Block Whittle Estimation of Time Varying Stochastic Regression Models with Long Memory*

Chris Toumping Fotso[1] and Philipp Sibbertsen[1]

[1]Leibniz University Hannover

November 11, 2024

Abstract

This paper proposes an estimator that accounts for time variation in a regression relationship with stochastic regressors exhibiting long-range dependence, covering weak fractional cointegration as a special case. An interesting application of this estimator is its ability to handle situations where the regression coefficient changes abruptly. The parametric formulation of this estimator is introduced using the Block-Whittle-based estimation. We analyze the asymptotic properties of this estimator, including consistency and asymptotic normality. Furthermore, we examine the finite sample behavior of the estimator through Monte Carlo simulations. Additionally, we consider a real-life application to demonstrate its advantages over the constant case.

1 Introduction

Linear regression is a widely used technique across various scientific fields. Its popularity comes from its ability to model and estimate the strength and the direction of the relationship between two or more variables. A standard regression model between two variables, Y and X is defined as:

$$Y_t = \beta X_t + \varepsilon_t \tag{1}$$

where Y_t represents the response variable, X_t is the predictor, and ε_t is the errors. Typically, a regression model relies on six key assumptions, that must be satisfied. Among these, the regressor X_t must be nonstochastic, and the error term should follow a normal distribution. Additionally, it assumes that the relationship between the response variable Y_t and the regressor X_t remains constant over time. However, this condition is often violated in practice, especially when considering structural breaks and regime shifts. For instance, Brown et al. (1997) modeled the change of house prices in the United Kingdom using time-varying coefficients, resulting in more accurate forecasts than those obtained with constant parameters.

The survey of stochastic parameter regression of Rosenberg (1973) highlights the importance and the advantages of considering stochastic parameter regression models. For instance, Newbold and Bos (1985) developed a stochastic regression model later generalized by Robinson and Hidalgo (1997) to allow the error ε_t to exhibit long memory. Furthermore, Yajima (1991) explored the properties of the linear least squares estimator in the context of regression with long-memory stationary errors. However, Yajima's study confined the long memory to the errors and did not allow the regressors to be stochastic.

*Contact: toumping@statistik.uni-hannover.de (C. Fotso), sibbertsen@statistik.uni-hannover.de (P. Sibbertsen)

In this paper, we allow both the response Y_t and the regressor X_t to be stochastic with the same long memory, which can also be time-varying, denoted as δ_1. The error term ε_t also exhibits long memory, denoted as δ_2, where $\delta_2 \leq \delta_1$ and the regression coefficient β_t is time-varying. Particularly, when $\delta_2 < \delta_1$, (1) represents a typical fractional cointegration relationship (Engle and Granger, 1987). Fractional cointegration has been applied to various fields of economics, such as exchange rate dynamics (Baillie and Bollerslev, 1994; Cheung and Lai, 1993) and financial volatility (da Silva and Robinson, 2008; Christensen and Nielsen, 2002). Additionally, recent research has explored its applications in the field of climate change (Carlini et al., 2023; Yaya et al., 2023). Some studies have addressed time variation in standard cointegrating relationships, such as the work by Park and Park (2015), which applied this concept to exchange rate predictability in Korea, and Kapetanios et al. (2020), which focused on the UK's great ratios. To the best of our knowledge, there has been limited research specifically on time-varying fractional cointegration.

Indeed, the concept of time-varying dependence models is a well-known problem in time series analysis. Priestley (1965) first discussed evolutionary spectral techniques that is, time-dependent spectral functions. Subsequently, Dahlhaus (1996, 1997) developed locally stationary processes which have gained significant attention from researchers over the years (Dahlhaus, 2000; Dahlhaus and Polonik, 2009; Beran, 2009, etc.). Recently, Palma and Olea (2010a) addressed locally stationary Gaussian processes that exhibit long-range dependence based, on the block-Whittle based estimation technique proposed in Dahlhaus (1997). Thus, this paper builds on their work to propose a new framework for time-varying stochastic regression parameter between (locally) stationary times series, and we analyze its asymptotic properties. In particular, we also show that the sigmoid function can be used, to estimate abrupt changes in the relationship using the estimator.

The remainder of this paper is organized as follows. Section 2 provides a reminder about the concept block Whittle-based estimation for long memory locally stationary process and also the model of time-varying fractional cointegration. Next, Section 3 introduces the new block Whittle function, states several important assumptions that will be used to investigate the large sample properties of the estimator, and presents the main results for consistency and asymptotic normality. Section 4 shows the finite sample behavior of the estimator through Monte Carlo simulation. Finally, Section 5 demonstrates a real-life application of our estimator, and the conclusion of the paper is presented in Section 6. Technical lemmas and all proofs can be found in the appendices A.2 and A.1 respectively.

2 Preliminaries and fractional cointegration model

2.1 Regression model

Let's consider a sequence of Gaussian bivariate stochastic processes $Z_{t,T} = (Y_{t,T}, X_{t,T})'$ $(t = 1, \ldots, T)$ locally stationary, having the same long memory parameter $\delta_1 < 0.5$ [1] and each defined by the spectral representation:

$$Y_t = \int_{-\pi}^{\pi} A^0_{t,T}(\lambda) e^{i\lambda t} d\zeta_1(\lambda), \tag{2}$$

$$X_t = \int_{-\pi}^{\pi} B^0_{t,T}(\lambda) e^{i\lambda t} d\zeta_2(\lambda) \tag{3}$$

for $t = 1, \ldots, T$, where $(\zeta_1(\lambda), \zeta_2(\lambda))$ is a Brownian motion on $[-\pi, \pi]$ with covariance matrix assumed without loss of generality to be $\Sigma = \mathbb{1}$. So, there exists a constant K and $2\pi-$periodic functions $A, B : (0,1] \times \mathbb{R} \longrightarrow \mathbb{C}$ with $\overline{A(u,\lambda)} = A(u,-\lambda)$, $\overline{B(u,\lambda)} = B(u,-\lambda)$ and

$$\sup_{t,\lambda} \left| A^0_{t,T}(\lambda) - A\left(\frac{t}{T}, \lambda\right) \right| \leq KT^{-1}, \quad \sup_{t,\lambda} \left| B^0_{t,T}(\lambda) - B\left(\frac{t}{T}, \lambda\right) \right| \leq KT^{-1},$$

[1] Our results remain valid even if the long memory is not constant.

for all $T \in \mathbb{N}$.

If Y_t and X_t exhibit a time-varying relationship in regression, it suggests that there exists $\beta(t/T)$ and a process ε_t with long memory parameter $\delta_2 \leq \delta_1$ so that:

$$Y_t - \beta\left(\frac{t}{T}\right) X_t = \varepsilon_t. \tag{4}$$

If one considers by

$$B_t = \left(1, -\beta\left(\frac{t}{T}\right)\right)', \tag{5}$$

then (4) can be written as

$$B_t' Z_t = \varepsilon_t, \tag{6}$$

and the spectral density and the periodogram of ε_t are equal to:

$$f_\varepsilon(t/T, \lambda) = B_t' f_Z(\lambda) B_t = f_{\delta_2}(t/T, \lambda) \tag{7}$$

$$I_\varepsilon(t/T, \lambda) = B_t' I_Z(\lambda) B_t \tag{8}$$

2.2 Block Whittle-based estimation

The Block Whittle likelihood function, which is a generalization of the standard likelihood (Dahlhaus, 2000), was first introduced by Dahlhaus (1997) for locally stationary processes.

A process $\varepsilon_{t,T}$ with transfer function G^0 is locally stationary if there exists a representation:

$$\varepsilon_{t,T} = \int_{-\pi}^{\pi} G^0_{t,T}(\lambda) e^{i\lambda t} d\zeta(\lambda), \tag{9}$$

for $t = 1 \dots T$, where $\zeta(\lambda)$ is a Brownian motion on $[-\pi, \pi]$ and there exists a constant K and $2\pi-$periodic functions $G : (0,1] \times \mathbb{R} \to \mathbb{C}$ with $\overline{G(u,\lambda)} = G(u, -\lambda)$ such that

$$\sup_{t,\lambda} \left| G^0_{t,T}(\lambda) - G\left(\frac{t}{T}, \lambda\right) \right| \leq K T^{-1},$$

for all $T \in \mathbb{N}$.

If we let $\theta \in \Theta$ be a parameter vector specifying model (9) where the parameter space Θ is a subset of a finite-dimensional Euclidean space. Given a sample $\{\varepsilon_1, \dots, \varepsilon_T\}$ of (9), we can estimate θ by minimizing the Block Whittle log-likelihood function:

$$\mathcal{L}_T(\theta) = \frac{1}{4\pi} \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \left\{ \log f_\theta(u_j, \lambda) + \frac{I_N(u_j, \lambda)}{f_\theta(u_j, \lambda)} \right\} d\lambda, \tag{10}$$

where $I_N(u, \lambda)$ is the local tapered periodogram over a segment of length N with midpoint $[uT]$ defined as

$$I_N(u, \lambda) = \frac{|D_N(u, \lambda)|^2}{2\pi H_{2,N}(0)}, \tag{11}$$

with

$$D_N(u, \lambda) = \sum_{s=0}^{N-1} h\left(\frac{s}{n}\right) \varepsilon_{[uT]-N/2+s+1} e^{-i\lambda s}, H_{k,N} = \sum_{s=0}^{N-1} h\left(\frac{s}{n}\right)^k e^{-i\lambda s},$$

$T = S(M-1) + N$, $u_j = t_j/T$, $t_j = S(j-1) + N/2$, $j = 1, \dots, M$ and $h(\cdot)$ is a data taper and it helps to reduce the bias due to nonstationarity on a segment.

