the predicted values (conditional means) of the proposed model track the positioning errors better than the VAR model; indeed, over the 100 observations, the root mean squared prediction error (mean absolute prediction error) of the proposed model was 1.84×10^{-4} (1.51×10^{-4}) as compared to the 3.41×10^{-4} (2.82×10^{-4}) of the VAR model. The prediction gain is substantial, especially in view of the fact that the proposed model has only six parameters and much more parsimonious than the VAR(19) model. When compared to VAR(p) models with $p < 19$ the prediction gains of the proposed method were even larger. The confidence bands of the proposed method were, as expected, slightly wider than those of the VAR model, and the individual 95% confidence ellipsoids from both models missed 3% of the 100 "future" values.

The p-value of the test for equality of the Hurst parameters using the first $N = 2{,}603$ data points turned out to be 0.0000. Therefore, we conclude that $H_1 > H_2$. The estimated Hurst parameters were $\widehat{H}_1 = 0.8697$, $\widehat{H}_2 = 0.6904$, and

$$\widehat{A} \;=\; 10^{-5} \times \begin{bmatrix} 5.7851 & 4.9164 \\ -1.1850 & 9.3053 \end{bmatrix}.$$

Therefore,

$$\widehat{A}^{-1} \;=\; 10^{3} \times \begin{bmatrix} 15.60 & -8.241 \\ 1.986 & 9.697 \end{bmatrix}.$$

Our fitted model implies that $A^{-1}(Y_{1t}, Y_{2t})'$ are uncorrelated series, which provides us with one way to do model diagnostics. Figure 5 displays the cross correlations between the transformed X and Y, after prewhitening. Based on the Bonferroni correction, a sample correlation is significant at a 5% family error rate, if it exceeds 0.0811 in

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

magnitude. Consequently, none of the cross correlations is significant, suggesting that the model is correctly specified.

7 Conclusion

In this paper, we consider the problem of modeling long-memory bivariate time series that are aggregates from an underlying long-memory continuous-time process by an instantaneous linear transformation of two independent fractional Gaussian noises. We provide theoretical and numerical justifications for our approach, and derive the Whittle likelihood estimator of the limiting model and the likelihood ratio test for testing the equality of the two underlying Hurst parameters within the framework of Whittle likelihood. An application illustrates that the proposed model can provide a parsimonious model that outperforms the VAR model in terms of forecasting. A systematic study of extending the proposed approach to multivariate long-memory aggregate time series of arbitrary dimension is an interesting research problem.

Supplementary Materials

The online Supplement file contains the proofs of (4) and Claim 2, and further details of proofs of Theorem 1.

Acknowledgements

Henghsiu Tsai thanks Academia Sinica, the National Science Council (NSC 98-2118-M-001-023-MY2), R.O.C., Heiko Rachinger thanks the Spanish Ministery for Science and Innovation (EEBB-2011-43610) and Kung-Sik Chan thanks U.S. National Science Foundation (DMS-1021292), for partial support. Part of this paper was written while Heiko Rachinger was visiting Academia Sinica. The authors thank two anonymous referees, an associate editor, and the Co-Editor for their helpful comments and suggestions.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

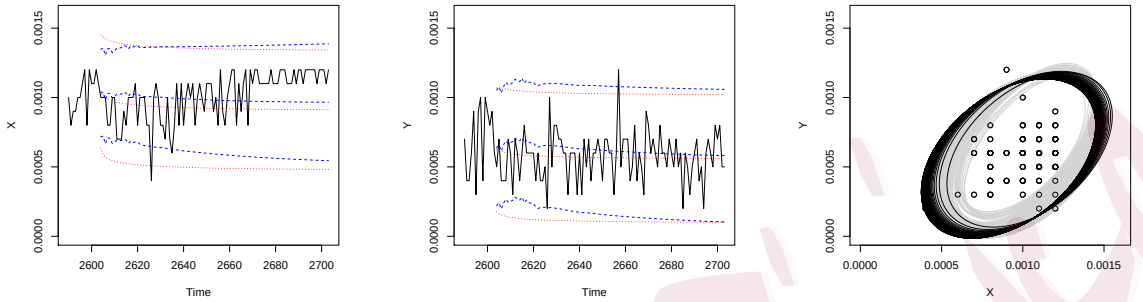

Figure 4: First two figures: black solid line – observations, red dotted line – forecasts from proposed model, blue dashed line – VAR model; last figure: circles – observations, black lines: prediction ellipsoids from proposed model, gray lines: VAR model.

