

***Multivariate Non-Linear Regression with
Applications: A Frequency Domain Approach***

Terdik, G and Subba Rao, T and Rao Jammalamadaka,
S

2006

MIMS EPrint: **2006.77**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: <http://eprints.maths.manchester.ac.uk/>

And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

Multivariate Non-Linear Regression with Applications: A Frequency Domain Approach

Gy. Terdik, T. Subba Rao & S. Rao Jammalamadaka

First version: 14 November 2005

Research Report No. 25, 2005, Probability and Statistics Group
School of Mathematics, The University of Manchester

Multivariate Non-Linear Regression with Applications: A Frequency Domain Approach

Gy. Terdik¹, T. Subba Rao² and S. Rao Jammalamadaka³

¹Department of Information Technology
Faculty of Informatics, University of Debrecen
4010 Debrecen, Pf.12, HU

²School of Mathematics, University of Manchester
PO Box 88, Manchester M60 1QD, UK

³Department of Statistics and Applied Probability
University of California
Santa Barbara, Ca. 93106, USA

ABSTRACT

In this paper we consider estimating the parameters of a multivariate multiple nonlinear regression model with correlated errors, through the use of Finite Fourier Transforms. Consistency and asymptotic normality of the weighted least squares estimates are established under various conditions on the regressor variables. These conditions involve different types of scalings, and such scaling factors are obtained explicitly for various nonlinear regression models including an interesting model which requires estimating the frequencies. This is a very classical problem in signal processing and is also of great interest in many other areas. We illustrate our techniques on the time-series data of polar motion (which is now widely known as "Chandlers Wobble") where one has to estimate the drift parameters, the offset parameters and the two periodicities associated with elliptical motion. The data was first analyzed by Arato, Kolmogorov and Sinai who treat it as bivariate time series data satisfying a finite order time series model. They estimate the periodicities using the coefficients of the models. Our analysis shows that the two dominant frequencies are 12 hours and 410 days and that the errors exhibit some long-range dependence.

1 Introduction

One of the classical problems in statistical analysis is to find a suitable relationship between a response variable Y_t and a set of p regressor variables x_1, x_2, \dots, x_p under suitable assumptions on the errors. The usual assumption is that the errors are independent, identically distributed random variables. This was later generalized to the case when the errors are correlated. In many situations the response function is a nonlinear function in both regressor variables and the parameters. The asymptotic results on the estimators of the parameters are now well known (see for example Jennrich, 1969). The results were later extended to the case of nonlinear multiple regression with correlated errors. Hannan (1971) proves results when the errors satisfy a linear stationary process. The methods used by Hannan are frequency domain methods, which heavily depend on the properties of Finite Fourier Transforms. The set of conditions imposed on the regressor variables (see Grenander and Rosenblatt, 1957) depends on the nature of the nonlinear parameters to be estimated. The central limit theorems associated with the estimated and the scaling factors are also dependent on the parameters. If the parameters to be estimated are frequencies, even though the model looks like linear, but contain sine and cosine terms, one has to impose a different set of conditions than the usual conditions. A brief discussion of this important aspect was discussed by Hannan (1971). Robinson (1972) extended the results of Hannan to the multivariate nonlinear regression situation, when the regression matrix is not of full rank and the parameters satisfy some constraints. The methods and the asymptotic theory of Robinson does not include the situation when the parameters are frequencies and also the variance-covariance matrices of the estimated parameters given are not explicit enough to compute. In this paper our objective is to consider various forms of the relationships (linear in parameters and nonlinear in parameters and a mixture of both) and find suitable scaling factors for establishing asymptotic properties of the estimates. Our main result is the exact form for the asymptotic variance of the estimator in terms of the regression spectral density function and the weighting matrix. We introduce new scaling factors which are required for proving the central limit theorems of the parameters of the mixed type models (linear and/or nonlinear in parameters)

An interesting problem one encounters in practice is to compare the features (such as long time trends, common periodicities etc) between two or more of the response variables when the errors are correlated and may have different marginal distributions. The frequency domain approach is extremely useful in such contexts as we do not need to know the distributions of the errors. We do not go into details on testing hypotheses on these parameters.

More specifically we consider in this paper a multivariate multiple non-linear regression model with multivariate correlated stationary random errors satisfying some conditions. The minimum contrast estimate ([12], [16]) of the unknown parameters are constructed in frequency domain. The mixed model containing the linear regression and linear combinations of the nonlinear regressions is considered in detail.

To illustrate our methods we consider the analysis of time series of Polar motion (in geophysics this is widely described as "Chandlers Wobble"). One of the first papers dealing with statistical analysis of this data is due to Arato, Kolmogorov and Sinai (1962) who estimate the parameters of the polar motion, such as offset (trend), drift and periodicities. Using high resolution GPS data it is shown in

this paper that besides the well-known Chandler period of 410 days, a secondary period of 12 hours is also present and the residual series seems to exhibit long range dependency. The motion of the pole (as a function of time) can be described by two polar coordinates which are strongly correlated. This motion is very similar to that of a rotating spinning top that is used by children as a toy.

2 Non-Linear Time Series Regression

2.1 Model

Consider a d dimensional observational vector \underline{Y}_t , the random disturbances \underline{Z}_t and the function $\underline{X}_t(\vartheta_0)$ satisfying the usual model

$$\underline{Y}_t = \underline{X}_t(\vartheta_0) + \underline{Z}_t. \quad (1)$$

The function $\underline{X}_t(\vartheta)$ can be a nonlinear function of both regressors variables and the p -dimensional parameter vector $\vartheta \in \Theta \subset \mathbb{R}^p$, while \underline{Z}_t is a d -dimensional stationary time series. The set Θ of admissible parameters ϑ is defined by a number of possibly nonlinear equations, see [27], for more details. We shall assume that the set Θ is chosen suitably in each case. For convenience we consider $\underline{X}_t(\vartheta)$ as a regressor vector, although in particular cases we may have to separate the regressors and the parameters. The regressor variables may depend on the parameters nonlinearly. One specific model, we have in mind, for $\underline{X}_t(\vartheta)$ can be written in the form

$$\begin{aligned} \underline{X}_t(\vartheta) &= \mathbf{B}_1 \underline{X}_{1,t} + \mathbf{B}_2 \underline{X}_{2,t}(\underline{\lambda}) \\ &= [\mathbf{B}_1, \mathbf{B}_2] \begin{bmatrix} \underline{X}_{1,t} \\ \underline{X}_{2,t}(\underline{\lambda}) \end{bmatrix} \\ &= \mathbf{B} \underline{X}_{3,t}(\underline{\lambda}). \end{aligned}$$

This can be considered as a mixed model since it is both linear as well as non-linear in the parameters at the same time. The parameter vector ϑ contains both \mathbf{B}_1 and \mathbf{B}_2 and also the vector $\underline{\lambda}$. The admissible set Θ is the union of three subsets. There is no restriction on the entries of matrix \mathbf{B}_1 with size $d \times p_1$. However the matrix \mathbf{B}_2 and the vector $\underline{\lambda}$ may have to satisfy some identifiability conditions. The parameter $\underline{\lambda}$ is identified unless some particular entries of \mathbf{B}_2 annihilate an entry, say λ_k , from the model. If λ are set of frequencies to be estimated we may have to put some constraints so that they lie within a compact set, for harmonic regressors $\lambda \in [-1/2, 1/2]^2$, see Section 5.3. We assume that \underline{Z}_t is a stationary linear processes, and has the moving average representation

$$\begin{aligned} \underline{Z}_t &= \sum_{k=-\infty}^{\infty} \mathbf{A}_k \underline{W}_{t-k}, \\ \text{with } \sum_{k=-\infty}^{\infty} \text{Tr}(\mathbf{A}_k \mathbf{C}_W \mathbf{A}_k^*) &< \infty. \end{aligned}$$

Here we assume that \underline{W}_t is a sequence of independent identically distributed random vectors (**i.i.d.vectors**) and $\mathbf{C}_W = \text{Var} \underline{W}_t$, is non-singular. We also assume that \underline{Z}_t has a (element by element) **continuous** spectral density matrix $\mathbf{S}_Z(\omega)$, as in [13], [5]. The model is **feedback free**, i.e. \underline{Z}_t does not depend on \underline{X}_t .

2.2 The regression spectrum

Consider the regressor function $\underline{X}_t(\vartheta)$ which is a function of t which may depend nonlinearly on parameters ϑ belonging to the compact set Θ . It is widely known that the Grenander's conditions (see [10], [9]), for the regressor $\underline{X}_t(\vartheta)$ are sufficient and in some situations ([32]), also necessary to establish the consistency of the least squares (LS) estimators. They are as follows. Let

$$\|X_{k,t}(\vartheta)\|_T^2 = \sum_{t=1}^T X_{k,t}^2(\vartheta).$$

denote the Euclidean norm of the k th regressor of the vector $X(\vartheta)$

Condition 1 (G1) For all $k = 1, 2, \dots, d$,

$$\lim_{T \rightarrow \infty} \|X_{k,t}(\vartheta)\|_T^2 = \infty.$$

Condition 2 (G2) For all $k = 1, 2, \dots, d$,

$$\lim_{T \rightarrow \infty} \frac{X_{k,T+1}^2(\vartheta)}{\|X_{k,t}(\vartheta)\|_T^2} = 0.$$

Without any loss of generality we can assume that the regressor $\underline{X}_t(\vartheta)$ is *scaled*: $\|X_{k,t}(\vartheta)\|_T^2 \simeq T$, for all k , see Definition 15 and a note therein. Define the following matrices. Let for each integer $h \in [0, T)$,

$$\begin{aligned} \widehat{\mathbf{C}}_{\underline{X},T}(h, \vartheta_1, \vartheta_2) &= \frac{1}{T} \sum_{t=1}^{T-h} \underline{X}_{t+h}(\vartheta_1) \underline{X}_t^\top(\vartheta_2), \\ \widehat{\mathbf{C}}_{\underline{X},T}(-h, \vartheta_1, \vartheta_2) &= \widehat{\mathbf{C}}_{\underline{X},T}^\top(h, \vartheta_2, \vartheta_1). \end{aligned} \quad (2)$$

Whenever $\vartheta_1 = \vartheta_2 = \vartheta$, throughout our paper we use the shorter notation $\widehat{\mathbf{C}}_{\underline{X},T}(h, \vartheta_1, \vartheta_2)_{\vartheta_1=\vartheta_2=\vartheta} = \widehat{\mathbf{C}}_{\underline{X},T}(h, \vartheta)$. The next condition we impose essentially means that the regressor $\underline{X}_t(\vartheta)$ is changing 'slowly' in the following sense. For each integer h , $\|X_{k,t}(\vartheta)\|_{T+h}^2 \simeq T$.

Condition 3 (G3) For each integer h ,

$$\lim_{T \rightarrow \infty} \widehat{\mathbf{C}}_{\underline{X},T}(h, \vartheta) = \mathbf{C}_{\underline{X}}(h, \vartheta).$$

Condition 4 (G4) $\mathbf{C}_{\underline{X}}(0, \vartheta)$ is non-singular.

One can use the Bochner's theorem for the limit $\mathbf{C}_{\underline{X}}$ such that

$$\mathbf{C}_{\underline{X}}(h, \vartheta) = \int_{-1/2}^{1/2} \exp(i2\pi\lambda h) d\mathbf{F}(\lambda, \vartheta),$$

where \mathbf{F} is a *spectral distribution matrix function of the regressors* (from now on abbreviated, **SDFR**) whose entries are of bounded variations. The SDFR can be obtained as the limit of the vector of periodogram ordinates (see [5] for details). We now introduce the discrete Fourier transform

$$\underline{d}_{\underline{X},T}(\omega, \vartheta) = \sum_{t=0}^{T-1} \underline{X}_t(\vartheta) z^{-t}, \quad z = \exp(2\pi i\omega), \quad -\frac{1}{2} \leq \omega < \frac{1}{2},$$

and the periodogram of the non-random vector $\underline{X}_t(\vartheta)$ given by

$$\mathbf{I}_{\underline{X},T}(\omega, \vartheta_1, \vartheta_2) = \frac{1}{T} \underline{d}_{\underline{X},T}(\omega, \vartheta_1) \underline{d}_{\underline{X},T}^*(\omega, \vartheta_2),$$

where $*$ denotes the transpose and complex conjugate. Both $\underline{d}_{\underline{X},T}$ and $\mathbf{I}_{\underline{X},T}$ may depend on some parameters. We have the well known relation

$$\begin{aligned} \widehat{\mathbf{C}}_{\underline{X},T}(h, \vartheta_1, \vartheta_2) &= \int_{-1/2}^{1/2} \exp(i2\pi\lambda h) \mathbf{I}_{\underline{X},T}(\lambda, \vartheta_1, \vartheta_2) d\lambda, \\ \mathbf{I}_{\underline{X},T}(\omega, \vartheta_1, \vartheta_2) &= \sum_{|h| < T} \widehat{\mathbf{C}}_{\underline{X},T}(h, \vartheta_1, \vartheta_2) e^{-i2\pi\omega h}, \\ \mathbf{I}_{\underline{X},T}^\top(\omega, \vartheta_1, \vartheta_2) &= \overline{\mathbf{I}_{\underline{X},T}(\omega, \vartheta_2, \vartheta_1)}. \end{aligned}$$

between $\widehat{\mathbf{C}}_{\underline{X},T}$ and the periodogram. The definition (2), which Jennrich ([21]) calls tail product, reminds us of the empirical cross-covariance matrix of a stationary time series, usually scaled by $1/T$ (which might not work in some particular cases of the regressors without some additional scaling). This implies that the series \underline{X}_t does not belong to L_2 , i.e.

$$\lim_{T \rightarrow \infty} \underline{d}_{\underline{X},T}(\omega, \vartheta)^2 = \infty.$$

For the univariate case, we refer the reader to the classical books of Grenander-Rosenblatt [9], T. W. Anderson [3], and for vector-valued case, to Hannan [13], Brillinger [5]. Now, define the empirical SDFR \mathbf{F}_T as

$$\mathbf{F}_T(\omega, \vartheta_1, \vartheta_2) = \int_0^\omega \mathbf{I}_{\underline{X}, T}(\lambda, \vartheta_1, \vartheta_2) d\lambda,$$

then it follows from the Grenander's conditions stated above that \mathbf{F} is the weak limit of \mathbf{F}_T and this is condition we really need later. See also Chapter 7 of Ibragimov and Rozanov [18].

Condition 5 (I-R) *The matrix function \mathbf{F}_T converges to \mathbf{F} weakly. More precisely, for each continuous bounded function $\varphi(\omega)$ the limit*

$$\lim_{T \rightarrow \infty} \int_{-1/2}^{1/2} \varphi(\omega) d\mathbf{F}_T(\omega, \vartheta_1, \vartheta_2) = \int_{-1/2}^{1/2} \varphi(\omega) d\mathbf{F}(\omega, \vartheta_1, \vartheta_2), \quad (3)$$

holds.

Note, if \mathbf{F}_T converges to \mathbf{F} weakly then (3) is valid not only for continuous bounded functions but also for some wider classes of functions such as piecewise continuous functions having discontinuity at finitely many ω -points with \mathbf{F} -measure zero, $d\mathbf{F}(\omega, \vartheta_1, \vartheta_2) = 0$. This is very important, in particular for disturbances with long memory, see [34]. The matrix function \mathbf{F} is Hermite symmetric since \mathbf{F}_T satisfies following

$$\mathbf{F}_T^T(\omega, \vartheta_1, \vartheta_2) = \overline{\mathbf{F}_T(\omega, \vartheta_2, \vartheta_1)} = \mathbf{F}_T(-\omega, \vartheta_2, \vartheta_1).$$

The regressor $\underline{X}_t(\vartheta)$ depends on the parameter $\vartheta \in \Theta$, therefore we require all **Grenander's conditions to hold uniformly in ϑ** .