3 Estimation procedure, assumptions and main results

3.1 Estimation and assumptions

Let's consider 2 samples $\{Y_1, \ldots, Y_T\}$ and $\{X_1, \ldots, X_T\}$ of the processes (2) and (3). Let $\theta \in \Theta$ also be a parameter vector specifying respectively $\beta(t/T)$ in model (6) where the parameter space Θ is a subset of a finite-dimensional Euclidean space. The spectral density and the periodogram of ε_t are now equal to:

$$f_\theta(u,\lambda) = B_\theta(u)' f_Z(u,\lambda) B_\theta(u) = f_{\delta_2}(\lambda), \tag{12}$$

$$I_\varepsilon^\theta(u,\lambda) = B_\theta(u)' I_Z(\lambda) B_\theta(u), \tag{13}$$

with $B_\theta(u) = (1, -\beta_\theta(u))'$ and $u = t/T$.

If we replace the new spectral density (12) and the periodogram (13) in the original block-Whittle likelihood (10) will get a new estimator of θ and δ_2 given by the following function:

$$\mathcal{L}_T(\theta, \delta_2) = \frac{1}{4\pi}\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\left\{\log f_{\delta_2}(u,\lambda) + \frac{I_N^\theta(u_j,\lambda)}{f_{\delta_2}(u,\lambda)}\right\} d\lambda, \tag{14}$$

where

$$I_N^\theta(u,\lambda) = B_\theta(u)' I_N^Z(u,\lambda) B_\theta(u), \tag{15}$$

with $B_\theta(u) = (1, -\beta_\theta(u))'$, and

$$I_N^Z(u,\lambda) = \begin{pmatrix} I_N^Y(u,\lambda) & I_N^{YX}(u,\lambda) \\ I_N^{XY}(u,\lambda) & I_N^X(u,\lambda) \end{pmatrix}, \tag{16}$$

$$f_Z(u,\lambda) = \begin{pmatrix} f_Y(u,\lambda) & f_{YX}(u,\lambda) \\ f_{XY}(u,\lambda) & f_X(u,\lambda) \end{pmatrix}, \tag{17}$$

$I_N^{(\cdot,\cdot)}(u,\lambda)$ is the tapered periodogram and it's defined by

$$I_N^{YX}(u,\lambda) = \frac{D_N^1(u,\lambda)\overline{D_N^2(u,\lambda)}}{2\pi H_{2,N}(0)} = \overline{I_N^{XY}}(u,\lambda), \tag{18}$$

with

$$D_N^1(u,\lambda) = \sum_{s=0}^{N-1} h\left(\frac{s}{n}\right) Y_{[uT]-N/2+s+1,T} e^{-i\lambda s},$$

$$D_N^2(u,\lambda) = \sum_{s=0}^{N-1} h\left(\frac{s}{n}\right) X_{[uT]-N/2+s+1,T} e^{-i\lambda s}.$$

Note that we only considered Gaussian time series in this paper. However, as shown by Chan and Palma (2020), the Block Whittle-based estimation can be applied to non-Gaussian time series with additional assumptions on the higher-order cumulants of the process.

We introduce the following assumptions. The first assumption concerns the spectral density of Z and ε. The second assumption pertains to the tapering function and the cointegrating function $\beta(u)$. The last assumption is on the block sampling scheme.

A1. Z is Gaussian and its spectral density matrix $f_Z(\lambda)$ satisfies:

$$f_Z(u,\lambda) \sim C_{f_Z}(\theta,u)|\lambda|^{-2\delta_1}, \text{ as } |\lambda| \to 0,$$

where $C_{f_Z}(\theta,u)$ is a 2 by 2 matrix. Also, $f_{\delta_2}(u,\lambda)$ satisfies:

$$f_{\delta_2}(u,\lambda) \sim C(\theta,u)|\lambda|^{-2\delta_2}, \text{ as } |\lambda| \to 0,$$

There is an integrable function $g(\lambda)$ such that $|\nabla \log f_\theta(u,\lambda)| \leq g(\lambda)$ for all $\theta \in \Theta$, $u \in [0,1]$ and $\lambda \in [-\pi,\pi]$. The functions $A(\lambda)$, $B(\lambda)$, $G(\lambda)$ satisfy:

$$\begin{cases} \int_{-\pi}^{\pi} A(u,\lambda)B(v,-\lambda)\exp(ik\lambda) & \sim & C(u,v)k^{2\delta_1-1}, \\ \int_{-\pi}^{\pi} A(u,\lambda)A(v,-\lambda)\exp(ik\lambda) & \sim & C(u,v)k^{2\delta_1-1}, \\ \int_{-\pi}^{\pi} B(u,\lambda)B(v,-\lambda)\exp(ik\lambda) & \sim & C(u,v)k^{2\delta_1-1}, \\ \int_{-\pi}^{\pi} G(u,\lambda)G(v,-\lambda)\exp(ik\lambda) & \sim & C(u,v)k^{2\delta_2-1}, \end{cases}$$

as $k \to \infty$, where $|C(u,v)| \leq K$. The function $f_{\delta_2}(u,\lambda)^{-1}$ is twice differentiable with respect to δ_2 and λ.

A2. The data taper $h(u)$ is a positive, bounded function for $u \in [0,1]$ and symmetric around $\frac{1}{2}$ with bounded derivative. We also assume that $\beta : [0,1] \to \mathbb{R}$ is differentiable with uniformly bounded derivatives.

A3. The sample size T and the subdivisions integers N, S and M tend to infinity satisfying:

$$\frac{S}{N} \to 0, \frac{\sqrt{T}\log^2 N}{N} \to 0, \frac{\sqrt{T}}{M} \to 0, \frac{N^3 \log^2 N}{T^2} \to 0.$$

3.2 Main results

The asymptotic properties of $\widehat{\delta}_2$ have already been established in Theorems 2.1, 2.2 and 2.3 of Palma and Olea (2010a). In this section, we explore the large sample properties of $\widehat{\theta}$ including consistency and normality.

Theorem 1 (Consistency). *Let θ_0 and δ_{20} be the true values of the parameters θ and δ_2. Under assumptions A1-A3, the estimators $\widehat{\theta}_T$ and $\widehat{\delta}_2$ satisfy $\widehat{\theta}_T \to \theta_0$, and $\widehat{\delta}_2 \to \delta_{20}$ in probability, as $T \to \infty$.*

Theorem 2 (Normality). *Let θ_0 be the true value of the parameter θ. If assumptions A1-A3 hold, then the estimators $\widehat{\theta}_T$ satisfies*

$$\sqrt{T^{1-2(\delta_1-\delta_2)}}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N[0, \Lambda(\theta_0)^{-1}V(\theta_0)\Lambda(\theta_0)^{-1}], \quad as\ T \to \infty,$$

where $\xrightarrow{d}$ denotes the convergence in distribution and Λ is given by

$$\Lambda(\theta_0) = \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \nabla^2 f_{\theta_0}(u,\lambda) f_{\delta_{20}}^{-1}(u,\lambda) d\lambda, \tag{19}$$

and

$$V(\theta_0) = \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \left[(\nabla f_{\theta_0}(u,\lambda))^2 + 2\nabla\beta_{\theta_0}(u,\lambda)\beta_{\theta_0}^2(u,\lambda) \times f_{YX}(u,\lambda) f_Y(u,\lambda))\right] (f_{\delta_{20}}^{-1}(u,\lambda))^2 \tag{20}$$

4 Simulations

In this section, we report some simulations to examine the finite sample performance of the estimator. We conducted some Monte Carlo experiments for the ARFIMA model. Specifically, Y_t and X_t are both $\mathrm{ARFIMA}(0, \delta_1, 0)$ processes with long memory $\delta_1 = 0.35$, while ε_t is also $\mathrm{ARFIMA}(0, \delta_2, 0)$ with $\delta_2 = 0.1$. Two cases are investigated for the cointegrating coefficient β: one representing an abrupt change and the other representing a smooth change, i.e.

$$\beta(u) = \begin{cases} \beta_0 & \text{for} \quad u \leq u_0 \\ \beta_1 & \text{for} \quad u > u_0 \end{cases} \tag{21}$$

$$\beta(u) = \beta_1 u + \beta_0, \tag{22}$$

with $u_0 = \{64/T, 200/T\}$. Due to assumption A2, $\beta(u)$ has to be a continuous function. Therefore, we approximated (21) by a continuous function using the sigmoid function as follows:

$$\beta(u) = \beta_0 + (\beta_1 - \beta_0)\mathrm{Sigmoid}\left(\frac{(u - u_0)}{0.00005}\right), \tag{23}$$

where

$$\mathrm{Sigmoid(x)} = \frac{1}{1 + e^{-x}}.$$

Figure 1 shows the approximation of the function (21) for $u_0 = 200/512$. It demonstrates that (23) provides a good approximation, which can be effectively used in practice.

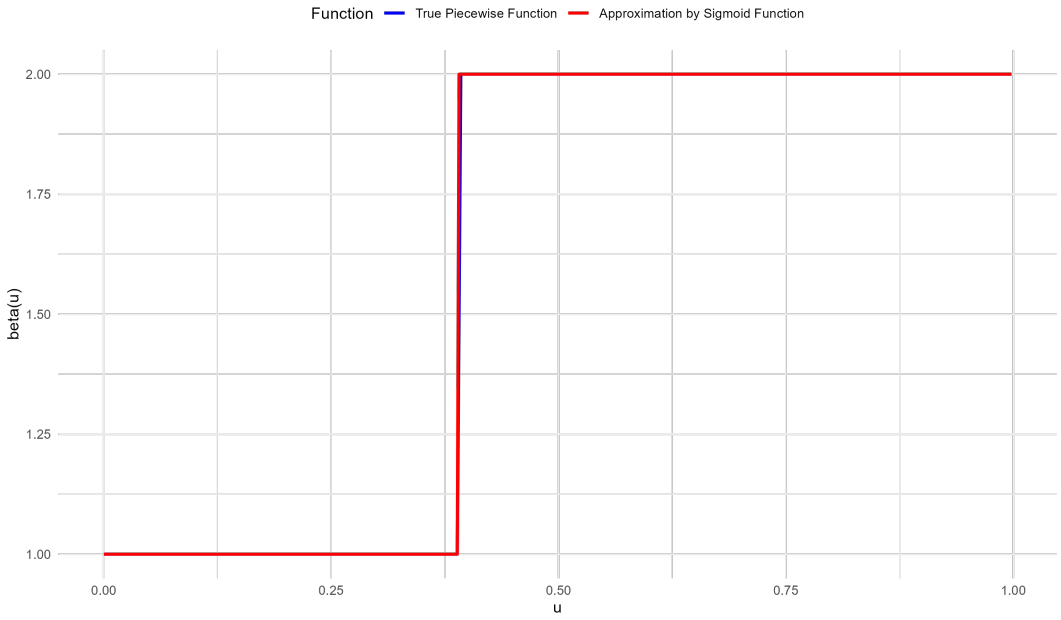

Figure 1: Approximation of (21) with Sigmoid Function.