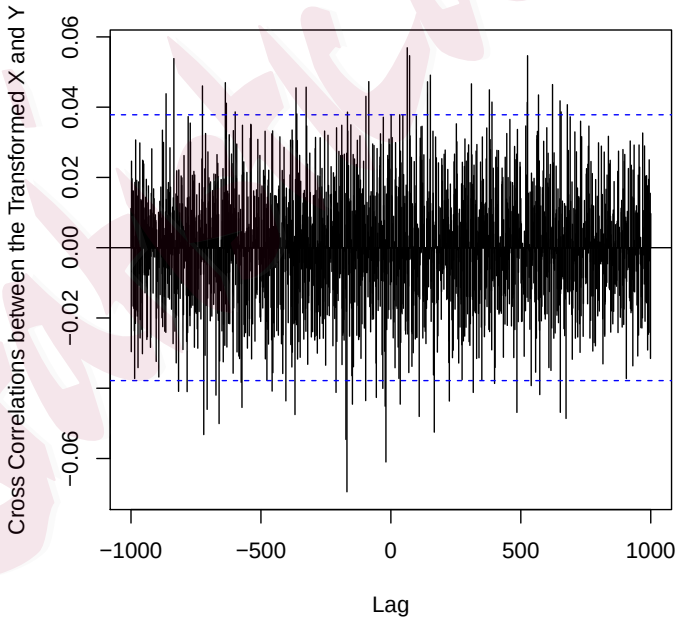

Figure 5: Cross correlations between the transformed X and Y, after prewhitening; cut-off for 5% family error rate is 0.0811.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

Appendix

Proof of Theorem 1

We prove Theorem 1 for the case that $H_1 > H_2$, and $\theta = (\theta_1, \cdots, \theta_6)$ $= (H_1, H_2, a_{11}, a_{12}, a_{21}, a_{22})$. The case that $H_1 = H_2 = H$, and $\theta = (\theta_1, \cdots, \theta_4) = (H, b_{11}, b_{12}, b_{22})$ can be proved similarly, and is omitted.

As in Chan and Tsai (2012), the approximation of the Whittle likelihood (5) by its truncated counterpart (7) preserves the asymptotic distribution of the estimator, under the stated growth rate of M. In particular, from adapting Tsai (2006), the approximation (5) by (7) is then asymptotically negligible. Further, due to compactness of the parameter space, the partial derivatives of the approximate Whittle likelihood differ from the ones of the true one by an error of order $o_p(1)$, uniformly over the parameter space. Finally, from Theorem 1 and Lemma 1 in Hosoya (1996), and from the approximation error of the partial derivatives of the approximate Whittle likelihood of order $O_p(M^{-2H})$, the asymptotic distribution of the QMLE from the approximated Whittle corresponds to the one of the true Whittle likelihood. Hence, with no loss of generality, we assume that the estimator is the exact QMLE, with which we can apply results in Hosoya (1996) to derive its large-sample distribution.

Theorem 1 follows from Theorem 2 of Hosoya (1996) if we can verify Conditions A, C, and D listed there. In the *Supplemental Materials*, we verify these conditions.

Proof of Theorem 2

Part (a)

The parameter space Θ for $\theta = (H_1, H_2, a_{11}, a_{12}, a_{21}, a_{22})'$ admits the restrictions that $0 < H_2 \leq H_1 < 1$, $1/2 < H_1 < 1$, and A is invertible. Under the constraint

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

$H_1 = H_2$, the matrix A is not identifiable but $B = AA'$ is. Here the entries are

$$\begin{aligned} b_{11} &= a_{11}^2 + a_{12}^2 \\ b_{12} &= a_{11}a_{21} + a_{12}a_{22} \\ b_{22} &= a_{21}^2 + a_{22}^2 \end{aligned}$$