2.3 The Objective Function

The frequency domain analysis has a number of advantages. First all the Fourier transforms of a large stationary sample behave like i.i.d. complex Gaussian random variables under some broad assumptions, see [5]. The FFT is technically simple to use. For example, it turns the data \underline{Y}_t , $t = 1, 2, \dots, T$, from time domain into frequency domain $\underline{d}_{\underline{Y}, T}(\omega_k)$, (here we define the Fourier frequencies $\omega_k = k/T \in [-1/2, 1/2]$, $k = -T_1, \dots, -1, 0, 1, \dots, T_1$, where $T_1 = \text{Int}[(T-1)/2]$, only). It is obvious that

$$\underline{d}_{\underline{Y}, T}(\omega) = \underline{d}_{\underline{X}, T}(\omega, \vartheta_0) + \underline{d}_{\underline{Z}, T}(\omega),$$

The parameter vector ϑ_0 denotes the true unknown value and we would like to adjust the regressor $\underline{X}_t(\vartheta)$ in the model such that the distance

$$\underline{d}_{\underline{Y}, T}(\omega) - \underline{d}_{\underline{X}, T}(\omega, \vartheta) = \underline{d}_{\underline{X}, T}(\omega, \vartheta_0) - \underline{d}_{\underline{X}, T}(\omega, \vartheta) + \underline{d}_{\underline{Z}, T}(\omega), \quad (4)$$

is minimal, in some sense. The Euclidean distance, for instance, is

$$\sum_{k=-T_1}^{T_1} \left(\underline{d}_{\underline{Y}, T}(\omega_k) - \underline{d}_{\underline{X}, T}(\omega_k, \vartheta) \right)^2 = \sum_{k=-T_1}^{T_1} \left(\underline{d}_{\underline{X}, T}(\omega_k, \vartheta_0) - \underline{d}_{\underline{X}, T}(\omega_k, \vartheta) + \underline{d}_{\underline{Z}, T}(\omega_k) \right)^2,$$

which, by Parseval Theorem, actually corresponds to the sum of squares in time domain

$$\sum_{t=0}^{T-1} \|\underline{Y}_t - \underline{X}_t(\vartheta)\|^2 = \sum_{t=0}^{T-1} \|\underline{X}_t(\vartheta_0) - \underline{X}_t(\vartheta) + \underline{Z}_t\|^2.$$

Therefore minimizing either of the above two expressions leads to the same result. The \underline{Z}_t itself is not necessarily an i.i.d. hence we are facing a generalized non-linear regression problem with stationary residuals. The quadratic function we minimize, somewhat parallel to that suggested by Hannan [14] for scalar valued case, is

$$\begin{aligned} Q_T(\vartheta) &= \frac{1}{T^2} \sum_{k=-T_1}^{T_1} \left(\underline{d}_{\underline{Y}, T}(\omega_k) - \underline{d}_{\underline{X}, T}(\omega_k, \vartheta) \right)^* \Phi(\omega_k) \left(\underline{d}_{\underline{Y}, T}(\omega_k) - \underline{d}_{\underline{X}, T}(\omega_k, \vartheta) \right) \\ &= \frac{1}{T} \sum_{k=-T_1}^{T_1} \text{Tr}(\mathbf{I}_{\underline{Y}, T}(\omega_k) \Phi(\omega_k)) + \text{Tr}(\mathbf{I}_{\underline{X}, T}(\omega_k, \vartheta) \Phi(\omega_k)) - 2 \text{Re} \text{Tr}(\mathbf{I}_{\underline{Y}, \underline{X}, T}(\omega_k, \vartheta) \Phi(\omega_k)), \end{aligned} \quad (5)$$

where $\Phi(\omega_k)$ is a series of matrix weights, originated from a continuous, Hermitian matrix function Φ , satisfying $\Phi(\omega) \geq 0$. The equation (4) can be rewritten as

$$Q_T(\underline{\vartheta}) = \frac{1}{T} \sum_{k=-T_1}^{T_1} \text{Tr}(\mathbf{I}_{\underline{X},T}(\omega_k, \underline{\vartheta}_0) \Phi(\omega_k)) + \text{Tr}(\mathbf{I}_{\underline{X},T}(\omega_k, \underline{\vartheta}) \Phi(\omega_k)) + \text{Tr}(\mathbf{I}_{\underline{Z},T}(\omega_k) \Phi(\omega_k)) \\ + \frac{2}{T} \sum_{k=-T_1}^{T_1} \text{Tr}([\mathbf{I}_{\underline{X},\underline{Z},T}(\omega_k, \underline{\vartheta}_0) - \mathbf{I}_{\underline{X},\underline{Z},T}(\omega_k, \underline{\vartheta})] \Phi(\omega_k)) - \text{Tr}(\mathbf{I}_{\underline{X},T}(\omega_k, \underline{\vartheta}, \underline{\vartheta}_0) \Phi(\omega_k)).$$

The proof of $\mathbf{I}_{\underline{X},\underline{Z},T}(\omega_k, \underline{\vartheta}) \rightarrow 0$, a.s. and uniformly in $\underline{\vartheta}$ is given by Robinson [27], Lemma 1. Now, suppose Conditions I-R, (or G1-G4) hold and we take the limit

$$Q(\underline{\vartheta}) = \lim_{T \rightarrow \infty} Q_T(\underline{\vartheta}) \\ = \int_{-1/2}^{1/2} \text{Tr}(\Phi(\omega) d[\mathbf{F}(\omega, \underline{\vartheta}_0) + \mathbf{F}(\omega, \underline{\vartheta}) - \mathbf{F}(\omega, \underline{\vartheta}_0, \underline{\vartheta}) - \mathbf{F}(\omega, \underline{\vartheta}, \underline{\vartheta}_0)]) \\ + \int_{-1/2}^{1/2} \text{Tr}[\mathbf{S}_{\underline{Z}}(\omega) \Phi(\omega)] d\omega \\ = R(\underline{\vartheta}, \underline{\vartheta}_0) + \int_{-1/2}^{1/2} \text{Tr}[\Phi(\omega) \mathbf{S}_{\underline{Z}}(\omega)] d\omega. \quad (6)$$

The function

$$R(\underline{\vartheta}, \underline{\vartheta}_0) = \int_{-1/2}^{1/2} \text{Tr}(\Phi(\omega) d[\mathbf{F}(\omega, \underline{\vartheta}_0) + \mathbf{F}(\omega, \underline{\vartheta}) - \mathbf{F}(\omega, \underline{\vartheta}_0, \underline{\vartheta}) - \mathbf{F}(\omega, \underline{\vartheta}, \underline{\vartheta}_0)]),$$

is the only part of $Q(\underline{\vartheta})$ which depends on $\underline{\vartheta}$. We shall require the following condition to ensure the existence of the minimum, (see [27])

Condition 6 (R)

$$R(\underline{\vartheta}, \underline{\vartheta}_0) > 0, \quad \underline{\vartheta} \in \Theta, \quad \underline{\vartheta} \neq \underline{\vartheta}_0.$$

Then we have the *contrast function* R for $\underline{\vartheta}_0$

$$R(\underline{\vartheta}, \underline{\vartheta}_0) > R(\underline{\vartheta}_0) = 0.$$

Therefore we minimize the *contrast process* $Q_T(\underline{\vartheta})$ for $R(\underline{\vartheta}_0, \underline{\vartheta})$. Obviously

$$\lim_{T \rightarrow \infty} [Q_T(\underline{\vartheta}) - Q_T(\underline{\vartheta}_0)] = R(\underline{\vartheta}_0, \underline{\vartheta}).$$

The *minimum contrast estimator* $\hat{\underline{\vartheta}}_T$ is the value which realizes that minimum value of $Q_T(\underline{\vartheta})$

$$\hat{\underline{\vartheta}}_T = \arg \min_{\underline{\vartheta} \in \Theta} Q_T(\underline{\vartheta}).$$

One can easily see (using [25] Theorem 7, Ch. 7) under some additional assumptions given below, that $Q_T(\underline{\vartheta})$ is convex since the Hessian $\mathbf{H}Q(\underline{\vartheta}_0)$ is nonnegative definite. Therefore the next Theorem, due to Robinson [27], is valid not only for a compact set Θ , but also for more general classes of parameters, for example which belong to convex parameter set Θ as well. The minimum contrast method is also called the quasi-likelihood and it is very efficient in several cases even in non-Gaussian situations, for instance see [2].

Theorem 7 Under the assumptions I-R (or G1-4), and R, the minimum contrast estimator $\hat{\underline{\vartheta}}_T$ converges a.s. to $\underline{\vartheta}_0$.

3 Asymptotic Normality

For the asymptotic normality it is necessary to consider the second derivatives of the SDFR and their limits for the objective function as usual, see [27]. The matrix of the second derivatives of $\hat{\mathbf{C}}_{\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)$ can be calculated, by using the matrix differential calculus, ([25])

$$\frac{\partial^2 \hat{\mathbf{C}}_{\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_2^T \partial \underline{\vartheta}_1^T} = \frac{\partial}{\partial \underline{\vartheta}_1^T} \text{Vec} \left(\frac{\partial \text{Vec} \hat{\mathbf{C}}_{\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_2^T} \right).$$

Here the differentiation of the RHS can be carried out directly, see Section 7.2 in the Appendix. Notice, the order of the variables ϑ_1, ϑ_2 in $\widehat{\mathbf{C}}_{X,T}$, is opposite to the order of the partial derivatives: $\partial\vartheta_2^\top\partial\vartheta_1^\top$. The later means that one differentiates first by ϑ_2 and then by ϑ_1 which operates on the right hand side Starting the differentiating by ϑ_1 , and then followed by ϑ_2 , it can be written as ‘direct’ one

$$\frac{\partial^2 \widehat{\mathbf{C}}_{X,T}(h, \vartheta_1, \vartheta_2)}{\partial\vartheta_1^\top\partial\vartheta_2^\top} = (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \mathbf{K}_{d \cdot dp} \frac{\partial^2 \widehat{\mathbf{C}}_{X,T}(-h, \vartheta_2, \vartheta_1)}{\partial\vartheta_1^\top\partial\vartheta_2^\top},$$

where we used the commutation matrix $\mathbf{K}_{p \cdot d}$, see (20), \otimes denotes the Kronecker product, and \mathbf{U}_d is the $d \times d$ identity matrix. Following Hannan [14] we assume the following condition

Condition 8 (H) *All the second partial derivatives of the regressor $\underline{X}_t(\vartheta)$ exist and $\frac{\partial^2 \widehat{\mathbf{C}}_{X,T}(h, \vartheta_1, \vartheta_2)}{\partial\vartheta_2^\top\partial\vartheta_1^\top}$ converges to some limit, denote it by $\frac{\partial^2 \mathbf{C}_X(h, \vartheta_1, \vartheta_2)}{\partial\vartheta_2^\top\partial\vartheta_1^\top}$.*

It is necessary to emphasize that Condition H is

$$\frac{\partial^2 \mathbf{C}_X(h, \vartheta_1, \vartheta_2)}{\partial\vartheta_2^\top\partial\vartheta_1^\top} \stackrel{\bullet}{=} \lim_{T \rightarrow \infty} \frac{\partial^2 \widehat{\mathbf{C}}_{X,T}(h, \vartheta_1, \vartheta_2)}{\partial\vartheta_2^\top\partial\vartheta_1^\top},$$

where the left hand side is defined by the limit(which is the derivative of \mathbf{C}_X .) From now on we shall use the symbol $\stackrel{\bullet}{=}$ for the definition of the left side of an expression.

The above notation is used for the regression spectrum as well.

Condition 9 (I-R-H) *The derivative $\frac{\partial^2 \mathbf{F}_T(\omega, \vartheta_1, \vartheta_2)}{\partial\vartheta_1^\top\partial\vartheta_2^\top}$, of the matrix function \mathbf{F}_T converges weakly to some function denoted by $\frac{\partial^2 \mathbf{F}(\omega, \vartheta_1, \vartheta_2)}{\partial\vartheta_2^\top\partial\vartheta_1^\top}$.*

Again

$$\frac{\partial^2 \mathbf{F}(h, \vartheta_1, \vartheta_2)}{\partial\vartheta_2^\top\partial\vartheta_1^\top} \stackrel{\bullet}{=} \lim_{T \rightarrow \infty} \frac{\partial^2 \mathbf{F}_T(\omega, \vartheta_1, \vartheta_2)}{\partial\vartheta_1^\top\partial\vartheta_2^\top},$$

by definition. Using the above formulae for the derivatives we calculate the Hessian \mathbf{HF} for the SDFR \mathbf{F} as well, see Section 7.2 in the Appendix for the proof.

Lemma 10 *Assume Condition I-R-H, then*

$$\begin{aligned} \mathbf{HF}(\omega, \vartheta) &= \mathbf{H}_{\vartheta_1} \mathbf{F}(\omega, \vartheta_1, \vartheta_2) + \frac{\partial^2 \mathbf{F}(\omega, \vartheta_1, \vartheta_2)}{\partial\vartheta_1^\top\partial\vartheta_2^\top} \\ &\quad + \mathbf{H}_{\vartheta_2} \mathbf{F}(\omega, \vartheta_1, \vartheta_2) + \frac{\partial^2 \mathbf{F}(\omega, \vartheta_1, \vartheta_2)}{\partial\vartheta_2^\top\partial\vartheta_1^\top} \Big|_{\vartheta_1 = \vartheta_2 = \vartheta}, \end{aligned} \quad (7)$$

where the indirect derivative satisfies

$$\frac{\partial^2 \mathbf{F}(\omega, \vartheta_1, \vartheta_2)}{\partial\vartheta_1^\top\partial\vartheta_2^\top} = (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \mathbf{K}_{d \cdot dp} \frac{\partial^2 \mathbf{F}(-\omega, \vartheta_2, \vartheta_1)}{\partial\vartheta_1^\top\partial\vartheta_2^\top}.$$

3.1 Asymptotic Variance

For the variance of $\text{Vec} \frac{\partial Q_T(\vartheta_0)}{\partial\vartheta^\top}$ consider the expression

$$\begin{aligned} \text{Vec} \frac{\partial Q_T(\vartheta)}{\partial\vartheta^\top} &= \frac{1}{T} \sum_{k=-T_1}^{T_1} \frac{\partial \text{Vec} \mathbf{I}_{X,T}(\omega_k, \vartheta)}{\partial\vartheta^\top} - \frac{\partial (\text{Vec} \mathbf{I}_{Y,X,T}(\omega_k, \vartheta) + \text{Vec} \mathbf{I}_{X,Y,T}(\omega_k, \vartheta))^\top}{\partial\vartheta^\top} \\ &\quad \times [\text{Vec} \Phi^\top(\omega_k)]. \end{aligned} \quad (8)$$

Let Ψ be some matrix function of appropriate dimension, and introduce the following notation, which will be frequently used,

$$\mathbf{J}(\Psi, \mathbf{F}) = \int_{-1/2}^{1/2} (\mathbf{U}_p \otimes [\text{Vec}(\Psi^\top(\omega_k))]^\top) d \frac{\partial^2 \mathbf{F}(\omega, \vartheta_1, \vartheta_2)}{\partial\vartheta_2^\top\partial\vartheta_1^\top} \Big|_{\vartheta_1 = \vartheta_2 = \vartheta_0},$$

where \mathbf{U}_p denotes the identity matrix of order p .

Lemma 11

$$\lim_{T \rightarrow \infty} \text{Var} \sqrt{T} \text{Vec} \frac{\partial Q_T(\underline{\vartheta}_0)}{\partial \underline{\vartheta}^\top} = 4\mathbf{J}(\Phi \mathbf{S}_Z \Phi, \mathbf{F}).$$

See Section 7.3 in the Appendix for the proof. The limit of the Hessian is calculated from (8). The Hessians according to the $\mathbf{H}_{\underline{\vartheta}_1} \mathbf{I}_{X,T}(\omega_k, \underline{\vartheta}_1, \underline{\vartheta}_2)$ and $\mathbf{H}_{\underline{\vartheta}_2} \mathbf{I}_{X,T}(\omega_k, \underline{\vartheta}_1, \underline{\vartheta}_2)$ of the terms in (8) at $\underline{\vartheta}_1 = \underline{\vartheta}_2 = \underline{\vartheta}_0$ will cancel with $\mathbf{H}_{\underline{\vartheta}} \mathbf{I}_{Y,X,T}(\omega_k, \underline{\vartheta})$ and $\mathbf{H}_{\underline{\vartheta}} \mathbf{I}_{X,Y,T}(\omega_k, \underline{\vartheta})$ respectively, so we have to deal only with the mixed derivatives of $\mathbf{I}_{X,T}(\omega_k, \underline{\vartheta})$. See Section 7.4 in Appendix. Hence the Hessian of $Q(\underline{\vartheta})$ at $\underline{\vartheta} = \underline{\vartheta}_0$ follows.