All the simulations were carried out using the Constrained Optimization BY Linear Approximations (COBYLA) algorithm from the NLopt library, an open-source library for nonlinear optimization. This algorithm proved to be the most effective for optimizing β in the form of (23). The Figures below display the contour curves for the empirical mean squared error (MSE) for the estimator $(\widehat{\beta}, \widehat{\delta}_2)$. The MSE is defined in this case as the average of $\left\|(\widehat{\beta}, \widehat{\delta}_2) - (\beta, \delta_2)\right\|^2$ over 100 replications of $(\widehat{\beta}, \widehat{\delta}_2)$. In these figures, the aim is to select the optimal bandwidth parameters N and S that result in the smallest MSE for sample sizes $T = 512$ and $T = 1024$, respectively. The darkest regions represent minimal empirical MSE, while clearer regions indicate greater MSE values.

Figure 2 shows the contour curves for $u_0 = 200/T$, $\beta_0 = 1$, $\beta_1 = 2$ and $T = 512$. This plot indicates that the region with minimal empirical MSE is approximately at $N \approx 60$ and $S \approx 16$ for $T = 512$. This same region was observed for $T = 1024$.

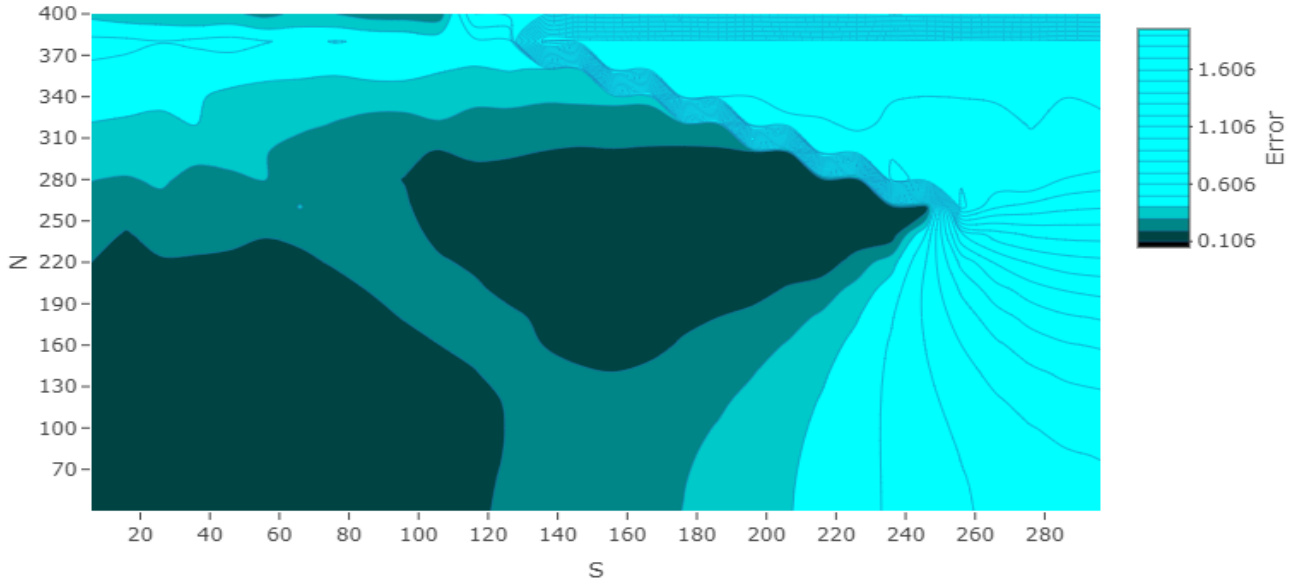

Figure 2: Contour curves of the empirical MSE with sample size T=512

We reported results from the Monte Carlo simulations for several parameter values based on 1000 replications in Table 1, using the optimal values we found.

$T = 512, \quad N = 60, \quad S = 16$								
Parameters			Estimates			Estimated SD		
β_0	β_1	δ_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\delta}_2$	$\hat{\sigma}(\hat{\beta}_0)$	$\hat{\sigma}(\hat{\beta}_1)$	$\hat{\sigma}(\hat{\delta}_2)$
1	2	0.1	1.01	1.985	0.11	0.067	0.053	0.041
2	4	0.1	2.02	3.97	0.126	0.073	0.055	0.048
0.5	-0.5	0.1	0.496	-0.491	0.124	0.068	0.053	0.042
$T = 1024, \quad N = 60, \quad S = 16$								
Parameters			Estimates			Estimated SD		
β_0	β_1	δ_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\delta}_2$	$\hat{\sigma}(\hat{\beta}_0)$	$\hat{\sigma}(\hat{\beta}_1)$	$\hat{\sigma}(\hat{\delta}_2)$
1	2	0.1	1.009	1.993	0.106	0.065	0.033	0.029
2	4	0.1	2.02	3.99	0.117	0.072	0.033	0.031
0.5	-0.5	0.1	0.499	-0.497	0.115	0.063	0.031	0.029

Table 1: Estimation for $u_0 = \frac{200}{T}$, $\beta(u) = \beta_0 + (\beta_1 - \beta_0)\text{Sigmoid}\left(\frac{(u-u_0)}{0.00005}\right)$.

In addition, we also reported the simulations results in Table 2 for $u_0 = 64/T$ across different samples sizes.

$T = 512, \quad N = 40, \quad S = 16$								
Parameters			Estimates			Estimated SD		
β_0	β_1	δ_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\delta}_2$	$\hat{\sigma}(\hat{\beta}_0)$	$\hat{\sigma}(\hat{\beta}_1)$	$\hat{\sigma}(\hat{\delta}_2)$
1	2	0.1	1.003	1.989	0.107	0.067	0.137	0.044
2	4	0.1	2.002	3.98	0.119	0.136	0.0458	0.0494
0.5	-0.5	0.1	0.5	-0.495	0.121	0.126	0.045	0.0438
$T = 1024, \quad N = 40, \quad S = 16$								
Parameters			Estimates			Estimated SD		
β_0	β_1	δ_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\delta}_2$	$\hat{\sigma}(\hat{\beta}_0)$	$\hat{\sigma}(\hat{\beta}_1)$	$\hat{\sigma}(\hat{\delta}_2)$
1	2	0.1	1.002	1.996	0.107	0.134	0.03	0.031
2	4	0.1	2	3.99	0.116	0.134	0.032	0.033
0.5	-0.5	0.1	0.507	-0.499	0.111	0.121	0.029	0.03

Table 2: Estimation for $u_0 = \frac{64}{T}$, $\beta(u) = \beta_0 + (\beta_1 - \beta_0)\text{Sigmoid}\left(\frac{(u-u_0)}{0.00005}\right)$.

For the smooth changes (22), we found that the region with minimal empirical MSE is approximately at $N \approx 80$ and $S \approx 16$ for $T = 512$. The Table 3 below reports the Monte Carlo simulations with 1000 replications for this case.

$T = 512, \quad N = 80, \quad S = 16$								
Parameters			Estimates			Estimated SD		
β_0	β_1	δ_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\delta}_2$	$\hat{\sigma}(\hat{\beta}_0)$	$\hat{\sigma}(\hat{\beta}_1)$	$\hat{\sigma}(\hat{\delta}_2)$
0.5	2	0.1	0.509	1.994	0.099	0.092	0.164	0.041
-0.4	2	0.1	-0.4	2.002	0.113	0.09	0.160	0.040
1	0.75	0.1	0.991	0.758	0.105	0.093	0.168	0.041
$T = 1024, \quad N = 80, \quad S = 16$								
Parameters			Estimates			Estimated SD		
β_0	β_1	δ_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\delta}_2$	$\hat{\sigma}(\hat{\beta}_0)$	$\hat{\sigma}(\hat{\beta}_1)$	$\hat{\sigma}(\hat{\delta}_2)$
0.5	2	0.1	0.499	2.002	0.102	0.064	0.113	0.027
-0.4	2	0.1	-0.398	1.998	0.105	0.063	0.109	0.029
1	0.75	0.1	0.995	0.756	0.101	0.059	0.106	0.028

Table 3: Estimation for $\beta(u) = \beta_0 + \beta_1 * u$.

Furthermore, it is also possible to extend the approximation expression (23) to account for more than one changing point. For example, if we consider two changing points, u_0 and u_1, as follows:

$$\beta(u) = \begin{cases} \beta_0 & \text{for} \quad u \leq u_0 \\ \beta_1 & \text{for} \quad u_0 \leq u < u_1 \\ \beta_2 & \text{for} \quad u \leq u_1. \end{cases} \tag{24}$$

Then, this can be approximated using:

$$\beta(u) = \beta_0 + (\beta_1 - \beta_0)\text{Sigmoid}\left(\frac{(u-u_0)}{0.00005}\right) + (\beta_2 - \beta_1)\text{Sigmoid}\left(\frac{(u-u_1)}{0.00005}\right), \tag{25}$$

Figure 3 illustrates this approximation with $T = 512$, $\beta_0 = 1$, $\beta_1 = 2$ and $\beta_2 = 0.5$, where the changing points are $u_0 = 64/T$ and $u_1 = 200/T$.