Under the null hypothesis, let the true values of the b's be denoted by $b = b_0$ where $b = (b_{11}, b_{12}, b_{22})'$, where b is a function of $a = (a_{11}, a_{12}, a_{21}, a_{22})'$. We claim that the parameter space Θ with $H_1 \neq H_2$ is a manifold with an atlas consisting of charts where b is always part of the co-ordinate system. To see this, consider the first derivative of b w.r.t. $a = (a_{11}, a_{12}, a_{21}, a_{22})'$:

$$Db = \begin{pmatrix} 2a_{11} & 2a_{12} & 0 & 0 \\ a_{21} & a_{22} & a_{11} & a_{12} \\ 0 & 0 & 2a_{21} & 2a_{22} \end{pmatrix}.$$

The invertibility of A implies that Db is of full-rank. Let a_0 be the vectorized form of an arbitrary but fixed, invertible A, some of whose elements are non-zero. With no loss of generality, suppose $a_{11,0} \neq 0$, where $a_{11,0}$ is the value of a_{11} under the null. Then the first derivative matrix of the function $(b', a_{12})'$ is Db augmented by the row vector $(0, 1, 0, 0)$, and its determinant is $4a_{11}|A|$ which is clearly non-zero when it is evaluated at a_0. Hence there exists a locally 1-1 C^1 transformation ψ such that $a = \psi(b', a_{kl})$ over a neighbourhood of a_0 (where we can take $k = 1$ and $l = 2$ if $a_{11,0} \neq 0$). We augment this function to form a chart as $(H, h, b', a_{kl})' \to (H_1 = H + h, H_2 = H - h, a = \psi(b', a_{kl}))'$, and hence the claim on the manifold structure. (Clearly, $H = (H_1 + H_2)/2$ and $h = (H_1 - H_2)/2$.) Moreover, this shows that we can always re-parameterize the model locally in terms of H, h, b and one of the $a_{i,j}$'s. Since there are at most four such co-ordinate systems, for simplicity, we assume that the compact parameter space is such that it can be covered by the co-ordinate system where $k = 1$ and $l = 2$ and that $a_{11} \neq 0$ over the parameter space (for a technical reason that will be clear later

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

on). The following arguments can be readily modified to handle the general case. Under the null hypothesis, $H = H_0$, the true Hurst parameter, $h = 0$ and $b = b_0$ for some b_0, but the "extra" a_{11} co-ordinate is irrelevant. We can remove the arbitrariness of a_{11} under the null hypothesis by considering the local co-ordinate $\vartheta = (h, k, H, b')'$, where $k = ha_{11}$. It is clear that, for $h > 0$, ϑ bears a smooth, one-to-one relationship with the co-ordinate $(h, a_{11}, H, b')'$ but with $k = 0$ under the null hypothesis, thereby removing the arbitrariness of a_{11} under the null hypothesis. Let Ξ be the parameter space of ϑ which is then a compact set. Let $S_N(\theta)$ be the partial derivative of the objective function (8) w.r.t. θ. Then it follows from the chain rule that the partial derivative of the objective function (8) w.r.t. ϑ equals $S_N(\vartheta) = \{J'(\vartheta)\}^{-1} S_N(\theta)$ where

$$J(\vartheta) = \frac{\partial \vartheta}{\partial \theta} = \begin{pmatrix} 1/2 & -1/2 & 0 & 0 & 0 & 0 \\ a_{11}/2 & -a_{11}/2 & h & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2a_{11} & 2a_{12} & 0 & 0 \\ 0 & 0 & a_{21} & a_{22} & a_{11} & a_{12} \\ 0 & 0 & 0 & 0 & 2a_{21} & 2a_{22} \end{pmatrix}.$$

Let $S_N(\vartheta)$ be partitioned into $S_{i,N}(\vartheta), i = 1, 2, 3$ according to the partition of ϑ into $h, k, \vartheta^* = (H, b_{11}, b_{12}, b_{22})'$. The true model parameter lies on the 1-dimensional manifold defined by $H = H_0, h = 0, k = 0, b = b_0$, where a_{11} is the free parameter. By adapting the proof of Theorem 2.1 of Hosoya (1997), the following claim can be readily verified.

Claim 1:

Suppose that there exists a sequence of estimators $\tilde{\vartheta}_N$ such that $S_{3,N}(\tilde{\vartheta}_N) = 0$, so $\tilde{H}_N$ and $b(\tilde{\vartheta}_N)$ converge to H_0 and b_0 in probability.