Lemma 12

$$\mathbf{H}Q(\underline{\vartheta}_0) = \lim_{T \rightarrow \infty} [\mathbf{H}Q_T(\underline{\vartheta}_0)] = 2\mathbf{J}(\Phi, \mathbf{F}).$$

Notice that the matrix $\mathbf{J} = \mathbf{J}(\Phi \mathbf{S}_Z \Phi, \mathbf{F})$ and the Hessian $\mathbf{H}Q(\underline{\vartheta}_0)$ are the same except that the later one depends only on Φ , i.e. $\mathbf{H}Q(\underline{\vartheta}_0) = \mathbf{J}(\Phi, \mathbf{F})$.

Put

$$\mathbf{J}_T = \text{Var} \sqrt{T} \text{Vec} \frac{\partial Q_T(\underline{\vartheta}_0)}{\partial \underline{\vartheta}^\top},$$

and suppose the following condition holds.

Condition 13 (R) *The limit variance matrix $\mathbf{J}(\Phi \mathbf{S}_Z \Phi, \mathbf{F})$ of the $\text{Var} \sqrt{T} \text{Vec} \frac{\partial Q_T(\underline{\vartheta}_0)}{\partial \underline{\vartheta}^\top}$, be positive definite, for all admissible spectral density \mathbf{S}_Z and SDFR \mathbf{F} , moreover suppose that $\mathbf{J}(\Phi, \mathbf{F}) > \mathbf{0}$.*

Theorem 14 *Under assumptions I-R, I-R-H and R*

$$\sqrt{T} \mathbf{J}_T^{-1/2} \mathbf{H}Q_T \hat{\underline{\vartheta}} - \hat{\underline{\vartheta}}_T - \underline{\vartheta}_0 \xrightarrow{\mathcal{D}} N(0, \mathbf{U}_p),$$

where $\hat{\underline{\vartheta}}$ is closer to $\underline{\vartheta}_0$ than $\hat{\underline{\vartheta}}_T$. In other words

$$\lim_{T \rightarrow \infty} \text{Var} \left[\sqrt{T} \hat{\underline{\vartheta}}_T - \underline{\vartheta}_0 \right] = \mathbf{J}^{-1}(\Phi, \mathbf{F}) \mathbf{J}(\Phi \mathbf{S}_Z \Phi, \mathbf{F}) \mathbf{J}^{-1}(\Phi, \mathbf{F}) \Big|_{\underline{\vartheta}=\underline{\vartheta}_0}. \quad (9)$$

The optimal choice of $\Phi(\omega)$ is $\mathbf{S}_Z^{-1}(\omega)$ assuming $\mathbf{S}_Z(\omega) > \mathbf{0}$. The choice $\mathbf{S}_Z^{-1}(\omega)$ is appropriate since the "residual" series $\underline{d}_{Z,T}(\omega_k)$ are asymptotically independent Gaussian random vectors with variance matrix $T\mathbf{S}_Z(\omega_k)$. The covariance matrix in this case $\Phi = \mathbf{S}_Z^{-1}$ follows from (9)

$$\lim_{T \rightarrow \infty} \text{Var} \left[\sqrt{T} \hat{\underline{\vartheta}}_T - \underline{\vartheta}_0 \right] = \mathbf{J}^{-1} \mathbf{S}_Z^{-1}, \mathbf{F}, \quad (10)$$

where

$$\mathbf{J}^{-1} \mathbf{S}_Z^{-1}, \mathbf{F} = \left[\int_{-1/2}^{1/2} \left[\mathbf{U}_p \otimes \text{Vec} \mathbf{S}_Z^{-1}(\omega) \text{ }^\top \text{ }^\top \right] d \frac{\partial^2 \mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top} \Big|_{\underline{\vartheta}_1=\underline{\vartheta}_2=\underline{\vartheta}_0} \right]^{-1}.$$

4 Scaling

To assess the generality of scaling consider the linear case

$$\underline{Y}_t = \mathbf{B} \underline{X}_t + \underline{Z}_t,$$

In this case $\underline{\vartheta} = \text{Vec} \mathbf{B}$, so the regressor \underline{X}_t depends on the parameter $\underline{\vartheta}$ linearly, (\underline{X}_t depends only on t) Here \mathbf{B} is $d \times p$ and \underline{X}_t is $p \times 1$. If $\|\underline{X}_{k,t}\|_T \simeq D_k(T)$ which tends to infinity by the Grenander's Condition G1, then the

$$\hat{\mathbf{C}}_{\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2) = \frac{1}{T} \sum_{t=1}^{T-h} \underline{X}_{t+h}(\underline{\vartheta}_1) \underline{X}_t^\top(\underline{\vartheta}_2),$$

might not converge unless each $D_k(T) \simeq \sqrt{T}$. This condition is not satisfied when we consider the polynomial regression models. Grenander's condition can be interpreted in the following way. Define

the diagonal matrix $\mathbf{D}_T = \text{diag}(D_1, D_2, \dots, D_p)$, where $D_k = D_k(T) \simeq \|X_{k,t}\|_T$. Now, consider the linear regression problem

$$\underline{Y}_t = \tilde{\mathbf{B}}\underline{V}_t + Z_t,$$

where $\underline{V}_t = \sqrt{T}\mathbf{D}_T^{-1}\underline{X}_t$. Now consider the revised version of the model and observe the connection

$$\tilde{\mathbf{B}}\underline{V}_t = \frac{1}{\sqrt{T}}\mathbf{B}\mathbf{D}_T \sqrt{T}\mathbf{D}_T^{-1}\underline{X}_t,$$

between the original and the scaled equation. Therefore the asymptotic variance of the estimate of the parameter matrix \mathbf{B} is now related by

$$\lim_{T \rightarrow \infty} \text{Var} \sqrt{T} \hat{\mathbf{B}} - \tilde{\mathbf{B}}_0 = \lim_{T \rightarrow \infty} \text{Var} \left[\hat{\mathbf{B}} - \mathbf{B}_0 \mathbf{D}_T \right].$$

We call this type of transformation as 'primary' scaling and the result is the properly scaled regressor. Note here that the procedure of scaling opens the possibility of considering random regressors which are not necessarily weak stationary because of the second order moments do not exist, (see [22]) or it may be asymptotically stationary.

Definition 15 *The series \underline{X}_t is properly scaled if*

$$\|X_{k,t}\|_T^2 \simeq T,$$

as $T \rightarrow \infty$, for each $k = 1, 2, \dots, d$.

In general, let $D_k(T) \simeq \|X_{k,t}\|_T$, for each k and define $\mathbf{D}_T = \text{diag}(D_1, D_2, \dots, D_d)$, then it is easy to see that the new regressor vector $\sqrt{T}\mathbf{D}_T^{-1}\underline{X}_t$ is properly scaled. The primary scaling of the nonlinear regressors $\underline{X}_t(\vartheta)$ is possible if $D_k(T)$ does not depend on the unknown parameter ϑ . Even if the regressors $\underline{X}_t(\vartheta)$ are properly scaled, there may be some problems arising when we take the limit of the derivatives because there is no guarantee for the convergence of the sums involved. Therefore we have to introduce some further scaling to the properly scaled regressors $\underline{X}_t(\vartheta)$.

Consider first, the diagonal matrix $\mathbf{D}_T = \text{diag}(D_{X,k}(T), k = 1, 2, \dots, d)$ and apply the scaling which results in $\sqrt{T}\mathbf{D}_T^{-1}\underline{X}_t(\vartheta)$. Another type of scaling can be obtained by the process of differentiation.

We define the scaled partial derivative $\partial_{s,T}(\vartheta)$ of the diagonal matrix $\mathbf{D}_{1,T} = \text{diag}(D_k^{(1)}(T), k = 1, 2, \dots, p)$ by $\partial \mathbf{D}_{1,T}^{-1}\vartheta$, resulting

$$\frac{\partial}{\partial_{s,T}\vartheta^\top} \mathbf{D}_T^{-1}\underline{X}_t(\vartheta) = \frac{\partial}{\partial\vartheta^\top} \mathbf{D}_T^{-1}\underline{X}_t(\vartheta) \mathbf{D}_{1,T}^{-1}, \quad (11)$$

which gives

$$\frac{\partial}{\partial_{s,T}\vartheta^\top} \mathbf{D}_T^{-1}\underline{X}_t(\vartheta) = \mathbf{D}_T^{-1} \frac{\partial}{\partial\vartheta^\top} \underline{X}_t(\vartheta) \mathbf{D}_{1,T}^{-1}.$$

Notice, the entries of the scaled partial derivatives are $\left[D_{X,j}(T) D_k^{(1)}(T) \right]^{-1} \partial X_{j,t}(\vartheta) / \partial \vartheta_k$. The second scaled derivatives of $\hat{\mathbf{C}}_{\underline{X},T}(h, \vartheta_1, \vartheta_2)$ are of interest

$$\begin{aligned} \frac{\partial_{s,T}^2 \hat{\mathbf{C}}_{\mathbf{D}_T \underline{X}, T}(h, \vartheta_1, \vartheta_2)}{\partial_{s,T} \vartheta_2^\top \partial_{s,T} \vartheta_1^\top} &= \mathbf{D}_{1,T}^{-1} \otimes \mathbf{U}_{d^2} \frac{\partial^2 \hat{\mathbf{C}}_{\sqrt{T}\mathbf{D}_T^{-1}\underline{X}, T}(h, \vartheta_1, \vartheta_2)}{\partial \vartheta_2^\top \partial \vartheta_1^\top} \mathbf{D}_{1,T}^{-1} \\ &= T \mathbf{D}_{1,T}^{-1} \otimes \mathbf{D}_T^{-1} \otimes \mathbf{D}_T^{-1} \frac{\partial^2 \hat{\mathbf{C}}_{\underline{X}, T}(h, \vartheta_1, \vartheta_2)}{\partial \vartheta_2^\top \partial \vartheta_1^\top} \mathbf{D}_{1,T}^{-1}, \end{aligned}$$

see Section 7.5 in Appendix for the proof. Note that $1/T$ in the expression of $\hat{\mathbf{C}}_{\underline{X},T}$ is canceled and the role of scaling has been absorbed completely by the scaling matrices.

Condition 16 (H²) *All the second partial derivatives of the regressor vector $\underline{X}_t(\vartheta)$ exist. There exist diagonal matrices \mathbf{D}_T and $\mathbf{D}_{1,T}$ such that, uniformly in ϑ , the scaled derivative $\frac{\partial_{s,T}^2 \hat{\mathbf{C}}_{\sqrt{T}\mathbf{D}_T^{-1}\underline{X}, T}(h, \vartheta_1, \vartheta_2)}{\partial_{s,T} \vartheta_2^\top \partial_{s,T} \vartheta_1^\top}$ converges to some limit, which we denote it by $\frac{\partial_s^2 \mathbf{C}_X(h, \vartheta_1, \vartheta_2)}{\partial_s \vartheta_1 \partial_s \vartheta_2}$.*

The above condition H implies that

$$\frac{\partial_s^2 \mathbf{C}_X(h, \vartheta_1, \vartheta_2)}{\partial_s \vartheta_2^\top \partial_s \vartheta_1^\top} \stackrel{\bullet}{=} \lim_{T \rightarrow \infty} \frac{\partial_{s,T}^2 \widehat{\mathbf{C}}_{\mathbf{D}_T X, T}(h, \vartheta_1, \vartheta_2)}{\partial_{s,T} \vartheta_2^\top \partial_{s,T} \vartheta_1^\top}.$$

The diagonal matrices \mathbf{D}_T and $\mathbf{D}_{1,T}$ can be chosen directly. In cases when the entries of the partial derivatives separate, i.e. when

$$\|\partial X_{j,t}(\vartheta) / \partial \vartheta_k\|_T \simeq B_{X,j}(T) B_k^{(1)}(T),$$

then $\mathbf{D}_T = \text{diag}(B_{X,j}(T), k = 1, 2, \dots, d)$, and $\mathbf{D}_{1,T} = \text{diag}(B_k^{(1)}(T), k = 1, 2, \dots, p)$, say. Note here that the matrix \mathbf{D}_T contains the factors of primary scaling. There may be regressors $\underline{X}_t(\vartheta)$, with derivatives which may require other forms of scaling may be required. and the above scaling does not apply.

The above notation is used for the regression spectrum as well.

Condition 17 (I-R-H') *The scaled derivative $\frac{\partial_{s,T}^2 \mathbf{F}_T(\omega, \vartheta_1, \vartheta_2)}{\partial_{s,T} \vartheta_1^\top \partial_{s,T} \vartheta_2^\top}$, of matrix function \mathbf{F}_T converges weakly to some function which we denote by $\frac{\partial_s^2 \mathbf{F}(h, \vartheta_1, \vartheta_2)}{\partial_s \vartheta_2^\top \partial_s \vartheta_1^\top}$.*

Define

$$\mathbf{J}_T(\mathbf{D}_T, \Psi, \mathbf{F}) = \int_{-1/2}^{1/2} (\mathbf{U}_p \otimes [\text{Vec}(\mathbf{D}_T \Psi^\top(\omega_k) \mathbf{D}_T)]^\top) d \left(\frac{\partial_s^2 \mathbf{F}(\omega, \vartheta_1, \vartheta_2)}{\partial_s \vartheta_2^\top \partial_s \vartheta_1^\top} \Big|_{\vartheta_1 = \vartheta_2 = \vartheta_0} \right),$$

Theorem 18 *Under the conditions I-R and I-R-H' we have*

$$\sqrt{T} \mathbf{J}_T^{-1/2} \mathbf{H}_{s,T} \mathbf{Q}_T(\vartheta_0) \mathbf{D}_{1,T} \widehat{\vartheta}_T - \vartheta_0 \xrightarrow{\mathcal{D}} N(0, \mathbf{U}_p).$$

In other words the variance of $\widehat{\vartheta}_T - \vartheta_0$ can be approximated by

$$\mathbf{D}_{1,T}^{-1} \mathbf{J}_T^{-1}(\mathbf{D}_T, \Phi, \mathbf{F}) \mathbf{J}_T(\mathbf{D}_T, \Phi \mathbf{S}_Z \Phi, \mathbf{F}) \mathbf{J}_T^{-1}(\mathbf{D}_T, \Phi, \mathbf{F}) \mathbf{D}_{1,T}^{-1} \Big|_{\vartheta = \vartheta_0}.$$

Moreover if $\Phi = \mathbf{S}_Z^{-1}$, one obtains the asymptotic variance in the reduced and neat form

$$\mathbf{D}_{1,T} \mathbf{J}_T^{-1} \mathbf{D}_T \mathbf{S}_Z^{-1} \mathbf{F} \mathbf{D}_{1,T}.$$

We shall see in the next Section that the linear regressors are scaled directly.

Remark 19 *The spectrum \mathbf{S}_Z in general is not known, which then it leads to a semiparametric problem, and therefore one uses a recursive form for the estimation of the parameters. In such situations one notices that the additional term to the function R in the objective function is the Whittle likelihood up to a constant. As long as we restrict to rational spectral density functions, the methods of Hannan ([14]) apply and both the estimator of the unknown parameter ϑ and the estimator for the parameters of the spectrum are consistent. The details will be published in a later paper.*

5 Some particular cases of special interest

We now consider some particular cases of the regression function which are of interest.

5.1 Multiple Linear Regression with Stationary Errors

Consider the linear case

$$Y_t = \mathbf{B} \underline{X}_t + Z_t,$$

and in this case $\vartheta = \text{Vec} \mathbf{B}$, so the regressors depend on the parameter ϑ linearly, (\underline{X}_t depends only on t but ϑ). Here \mathbf{B} is $d \times p$ and \underline{X}_t is $p \times 1$. The 'primary' scaling, if it is necessary, is given by the

diagonal matrix $\sqrt{T}\mathbf{D}_T^{-1}$ with $\mathbf{D}_T = \text{diag}(D_1, D_2, \dots, D_p)$, where $D_k(T) \simeq \|X_{k,t}\|_T$. Since it is easy to see that

$$\widehat{\mathbf{C}}_{\underline{X},T}(h, \vartheta_1, \vartheta_2) = \frac{1}{T} \sum_{t=1}^{T-h} \underline{X}_{t+h}(\vartheta_1) \underline{X}_t^\top(\vartheta_2),$$

where $\underline{X}_t(\vartheta) = \mathbf{B}\underline{X}_t$, converges for all possible values of \mathbf{B} if and only if

$$\widehat{\mathbf{C}}_{\underline{X},T}(h) = \frac{1}{T} \sum_{t=1}^{T-h} \underline{X}_{t+h} \underline{X}_t^\top,$$

converges. Assume that \underline{X}_t is properly scaled (otherwise scale it, $\sqrt{T}\mathbf{D}_T^{-1}\underline{X}_t$). Observe that $\frac{\partial \underline{X}_t(\vartheta)}{\partial \vartheta_k} = \frac{\partial \mathbf{B}\underline{X}_t}{\partial \vartheta_k} = [0, \dots, 0, \underline{X}_{j_k, t}, 0, \dots, 0]^\top$, therefore the 'secondary' scaling is $\mathbf{D}_{1,T} = \mathbf{U}_{dp}$.