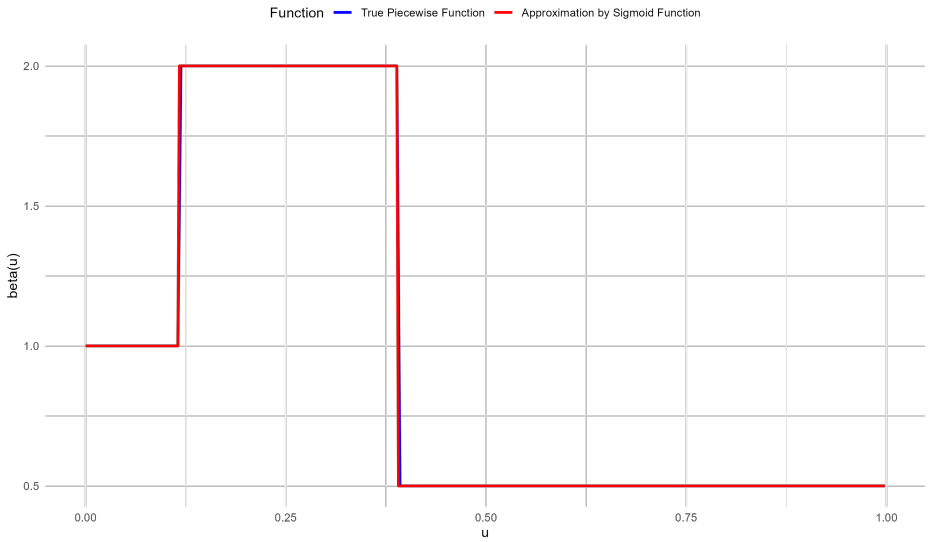

Figure 3: Approximation of (24) with Sigmoid Function.

We conducted simulations for this case using the same procedure as explained earlier. Our findings indicate that $N = 40$ and $S = 6$ are the optimal bandwidth values for this scenario. Table 4 presents then results of Monte Carlo simulations with 1000 replications:

$T = 512, \quad N = 40, \quad S = 6$											
Parameters				Estimates				Estimated SD			
β_0	β_1	β_2	δ_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\delta}_2$	$\hat{\sigma}(\hat{\beta}_0)$	$\hat{\sigma}(\hat{\beta}_1)$	$\hat{\sigma}(\hat{\beta}_2)$	$\hat{\sigma}(\hat{\delta}_2)$
1	2	0.5	0.1	1.038	1.97	0.506	0.122	0.141	0.080	0.056	0.045
4	2	-0.5	0.1	3.923	2.024	-0.492	0.125	0.149	0.084	0.057	0.052
-0.75	-0.3	2	0.1	-0.73	-0.294	1.993	0.12	0.143	0.081	0.059	0.049
$T = 1024, \quad N = 40, \quad S = 6$											
Parameters				Estimates				Estimated SD			
β_0	β_1	β_2	δ_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\delta}_2$	$\hat{\sigma}(\hat{\beta}_0)$	$\hat{\sigma}(\hat{\beta}_1)$	$\hat{\sigma}(\hat{\beta}_2)$	$\hat{\sigma}(\hat{\delta}_2)$
1	2	0.5	0.1	1.038	1.97	0.506	0.122	0.141	0.080	0.056	0.045
4	2	-0.5	0.1	3.923	2.024	-0.492	0.125	0.149	0.084	0.057	0.052
-0.75	-0.3	2	0.1	-0.73	-0.294	1.997	0.115	0.139	0.079	0.033	0.032

Table 4: Estimation for $u_0 = \frac{60}{T}$, $u_0 = \frac{200}{T}$, $\beta(u) = \beta_0 + (\beta_1 - \beta_0)\text{Sigmoid}\left(\frac{(u-u_0)}{0.00005}\right) + (\beta_2 - \beta_1)\text{Sigmoid}\left(\frac{(u-u_1)}{0.00005}\right)$.

Finally, all the simulation tables provided in this section demonstrate that the estimator exhibits very good finite sample performance, as the estimated parameters are close to their true values. Additional simulations with other model specifications can be found in the appendix A.3.

5 Empirical Application

We consider an application using high-frequency data of the DAX (Deutscher Aktienindex) stock index collected from the Thomson Reuters Tick History (TRTH) database. The data spans from January 2, 2014, to June 27, 2023, providing $T = 2043$ daily observations. We computed daily realized correlations using 51 intra-day returns sampled at a frequency of 10 minutes. For this analysis, we

selected three stocks: Adidas AG (ADS.DE), BASF (Baden Aniline and Soda Factory) SE (BAS.DE) and Bayer AG (BAYN.DE). These three companies are significant players in the German markets and have interconnected relationships. BASF and Bayer belongs both to the chemical industry one of the main industrial sectors in Germany and Adidas as a counterpart producing sports wear. Further, BASF has supported Adidas in developing the performance of its running shoes through innovative materials for the past 20 years. In addition, since 2018, BASF has acquired several businesses and assets from Bayer.

The realized correlations were calculated using realized covariance matrices, constructed using the methodology outlined in Chiriac and Voev (2011). After computing the correlations, we applied the Fisher-z-transformation, which redefined the correlations on the set of real numbers, removing the original constraint of being bounded within the interval $[-1, 1]$. More precisely, if $RC_t^{i,j}$ denotes the realized correlation between stocks i and j at time t, then the Fisher transformed realized correlation is defined by:

$$\widetilde{RC_t^{i,j}} = \frac{1}{2}\log\left(\frac{1+RC_t^{i,j}}{1-RC_t^{i,j}}\right)$$

We used BASF SE as the reference series to compute the realized correlation, as illustrated in Figure 4.

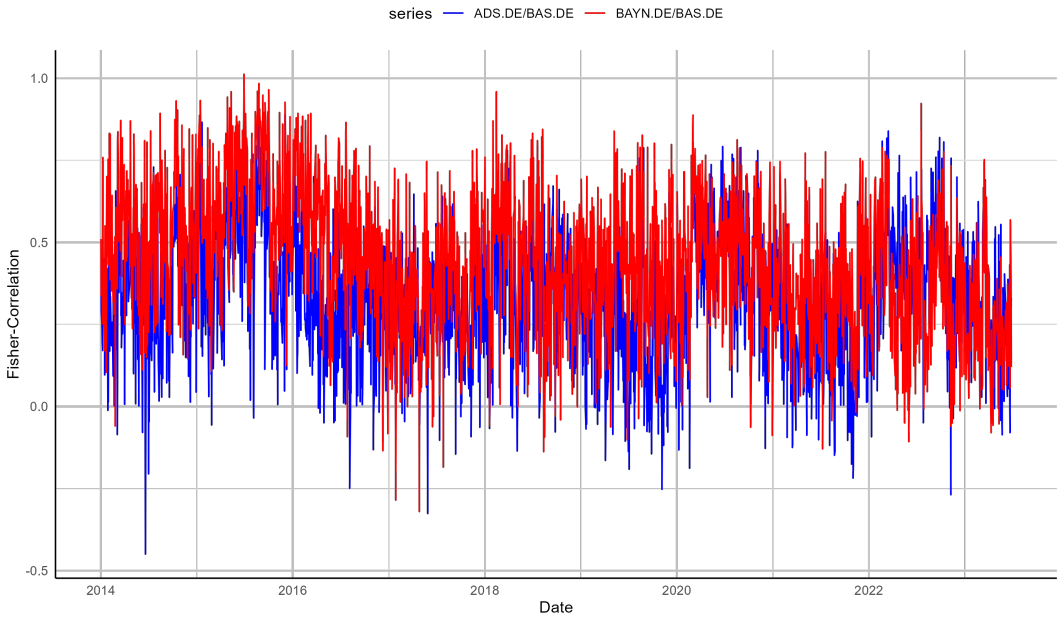

Figure 4: Realized correlations.

The regression model that we will estimate is presented as follows:

$$\widetilde{RC_t}^{BAYN,BAS} = \beta_t \widetilde{RC_t}^{ADS,BAS} + \varepsilon_t. \tag{26}$$

Applying the exact local Whittle estimator (ELW) with default bandwidth we get that $\widetilde{RC_t}^{BAYN,BAS}$ has a memory parameter of 0.36, while $\widetilde{RC_t}^{ADS,BAS}$ shows a memory parameter of 0.41.

We applied our estimator for both the constant and abrupt change scenarios. For the constant case, we explored various values of N and S, ultimately selecting $N = 600$ and $S = 360$. In contrast, for the time-varying case, we opted for $N = 535$ and $S = 400$. In the constant case, we found $\widehat{\beta} = 0.44$ and $\widehat{d} = 0.24$. Subsequently, we plotted the residuals $\widetilde{RC_t}^{BAYN,BAS} - \widehat{\beta}_t \widetilde{RC_t}^{ADS,BAS}$, depicted in Figure 5.

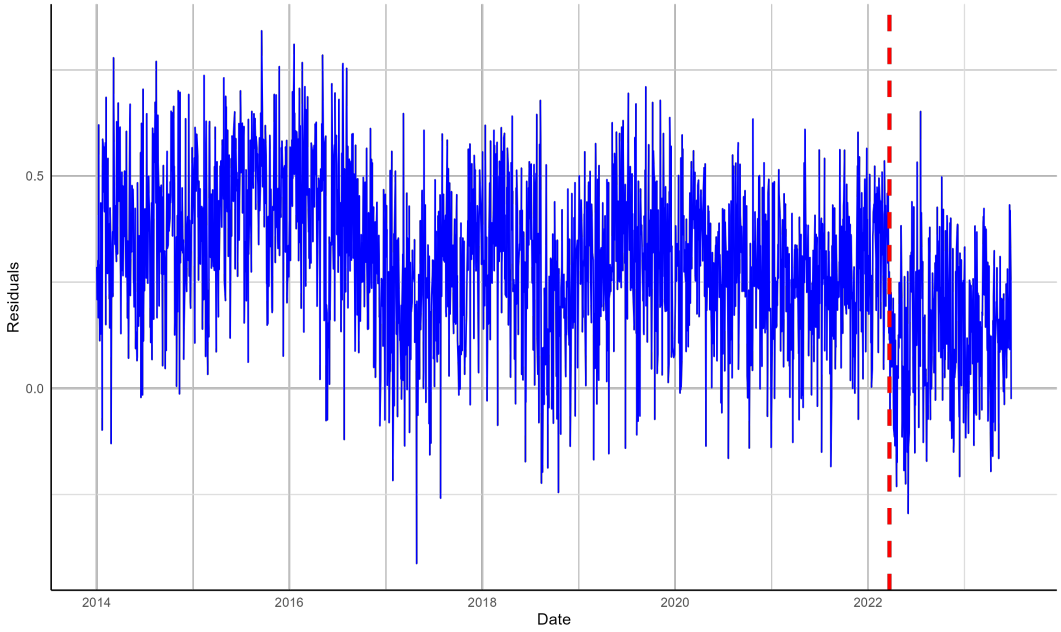

Figure 5: Residuals.