The spectral estimator under the general hypothesis, $\hat{\vartheta}_N$ converges to the set $\{\vartheta \in \Xi : b = b_0, H = H_0\}$, in probability. On the other hand, it can be shown that the constrained estimator $\hat{\vartheta}_{N,0}$ under the null hypothesis is also consistent. Here $\hat{\vartheta}_N = \hat{\vartheta}_{N,0}$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

is equivalent to the log-likelihood ratio test statistic being equal to 0, which obtains if the partial derivative of the objective function w.r.t. h, denoted by $S_{1,N}(\vartheta_N)$, is positive when evaluated at $\hat{\vartheta}_{N,0}$; the test statistic is positive if $S_{1,N}(\vartheta_N) < 0$ when evaluated at $\hat{\vartheta}_{N,0}$, by the Karush-Kuhn-Tucker (KKT) optimality condition for constrained optimization.

Because Ξ is a compact set, formula (3.15) of Hosoya (1997) can be extended to show that

$$\sqrt{N}S_N(\hat{\vartheta}_{N,0}) - \sqrt{N}S_N(\vartheta_0) - \sqrt{N}V(\hat{\vartheta}_{N,0})(1 + o_p(1)) = o_p(1)$$

where $V(\vartheta)$ is a vector consisting of the partial derivatives

$$V_j(\vartheta) = H_j(\vartheta) + \int_{-\pi}^{\pi} tr\{h_j(\omega, \vartheta)f(\omega)\}d\omega, \tag{15}$$

where $f(\omega)$ is the true spectral density function, $h_j(\omega, \vartheta) = \partial f^{-1}(\omega, \vartheta)/\partial\vartheta_j$, and $H_j(\vartheta) = \partial \int_{-\pi}^{\pi} \log\det f(\omega; \vartheta)d\omega/\partial\vartheta_j$.

The following claim is proved in the *Supplementary Material*.

Claim 2:

$V(\vartheta)$ is differentiable in a neighbourhood of ϑ_0, with its first derivative matrix denoted by $\frac{\partial V}{\partial\vartheta}(\vartheta)$ whose value at ϑ_0 is denoted by W. Moreover, as $\vartheta \to \vartheta_0$, $\frac{\partial V}{\partial\vartheta}(\vartheta) \to W$. Let W be partitioned into a 3×3 block matrix according to $h, k, \vartheta^* = (H, b_{11}, b_{12}, b_{22})'$ whose (i, j)-th block is denoted by $W_{i,j}$. (The true value of ϑ^* is denoted by ϑ_0^*.) Then, W_{33} is positive definite.

From a Taylor expansion,

$$\sqrt{N}S_{1,N}(\hat{\vartheta}_{N,0}) = \sqrt{N}S_{1,N}(\vartheta_0) + \sqrt{N}W_{1,1:3}\left(\hat{\vartheta}_{N,0} - \vartheta_0\right) + o_p(1), \tag{16}$$

where $W_{1,1:3} = [W_{1,1}, W_{1,2}, W_{1,3}]$ is the first row of blocks of W. Because the first two components of $\hat{\vartheta}_{N,0}$ are 0, it follows that

$$\sqrt{N}(\hat{\vartheta}^*_{N,0} - \vartheta^*_0) = -\sqrt{N}W^{-1}_{3,3}S_{3,N}(\vartheta_0) + o_p(1),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

upon noting that $S_{3,N}(\hat{\vartheta}_{N,0}) = 0$. Consequently,

$$\sqrt{N}S_{1,N}(\hat{\vartheta}_{N,0}) = \sqrt{N}S_{1,N}(\vartheta_0) - \sqrt{N}W_{1,3}W_{3,3}^{-1}S_{3,N}(\vartheta_0) + o_p(1).$$

The right side here converges weakly to a non-degenerate normal distribution of zero mean, hence, as $N \to \infty$, $P(S_{1,N}(\hat{\vartheta}_{N,0}) > 0) \to 1/2$ and $P(S_{1,N}(\hat{\vartheta}_{N,0}) \leq 0) \to 1/2$, proving that the log-likelihood ratio test statistic is asymptotically 0 with probability 1/2. Conditional on $\widehat{h} > 0$ so that $S_N(\hat{\vartheta}_N) = 0$, it can be shown by adapting the arguments in the proof of Theorem 2.4 of Hosoya (1997) that the likelihood ratio test is asymptotically χ^2 with two degrees of freedom under the null hypothesis. This completes the proof of Part (a).