For the above model the discrete Fourier transform reduces to

$$\underline{d}_{\underline{Y},T}(\omega) = \mathbf{B}\underline{d}_{\underline{X},T}(\omega) + \underline{d}_{\underline{Z},T}(\omega).$$

and the periodograms for each term of the above are given by

$$\begin{aligned} \mathbf{I}_{\underline{X},T}(\omega_k, \vartheta) &= \mathbf{B}\mathbf{I}_{\underline{X},T}(\omega_k)\mathbf{B}^\top, \\ \mathbf{I}_{\underline{Y},\underline{X},T}(\omega_k, \vartheta) &= \mathbf{I}_{\underline{Y},\underline{X},T}(\omega_k)\mathbf{B}^\top, \\ \mathbf{I}_{\underline{X},\underline{Y},T}(\omega_k, \vartheta) &= \mathbf{B}\mathbf{I}_{\underline{X},\underline{Y},T}(\omega_k), \end{aligned}$$

The normal equations are obtained by solving

$$\frac{\partial Q_T(\mathbf{B})}{\partial \mathbf{B}} = \mathbf{0},$$

and the estimates can be obtained. The expression for the variance covariance matrix of the estimate vector is (in the vectorized form)

$$\text{Vec } \widehat{\mathbf{B}} = \left(\sum_{k=-T_1}^{T_1} \mathbf{I}_{\underline{X},T}^\top(\omega_k) \otimes \Phi(\omega_k) \right)^{-1} \text{Vec } \sum_{k=-T_1}^{T_1} \Phi(\omega_k) \mathbf{I}_{\underline{Y},\underline{X},T}(\omega_k)$$

If the inverse does not exist, we can use the Generalized inverse. This estimate is *linear and unbiased* since

$$\mathbf{E} \sum_{k=-T_1}^{T_1} \Phi(\omega_k) \mathbf{I}_{\underline{Y},\underline{X},T}(\omega_k) = \Phi(\omega_k) \mathbf{B}_0 \mathbf{I}_{\underline{X},T}(\omega_k).$$

The Hessian of $Q_T(\mathbf{B})$ is

$$\begin{aligned} \text{H}Q_T(\mathbf{B}) &= \frac{1}{T} \sum_{k=-T_1}^{T_1} \mathbf{I}_{\underline{X},T}^\top(\omega_k) \otimes \Phi(\omega_k) + \mathbf{I}_{\underline{X},T}(\omega_k) \otimes \Phi^\top(\omega_k) \\ &= \int_{-1/2}^{1/2} d\mathbf{F}^\top(\omega) \otimes \Phi(\omega) + o(1). \end{aligned}$$

The variance matrix of the estimate $\widehat{\mathbf{B}}$

$$\lim_{T \rightarrow \infty} \frac{\text{Var}}{\sqrt{T}} \left(\frac{1}{\sqrt{T}} \text{Vec } \sum_{k=-T_1}^{T_1} \Phi(\omega_k) \mathbf{I}_{\underline{Y},\underline{X},T}(\omega_k) \right) = 4 \text{Vec } \int_{-1/2}^{1/2} d\mathbf{F}^\top(\omega) \otimes [\Phi(\omega) \mathbf{S}_{\underline{Z}}(\omega) \Phi(\omega)],$$

see (29). In particular we have the expressions in both situations when the estimates are the ordinary least squares estimates (when the errors are independent) and also in the case of weighted least squares (LS) estimates. In the first case $\Phi(\omega) = \mathbf{U}_d$ and in the later case $\Phi(\omega) = \mathbf{S}_{\underline{Z}}^{-1}(\omega)$ which leads to the best linear unbiased estimates, BLUE. Grenander(1954) shows that under some assumptions asymptotically the LS and BLUE are equivalent. When such conditions satisfy we have $\text{Var} \left[\sqrt{T} \text{Vec } \widehat{\mathbf{B}} \right] = \left[\int_{-1/2}^{1/2} d\mathbf{F}^\top(\omega) \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega) \right]^{-1}$. This limit does not depend on \mathbf{B}_0 . This result can also be obtained from the general formula (10) for the variance.

Remark 20 The estimation of the transpose of the matrix \mathbf{B} is used more often in time domain (see [13]). Then the variance of $\text{Vec}(\widehat{\mathbf{B}}^\top)$ follows from (12) easily

$$\begin{aligned}\text{Var}\left[\sqrt{T}\text{Vec}\widehat{\mathbf{B}}^\top\right] &= \text{Var}\left[\mathbf{K}_{p,d}\text{Vec}\widehat{\mathbf{B}}\right] \\ &= \mathbf{K}_{p,d}\text{Var}\text{Vec}\widehat{\mathbf{B}}\mathbf{K}_{d,p},\end{aligned}$$

hence

$$\lim_{T \rightarrow \infty} \text{Var}\left[\sqrt{T}\text{Vec}\widehat{\mathbf{B}}^\top\right] = \left[\int_{-1/2}^{1/2} \mathbf{S}_{\underline{Z}}^{-1}(\omega) \otimes d\mathbf{F}^\top(\omega)\right]^{-1}.$$

See Section 7.6 of the Appendix. In practice, we are interested in the asymptotic variance of $\widehat{\mathbf{B}}$ of the original unscaled regressors. Since we have an estimate of the matrix $\frac{1}{\sqrt{T}}\mathbf{B}\mathbf{D}_T$, writing

$$\mathbf{B}\underline{X}_t = \frac{1}{\sqrt{T}}\mathbf{B}\mathbf{D}_T\sqrt{T}\mathbf{D}_T^{-1}\underline{X}_t,$$

we get the asymptotic variance of $\text{Vec}\widehat{\mathbf{B}}$ to be

$$\left[\int_{-1/2}^{1/2} \mathbf{D}_T d\mathbf{F}^\top(\omega) \mathbf{D}_T \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega)\right]^{-1}, \quad (12)$$

(see also [13], Theorem 10, Chapter VII). For instance, if we consider the polynomial regressors of the form $\underline{X}_{j,t} = t^{j-1}$, $j = 1, \dots, p$, then the corresponding scaling factors we should use are $T_j(T) \simeq \sqrt{T^{2j-1}/(2j-1)}$ (this later one applies for any fractional $j > 1/2$ as well), and $\mathbf{D}_T = \text{diag}(T_1, T_2, \dots, T_d)$. In this case the SDFR \mathbf{F} is concentrated at zero with values $d\mathbf{F}_{j,k}(0) = \sqrt{(2k-1)(2j-1)/(k+j-1)}$, so the asymptotic variance of $\text{Vec}\widehat{\mathbf{B}}$ is

$$\mathbf{D}_T^{-1}d\mathbf{F}^{-1}(0)\mathbf{D}_T^{-1} \otimes \mathbf{S}_{\underline{Z}}(0),$$

(see [9] p. 247, for scalar valued case.)

5.2 A mixed model involving parameters both linearly and nonlinearly

A very realistic model is the following

$$\underline{Y}_t = \underline{X}_t(\vartheta_0) + \underline{Z}_t,$$

where the regressor is of the form

$$\begin{aligned}\underline{X}_t(\vartheta) &= \mathbf{B}_1\underline{X}_{1,t} + \mathbf{B}_2\underline{X}_{2,t}(\lambda) \\ &= [\mathbf{B}_1, \mathbf{B}_2] \begin{array}{c} \underline{X}_{1,t} \\ \underline{X}_{2,t}(\lambda) \end{array} \\ &= \mathbf{B}\underline{X}_{3,t}(\lambda),\end{aligned} \quad (13)$$

Here the unknown parameter $\vartheta = \text{Vec}(\text{Vec}\mathbf{B}_1, \text{Vec}\mathbf{B}_2, \lambda)$, where \mathbf{B}_1 is $d \times p$, \mathbf{B}_2 is $d \times q$, λ is $r \times 1$, $\underline{X}_{1,t}$ is of dimension p , $\underline{X}_{2,t}(\lambda)$ is of dimension q , $\mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2]$ and $\underline{X}_{3,t}(\lambda) = \begin{array}{c} \underline{X}_{1,t} \\ \underline{X}_{2,t}(\lambda) \end{array}$. First we consider the problem of estimation, by minimizing the objective function

$$\begin{aligned}Q_T(\mathbf{B}, \lambda) &= \frac{1}{T} \sum_{k=-T_1}^{T_1} \text{Tr}(\mathbf{I}_{\underline{Y},T}(\omega_k)\Phi(\omega_k)) + \text{Tr}\mathbf{B}\mathbf{I}_{\underline{X}_3,T}(\omega_k, \lambda)\mathbf{B}^\top\Phi(\omega_k) \\ &\quad - \text{Tr}\mathbf{I}_{\underline{Y},\underline{X}_3,T}(\omega_k, \lambda)\mathbf{B}^\top\Phi(\omega_k) - \text{Tr}\mathbf{B}\mathbf{I}_{\underline{X}_3,\underline{Y},T}(\omega_k, \lambda)\Phi(\omega_k).\end{aligned} \quad (14)$$

Now, differentiate with respect to $\mathbf{B}_1, \mathbf{B}_2$ and λ . We can apply the linear methods for \mathbf{B} in terms of $\underline{X}_{3,t}(\lambda)$. Suppose that the $\widehat{\mathbf{B}} = [\widehat{\mathbf{B}}_1, \widehat{\mathbf{B}}_2]$ and $\widehat{\lambda}$, satisfies the system of equations

$$\begin{aligned}\frac{\partial Q_T(\mathbf{B}, \lambda)}{\partial \mathbf{B}} &= \mathbf{0}, \\ \text{Vec}\frac{\partial Q_T(\lambda)}{\partial \lambda^\top} &= \mathbf{0}.\end{aligned}$$

The estimation of the linear parameters \mathbf{B}_1 and \mathbf{B}_2 can be carried out as in linear regression when the parameter $\underline{\lambda}$ is fixed. It leads to a recursive procedure. When we first set $\underline{\lambda} = \hat{\underline{\lambda}}$ (chosen), the normal equations result in

$$\text{Vec } \hat{\mathbf{B}} = \left(\sum_{k=-T_1}^{T_1} \mathbf{I}_{\underline{X}_{3,T}}^T \omega_k, \hat{\underline{\lambda}} \otimes \Phi(\omega_k) \right)^{-1} \text{Vec } \sum_{k=-T_1}^{T_1} \Phi(\omega_k) \mathbf{I}_{\underline{Y}, \underline{X}_{3,T}} \omega_k, \hat{\underline{\lambda}},$$

Now, to obtain the estimates for $\underline{\lambda}$, we keep $\mathbf{B} = \hat{\mathbf{B}}$ fixed and then minimize (14), i.e. find the solution to the equation

$$\sum_{k=-T_1}^{T_1} \frac{\partial \mathbf{B} \mathbf{I}_{\underline{X}_{3,T}}(\omega_k, \underline{\lambda}) \mathbf{B}^T}{\partial \underline{\lambda}^T} - \frac{\partial \mathbf{I}_{\underline{Y}, \underline{X}_{3,T}}(\omega_k, \underline{\lambda}) \mathbf{B}^T}{\partial \underline{\lambda}^T} - \frac{\partial \mathbf{B} \mathbf{I}_{\underline{X}_{3,Y,T}}(\omega_k, \underline{\lambda})}{\partial \underline{\lambda}^T} \Big|_{\underline{\lambda}=\hat{\underline{\lambda}}}^T \text{Vec } \Phi^T(\omega_k) = 0.$$

The primary scaling of $\underline{X}_{3,t}(\underline{\lambda}) = \begin{bmatrix} \underline{X}_{1,t} \\ \underline{X}_{2,t}(\underline{\lambda}) \end{bmatrix}$, is given by $\mathbf{D}_T = \text{diag}(\mathbf{D}_{X_1,T}, \mathbf{D}_{X_2,T})$ where $\mathbf{D}_{X_1,T} = \text{diag}(D_{X_1,k}(T), k=1, 2, \dots, p)$, and $\mathbf{D}_{X_2,T} = \text{diag}(D_{X_2,k}(T), k=1, 2, \dots, q)$. The secondary scaling of the regressors are $\mathbf{D}_{1,T} = \text{diag}(\mathbf{U}_{dp+dq}, \mathbf{D}_{3,T})$. Let us denote the limit of the variance covariance of the derivatives

$$\begin{bmatrix} \left[\frac{\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \underline{\lambda})}{\partial \text{Vec } \mathbf{B}_1} \right]^T \\ \left[\frac{\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \underline{\lambda})}{\partial \text{Vec } \mathbf{B}_2} \right]^T \\ \left[\frac{\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \underline{\lambda})}{\partial \underline{\lambda}^T} \right]^T \end{bmatrix}$$

by

$$\Sigma = 2 \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{1\lambda} \mathbf{D}_{3,T} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{2\lambda} \mathbf{D}_{3,T} \\ \mathbf{D}_{3,T} \Sigma_{\lambda 1} & \mathbf{D}_{3,T} \Sigma_{\lambda 2} & \mathbf{D}_{3,T} \Sigma_{\lambda \lambda} \mathbf{D}_{3,T} \end{bmatrix},$$

where the blocks of Σ contain already the scaling \mathbf{D}_T of the regressor, (this includes the case when $\Phi = \mathbf{S}_{\underline{Z}}^{-1}$). The part that is linear in parameters results in

$$\begin{aligned} \Sigma_{11} &= \int_{-1/2}^{1/2} \mathbf{D}_{X_1,T} d\mathbf{F}_{11}^T(\omega) \mathbf{D}_{X_1,T} \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega), \\ \Sigma_{12} &= \int_{-1/2}^{1/2} \mathbf{D}_{X_2,T} d\mathbf{F}_{12}^T(\omega, \underline{\lambda}_0) \mathbf{D}_{X_1,T} \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega) \\ \Sigma_{22} &= \int_{-1/2}^{1/2} \mathbf{D}_{X_2,T} d\mathbf{F}_{22}^T(\omega, \underline{\lambda}_0) \mathbf{D}_{X_2,T} \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega). \end{aligned}$$

In the mixed context we get

$$\begin{aligned} \Sigma_{1\lambda} &= \int_{-1/2}^{1/2} \mathbf{D}_{X_1,T} \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega) \mathbf{B}_{2,0} \mathbf{D}_{X_2,T} d \frac{\partial \mathbf{F}_{1,2}(\omega, \underline{\lambda}_0)}{\partial \underline{\lambda}^T}, \\ \Sigma_{2\lambda} &= \int_{-1/2}^{1/2} \mathbf{D}_{X_2,T} \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega) \mathbf{B}_{2,0} \mathbf{D}_{X_2,T} \frac{\partial \mathbf{F}_{2,2}(\omega, \underline{\lambda}_0)}{\partial \underline{\lambda}^T}. \end{aligned}$$

The following nonlinear block matrix $\Sigma_{\lambda\lambda}$ comes from the general result (10)

$$\Sigma_{\lambda\lambda} = 2 \int_{-1/2}^{1/2} \mathbf{U}_r \otimes \text{Vec } \mathbf{D}_{X_2,T} \mathbf{B}_{2,0}^T \mathbf{S}_{\underline{Z}}^{-1}(\omega) \mathbf{B}_{2,0} \mathbf{D}_{X_2,T} \Big|_{\underline{\lambda}_1=\underline{\lambda}_2=\underline{\lambda}_0}^T d \frac{\partial^2 \mathbf{F}_{2,2}(\omega, \underline{\lambda}_1, \underline{\lambda}_2)}{\partial \underline{\lambda}_2^T \partial \underline{\lambda}_1^T}.$$

Finally the variance matrix of the estimates $\text{Vec } \text{Vec } \hat{\mathbf{B}}_1, \text{Vec } \hat{\mathbf{B}}_2, \hat{\underline{\lambda}}$ is

$$\text{Var} \left[\text{Vec } \text{Vec } \hat{\mathbf{B}}_1, \text{Vec } \hat{\mathbf{B}}_2, \hat{\underline{\lambda}} \right] \simeq \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{1\lambda} \mathbf{D}_{3,T} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{2\lambda} \mathbf{D}_{3,T} \\ \mathbf{D}_{3,T} \Sigma_{\lambda 1} & \mathbf{D}_{3,T} \Sigma_{\lambda 2} & \mathbf{D}_{3,T} \Sigma_{\lambda \lambda} \mathbf{D}_{3,T} \end{bmatrix}^{-1}.$$

See Section [?] of the Appedix for details.