This visualization helped identify $u_0 = 2080/T$ as the change point for the abrupt change case corresponding to the date of March 23, 2022. For this latter case, we obtained $\widehat{\beta}_0 = 0.49$, $\widehat{\beta}_1 = 0.09$ and $\widehat{d} = 0.198$.

We also compared our results with those obtained using the narrow band frequency domain least squares (FDLS) method to estimate the cointegration coefficient and the exact local Whittle (ELW) to estimate the long memory parameter denoted as $\widehat{d}_{ELW}$. Table 5 shows the results for all these 3 scenarios. We also used the ELW to estimate $\widehat{d}_{ELW}$ with the coefficient we computed from our estimator. This table shows that our estimator provided a new cointegrating relationship, and the one from the time-varying case led to a smaller long memory.

	N	S	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{d}$	$\widehat{d}_{ELW}$
Constant	600	400	0.44	-	0.242	0.344
Time-Varying	535	360	0.49	0.09	0.198	0.298
FDLS	-	-	0.69	-	-	0.34

Table 5: Estimation of the coefficients

The new residuals from the abrupt changes case in Figure 6 appear also to be more stable and stationary compared to those from the constant case.

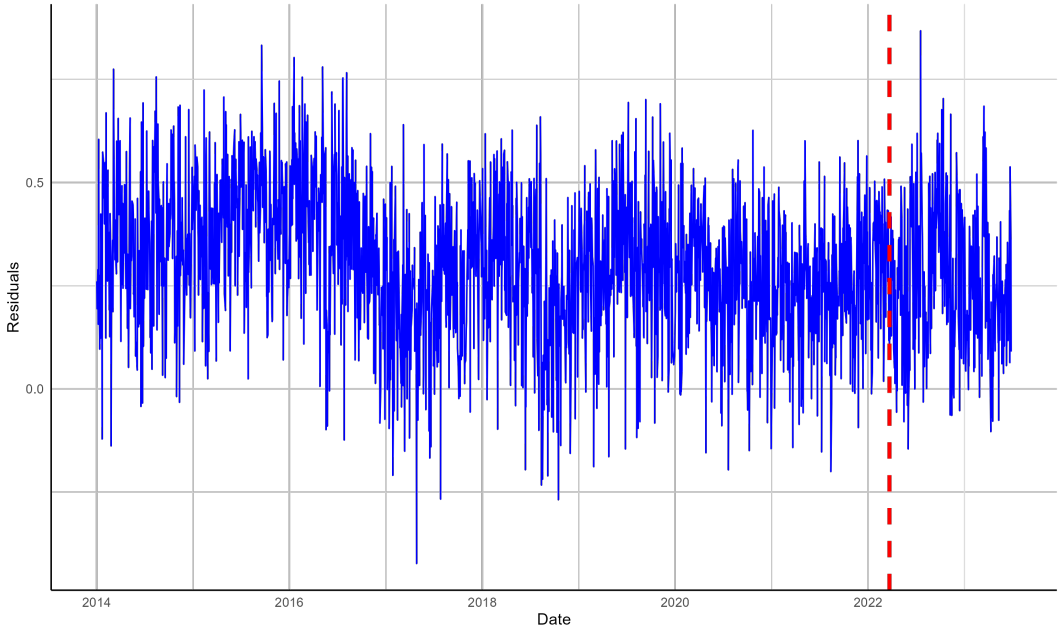

Figure 6: Realized correlations.

In March 2022, significant events occurred for these companies, impacting their stock prices and therefore the realized correlations as well. Our results suggest that after March 23, 2022, the cointegrating coefficient between both variables became small. If we had used constant cointegration, these changes would not have been captured.

6 Conclusion

In this paper, we examined the use of block Whittle-based estimation to detect changes within a stochastic regression framework. We derived the consistency of the estimator and established the convergence rate of its asymptotic normality. The regression parameter β_t can take various forms such as polynomial and Fourier expansions, or a mixture such as the sigmoid approximation.

Through extensive Monte Carlo simulations, we evaluated the estimator's performance under various settings involving both continuous and abrupt changes. The results demonstrated that the estimator performs well even in the presence of multiple change points.

Furthermore, we applied the methodology to realized correlations using stocks from the DAX index and found that the estimator could detect abrupt changes, resulting in a new cointegrating relationship. This new cointegration vector led to more stable residuals compared to the constant case, and the associated long memory value was reduced. This application highlights the importance and the advantage of accounting for time variation in parameters, rather than assuming it to be constant.

In conclusion, the block Whittle-based estimation proves to be an effective method for analyzing changes in a stochastic regression, and future research could explore bandwidth selection criteria for N and S, as this remains an open problem.

Appendix. Proofs

A.1 Technical Lemmas

Lemma 1. *Consider d_1, d_2, and $d_3 \in [0, 1/2)$ and for any $l \in \mathbb{Z}$ define the integral*

$$I(l) = \int_1^\infty \left[(x-1)^{-2d_1} - x^{-2d_1}\right] |l+x|^{d_2+d_3-1} dx.$$

Then $I(l) = O\left(|l|^{d_2+d_3-1}\right)$.

Proof. The proof is similar to the one of Lemma 3 in Palma and Olea (2010b). □

Lemma 2. *Let $f(\lambda)$ and $\phi(\lambda)$ be two functions defined over $\lambda \in [-\pi, \pi]$ with Fourier coefficients $\widehat{f}(k)$ and $\widehat{\phi}(k)$, respectively, satisfying*

$$\left|\widehat{f}(k)\widehat{\phi}(k)\right| \leq \frac{K \log k}{k^{2-2(d_1-d_2)}},$$

for some positive constant K and $|k| > 0$.

Let $C(N)$ be given by

$$C(N) = \sum_{t=0}^{N-1} h^2\left(\frac{t}{N}\right) \sum_{k=N-t}^{N-1} \widehat{f}(k)\widehat{\phi}(k),$$

with bounded data taper, $|h(u)| \leq K$, for all $u \in [0,1]$. Then there exists a positive constant K such that

$$|C(N)| \leq K \log N$$

Proof. The proof of this lemma is similar to the proof of Lemma 7 presented in Palma and Olea (2010b).

Let's consider a real number $\alpha \in [0,1]$ satisfying $1 - \alpha = O(1/N)$. Combined with the fact that $h(u)$ is bounded, we can write

$$|C(N)| \leq K \log N \left\{ \sum_{t=1}^{\alpha N} \sum_{k=N-t}^{N-1} \frac{1}{k^{2-2(d_1-d_2)}} + \sum_{t=\alpha N+1}^{N-1} \sum_{k=N-t}^{N-1} \frac{1}{k^{2-2(d_1-d_2)}} \right\}. \tag{27}$$

For $\frac{t}{N} < \alpha < 1$ we have

$$\begin{aligned}
\sum_{k=N-t}^{N-1} \frac{1}{k^{2-2(d_1-d_2)}} &= \frac{1}{N^{1-2(d_1-d_2)}} \sum_{k=N-t}^{N-1} \left(\frac{N}{k}\right)^{2-2(d_1-d_2)} \frac{1}{N} \\
&\leq \frac{K}{N^{1-2(d_1-d_2)}} \int_{1-\frac{t}{N}}^{1} \frac{dx}{x^{2-2(d_1-d_2)}} \\
&\leq \frac{K}{N^{1-2(d_1-d_2)}} \left(\frac{1}{(1-t/N)^{1-2(d_1-d_2)} - 1} \right).
\end{aligned}$$

Thus,

$$\begin{aligned}
\sum_{t=1}^{\alpha N} \sum_{k=N-t}^{N-1} \frac{1}{k^{2-2(d_1-d_2)}} &\leq \frac{K}{N^{2(d_1-d_2)}} \frac{1}{N} \sum_{t=1}^{\alpha N} \left(\frac{1}{(1-t/N)^{1-2(d_1-d_2)}} - 1 \right) \\
&\leq \frac{K}{N^{2(d_1-d_2)}} \int_0^{\alpha} \left(\frac{1}{(1-x)^{1-2(d_1-d_2)}} - 1 \right) dx \\
&\leq \frac{K}{N^{2(d_1-d_2)}} \frac{1}{(1-\alpha)^{-2(d_1-d_2)}} = K(N(1-\alpha))^{2(d_1-d_2)} \\
&= O(1), \qquad (28)
\end{aligned}$$

because $1-\alpha = \mathcal{O}(1/N)$, and therefore $N(1-\alpha) = \mathcal{O}(1)$.