Part (b)

Under the general alternative hypothesis $H_A : H_1 > H_2$, it follows from Theorem 1 that the first two elements of $\sqrt{N}(\hat{\vartheta}_{N,0} - \vartheta_0)$ equal $\left(-\sqrt{N}\frac{H_1-H_2}{2}, \ -\sqrt{N}\frac{H_1-H_2}{2}a_{11} \right)' + O_p(1)$, where the parameters are evaluated at their true values; hence they diverge to $-\infty$. In (16), the first two terms in the second summand dominate as $-\sqrt{N}\frac{H_1-H_2}{2}W_{1,1} - \sqrt{N}\frac{H_1-H_2}{2}a_{11}W_{1,2}$ which, noting that $W_{1,2} = \frac{W_{1,1}}{a_{11}}$, equals $\sqrt{N}\,(H_1 - H_2)\,W_{1,1}$, and, thus, (16) diverges to $-\infty$. Since, then, $P(S_{1,N}(\hat{\vartheta}_{N,0}) \leq 0) \to 1$, the test statistic becomes arbitrarily large, establishing the consistency of the test under H_A.

References

Andrews, D. W. K. (1998). Hypothesis testing with a restricted parameter space. *Journal of Econometrics* **84**, 155-199.

Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* **69**, 683-734.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

Asai, M. and McAleer, M. (2013). A fractionally integrated Wishart stochastic volatility model. Available at `http://eprints.ucm.es/18068/`

Barndorff-Nielsen, O. E. and Stelzer, R. (2011). Multivariate supOU processes. *The Annals of Applied Probability* **21**, 140-182.

Beran, J. (1994). *Statistics for Long-Memory Processes*. New York: Chapman and Hall.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. New York: Springer.

Brockwell, P. J. and Marquardt, T. (2005). Lévy-driven and fractionally integrated ARMA processes with continuous time parameter. *Statist. Sinica* **15**, 477-494.

Casas, I. and Gao, J. (2008). Econometric estimation in long-range dependent volatility models: theory and practice. *Journal of Econometrics* **147**, 72-83.

Chambers, M. J. (1995). The simulation of random vector time series with given spectrum. *Math. Comput. Modelling* **22**, 1-6.

Chambers, M. J. (1996). The estimation of continuous parameter long-memory time series models. *Econometric Theory* **12**, 374-390.

Chambers, M. J. (1998). Long memory and aggregation in macroeconomic time series. *International Economic Review* **39**, 1053-1072.

Chan, K.S. and Tsai, H. (2012). Inference of seasonal long-memory aggregate time series. *Bernoulli* **18**, 1448-1464.

Cheridito, P., Kawaguchi, H. and Maejima, M. (2003). Fractional Ornstein-Uhlenbeck processes. *Electronic Journal of Probability* **8**, 1-14.

Comte, F. (1996). Simulation and estimation of long memory continuous time models. *Journal of Time Series Analysis* **17**, 19-36.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

Comte, F. and Renault, E. (1996). Long memory continuous time models. *Journal of Econometrics* **73**, 101-149.

Comte, F. and Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance* **8**, 291-323.

Cryer, J. D. and Chan, K.-S. (2008). *Time Series Analysis With Applications in R*. New York: Springer.

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, New York.

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

Hosoya, Y. (1997). A limit theory for long-range dependence and statistical inference on related models. *Annals of Statistics* **25**, 105-137.

Mandelbrot, B.B. and Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review* **10**, 422-437.

Marquardt, T. (2007). Multivariate fractionally integrated CARMA processes. *Journal of Multivariate Analysis* **98**, 170-1725.

Tsai, H. (2006). Quasi-maximum likelihood estimation of long-memory limiting aggregate processes. *Statistica Sinica* **16**, 213-226.

Tsai, H. and Chan, K. S. (2005a). Maximum likelihood estimation of linear continuous time long memory processes with discrete time data. *J. Roy. Statist. Soc. Ser. B* **67**, 703-716.

Tsai, H. and Chan, K. S. (2005b). Quasi-maximum likelihood estimation for a class of continuous-time long-memory processes. *Journal of Time Series Analysis* **26**, 691-713.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0112

Tsai, H. and Chan, K. S. (2005c). Temporal aggregation of stationary and nonstationary continuous-time processes. *Scandinavian Journal of Statistics* **32**, 583-597.