5.3 Linear trend with harmonic components, a worked example

Here we consider a special case of the mixed model considered above. This model is later used to illustrate our analysis of Chandler's Wobble. Let

$$\underline{Y}_t = \underline{X}_t(\vartheta_0) + \underline{Z}_t,$$

where

$$\underline{X}_t(\vartheta) = \mathbf{B} \begin{bmatrix} 1 \\ t \end{bmatrix} + \mathbf{A} \begin{bmatrix} \cos(2\pi t\lambda_1) \\ \sin(2\pi t\lambda_1) \\ \cos(2\pi t\lambda_2) \\ \sin(2\pi t\lambda_2) \end{bmatrix}.$$

The parameter is $\vartheta^\top = ([\text{Vec } \mathbf{B}_1]^\top, [\text{Vec } \mathbf{B}_2]^\top, [\lambda_1, \lambda_2])$, $|\lambda_i| \leq \pi$, $\lambda_1 \neq \lambda_2$, $\lambda_i \neq 0, \pm 1/2$. It is readily seen that the estimation of the coefficient matrix \mathbf{B} of the linear regression is given by

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

We see later that we can estimate B independently of A given by

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix}.$$

The primary scaling for $\underline{X}_{1,t}$ is $\mathbf{D}_{\underline{X}_1, T} = \text{diag } T^{1/2}, T^{3/2}/\sqrt{3}$, and for $\underline{X}_{2,t}(\lambda)$ is $\mathbf{D}_{\underline{X}_2, T} = T^{1/2}/\sqrt{2}\mathbf{U}_4$, since $\underline{X}_{1,t} = [1, t]^\top$ and $\underline{X}_{2,t}(\lambda) = [\cos(2\pi t\lambda_1), \sin(2\pi t\lambda_1), \cos(2\pi t\lambda_2), \sin(2\pi t\lambda_2)]$. The secondary scaling for the linear part $\underline{X}_{1,t}$ as we have already seen is \mathbf{U}_2 , and the secondary one for the nonlinear part is \mathbf{U}_4 and the scaled partial derivatives corresponding to λ is $\mathbf{D}_{3, T} = 2\pi T/\sqrt{3}\mathbf{U}_2$ since the primary scaling $\sqrt{T/2}$ has already been applied. Therefore the scaling matrix \mathbf{D}_T of the regressors $\underline{X}_{1,t}^\top, \underline{X}_{2,t}^\top(\lambda)^\top$ is $\mathbf{D}_T = \text{diag } \mathbf{D}_{\underline{X}_1, T}, \mathbf{D}_{\underline{X}_2, T}$, and $\mathbf{D}_{1, T} = \text{diag } (\mathbf{U}_{12}, \mathbf{D}_{3, T})$. The asymptotic variance is therefore given by

$$\mathbf{D}_{1, T}^{-1} \mathbf{J}^{-1} \mathbf{D}_T \mathbf{S}_{\underline{Z}}^{-1} \mathbf{D}_T \mathbf{F} \mathbf{D}_{1, T}^{-1}.$$

In general, the proper scaling for the term $X_{k,t} X_{m,t+h}$ in $\hat{\mathbf{C}}_{\underline{X}, T}$ is $1/ \|X_{k,t}\|_T \|X_{m,t}\|_T$. Here it can be changed into an equivalent function of T , instead of (2) we have

$$\hat{\mathbf{C}}_{\underline{X}, T}(h, \vartheta_1, \vartheta_2) = \mathbf{D}_T^{-1} \sum_{t=1}^{T-h} \underline{X}_{t+h}(\vartheta_1) \underline{X}_t^\top(\vartheta_2) \mathbf{D}_T^{-1}.$$

Let us partition the second derivative of SDFR according to the parameters, using the obvious notation

$$\frac{\partial^2 \mathbf{F}(\omega, \lambda_1, \lambda_2)}{\partial \lambda_2^\top \partial \lambda_1^\top} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} & \mathbf{F}_{1\lambda} \\ \mathbf{F}_{21} & \mathbf{F}_{22} & \mathbf{F}_{2\lambda} \\ \mathbf{F}_{\lambda 1} & \mathbf{F}_{\lambda 2} & \mathbf{F}_{\lambda\lambda} \end{bmatrix},$$

Assume $\lambda_1 \neq \lambda_2, \lambda_i \neq 0, \pm 1/2$.

1. The regression spectrum of the linear part

$$d\mathbf{F}_{11}(\omega) = \begin{bmatrix} 1 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1 \end{bmatrix} d\delta_{\omega \geq 0}$$

$\delta_{\omega \geq 0}$ denotes the Kronecker delta. Hence the block Σ_{11} reduces to

$$\Sigma_{11} = \mathbf{D}_{\underline{X}_1, T} d\mathbf{F}_{11}(0) \mathbf{D}_{\underline{X}_1, T} \otimes \mathbf{S}_{\underline{Z}}^{-1}(0).$$

2. It is seen here that the mixed model parameters has no effect, $\mathbf{F}_{12}(\omega, \lambda_0) = \mathbf{0}$, $\Sigma_{12} = \mathbf{0}$, and $\mathbf{F}_{1\lambda}(\omega, \lambda_0) = \mathbf{0}$, $\Sigma_{1\lambda} = \mathbf{0}$.
3. The $\mathbf{F}_{22}(\omega, \lambda_0)$ corresponds to the coefficient \mathbf{A} . Let

$$\mathbf{H}_{1h}(\lambda) = \begin{bmatrix} \cos(2\pi\lambda h) & -\sin(2\pi\lambda h) \\ \sin(2\pi\lambda h) & \cos(2\pi\lambda h) \end{bmatrix},$$

Notice

$$\begin{aligned}\widehat{\mathbf{C}}_{\underline{X}_2, T} \ h, \underline{\lambda}, \underline{\mu} &= \mathbf{D}_{\underline{X}_2, T}^{-1} \sum_{t=1}^{T-h} \underline{X}_{2, t+h}(\underline{\lambda}) \underline{X}_{2, t}^T \underline{\mu} \mathbf{D}_{\underline{X}_2, T}^{-1} \\ &\rightarrow \begin{bmatrix} \delta_{\lambda_1=\mu_1} \mathbf{H}_{1h}(\lambda_1) & \delta_{\lambda_1=\mu_2} \mathbf{H}_{1h}(\lambda_1) \\ \delta_{\lambda_2=\mu_1} \mathbf{H}_{1h}(\lambda_2) & \delta_{\lambda_2=\mu_2} \mathbf{H}_{1h}(\lambda_2) \end{bmatrix},\end{aligned}$$

where $\delta_{\lambda=\omega}$ denotes the Kronecker delta. Define the step functions

$$g_{c\lambda}(\omega) = \begin{cases} 0, & \omega < -\lambda, \\ 1/2, & -\lambda \leq \omega < \lambda, \\ 1, & \lambda \leq \omega, \end{cases} \quad g_{s\lambda}(\omega) = \begin{cases} 0, & \omega < -\lambda, \\ i/2, & -\lambda \leq \omega < \lambda, \\ 0, & \lambda \leq \omega, \end{cases}$$

and

$$\mathbf{G}_{1\lambda}(\omega) = \begin{bmatrix} g_{c\lambda}(\omega) & -g_{s\lambda}(\omega) \\ g_{s\lambda}(\omega) & g_{c\lambda}(\omega) \end{bmatrix}.$$

Now we have

$$\lim_{T \rightarrow \infty} \widehat{\mathbf{C}}_{\underline{X}_2, T} \ h, \underline{\lambda}, \underline{\mu} = \int_{-1/2}^{1/2} \exp(i2\pi\omega h) d\mathbf{F}_{22} \ \omega, \underline{\lambda}, \underline{\mu},$$

where

$$\mathbf{F}_{22} \ \omega, \underline{\lambda}, \underline{\mu} = \begin{bmatrix} \delta_{\lambda_1=\mu_1} \mathbf{G}_{1\lambda_1}(\omega) & \delta_{\lambda_1=\mu_2} \mathbf{G}_{1\lambda_1}(\omega) \\ \delta_{\lambda_2=\mu_1} \mathbf{G}_{1\lambda_2}(\omega) & \delta_{\lambda_2=\mu_2} \mathbf{G}_{1\lambda_2}(\omega) \end{bmatrix}.$$

The scaled version of the block is

$$\begin{aligned}\Sigma_{22} &= \int_{-1/2}^{1/2} (\mathbf{D}_{\underline{X}_2, T} d\mathbf{F}_{22}^T(\omega, \underline{\lambda}_0) \mathbf{D}_{\underline{X}_2, T}) \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega) \\ &= T/2 \begin{bmatrix} \operatorname{Re} \mathbf{S}_{\underline{Z}}^{-1}(\lambda_1) & \operatorname{Im} \mathbf{S}_{\underline{Z}}^{-1}(\lambda_1) & & \mathbf{0} \\ -\operatorname{Im} \mathbf{S}_{\underline{Z}}^{-1}(\lambda_1) & \operatorname{Re} \mathbf{S}_{\underline{Z}}^{-1}(\lambda_1) & & \\ & \mathbf{0} & \operatorname{Re} \mathbf{S}_{\underline{Z}}^{-1}(\lambda_2) & \operatorname{Im} \mathbf{S}_{\underline{Z}}^{-1}(\lambda_2) \\ & & -\operatorname{Im} \mathbf{S}_{\underline{Z}}^{-1}(\lambda_2) & \operatorname{Re} \mathbf{S}_{\underline{Z}}^{-1}(\lambda_2) \end{bmatrix}.\end{aligned}$$

4. For $\mathbf{F}_{2\lambda}(\omega, \underline{\lambda}_0)$, define the matrices

$$\begin{aligned}\mathbf{U}_2(1) &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \\ \mathbf{U}_2(2) &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},\end{aligned}$$

then we have

$$\frac{\sqrt{3}}{2\pi T} \frac{\partial \operatorname{Vec} \widehat{\mathbf{C}}_{\underline{X}_2, \underline{X}_2, T} \ h, \underline{\lambda}, \underline{\mu}}{\partial \underline{\mu}^T} \rightarrow \frac{\sqrt{3}}{2} \begin{bmatrix} \delta_{\lambda_1=\mu_1} \mathbf{U}_2(1) \otimes \begin{bmatrix} -\sin(2\pi\lambda_1 h) \\ \cos(2\pi\lambda_1 h) \end{bmatrix} \\ \delta_{\lambda_1=\mu_2} \mathbf{U}_2(1) \otimes \begin{bmatrix} -\cos(2\pi\lambda_1 h) \\ -\sin(2\pi\lambda_1 h) \end{bmatrix} \\ \delta_{\lambda_2=\mu_1} \mathbf{U}_2(2) \otimes \begin{bmatrix} -\sin(2\pi\lambda_2 h) \\ \cos(2\pi\lambda_2 h) \end{bmatrix} \\ \delta_{\lambda_2=\mu_2} \mathbf{U}_2(2) \otimes \begin{bmatrix} -\cos(2\pi\lambda_2 h) \\ -\sin(2\pi\lambda_2 h) \end{bmatrix} \end{bmatrix},$$

Notice, if $\underline{\lambda} = \underline{\mu}$ and $\lambda_1 \neq \lambda_2$ then this later matrix is written

$$\operatorname{Vec} [\mathbf{U}_2(1) \otimes \mathbf{H}_{2h}(\lambda_1)], \quad \operatorname{Vec} [\mathbf{U}_2(2) \otimes \mathbf{H}_{2h}(\lambda_2)],$$

where

$$\mathbf{H}_{2h}(\lambda) = \begin{bmatrix} -\sin(2\pi\lambda h) & -\cos(2\pi\lambda h) \\ \cos(2\pi\lambda h) & -\sin(2\pi\lambda h) \end{bmatrix}.$$

Notice that for the three frequencies $\underline{\lambda} = [\lambda_1, \lambda_2, \lambda_3]$ we would have

$$\operatorname{Vec} [\mathbf{U}_3(1) \otimes \mathbf{H}_{2h}(\lambda_1)], \quad \operatorname{Vec} [\mathbf{U}_3(2) \otimes \mathbf{H}_{2h}(\lambda_2)], \quad \operatorname{Vec} [\mathbf{U}_3(3) \otimes \mathbf{H}_{2h}(\lambda_3)],$$

where $\mathbf{U}_3(j)$ is a 3×3 matrix with zero elements except the j^{th} entry in the diagonal which is 1.

$$\mathbf{F}_{2\lambda}(\omega, \underline{\lambda}) = \frac{\sqrt{3}}{2} \text{Vec}[\mathbf{U}_2(1) \otimes \mathbf{G}_{2\lambda_1}(\omega)], \quad \text{Vec}[\mathbf{U}_2(2) \otimes \mathbf{G}_{2\lambda_2}(\omega)] ,$$

where

$$\mathbf{G}_{2\lambda}(\omega) = \begin{pmatrix} -g_{s\lambda}(\omega) & -g_{c\lambda}(\omega) \\ g_{c\lambda}(\omega) & -g_{s\lambda}(\omega) \end{pmatrix} .$$

Applying the general formula for $\Sigma_{2\lambda}$,

$$\begin{aligned} \Sigma_{2\lambda} &= \int_{-1/2}^{1/2} \mathbf{U}_4 \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega) \mathbf{A}_0 \mathbf{D}_{\underline{X}_2, T} \quad \mathbf{D}_{\underline{X}_2, T} \otimes \mathbf{U}_4 \quad d\mathbf{F}_{2\lambda}(\omega, \underline{\lambda}_0) \\ &= \frac{T}{2} \int_{-1/2}^{1/2} \mathbf{U}_4 \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega) \mathbf{A}_0 \quad d\mathbf{F}_{2\lambda}(\omega, \underline{\lambda}_0) . \end{aligned}$$

Put

$$\Gamma_2 = \begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix}$$

$$\Lambda_2(\omega) = \mathbf{U}_4 \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega) \mathbf{A}_0$$

$$\Sigma_{2\lambda} = \frac{\sqrt{3}T}{4} \Lambda_2(\lambda_1) \text{Vec}[\mathbf{U}_2(1) \otimes \Gamma_2], \quad \Lambda_2(\lambda_2) \text{Vec}[\mathbf{U}_2(2) \otimes \Gamma_2] .$$

5. Finally, $\mathbf{F}_{\lambda\lambda}(\omega, \underline{\lambda}_0)$ reduces to

$$\frac{3}{(2\pi)^2 T^2} \frac{\partial^2 \text{Vec} \hat{\mathbf{C}}_{\underline{X}_2, T} h, \lambda, \mu}{\partial \underline{\mu}^\top \partial \underline{\lambda}^\top} \rightarrow \begin{bmatrix} \delta_{\lambda_1=\mu_1} \mathbf{U}_2(1) \otimes \begin{matrix} \cos(2\pi\lambda_1 h) \\ \sin(2\pi\lambda_1 h) \end{matrix} \\ \delta_{\lambda_1=\mu_1} \mathbf{U}_2(1) \otimes \begin{matrix} -\sin(2\pi\lambda_1 h) \\ \cos(2\pi\lambda_1 h) \end{matrix} \\ \mathbf{0}_{16 \times 2} \\ \delta_{\lambda_2=\mu_2} \mathbf{U}_2(2) \otimes \begin{matrix} \cos(2\pi\lambda_2 h) \\ \sin(2\pi\lambda_2 h) \end{matrix} \\ \delta_{\lambda_2=\mu_2} \mathbf{U}_2(2) \otimes \begin{matrix} -\sin(2\pi\lambda_2 h) \\ \cos(2\pi\lambda_2 h) \end{matrix} \end{bmatrix} .$$

Define now the matrix $\mathbf{U}_{2,4}(1,1)$ of 2×4 with all element zero except the entry (1,1) which is one. Then we have

$$\text{Vec}[\mathbf{U}_{2,4}(1,1) \otimes \mathbf{H}_{3h}(\lambda_1)], \quad \text{Vec}[\mathbf{U}_{2,4}(2,4) \otimes \mathbf{H}_{3h}(\lambda_2)] ,$$

where

$$\mathbf{H}_{3h}(\lambda) = \begin{pmatrix} \cos(2\pi\lambda h) & -\sin(2\pi\lambda h) \\ \sin(\lambda 2\pi h) & \cos(2\pi\lambda h) \end{pmatrix} .$$