On the other hand, for $\alpha N + 1 \leq t \leq N-1$ we can write using Bertrand's series convergence theorem that

$$\sum_{k=N-t}^{N-1} \frac{1}{k^{2-2(d_1-d_2)}} \leq \sum_{k=1}^{\infty} \frac{1}{k^{2-2(d_1-d_2)}} < \infty.$$

Therefore

$$\sum_{t=\alpha N+1}^{N-1} \sum_{k=N-t}^{N-1} \frac{1}{k^{2-2(d_1-d_2)}} \leq K \sum_{t=\alpha N+1}^{N-1} \leq KN(1-\alpha) = \mathcal{O}(1). \tag{29}$$

By combining (27), (28), and (29) we obtain the result. □

Lemma 3. *Let's define*

$$D(N,T) = \frac{1}{N} \sum_{t=0}^{N-1} \sum_{k=N-t+1}^{N-1} \frac{\psi(k)}{k^{2-2(d_1-d_2)}} \left(\frac{t+1-N/2}{T} \right),$$

with a function $|\psi(k)| \leq C \log N$ *for all* $0 \leq k \leq N$, $N > 1$, *where* C *is a positive constant. Then there exists a constant* $K > 0$ *such that*

$$|D(N,T)| \leq K \frac{\log N}{T}.$$

Proof. The proof follows a similar structure to that of the previous Lemma 2, and the steps parallel those of Lemma 8 in Palma and Olea (2010b). □

A.2 Technical proofs

In this section, we provided the proofs of the asymptotic results stated in Theorems 1 and 2. Before proving these theorems, we present propositions and their proofs that aid in establishing the main results. Additionally, auxiliary lemmas and their proofs are included in the Appendix. Throughout this section, K is a generic constant that can vary from line to line. We also consider the function function $\phi : [0,1] : \times [-\pi, \pi] \longrightarrow \mathbb{R}$ and introduce the corresponding functional operator:

$$J_f(\phi) = \int_0^1 \int_{-\pi}^{\pi} \phi(u,\lambda) P(\lambda), \tag{30}$$

where $f(\lambda)$ can represent successively $f_Y(\lambda)$, $f_X(\lambda)$, $f_{XY}(\lambda)$, $f_{YX}(\lambda)$ and we define its sample version as:

$$J_T^I(\phi) = \frac{1}{M} \sum_{j=1}^{M} \int_{-\pi}^{\pi} \phi(u_j,\lambda) I_N(u_j,\lambda), \tag{31}$$

where $I_N(u,\lambda)$ is either $I_N^Y(u,\lambda)$, $I_N^X(u,\lambda)$, $I_N^{XY}(u,\lambda)$ or $I_N^{YX}(u,\lambda)$.

We also denote the Fourier transforms as follows:

$$\widehat{f}(u,\cdot) = \int_{-\pi}^{\pi} f(u,\lambda) e^{i\lambda\cdot} d\lambda.$$

$$\widehat{f}(u,v,\cdot) = \int_{-\pi}^{\pi} A(u,\lambda) A(v,\lambda) e^{i\lambda\cdot} d\lambda.$$

Moreover, we specify the matrix

$$Q(u) = \left(\int_{-\pi}^{\pi} \phi(u,\lambda) e^{i\lambda(s-t)} d\lambda \right)_{s,t=1,\ldots,N}, \tag{32}$$

and the block diagonal matrix $Q(\phi) = diag[Q(u_1), \ldots, Q(u_M)]$.

A.2.1 Propositions

We extend propositions 1 and 2 in Palma and Olea (2010a) to the case of the cross-spectral density and the cross-periodogram. All the proofs are straightforward, closely following the steps outlined in Palma and Olea (2010a).

Proposition 1 (Proposition 1, Palma and Olea (2010a)). *Let $f_{\delta_2}(\lambda)$ be the spectral density of ε_t satisfying assumption A1 and assume that $\phi(u,\lambda)$ appearing in (30) is symmetric in λ and twice differentiable with respect to u. Let $\widehat{f}_{\delta_2}(k)$ and $\widehat{\phi}(u,k)$ be their Fourier coefficients respectively. If there is a positive constant K such that*

$$|\widehat{f}_{\delta_2}(u,k)\widehat{\phi}(u,k)| \leq K\left(\frac{\log k}{k^2}\right),$$

for all $u \in [0,1]$ and $k > 1$, then, under assumptions A1 and A3 we have that

$$\mathbb{E}[J_T^{I_N^\theta}(\phi)] = J_{f_{\delta_2}}(\phi) + O\left(\frac{\log^2 N}{N}\right) + O\left(\frac{1}{M}\right), \tag{33}$$

The following proposition is similar to the previous one.

Proposition 2. *Let $f(u,\lambda)$ and $I_N(u_j,\lambda)$ be successively $\{f_Y(u,\lambda), I_N^Y(u,\lambda)\}$, $\{f_X(u,\lambda), I_N^X(u,\lambda)\}$, $\{f_{YX}(u,\lambda), I_N^{YX}(u,\lambda)\}$, $\{f_{XY}(u,\lambda), I_N^{XY}(u,\lambda)\}$ and assume that $\phi(u,\lambda)$ appearing in (30) is symmetric in λ and twice differentiable with respect to u. Let $\widehat{f}(u,k)$ and $\widehat{\phi}(u,k)$ be their Fourier coefficients respectively. If there is a positive constant K such that*

$$|\widehat{f}(u,k)\widehat{\phi}(u,k)| \leq K\left(\frac{\log k}{k^{2-2(\delta_1-\delta_2)}}\right),$$

for all $u \in [0,1]$ and $k > 1$, then, under assumptions A1 and A3 we have that

$$\mathbb{E}[J_T^{I_N}(\phi)] = J_f(\phi) + O\left(\frac{\log N}{N^{1-2(\delta_1-\delta_2)}}\right) + O\left(\frac{1}{M}\right), \tag{34}$$

for $\{f, I_N\}$ described as above.

Proof. The proof of this proposition follows exactly the steps outlined in proposition 1. The only things that have slightly changed are that instead of applying Lemma 7 and 8 of Palma and Olea (2010a), we now apply Lemma 2 and Lemma 3. □

Proposition 3. *[Proposition 2, Palma and Olea (2010a)] Let $f_{\delta_2}(\lambda)$ be the spectral density of ε_t satisfying assumption A1. Let $\phi_1, \phi_2 : [0,1] \to \mathbb{R}$ be two functions such that $\phi_1(u,\lambda)$ and $\phi_2(u,\lambda)$ are symmetric in λ, and twice differentiable with respect to u and their Fourier coefficients satisfy $\widehat{\phi}_1(u,k), \widehat{\phi}_2(u,k) \leq K|k|^{-2\delta_2-1}$ for $u \in [0,1]$ and $|k| > 1$.*

If assumptions A2 and A3 hold, then

$$\lim_{T\to\infty} T^{1-2(\delta_1-\delta_2)}\mathrm{Cov}\left[J_T^{I_N^\theta}(\phi_1), J_T^{I_N^\theta}(\phi_2)\right] = 4\pi\int_0^1\int_{-\pi}^{\pi}\phi_1(u,\lambda)\phi_2(u,\lambda)f_{\delta_2}^2(u,\lambda)d\lambda du.$$

Proposition 4. *Let $f_1(u,\lambda)$, $I_{N,1}(u_j,\lambda)$, $f_2(u,\lambda)$ and $I_{N,2}(u_j,\lambda)$ be as described in Proposition 2. Let $\phi_1, \phi_2 : [0,1] \to \mathbb{R}$ be two functions such that $\phi_1(u,\lambda)$ and $\phi_2(u,\lambda)$ are symmetric in λ, and twice differentiable with respect to u and their Fourier coefficients satisfy $\widehat{\phi}_1(u,k), \widehat{\phi}_2(u,k) \leq K|k|^{-2\delta_1-1}$ for $u \in [0,1]$ and $|k| > 1$.*

If assumptions A2 and A3 hold, then

$$\lim_{T\to\infty} T^{1-2(\delta_1-\delta_2)}\mathrm{Cov}\left[J_T^{I_{N,1}}(\phi_1), J_T^{I_{N,2}}(\phi_2)\right] = 4\pi\int_0^1\int_{-\pi}^{\pi}\phi_1(u,\lambda)\phi_2(u,\lambda) \times f_1(u,\lambda)f_2(u,\lambda)d\lambda du.$$

Proof. The proof is similar to Proposition 2 in Palma and Olea (2010a). In this proof, we apply Lemma 1 instead of Lemma 3 from Palma and Olea (2010a). It's worth noting that Lemma 1 is a more generalized version. The expression of C_N, as depicted in equation 26 of Palma and Olea (2010a), takes on the following form in our case:

$$\begin{aligned} C_N &= O\left(\left[\frac{\log M}{ST^{2(\delta_1-\delta_2)}} + \frac{T^{1-2(\delta_1-\delta_2)}}{N^2}\right] N^{4(\delta_1-\delta_2)} + NT^{\delta_1-1-2(\delta_1-\delta_2)}\right) \\ &= O\left(\frac{\log M\, N^{4(\delta_1-\delta_2)}}{ST^{2(\delta_1-\delta_2)}} + \left(\frac{T}{N^2}\right)^{1-2(\delta_1-\delta_2)} + NT^{\delta_1-1-2(\delta_1-\delta_2)}\right) \end{aligned}$$

We conclude with assumption A3 that $C_N = o(1)$. □

Proposition 5. *Let* $\mathrm{Cum}_p(\cdot)$ *be the* pth *order cumulant with* $p \geq 3$. *Then,*

$$T^{(1-2(\delta_1-\delta_2))\frac{p}{2}}\mathrm{Cum}_p[\nabla_\theta \mathcal{L}_T(\theta,\delta_2)] \to 0,$$

as $T \to \infty$.

Proof. For notational simplicity, we will omit θ from $\nabla_\theta \mathcal{L}_T(\theta,\delta_2)$ so that it becomes $\nabla \mathcal{L}_T(\theta,\delta_2)$.