The SDFR is given by

$$\mathbf{F}_{\lambda\lambda}(\omega, \underline{\lambda}) = (2\pi)^2 \text{Vec}[\mathbf{U}_{2,4}(1,1) \otimes \mathbf{G}_{3\lambda_1}(\omega)], \quad \text{Vec}[\mathbf{U}_{2,4}(2,4) \otimes \mathbf{G}_{3\lambda_2}(\omega)] ,$$

where

$$\mathbf{G}_{3\lambda}(\omega) = \begin{pmatrix} g_{c\lambda}(\omega) & -g_{s\lambda}(\omega) \\ g_{s\lambda}(\omega) & g_{c\lambda}(\omega) \end{pmatrix} .$$

The corresponding variance matrix is

$$\begin{aligned} \Sigma_{\lambda\lambda} &= \frac{T}{2} \int_{-1/2}^{1/2} \mathbf{U}_2 \otimes \text{Vec} \left[\mathbf{A}^\top \mathbf{S}_{\underline{Z}}^{-1}(\omega) \mathbf{A} \right]^\top d\mathbf{F}_{\lambda\lambda}(\omega, \underline{\lambda}) \\ &= \frac{T}{2} \int_{-1/2}^{1/2} \begin{bmatrix} \text{Vec} \mathbf{A}^\top \left[\mathbf{S}_{\underline{Z}}^{-1}(\omega) \right]^\top \mathbf{A} \right]^\top & \mathbf{0}_{1 \times 16} \\ \mathbf{0}_{1 \times 16} & \left[\text{Vec} \mathbf{A}^\top \left[\mathbf{S}_{\underline{Z}}^{-1}(\omega) \right]^\top \mathbf{A} \right]^\top \end{bmatrix} d\mathbf{F}_{\lambda\lambda}(\omega, \underline{\lambda}) . \end{aligned}$$

For computational purposes, set

$$\Gamma_3 = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} ,$$

$$\Lambda(\omega) = \mathbf{U}_2 \otimes \left[\text{Vec} \mathbf{A}^\top \mathbf{S}_{\underline{Z}}^{-1}(\omega) \mathbf{A} \right]^\top ,$$

then the variance matrix has the form

$$\Sigma_{\lambda\lambda} = (2\pi)^2 \frac{T}{2} \text{Re} \Lambda(\lambda_1) \text{Vec} [\mathbf{U}_{2,4}(1, 1) \otimes \Gamma_3], \quad \Lambda(\lambda_2) \text{Vec} [\mathbf{U}_{2,4}(2, 4) \otimes \Gamma_3] .$$

It simplifies further

$$\Sigma_{\lambda\lambda} = \frac{T}{2} \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} ,$$

the entries are given in terms of the entries $A_{mn}(\omega) = \left[\mathbf{A}^\top \left[\mathbf{S}_{\underline{Z}}^{-1} \right]^\top \mathbf{A} \right]_{mn}$,

$$\begin{aligned} \sigma_{11} &= \text{Re} A_{11}(\lambda_1) + \text{Im} A_{21}(\lambda_1) - \text{Im} A_{12}(\lambda_1) + \text{Re} A_{22}(\lambda_1), \\ \sigma_{22} &= \text{Re} A_{33}(\lambda_2) + \text{Im} A_{43}(\lambda_2) - \text{Im} A_{34}(\lambda_2) + \text{Re} A_{44}(\lambda_2). \end{aligned}$$

Now we return to the asymptotic variance matrix of the parameters. Let us collect the blocks of the variance matrix

$$\mathbf{D}_{1,T} \begin{bmatrix} \Sigma_{11} & 0 & 0 \\ 0 & \Sigma_{22} & \Sigma_{2\lambda} \\ 0 & \Sigma_{\lambda 2} & \Sigma_{\lambda\lambda} \end{bmatrix} \mathbf{D}_{1,T},$$

where $\mathbf{D}_{1,T} = \text{diag}(\mathbf{U}_{12}, \mathbf{D}_{2,T})$. The variance matrix of the coefficient $\hat{\mathbf{A}}$

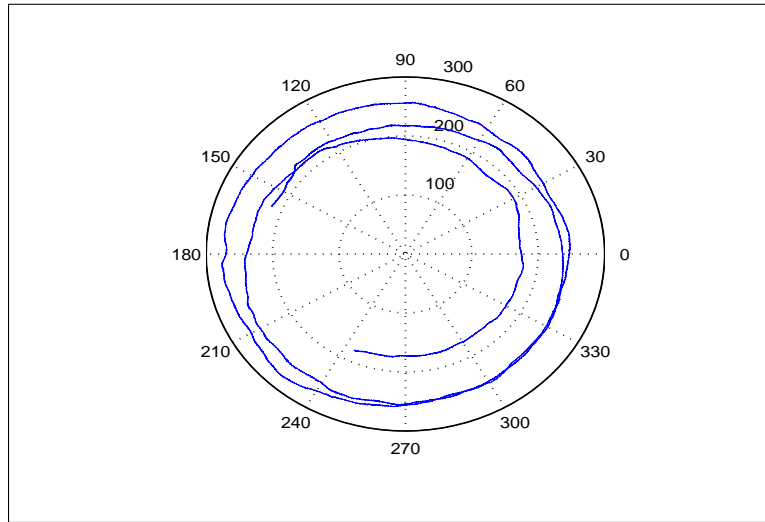
$$\frac{2}{T} \Sigma'_{22} - \Sigma'_{2\lambda} \Sigma'^{-1}_{\lambda\lambda} \Sigma'^{-1}_{\lambda 2} ,$$

and of $\hat{\lambda}$

$$\mathbf{D}_{3,T}^{-1} \Sigma_{\lambda\lambda} - \Sigma_{\lambda 2} \Sigma_{22}^{-1} \Sigma_{2\lambda}^{-1} \mathbf{D}_{3,T}^{-1} = \frac{6}{(2\pi)^2 T^3} \Sigma'_{\lambda\lambda} - \Sigma'_{\lambda 2} \Sigma'^{-1}_{22} \Sigma'^{-1}_{2\lambda} ,$$

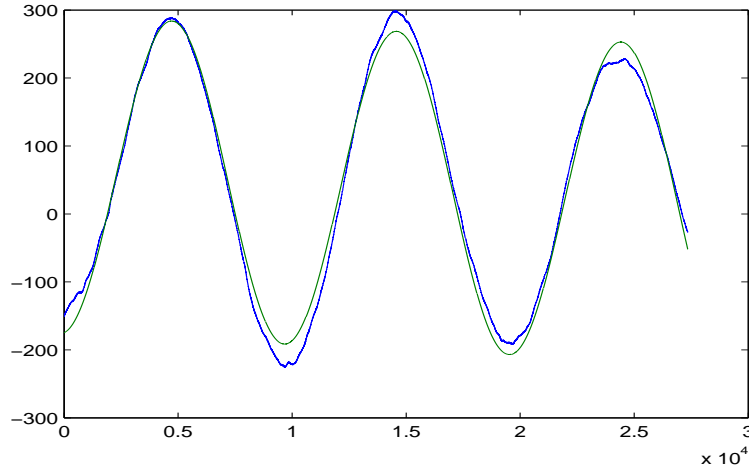
where $\mathbf{D}_{3,T} = 2\pi T / \sqrt{3} \mathbf{U}_{12}$, and Σ' denotes the covariance matrix without scaling. The speed of convergence of the variance matrix of the coefficient $\hat{\mathbf{A}}$ is $T/2$ and the frequency $\hat{\lambda}$ is $(2\pi)^2 T^3 / 6$.

6 Chandler wobble



Centralized wobbling motion in polar coordinates

Data Description: The Chandler wobble, named after its 1891 discoverer, Seth Carlo Chandler, Jr., is one of several wobbling motions exhibited by Earth as it rotates on its vertical axis, similar to a spinning top. The period of this wobble is approximately between 430-435 days. It has been estimated by several researchers using various models and methods (see, for instance Brillinger [4] and Arató-Kolmogorov-Sinay [1].) Some properties of the monthly data has been described in [19].



x -values and the fitted ones in MAS vs. hours

Since 1995, an integrated solution to the various GPS (Global Positioning System) series has been available. For our current analysis here, we use the hourly measurements between Modified Julian Day (MJD) 49719 (corresponding to JAN 1, '95) and MJD 50859 (corresponding to FEB 15, '98). The values of the data are given in milli-arcseconds or MAS, where 1 arcsec \sim 30m. The number of data points are $T = 27,361$.

Rotational variations of polar motion are due to the superposition of the influences of 6 partial tides. Different techniques suggest that these are real oscillations of polar motion. Rapid oscillations with periods of 12hours has already been considered, see IVS 2004 General Meeting Proceedings, [11].

The aim of our investigation is to provide statistical evidence for the presence of 12h oscillation, i.e. to show that the frequency $2\pi/12$ has statistically significant non-zero weight. Also another question of interest is whether there is any significant shift in the position of the center.

The model, [28], to be fitted is combination of a linear trend with harmonic components (usually termed as drift, offset and elliptical periodic motions parameters),

$$\underline{Y}_t = \mathbf{B}_{2 \times 2} \begin{matrix} 1 \\ t \end{matrix} + \mathbf{A}_{4 \times 2} \begin{bmatrix} \cos(2\pi t \lambda_1) \\ \sin(2\pi t \lambda_1) \\ \cos(2\pi t \lambda_2) \\ \sin(2\pi t \lambda_2) \end{bmatrix} + \underline{Z}_t,$$

where \underline{Y}_t is the measurement vector corresponding to the polar coordinates of the position. The matrices \mathbf{A} and \mathbf{B} together with the frequencies λ_i , ($|\lambda_i| \leq \pi$) are unknown and are to be estimated. This model is a special case of the nonlinear model we considered in this paper. We started the computations with the initial values $\lambda_1 = 2\pi/410/24$, and $\lambda_2 = 2\pi/12$, and the number of Fourier frequencies we used are 2^{13} . The estimates of the parameters are found to be

$$\hat{\mathbf{B}} = \begin{matrix} 41.6043 & 0.0003 \\ 323.4485 & -0.0007 \end{matrix},$$

$$\hat{\mathbf{A}} = \begin{matrix} -244.8065 & 16.5279 & 0.1248 & -0.0521 \\ 25.3854 & 256.5682 & 0.0166 & 0.1064 \end{matrix},$$

and

$$\hat{\lambda} = \begin{matrix} 0.0001 \\ 0.0833 \end{matrix}.$$

The estimated frequencies correspond to the periods 410.5626 days and 11.9999 hours which are close to the estimates obtained by geophysicists. Analyzing the residual series \underline{Z}_t we found some evidence of long range memory behavior in the residuals. In the following subsection we discuss the effect of this on our analysis.

6.1 Disturbance with Long Memory

Let \underline{Z}_t be a stationary time series with piecewise **continuous** spectral density $\mathbf{S}_Z(\omega)$

$$\mathbf{S}_Z(\omega) = \Lambda(\omega) \mathbf{S}_2^\#(\omega) \Lambda^*(\omega),$$

where the $\Lambda(\omega) = \text{diag } 1 - e^{i2\pi\omega^{-h_1}}, 1 - e^{i2\pi\omega^{-h_2}}, \dots, 1 - e^{i2\pi\omega^{-h_d}}$, $h_k \in [0, 1/2)$, $k = 1, 2, \dots, d$, and the matrix $\mathbf{S}_2^\#(\omega)$ is a positive continuous spectral density matrix (we have often in mind a stationary, physically realizable, vector-ARIMA time series). The Hurst exponents (h_1, h_2, \dots, h_d) are not necessarily different, denote them $\underline{h} = (h_1, h_2, \dots, h_d)$. Following Yajima(1991) we can define for each fixed h_k the regressors can be classified according to the discontinuity of their spectrum at zero. [33]. In the present case we consider that the frequencies are known, thus reducing the nonlinear regression to linear regression. We introduce the scaling according to the long memory. Let $\mathbf{D}_{L,T} = \text{diag } T^{h_k}, k = 1, 2, \dots, d$ be the diagonal matrix then

$$\Sigma_{11} = \mathbf{D}_{\underline{X},T} d\mathbf{F}_{11}(0) \mathbf{D}_{\underline{X},T} \otimes \mathbf{D}_{L,T} \mathbf{S}_{\underline{Z}}^{-1}(0) \mathbf{D}_{L,T}.$$

Robinson-Hidalgo ([26] Theorem 5) have shown that the weights $\mathbf{S}_{\underline{Z}}^{-1}$ are consistently estimated via recursion even if the data are long range dependent.

The technique of estimation we follow is based on multiple recursion. First we set $\mathbf{S}_{\underline{Z}}$, equal to a known matrix and estimate parameters $\underline{\vartheta}$. We keep the initial value of $\underline{\vartheta}$ fixed and estimate residual spectrum $\mathbf{S}_{\underline{Z}}$. Then fix the estimated linear parameter and find the estimate of the nonlinear parameter $\underline{\vartheta}$ through a weighted nonlinear least squares procedure. The procedure is repeated until the estimates converge.

Conclusions

- The estimation of the Hurst parameter is done by using the methods of Terdik and Igloi [20], which is based on the higher order cumulants (up to order 5). We found that both Hurst estimates are very close to 1/2, i.e. $h_1 = 0.4986$ and $h_2 = 0.4860$. Therefore we used the marginal Hurst parameters for our estimation of the entire model considered earlier.
- As expected (see the models fitted by geophysicists), there is no real information on the location parameter (the constant in the model) because the estimated variances of parameters and b_{21} so large. Some improvement can be obtained by using Dahlhaus's method, [8]. The diagonals of the variance matrix of $\text{Vec } \hat{\mathbf{B}}$ are $1.1733 * 10^6, 0.7725 * 10^6, 0.2097, 0.1381$
- The Standard Error (SE) for the parameters b_{12} , and b_{22} are 0.4579 and 0.3716, hence there is no evidence of the shifting of either coordinates, at least with larger than 95% confidence.
- Actually, we have only 'two observations' of the period ~ 410 days therefore it is not surprising that the SE of the parameters $a_{1:2,1:2}$ again, are large enough showing no information on the values. Specifically, the SEs are [397.1890, 481.8903, 436.7575, 442.9037].
- Now, the main interest is the SE of the parameters λ_2 and $a_{1:2,3:4}$. The SE of the estimates $\hat{a}_{1:2,3:4}$ is [0.0154, 0.0218, 0.0233, 0.0146] so we conclude that all of them are significantly different from zero (except $a_{1,3}$). There is some empirical evidence to justify fitting a new model with an additional frequency $\lambda_3 = 30$, $\lambda_4 = 2\lambda_2$. The estimation of frequencies such as λ_4 can create problems (similar problems do arise in biology -see eg. [6]).

Acknowledgement 21 We thank Professor József Závoti for providing us the High resolution Chandler Wobbler data. This research was partially supported by the Hungarian NSF, OTKA No. T047067

7 Appendix

7.1 Some Useful Matrix Relations

$$\text{Vec}(ab^T) = \underline{b} \otimes \underline{a}, \quad (15)$$

$$\underline{a}^T \otimes \underline{b} = \underline{ba}^T = \underline{b} \otimes \underline{a}^T. \quad (16)$$

see ([25]), p.28.

$$(\text{Vec } \mathbf{A})^T \text{Vec } \mathbf{B} = \text{Tr}(\mathbf{A}^T \mathbf{B}), \quad (17)$$

see ([25]), p30. The vectors $\underline{a}, \underline{b}$ and \underline{c} fulfil

$$(\underline{a} \otimes \underline{b}) \underline{c}^T = \underline{a} \otimes (\underline{bc}^T), \quad (18)$$

$$(\underline{a} \otimes \underline{b}) \underline{c}^\top = (\underline{a} \underline{c}^\top) \otimes \underline{b}, \quad (19)$$

The commutation matrix $\mathbf{K}_{m \cdot n}$ is defined by the relation

$$\mathbf{K}_{m \cdot n} \text{Vec } \mathbf{A} = \text{Vec } \mathbf{A}^\top, \quad (20)$$

for any matrix A with dimension $m \times n$. Next identity is

$$(\underline{a} \otimes \underline{b}) \underline{c}^\top = \mathbf{K}_{d \cdot d} (\underline{b} \otimes \underline{a}) \underline{c}^\top. \quad (21)$$

We have, see ([25]), p.47., if A is $m \times n$ and B is $p \times q$ then

$$\text{Vec}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{U}_n \otimes \mathbf{K}_{q \cdot m} \otimes \mathbf{U}_p) (\text{Vec } \mathbf{A} \otimes \text{Vec } \mathbf{B}) \quad (22)$$

$$\mathbf{K}_{p \cdot m} (\mathbf{A} \otimes \mathbf{B}) \mathbf{K}_{n \cdot q} = \mathbf{B} \otimes \mathbf{A}. \quad (23)$$

One can prove the following identity

$$\mathbf{A} \mathbf{B} \underline{b} \underline{a}^\top = [\mathbf{U} \otimes (\text{Vec } \mathbf{B})^\top] (\mathbf{K}_{n \cdot d} \otimes \mathbf{U}_m) \mathbf{K}_{d \cdot mn} [\underline{b} \underline{a}^\top \otimes \text{Vec } \mathbf{A}^\top], \quad (24)$$

where the only assumption for the matrices \mathbf{A} and \mathbf{B} , vectors \underline{b} and \underline{a} , is that the matrix product on the left side should be valid, \mathbf{U} is the identity matrix with appropriate order. We also have the following

$$\mathbf{K}_{dp \cdot d} (\mathbf{K}_{d \cdot p} \otimes \mathbf{U}_d) = \mathbf{U}_p \otimes \mathbf{K}_{d \cdot d}. \quad (25)$$

7.2 Jacobian and Hessian of SDFR, Proofs

Consider the Jacobian $\partial \underline{X}_t(\vartheta) / \partial \vartheta^\top$ of the regressor $\underline{X}_t(\vartheta)$ then the Jacobian of $\widehat{\mathbf{C}}_{\underline{X}, T}(h, \vartheta)$ is

$$\begin{aligned} \frac{\partial \widehat{\mathbf{C}}_{\underline{X}, T}(h, \vartheta)}{\partial \vartheta^\top} &= \frac{1}{T} \sum_{t=1}^{T-h} \underline{X}_t(\vartheta) \otimes \frac{\partial \underline{X}_{t+h}(\vartheta)}{\partial \vartheta^\top} + \frac{\partial \underline{X}_t(\vartheta)}{\partial \vartheta^\top} \otimes \underline{X}_{t+h}(\vartheta) \\ &= \frac{\partial \widehat{\mathbf{C}}_{\underline{X}}(h, \vartheta_1, \vartheta_2)}{\partial \vartheta_1^\top} + \frac{\partial \widehat{\mathbf{C}}_{\underline{X}}(h, \vartheta_1, \vartheta_2)}{\partial \vartheta_2^\top} \Big|_{\vartheta_1 = \vartheta_2 = \vartheta}, \end{aligned}$$

see (18, 19) and (16). Now take the limit of $\partial \widehat{\mathbf{C}}_{\underline{X}, T}(h, \vartheta) / \partial \vartheta^\top$ and define the Jacobian $\partial \mathbf{F}(\lambda, \vartheta) / \partial \vartheta^\top$ for SDFR \mathbf{F} by

$$\frac{\partial \mathbf{C}_{\underline{X}}(h, \vartheta)}{\partial \vartheta^\top} = \int_{-1/2}^{1/2} \exp(i2\pi\lambda h) d \frac{\partial \mathbf{F}(\lambda, \vartheta)}{\partial \vartheta^\top},$$

i.e. the $\partial \mathbf{C}_{\underline{X}}(h, \vartheta) / \partial \vartheta^\top$ is the inverse Fourier transform of $\partial \mathbf{F}(\lambda, \vartheta) / \partial \vartheta^\top$. If the limit of the Jacobian

$$\frac{\partial \mathbf{F}_T(\omega, \vartheta)}{\partial \vartheta^\top} = \int_0^\omega \frac{\partial \mathbf{I}_{\underline{X}, T}(\lambda, \vartheta)}{\partial \vartheta^\top} d\lambda,$$

exists and the differential operator and the limit are exchangeable then we have

$$\lim_{T \rightarrow \infty} \frac{\partial \mathbf{F}_T(\omega, \vartheta)}{\partial \vartheta^\top} = \frac{\partial}{\partial \vartheta^\top} \lim_{T \rightarrow \infty} \mathbf{F}_T(\omega, \vartheta) = \frac{\partial \mathbf{F}(\omega, \vartheta)}{\partial \vartheta^\top}.$$

This is not always the case of course. Notice

$$\begin{aligned} \mathbf{F}_T(\omega, \vartheta) &= \int_0^\omega \mathbf{I}_{\underline{X}, T}(\lambda, \vartheta) d\lambda \\ &= \int_0^\omega \frac{1}{T} \underline{d}_{\underline{X}, T}(\lambda, \vartheta_1) \underline{d}_{\underline{X}, T}^*(\lambda, \vartheta_2) d\lambda \Big|_{\vartheta_1 = \vartheta_2 = \vartheta} \\ &= \mathbf{F}_T(\omega, \vartheta_1, \vartheta_2) \Big|_{\vartheta_1 = \vartheta_2 = \vartheta}, \end{aligned}$$

therefore

$$\frac{\partial \mathbf{F}(\omega, \vartheta)}{\partial \vartheta^\top} = \frac{\partial \mathbf{F}(\omega, \vartheta_1, \vartheta_2)}{\partial \vartheta_1^\top} + \frac{\partial \mathbf{F}(\omega, \vartheta_1, \vartheta_2)}{\partial \vartheta_2^\top} \Big|_{\vartheta_1 = \vartheta_2 = \vartheta}.$$

This corresponds to the Jacobian

$$\frac{\partial \mathbf{C}_{\underline{X}}(h, \vartheta)}{\partial \vartheta^\top} = \frac{\partial \mathbf{C}_{\underline{X}}(h, \vartheta_1, \vartheta_2)}{\partial \vartheta_1^\top} + \frac{\partial \mathbf{C}_{\underline{X}}(h, \vartheta_1, \vartheta_2)}{\partial \vartheta_2^\top} \Big|_{\vartheta_1 = \vartheta_2 = \vartheta}.$$

The Hessian \mathbf{HF} of $\mathbf{F}(\lambda, \underline{\vartheta})$ is defined similarly, first the Hessian of $\widehat{\mathbf{C}}_{\underline{X}, T}(h, \underline{\vartheta})$

$$\begin{aligned} \mathbf{H}\widehat{\mathbf{C}}_{\underline{X}, T}(h, \underline{\vartheta}) &= \frac{1}{T} \sum_{t=1}^{T-h} \frac{\partial \text{Vec} \left[\underline{X}_t(\underline{\vartheta}) \otimes \frac{\partial \underline{X}_{t+h}(\underline{\vartheta})}{\partial \underline{\vartheta}^\top} + \frac{\partial \underline{X}_t(\underline{\vartheta})}{\partial \underline{\vartheta}^\top} \otimes \underline{X}_{t+h}(\underline{\vartheta}) \right]}{\partial \underline{\vartheta}^\top} \\ &= \frac{1}{T} \sum_{t=1}^{T-h} \frac{\partial (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \underline{X}_t(\underline{\vartheta}) \otimes \text{Vec} \frac{\partial \underline{X}_{t+h}(\underline{\vartheta})}{\partial \underline{\vartheta}^\top}}{\partial \underline{\vartheta}^\top} \\ &\quad + \frac{\partial \text{Vec} \frac{\partial \underline{X}_t(\underline{\vartheta})}{\partial \underline{\vartheta}^\top} \otimes \underline{X}_{t+h}(\underline{\vartheta})}{\partial \underline{\vartheta}^\top} \\ &= \frac{1}{T} \sum_{t=1}^{T-h} (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \frac{\partial \underline{X}_t(\underline{\vartheta})}{\partial \underline{\vartheta}^\top} \otimes \text{Vec} \frac{\partial \underline{X}_{t+h}(\underline{\vartheta})}{\partial \underline{\vartheta}^\top} + \underline{X}_t(\underline{\vartheta}) \otimes \mathbf{H}\underline{X}_{t+h}(\underline{\vartheta}) \\ &\quad + \text{Vec} \frac{\partial \underline{X}_t(\underline{\vartheta})}{\partial \underline{\vartheta}^\top} \otimes \frac{\partial \underline{X}_{t+h}(\underline{\vartheta})}{\partial \underline{\vartheta}^\top} + \mathbf{H}\underline{X}_t(\underline{\vartheta}) \otimes \underline{X}_{t+h}(\underline{\vartheta}) . \end{aligned}$$

Notice

$$\begin{aligned} (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \frac{\partial \underline{X}_t(\underline{\vartheta})}{\partial \underline{\vartheta}^\top} \otimes \text{Vec} \frac{\partial \underline{X}_{t+h}(\underline{\vartheta})}{\partial \underline{\vartheta}^\top} &= (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \mathbf{K}_{d \cdot dp} \text{Vec} \frac{\partial \underline{X}_{t+h}(\underline{\vartheta})}{\partial \underline{\vartheta}^\top} \otimes \frac{\partial \underline{X}_t(\underline{\vartheta})}{\partial \underline{\vartheta}^\top} , \\ (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \underline{X}_t(\underline{\vartheta}) \otimes \mathbf{H}\underline{X}_{t+h}(\underline{\vartheta}) &= (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \mathbf{K}_{d \cdot dp} \mathbf{H}\underline{X}_{t+h}(\underline{\vartheta}) \otimes \underline{X}_t(\underline{\vartheta}) , \end{aligned}$$

see (23). Let us denote the limit of $\mathbf{H}\widehat{\mathbf{C}}_{\underline{X}, T}(h, \underline{\vartheta})$ by $\mathbf{H}\mathbf{C}_{\underline{X}}(h, \underline{\vartheta})$, and its inverse Fourier transform by $\mathbf{HF}(\lambda, \underline{\vartheta})$, i.e.

$$\mathbf{H}\mathbf{C}_{\underline{X}}(h, \underline{\vartheta}) = \int_{-1/2}^{1/2} \exp(i2\pi\lambda h) d\mathbf{HF}(\lambda, \underline{\vartheta}) .$$

Similarly

$$\begin{aligned} \mathbf{H}\mathbf{C}_{\underline{X}}(h, \underline{\vartheta}) &= (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \mathbf{K}_{d \cdot dp} \mathbf{H}_{\underline{\vartheta}_1} \mathbf{C}_{\underline{X}}(-h, \underline{\vartheta}_2, \underline{\vartheta}_1) + \frac{\partial^2 \mathbf{C}_{\underline{X}}(-h, \underline{\vartheta}_2, \underline{\vartheta}_1)}{\partial \underline{\vartheta}_1^\top \partial \underline{\vartheta}_2^\top} \Big|_{\underline{\vartheta}_1 = \underline{\vartheta}_2 = \underline{\vartheta}} \\ &\quad + \mathbf{H}_{\underline{\vartheta}_2} \mathbf{C}_{\underline{X}}(h, \underline{\vartheta}_1, \underline{\vartheta}_2) + \frac{\partial^2 \mathbf{C}_{\underline{X}}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top} \Big|_{\underline{\vartheta}_1 = \underline{\vartheta}_2 = \underline{\vartheta}} \end{aligned}$$

where we used the short notation

$$\frac{\partial^2 \mathbf{C}_{\underline{X}}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top} = \frac{\partial}{\partial \underline{\vartheta}_1^\top} \text{Vec} \frac{\partial \text{Vec} \mathbf{C}_{\underline{X}}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_2^\top} ,$$

here the partial derivative of the right side can be carried out directly (the order of the variables $\underline{\vartheta}_1, \underline{\vartheta}_2$ is opposite to the order of the derivatives, $\partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top$ means that first by $\underline{\vartheta}_2$ then by $\underline{\vartheta}_1$, i.e. the operator acting by right hand side). Starting by $\underline{\vartheta}_1$ then followed by $\underline{\vartheta}_2$ is 'indirect', since

$$\frac{\partial^2 \mathbf{C}_{\underline{X}}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_1^\top \partial \underline{\vartheta}_2^\top} = (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \mathbf{K}_{d \cdot dp} \frac{\partial^2 \mathbf{C}_{\underline{X}}(-h, \underline{\vartheta}_2, \underline{\vartheta}_1)}{\partial \underline{\vartheta}_1^\top \partial \underline{\vartheta}_2^\top} ,$$

for the reason of this see (22). Note that

$$\mathbf{C}_{\underline{X}}(-h, \underline{\vartheta}_2, \underline{\vartheta}_1) = \mathbf{C}_{\underline{X}}^\top(h, \underline{\vartheta}_1, \underline{\vartheta}_2) .$$

Similarly the Hessian by $\underline{\vartheta}_2$ is direct and by $\underline{\vartheta}_1$ is indirect, i.e.

$$\mathbf{H}_{\underline{\vartheta}_1} \mathbf{C}_{\underline{X}}(h, \underline{\vartheta}_1, \underline{\vartheta}_2) = (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \mathbf{K}_{d \cdot dp} \mathbf{H}_{\underline{\vartheta}_1} \mathbf{C}_{\underline{X}}(-h, \underline{\vartheta}_2, \underline{\vartheta}_1) .$$

According to the above notations we write

$$\begin{aligned} \mathbf{HF}(\omega, \underline{\vartheta}) &= \mathbf{H}_{\underline{\vartheta}_1} \mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2) + \frac{\partial^2 \mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_1^\top \partial \underline{\vartheta}_2^\top} \\ &\quad + \mathbf{H}_{\underline{\vartheta}_2} \mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2) + \frac{\partial^2 \mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top} \Big|_{\underline{\vartheta}_1 = \underline{\vartheta}_2 = \underline{\vartheta}} , \end{aligned}$$

Again here, for instance

$$\frac{\partial^2 \mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_1^\top \partial \underline{\vartheta}_2^\top} = (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \mathbf{K}_{d \cdot dp} \frac{\partial^2 \mathbf{F}(-\omega, \underline{\vartheta}_2, \underline{\vartheta}_1)}{\partial \underline{\vartheta}_1^\top \partial \underline{\vartheta}_2^\top} .$$

7.3 Variance of the derivatives

The summands in

$$\text{Vec} \frac{\partial Q_T(\underline{\vartheta})}{\partial \underline{\vartheta}^\top} = \frac{1}{T} \sum_{k=-T_1}^{T_1} \frac{\partial \text{Vec} \mathbf{I}_{X,T}(\omega_k, \underline{\vartheta})}{\partial \underline{\vartheta}^\top} - \frac{\partial (\text{Vec} \mathbf{I}_{Y,X,T}(\omega_k, \underline{\vartheta}) + \text{Vec} \mathbf{I}_{X,Y,T}(\omega_k, \underline{\vartheta}))}{\partial \underline{\vartheta}^\top}^\top \times [\text{Vec} \mathbf{\Phi}^\top(\omega_k)],$$

are asymptotically independent therefore we are interested in the variance separately. Notice

$$\begin{aligned} \frac{\partial (\text{Vec} \mathbf{I}_{Z,X,T}(\omega_k, \underline{\vartheta}) + \text{Vec} \mathbf{I}_{X,Z,T}(\omega_k, \underline{\vartheta}))}{\partial \underline{\vartheta}^\top}^\top \text{Vec} \mathbf{\Phi}^\top(\omega_k) \\ = 2 \frac{\partial \text{Vec} \mathbf{I}_{Z,X,T}(\omega_k, \underline{\vartheta})}{\partial \underline{\vartheta}^\top}^\top \text{Vec} \mathbf{\Phi}^\top(\omega_k), \end{aligned}$$

indeed

$$\begin{aligned} \frac{\partial \text{Vec} \mathbf{I}_{Z,X,T}(\omega_k, \underline{\vartheta})}{\partial \underline{\vartheta}^\top}^\top [\text{Vec} \mathbf{\Phi}^\top(\omega_k)] &= \text{Vec} \frac{\partial \text{Tr} \mathbf{\Phi}^\top(\omega_k) \mathbf{I}_{X,Z,T}^\top(\omega_k, \underline{\vartheta})}{\partial \underline{\vartheta}^\top} \\ &= \frac{\partial \text{Vec} \mathbf{I}_{X,Z,T}(\omega_k, \underline{\vartheta})}{\partial \underline{\vartheta}^\top}^\top \text{Vec} \mathbf{\Phi}^\top(\omega_k), \end{aligned}$$

and

$$T \sum_{k=-T_1}^{T_1} \frac{\partial \text{Vec} \mathbf{I}_{Z,X,T}(\omega_k, \underline{\vartheta})}{\partial \underline{\vartheta}^\top} \text{Vec} \mathbf{\Phi}^\top(\omega_k) = \sum_{k=-T_1}^{T_1} \left[\text{Vec} \frac{\partial \underline{d}_{X,T}(\omega_k, \underline{\vartheta})}{\partial \underline{\vartheta}^\top} \right] \otimes \underline{d}_{Z,T}(\omega_k) \text{Vec} \mathbf{\Phi}^\top(\omega_k),$$

therefore we consider the variance matrix of the complex random variables, see [5] p. 89.,

$$\begin{aligned} \text{Var} \frac{\partial \text{Vec} \mathbf{I}_{X,Z,T}(\omega_k, \underline{\vartheta})}{\partial \underline{\vartheta}^\top}^\top [\text{Vec} \mathbf{\Phi}^\top(\omega_k)] \\ = \frac{1}{T^2} \text{Var} \left(\frac{\partial \underline{d}_{X,T}(\omega_k, \underline{\vartheta})}{\partial \underline{\vartheta}^\top}^\top \mathbf{\Phi}^\top(\omega_k) \overline{\underline{d}_{Z,T}(\omega_k)} \right) \\ = \frac{1}{T} \frac{\partial \underline{d}_{X,T}(\omega_k, \underline{\vartheta})}{\partial \underline{\vartheta}^\top}^\top \mathbf{\Phi}^\top(\omega_k) \mathbf{S}_Z^\top(\omega_k) \mathbf{\Phi}^\top(\omega_k) \left[\frac{\partial \overline{\underline{d}_{X,T}(\omega_k, \underline{\vartheta})}}{\partial \underline{\vartheta}^\top} \right] + o(1). \end{aligned}$$