First, the expression for $\nabla \mathcal{L}_T(\theta,\delta_2)$ is provided as:

$$\begin{aligned} \nabla \mathcal{L}_T(\theta,\delta_2) &= \frac{1}{4\pi}\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi} \nabla I_N^\theta(u_j,\lambda) f_{\delta_2}^{-1}(u,\lambda)d\lambda \\ &= \frac{1}{4\pi}\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi} \Big(-\nabla\beta_\theta(u_j) I_N^{YX}(u_j,\lambda) - \nabla\beta_\theta(u_j) I_N^{XY}(u_j,\lambda) \\ &\qquad + \nabla\beta_\theta^2(u_j) I_N^{Y}(u_j,\lambda)\Big) \times f_{\delta_2}^{-1}(u,\lambda)d\lambda \qquad (35) \\ &= \frac{1}{2\pi M H_{2,N}(0)}\left[Y'Q(\phi_1)X + X'Q(\phi_1)Y + Y'Q(\phi_2)Y\right], \qquad (36) \end{aligned}$$

where $\phi_1(u,\lambda) = -\nabla\beta_\theta(u) f_{\delta_2}^{-1}(u,\lambda)$, $\phi_2(u,\lambda) = \nabla\beta_\theta(u)^2 f_{\delta_2}^{-1}(u,\lambda)$

$$\begin{cases} \phi_1(u,\lambda) &= -\nabla\beta_\theta(u) f_{\delta_2}^{-1}(u,\lambda) \\ \phi_2(u,\lambda) &= \nabla\beta_\theta(u)^2 f_{\delta_2}^{-1}(u,\lambda), \end{cases} \qquad (37)$$

the block-diagonal matrix $Q(\cdot)$ is defined in (32) and $Y, X \in \mathbb{R}^{NM}$ are Gaussian random vectors defined by $Y = (Y(u_1)', \ldots, Y(u_M)')$, $X = (X(u_1)', \ldots, X(u_M)')$, $Y(u) = (Y_1(u), \ldots, Y_N(u))$, $X(u) = (X_1(u), \ldots, X_N(u))$, $Y_t(u) = h\left(\frac{t}{n}\right) Y_{[uT]-N/2+t+1}$ and $X_t(u) = h\left(\frac{t}{n}\right) X_{[uT]-N/2+t+1}$. For simplicity, we write the matrix $Q(\phi_1)$ as Q_1, $Q(\phi_2)$ as Q_2 and (36) becomes as follow when we consider by $Z = (Y, X)'$:

$$\nabla \mathcal{L}_T(\theta,\delta_2) = \frac{1}{2\pi M H_{2,N}(0)} Z'QZ, \qquad (38)$$

where

$$Q = \begin{pmatrix} Q_2 & Q_1 \\ Q_1 & 0 \end{pmatrix}.$$

Since Z is Gaussian we have that:

$$\mathrm{Cum}_p(\nabla \mathcal{L}_T(\theta,\delta_2)) = \frac{2^{p-1}(p-1)!}{(2\pi M H_{2,N}(0))^p} tr(RQ)^p \leq K\frac{tr(RQ)^p}{M^p N^p}, \qquad (39)$$

where $R = Var(Z)$.

$$R = Var(Z) = \begin{pmatrix} Var(Y) & \text{Cov}(Y,X) \\ \text{Cov}(X,Y) & Var(X) \end{pmatrix} = \begin{pmatrix} R_Y & R_{YX} \\ R_{XY} & R_X \end{pmatrix} \tag{40}$$

Let denote by $|A| = [tr(A'A)]^{\frac{1}{2}}$ be the Euclidian norm of the matrix A and $||A|| = \sup_{||x||=1}(Ax)'Ax$ be the spectral norm of A. Since $|tr(QB)| \leq |Q||B|$ and $|QB| \leq ||Q|||B|$ we get

$$|tr(RQ)^p| \leq ||RQ||^{p-2}|RQ|^2 \tag{41}$$

Additionally, for fixed λ, decompose the functions $\phi_1(\cdot,\lambda)$ and $\phi_2(\cdot,\lambda)$ as $\phi_1(\cdot,\lambda) = \phi_{1,+}(\cdot,\lambda) - \phi_{1,-}(\cdot,\lambda)$ and $\phi_2(\cdot,\lambda) = \phi_{2,+}(\cdot,\lambda) - \phi_{2,-}(\cdot,\lambda)$ where $\phi_{1,+}(\cdot,\lambda)$, $\phi_{1,-}(\cdot,\lambda)$, $\phi_{2,+}(\cdot,\lambda)$, $\phi_{2,-}(\cdot,\lambda) \geq 0$. Thus, we can write $Q_1 = Q(\phi_1) = Q(\phi_{1,+}) - Q(\phi_{1,-}) = Q_{1,+} - Q_{1,-}$ and $Q_2 = Q(\phi_2) = Q(\phi_{2,+}) - Q(\phi_{2,-}) = Q_{2,+} - Q_{2,-}$. So, $Q = Q_+ - Q_-$ where

$$Q_+ = \begin{pmatrix} Q_{2,+} & Q_{1,+} \\ Q_{1,+} & 0 \end{pmatrix}, Q_- = \begin{pmatrix} Q_{2,-} & Q_{1,-} \\ Q_{1,-} & 0 \end{pmatrix}.$$

Now, using the fact that $||RQ|| \leq ||RQ_+|| + ||RQ_-||$ and (40) we can write $||RQ_+||$ as follows:

$$\begin{aligned} ||RQ_+|| \quad \leq \quad & K\left(||R_Y Q_{2,+}|| + ||R_Y Q_{1,+}|| + ||R_{YX} Q_{1,+}|| \right. \\ & \left. + ||R_{XY} Q_{2,+}|| + ||R_{XY} Q_{1,+}|| + ||R_X Q_{1,+}||\right). \end{aligned}$$

By Lemma 6 in Palma and Olea (2010a) we conclude that

$$||RQ_+|| \leq K(MN^{1-2\delta_2}T^{2\delta_1-1}),$$

and by Proposition 2 in Palma and Olea (2010a) we have that $|RQ|^2 \leq K\frac{M^2N^2}{T}$. Thus, (41) becomes

$$\begin{aligned} |tr(RQ)^p| \quad & \leq \quad K(MN^{1-2\delta_2}T^{2\delta_1-1})^{p-2}\frac{M^2N^2}{T} \\ & \leq \quad K(MN^{1-2\delta_2}T^{2\delta_2-1}T^{2(\delta_1-\delta_2)})^{p-2}\frac{M^2N^2}{T}. \end{aligned} \tag{42}$$

Consequently by combining (39) and (42) we have,

$$|T^{(1-2(\delta_1-\delta_2))\frac{p}{2}}\text{Cum}_p[\nabla_\theta \mathcal{L}_T(\theta,\delta_2)]| \quad \leq \quad \frac{K}{T^{2(\delta_2-\delta_1)}}\left(\frac{N}{T}\right)^{(1-2\delta_2)(p-2)}\left(\frac{\sqrt{T}T^{(\delta_1-\delta_2)}}{N}\right)^{p-2}$$

Since $p \geq 3$ and by assumption A3, $N/T \to 0$ and $(\sqrt{T}T^{\delta_1-\delta_2})/N \to 0$ as T, $N \to \infty$, we obtain the result. □

A.2.2 Proof of theorems

Proof of Theorem 1. The proof for the consistency of $\widehat{\delta}_2$ is already established in Theorem 2.1 of Palma and Olea (2010a). We can follow a similar approach to demonstrate the consistency of $\widehat{\theta}$. We need to show that

$$\sup_\theta |\mathcal{L}_T(\theta) - \mathcal{L}(\theta)| \xrightarrow{p} 0,$$

as $T \to \infty$, where

$$\mathcal{L}(\theta) = \frac{1}{4\pi}\int_0^1 \int_{-\pi}^{\pi} \left\{\log f_{\delta_2}(\lambda) + \frac{f_\theta(u,\lambda)}{f_{\delta_2}(\lambda)}\right\} d\lambda du, \tag{43}$$

with $f_\theta(u,\lambda)$ given in (12). The proof follows in the same way and we therefore obtain the consistency of $\widehat{\theta}$. □

Proof of Theorem 2. We first start by applying the mean value theorem which tells us that there exists a vector $\overline{\theta}$, such that

$$\nabla\mathcal{L}_T(\widehat{\theta}) - \nabla\mathcal{L}_T(\theta_0) = [\nabla^2\mathcal{L}_T(\overline{\theta})](\widehat{\theta} - \theta_0), \tag{44}$$

with $||\overline{\theta} - \theta_0|| \leq ||\widehat{\theta} - \theta_0||$. Therefore, if we prove:

a) $\nabla^2\mathcal{L}_T(\overline{\theta}) - \nabla^2\mathcal{L}_T(\theta_0) \xrightarrow{p} 0$ in probability, as $T \to \infty$,

b) $\nabla^2\mathcal{L}_T(\theta_0) \xrightarrow{p} \Lambda(\theta_0)$ in probability, as $T \to \infty$,

c) $\sqrt{T}\nabla\mathcal{L}_T(\theta_0) \xrightarrow{d} N[0, \Lambda(\theta_0)]$ in distribution, as $T \to \infty$,

then the result will follow.

1. The expression of $\nabla^2\mathcal{L}_T(\theta)$ is given by:

$$\begin{aligned}
\nabla^2\mathcal{L}_T(\theta) &= \frac{1}{4\pi}\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\nabla^2 I_N^{\theta}(u_j,\lambda)f_{\delta_2}^{-1}(u,\lambda)d\lambda \\
&= \frac{1}{4\pi}\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\Big(-\nabla^2\beta_\theta(u_j)I_N^{YX}(u_j,\lambda) - \nabla^2\beta_\theta(u_j)I_N^{XY}(u_j,\lambda) \\
&\qquad + \nabla^2\beta_\theta^2(u_j)I_N^{Y}(u_j,\lambda)\Big) \times f_{\delta_2}^{-1}(u,\lambda)d\lambda. \qquad (45) \\
&= \frac{1}{2\pi M H_{2,N}(0)}\left[Y'Q(\phi_3)X + X'Q(\phi_3)Y + Y'Q(\phi_4)Y\right],
\end{aligned}$$

   where $\phi_3(u,\lambda) = -\nabla^2\beta_\theta(u)f_{\delta_2}^{-1}(u,\lambda)$, $\phi_4(u,\lambda) = \nabla^2\beta_\theta(u)^2 f_{\delta_2}^{-1}(u,\lambda)$.
   To prove (a) we need to observe that $\nabla^2\mathcal{L}_T(\theta)$ is equicontinuous in probability which is obtained analogously with Lemma 2.7 in Dahlhaus (2000). Hence, (a) follows from the fact that $\overline{\theta} \xrightarrow{p} \theta_0$ and the equicontinuity of $\nabla^2\mathcal{L}_T(\theta)$.

2. On the other hand, (b) follows from the application of Proposition 2 and Proposition 4.

3. Finally, to prove (c) we use the method of cumulants.

   (i) We obtain from (35) using Proposition 2 ,

$$\mathrm{E}\left[\nabla\mathcal{L}_T(\theta_0)\right] = \nabla\mathcal{L}(\theta_0) + \mathcal{O}\left(\frac{\log N}{N^{1-2(\delta_1-\delta_2)}}\right) + \mathcal{O}\left(\frac{1}{M}\right),$$

   where $\mathcal{L}(\theta)$ is defined in 43 and

$$\nabla\mathcal{L}(\theta) = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\nabla f_\theta(u,\lambda)f_{\delta_2}^{-1}(u,\lambda)d\lambda du.$$

   Since $\nabla\mathcal{L}(\theta_0) = 0$ then by assumption A3, we get

$$\begin{aligned}
\sqrt{T^{1-2(\delta_1-\delta_2)}}\mathrm{E}\left[\nabla\mathcal{L}_T(\theta_0)\right] &= \mathcal{O}\left(\frac{\sqrt{T^{1-2(\delta_1-\delta_2)}}\log N}{N^{1-2(\delta_1-\delta_2)}}\right) \\
&\quad + \mathcal{O}\left(\frac{\sqrt{T^{1-2(\delta_1-\delta_2)}}}{M}\right) \to 0.
\end{aligned}$$

(ii) Furthermore, we obtain through (35) and Proposition 4 (with ϕ_1 and ϕ_2 given in (37)) that the second-order cumulant satisfies

$$\begin{aligned}
&\lim_{T\to\infty} T^{1-2(\delta_1-\delta_2)}\mathrm{Cov}[\nabla\mathcal{L}_T(\theta_0),\nabla\mathcal{L}_T(\theta_0)]\\
&\quad=\frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\Big[4(\nabla\beta_{\theta_0}(u,\lambda))^2 f_{YX}^2(u,\lambda)+(\nabla\beta_{\theta_0}^2(u,\lambda))^2 f_Y^2(u,\lambda)\\
&\quad\quad -2\nabla\beta_{\theta_0}(u,\lambda)\nabla\beta_{\theta_0}^2(u,\lambda)f_{YX}(u,\lambda)f_Y(u,\lambda)\Big](f_{\delta_{20}}^{-1}(u,\lambda))^2\\
&\quad=\frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\Big[(\nabla f_{\theta_0}(u,\lambda))^2+2\nabla\beta_{\theta_0}(u,\lambda)\beta_{\theta_0}^2(u,\lambda)d\lambda du\\
&\quad\quad\times f_{YX}(u,\lambda)f_Y(u,\lambda))\Big](f_{\delta_{20}}^{-1}(u,\lambda))^2 d\lambda du = V(\theta_0).
\end{aligned}$$

(iii) Finally, we obtain with Proposition 5 that the cumulants of order more than or equal to 3 are zero.

Therefore, the asymptotic normality of $\widehat{\theta}$ follows and the theorem is proved. □

A.3 Additional simulations

This appendix presents additional Monte Carlo simulation results in different tables, based on other model specifications. For all the tables below we choose $u_0 = 64/T$ and $\beta(u) = \beta_0 + (\beta_1 - \beta_0)\mathrm{Sigmoid}\left(\frac{(u-u_0)}{0.00005}\right)$.

1. $X_t \sim ARFIMA(0, 0.35, 1)$, and $\varepsilon_t \sim ARFIMA(0, 0.1, 1)$

$T=512,\quad N=380,\quad S=76$								
Parameters			Estimates			Estimated SD		
β_0	β_1	δ_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\delta}_2$	$\hat{\sigma}(\hat{\beta}_0)$	$\hat{\sigma}(\hat{\beta}_1)$	$\hat{\sigma}(\hat{\delta}_2)$
1	2	0.1	0.989	2.001	0.09	0.168	0.075	0.094
2	4	0.1	2.02	3.977	0.12	0.161	0.054	0.099
0.5	-0.5	0.1	0.502	-0.519	0.12	0.162	0.061	0.09
$T=1024,\quad N=380,\quad S=76$								
Parameters			Estimates			Estimated SD		
β_0	β_1	δ_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\delta}_2$	$\hat{\sigma}(\hat{\beta}_0)$	$\hat{\sigma}(\hat{\beta}_1)$	$\hat{\sigma}(\hat{\delta}_2)$
1	2	0.1	0.998	1.999	0.104	0.16	0.045	0.059
2	4	0.1	1.995	3.998	0.105	0.160	0.044	0.062
0.5	-0.5	0.1	0.509	-0.504	0.12	0.15	0.043	0.063

Table 6: Estimation for $X_t \sim ARFIMA(0, 0.35, 1)$, and $\varepsilon_t \sim ARFIMA(0, 0.1, 1)$.

2. $X_t \sim ARFIMA(1, 0.35, 1)$, and $\varepsilon_t \sim ARFIMA(0, 0.1, 1)$

$T = 512, \quad N = 100, \quad S = 86$								
Parameters			Estimates			Estimated SD		
β_0	β_1	δ_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\delta}_2$	$\hat{\sigma}(\hat{\beta}_0)$	$\hat{\sigma}(\hat{\beta}_1)$	$\hat{\sigma}(\hat{\delta}_2)$
1	2	0.1	1.045	2.000	0.105	0.151	0.069	0.096
2	4	0.1	2.07	4.001	0.11	0.155	0.072	0.098
0.5	-0.5	0.1	0.47	-0.5	0.1	0.157	0.072	0.1
$T = 1024, \quad N = 100, \quad S = 86$								
Parameters			Estimates			Estimated SD		
β_0	β_1	δ_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\delta}_2$	$\hat{\sigma}(\hat{\beta}_0)$	$\hat{\sigma}(\hat{\beta}_1)$	$\hat{\sigma}(\hat{\delta}_2)$
1	2	0.1	1.044	2.001	0.108	0.150	0.046	0.061
2	4	0.1	2.089	4.000	0.109	0.150	0.047	0.059
0.5	-0.5	0.1	0.47	-0.505	0.127	0.151	0.043	0.063

Table 7: Estimation for $X_t \sim ARFIMA(1, 0.35, 1)$, and $\varepsilon_t \sim ARFIMA(0, 0.1, 1)$.

References

Richard T Baillie and Tim Bollerslev. Cointegration, fractional cointegration, and exchange rate dynamics. *The Journal of Finance*, 49(2):737–745, 1994.

Janis Becker, Fabian Hollstein, Marcel Prokopczuk, and Philipp Sibbertsen. The memory of beta. *Journal of Banking & Finance*, 124:106026, 2021.

Jan Beran. On parameter estimation for locally stationary long-memory processes. *Journal of Statistical Planning and Inference*, 139(3):900–915, 2009.

Jane P Brown, Haiyan Song, and Alan McGillivray. Forecasting uk house prices: A time varying coefficient approach. *Economic Modelling*, 14(4):529–548, 1997.

Federico Carlini, Bent Jesper Christensen, Nabanita Datta Gupta, and Paolo Santucci de Magistris. Climate, wind energy, and co2 emissions from energy production in denmark. *Energy Economics*, 125:106821, 2023.

Ngai Hang Chan and Wilfredo Palma. On the estimation of locally stationary long-memory processes. *Statistica Sinica*, 30(1):111–134, 2020.

Yin-Wong Cheung and Kon S Lai. A fractional cointegration analysis of purchasing power parity. *Journal of Business & Economic Statistics*, 11(1):103–112, 1993.

Roxana Chiriac and Valeri Voev. Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947, 2011.

Bent Jesper Christensen and Morten Ørregaard Nielsen. Semiparametric analysis of stationary fractional cointegration and the implied-realized volatility relation. 2002.

Afonso Gonçalves da Silva and Peter M Robinson. Fractional cointegration in stochastic volatility models. *Econometric Theory*, 24(5):1207–1253, 2008.

Rainer Dahlhaus. On the kullback-leibler information divergence of locally stationary processes. *Stochastic processes and their applications*, 62(1):139–168, 1996.

Rainer Dahlhaus. Fitting time series models to nonstationary processes. *The annals of Statistics*, 25 (1):1–37, 1997.

Rainer Dahlhaus. A likelihood approximation for locally stationary processes. *The Annals of Statistics*, 28(6):1762–1794, 2000.

Rainer Dahlhaus and Wolfgang Polonik. Empirical spectral processes for locally stationary time series. 2009.

Robert F Engle and Clive WJ Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276, 1987.

Søren Johansen. Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2-3):231–254, 1988.

George Kapetanios, Stephen Millard, Katerina Petrova, and Simon Price. Time-varying cointegration with an application to the uk great ratios. *Economics Letters*, 193:109213, 2020.

Paul Newbold and Theodore Bos. Stochastic parameter regression models. *(No Title)*, 1985.

Wilfredo Palma and Ricardo Olea. An efficient estimator for locally stationary gaussian long-memory processes. 2010a.

Wilfredo Palma and Ricardo Olea. Supplement to "an efficient estimator for locally stationary gaussian long-memory processes.", 2010b.

Soo Kyung Park and Choel Beom Park. Time-varying cointegration models and exchange rate predictability in korea. *KDI Journal of Economic Policy*, 37(4):1–20, 2015.

Peter CB Phillips and Sam Ouliaris. Testing for cointegration using principal components methods. *Journal of Economic Dynamics and Control*, 12(2-3):205–230, 1988.

Maurice B Priestley. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 27(2):204–229, 1965.

Peter M Robinson and FJ Hidalgo. Time series regression with long-range dependence. *The Annals of Statistics*, 25(1):77–104, 1997.

Barr Rosenberg. A survey of stochastic parameter regression. In *Annals of Economic and Social Measurement, Volume 2, number 4*, pages 381–397. NBER, 1973.

Yoshihiro Yajima. Asymptotic properties of the lse in a regression model with long-memory stationary errors. *The Annals of Statistics*, pages 158–177, 1991.

OlaOluwa S Yaya, Oluwaseun A Adesina, Hammed A Olayinka, Oluseyi E Ogunsola, and Luis A Gil-Alana. Long memory cointegration in the analysis of maximum, minimum and range temperatures in africa: Implications for climate change. *Atmosphere*, 14(8):1299, 2023.

Jelmer Ypma. Introduction to nloptr: an r interface to nlopt. *R Package*, 2, 2014.