Because of (24), this limit is written

$$\begin{aligned} \frac{\partial \underline{d}_{X,T}(\omega_k, \underline{\vartheta})}{\partial \underline{\vartheta}_1^\top}^\top \mathbf{\Phi}^\top(\omega_k) \mathbf{S}_Z^\top(\omega_k) \mathbf{\Phi}^\top(\omega_k) \left[\frac{\partial \overline{\underline{d}_{X,T}(\omega_k, \underline{\vartheta})}}{\partial \underline{\vartheta}_2^\top} \right] \\ = T \mathbf{U}_p \otimes \left[\text{Vec} \mathbf{\Phi}^\top(\omega_k) \mathbf{S}_Z^\top(\omega_k) \mathbf{\Phi}^\top(\omega_k) \right]^\top \frac{\partial \overline{\underline{d}_{X,T}(\omega_k, \underline{\vartheta}_2)}}{\partial \underline{\vartheta}_2^\top} \otimes \text{Vec} \frac{\partial \underline{d}_{X,T}(\omega_k, \underline{\vartheta}_1)}{\partial \underline{\vartheta}_1^\top} \end{aligned}$$

The variance matrix of the derivative $\frac{\partial Q_T(\underline{\vartheta}_0)}{\partial \underline{\vartheta}^\top}$ has the limit

$$\begin{aligned} \lim_{T \rightarrow \infty} \text{Var} \sqrt{T} \text{Vec} \frac{\partial Q_T(\underline{\vartheta}_0)}{\partial \underline{\vartheta}^\top} &= \int_{-1/2}^{1/2} \mathbf{U}_p \otimes \left[\text{Vec} \mathbf{\Phi}^\top(\omega) \mathbf{S}_Z^\top(\omega) \mathbf{\Phi}^\top(\omega) \right]^\top \\ &\times d \left(\frac{\partial^2 \mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top} + \frac{\partial^2 \mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_1^\top \partial \underline{\vartheta}_2^\top} \right)_{\underline{\vartheta}_1 = \underline{\vartheta}_2 = \underline{\vartheta}_0} \end{aligned}$$

It is worth noting that

$$\mathbf{U}_p \otimes \left[\text{Vec} \mathbf{\Phi}^\top(\omega) \mathbf{S}_Z^\top(\omega) \mathbf{\Phi}^\top(\omega) \right]^\top (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \mathbf{K}_{d \cdot dp} = (\mathbf{U}_p \otimes [\text{Vec}(\mathbf{\Phi}(\omega) \mathbf{S}_Z(\omega) \mathbf{\Phi}(\omega))]^\top), \quad (26)$$

and

$$\frac{\partial^2 \mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_1^\top \partial \underline{\vartheta}_2^\top} = (\mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) \mathbf{K}_{d \cdot dp} \frac{\partial^2 \mathbf{F}(-\omega, \underline{\vartheta}_2, \underline{\vartheta}_1)}{\partial \underline{\vartheta}_1^\top \partial \underline{\vartheta}_2^\top},$$

hence the asymptotic variance is written

$$\begin{aligned} & \int_{-1/2}^{1/2} \mathbf{U}_p \otimes \left[\text{Vec } \mathbf{\Phi}^\top(\omega_k) \mathbf{S}_{\underline{Z}}^\top(\omega_k) \mathbf{\Phi}^\top(\omega_k) \right]^\top d \left(\frac{\partial^2 \mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top} + \frac{\partial^2 \mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_1^\top \partial \underline{\vartheta}_2^\top} \Big|_{\underline{\vartheta}_1 = \underline{\vartheta}_2 = \underline{\vartheta}_0} \right) \\ &= 2 \int_{-1/2}^{1/2} \mathbf{U}_p \otimes \left[\text{Vec } \mathbf{\Phi}^\top(\omega_k) \mathbf{S}_{\underline{Z}}^\top(\omega_k) \mathbf{\Phi}^\top(\omega_k) \right]^\top d \left(\frac{\partial^2 \mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top} \Big|_{\underline{\vartheta}_1 = \underline{\vartheta}_2 = \underline{\vartheta}_0} \right). \end{aligned}$$

7.4 Hessian of Q

We are interested in, for instance,

$$\frac{\partial}{\partial \underline{\vartheta}_2^\top} \frac{\partial \text{Vec } \mathbf{I}_{X,T}(\omega_k, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_1^\top} \Big|_{\underline{\vartheta}_1 = \underline{\vartheta}_2 = \underline{\vartheta}}^\top \text{Vec } \mathbf{\Phi}^\top(\omega_k),$$

Now, using the chain rule for the derivatives, we have

$$\begin{aligned} & \frac{\partial}{\partial \underline{\vartheta}_2^\top} \text{Vec } \left[\text{Vec } \mathbf{\Phi}^\top(\omega_k) \right]^\top \frac{\partial \overline{d_{X,T}(\omega_k, \underline{\vartheta}_2)}}{\partial \underline{\vartheta}_1^\top} \otimes \frac{\partial d_{X,T}(\omega_k, \underline{\vartheta}_1)}{\partial \underline{\vartheta}_1^\top} \\ &= (\mathbf{U}_p \otimes [\text{Vec } \mathbf{\Phi}^\top(\omega_k)]^\top) \frac{\partial}{\partial \underline{\vartheta}_2^\top} \text{Vec } \overline{d_{X,T}(\omega_k, \underline{\vartheta}_2)} \otimes \frac{\partial d_{X,T}(\omega_k, \underline{\vartheta}_1)}{\partial \underline{\vartheta}_1^\top} \\ &= (\mathbf{U}_p \otimes [\text{Vec } \mathbf{\Phi}^\top(\omega_k)]^\top) (\mathbf{K}_{p,d} \otimes \mathbf{U}_d) \mathbf{K}_{d,dp} \text{Vec } \left(\frac{\partial d_{X,T}(\omega_k, \underline{\vartheta}_1)}{\partial \underline{\vartheta}_1^\top} \otimes \frac{\partial \overline{d_{X,T}(\omega_k, \underline{\vartheta}_2)}}{\partial \underline{\vartheta}_2^\top} \right) \\ &= (\mathbf{U}_p \otimes [\text{Vec } \mathbf{\Phi}^\top(\omega_k)]^\top) \text{Vec } \left(\frac{\partial d_{X,T}(\omega_k, \underline{\vartheta}_1)}{\partial \underline{\vartheta}_1^\top} \otimes \frac{\partial \overline{d_{X,T}(\omega_k, \underline{\vartheta}_2)}}{\partial \underline{\vartheta}_2^\top} \right) \\ &= T(\mathbf{U}_p \otimes [\text{Vec } \mathbf{\Phi}^\top(\omega_k)]^\top) \frac{\partial^2 \mathbf{I}_{X,T}(-\omega, \underline{\vartheta}_2, \underline{\vartheta}_1)}{\partial \underline{\vartheta}_1^\top \partial \underline{\vartheta}_2^\top}, \end{aligned} \tag{27}$$

see (24) and (25). Similar steps lead to the mixed derivative $\partial^2 / \partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top$

$$\begin{aligned} & \frac{\partial}{\partial \underline{\vartheta}_1^\top} \text{Vec } \left[\text{Vec } \mathbf{\Phi}^\top(\omega_k) \right]^\top \left[\frac{\partial \overline{d_{X,T}(\omega_k, \underline{\vartheta}_2)}}{\partial \underline{\vartheta}_2^\top} \otimes d_{X,T}(\omega_k, \underline{\vartheta}_1) \right] \\ &= (\mathbf{U}_p \otimes [\text{Vec } \mathbf{\Phi}^\top(\omega_k)]^\top) \left[\text{Vec } \frac{\partial \overline{d_{X,T}(\omega_k, \underline{\vartheta}_2)}}{\partial \underline{\vartheta}_2^\top} \otimes \frac{\partial d_{X,T}(\omega_k, \underline{\vartheta}_1)}{\partial \underline{\vartheta}_1^\top} \right] \\ &= T(\mathbf{U}_p \otimes [\text{Vec } \mathbf{\Phi}^\top(\omega_k)]^\top) \frac{\partial^2 \mathbf{I}_{X,T}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top}, \end{aligned} \tag{28}$$

clearly at $\underline{\vartheta}_1 = \underline{\vartheta}_2 = \underline{\vartheta}$ the expressions (27) and (28) are complex conjugates of each other.

7.5 Scaled derivatives

Consider the second scaled derivative of $X_{t+h}(\underline{\vartheta}_1) X_t^\top(\underline{\vartheta}_2)$

$$\begin{aligned} \frac{\partial_{s,T}^2 X_{t+h}(\underline{\vartheta}_1) X_t^\top(\underline{\vartheta}_2)}{\partial_{s,T} \underline{\vartheta}_2^\top \partial_{s,T} \underline{\vartheta}_1^\top} &= \text{Vec } \frac{\partial_{s,T} X_t(\underline{\vartheta}_2)}{\partial_{s,T} \underline{\vartheta}_2^\top} \otimes \frac{\partial_{s,T} X_{t+h}(\underline{\vartheta}_1)}{\partial_{s,T} \underline{\vartheta}_1^\top} \\ &\quad + (\mathbf{K}_{p,d} \otimes \mathbf{U}_d) \mathbf{K}_{d,dp} \text{Vec } \frac{\partial_{s,T} X_{t+h}(\underline{\vartheta}_1)}{\partial_{s,T} \underline{\vartheta}_1^\top} \otimes \frac{\partial_{s,T} X_t(\underline{\vartheta}_2)}{\partial_{s,T} \underline{\vartheta}_2^\top}, \end{aligned}$$

at $\mathbf{D}_T X_t(\underline{\vartheta})$. The scaled derivative of each term is

$$\mathbf{D}_T \frac{\partial}{\partial \underline{\vartheta}^\top} X_t(\underline{\vartheta}) \mathbf{D}_{1,T},$$

by definition, hence

$$\begin{aligned} \text{Vec } \frac{\partial_{s,T} \mathbf{D}_T X_t(\underline{\vartheta}_2)}{\partial_{s,T} \underline{\vartheta}_2^\top} \otimes \frac{\partial_{s,T} \mathbf{D}_T X_{t+h}(\underline{\vartheta}_1)}{\partial_{s,T} \underline{\vartheta}_1^\top} &= (\mathbf{D}_{1,T} \otimes \mathbf{D}_T \otimes \mathbf{D}_T) \\ &\quad \times \text{Vec } \frac{\partial X_t(\underline{\vartheta}_2)}{\partial \underline{\vartheta}_2^\top} \otimes \frac{\partial X_{t+h}(\underline{\vartheta}_1)}{\partial \underline{\vartheta}_1^\top} \mathbf{D}_{1,T}. \end{aligned}$$

The matrices $(\mathbf{D}_{1,T} \otimes \mathbf{D}_T \otimes \mathbf{D}_T)$ and $(\mathbf{K}_{p,d} \otimes \mathbf{U}_d) \mathbf{K}_{d,dp}$ commute. We conclude

$$\frac{\partial_{s,T}^2 \mathbf{D}_T \underline{\mathbf{X}}_{t+h}(\vartheta_1) \mathbf{D}_T \underline{\mathbf{X}}_t^\top(\vartheta_2)}{\partial_{s,T} \vartheta_2^\top \partial_{s,T} \vartheta_1^\top} = (\mathbf{D}_{1,T} \otimes \mathbf{D}_T \otimes \mathbf{D}_T) \left[\frac{\partial^2 \underline{\mathbf{X}}_{t+h}(\vartheta_1) \underline{\mathbf{X}}_t^\top(\vartheta_2)}{\partial \vartheta_2^\top \partial \vartheta_1^\top} \right] \mathbf{D}_{1,T}.$$

7.6 Asymptotic Variance for the parameters of the linear model

The variance matrix of the complex vector $\text{Vec} \sum_{k=-T_1}^{T_1} \Phi(\omega_k) \mathbf{I}_{Y,X,T}(\omega_k)$, can be easily calculated as in Terdik (see [29]). Using the procedure outlined there, we obtain

$$\begin{aligned} & \lim_{T \rightarrow \infty} \text{Var} \left(\frac{1}{T} \text{Vec} \sum_{k=-T_1}^{T_1} \Phi(\omega_k) \mathbf{I}_{Y,X,T}(\omega_k) \right) \\ &= \text{Vec} \int_{-1/2}^{1/2} d\mathbf{F}^\top(\omega) \otimes [\Phi(\omega) \mathbf{S}_{\underline{Z}}(\omega) \Phi(\omega)] \end{aligned} \quad (29)$$

Setting $\Phi(\omega) = \mathbf{S}_{\underline{Z}}^{-1}(\omega)$, we can derive the variance (12) directly from (10). The mixed derivative

$$\frac{\partial^2 \mathbf{F}(\omega, \vartheta_1, \vartheta_2)}{\partial \vartheta_2^\top \partial \vartheta_1^\top},$$

is the inverse Fourier transform of the same mixed derivative of $\mathbf{C}_{\underline{X}}(h, \vartheta)$ which is the limit of the $\frac{\partial^2 \mathbf{I}_{X,T}(\omega, \vartheta_1, \vartheta_2)}{\partial \vartheta_2^\top \partial \vartheta_1^\top}$ at $\vartheta_1 = \vartheta_2 = \vartheta$, in our case,

$$\begin{aligned} \frac{\partial^2 \mathbf{I}_{X,T}(\omega, \vartheta_1, \vartheta_2)}{\partial \vartheta_2^\top \partial \vartheta_1^\top} &= \text{Vec} \frac{\overline{\partial \underline{d}_{X,T}(\omega_k, \vartheta_2)}}{\partial \vartheta_2^\top} \otimes \frac{\partial \underline{d}_{X,T}(\omega_k, \vartheta_1)}{\partial \vartheta_1^\top} \\ &= \text{Vec} \frac{\overline{\partial \mathbf{B}_2 \underline{d}_{X,T}(\omega_k)}}{\partial (\text{Vec } \mathbf{B}_2)^\top} \otimes \frac{\partial \mathbf{B}_1 \underline{d}_{X,T}(\omega_k)}{\partial (\text{Vec } \mathbf{B}_1)^\top} \\ &= \overline{\underline{d}_{X,T}(\omega_k)} \otimes \text{Vec } \mathbf{U}_d \otimes \underline{d}_{X,T}^\top(\omega_k) \otimes \mathbf{U}_d, \end{aligned} \quad (30)$$

and the product

$$(\mathbf{U}_{pd} \otimes [\text{Vec } \Phi^\top(\omega_k)]^\top) \overline{\underline{d}_{X,T}(\omega_k)} \otimes \text{Vec } \mathbf{U}_d \otimes \underline{d}_{X,T}^\top(\omega_k) \otimes \mathbf{U}_d$$

equals to

$$[\underline{d}_{X,T}(\omega_k) \underline{d}_{X,T}^*(\omega_k)]^\top \otimes \Phi(\omega_k),$$

and (12) follows.

7.7 The Variance matrix for the parameters of the mixed model

Rewrite the objective function in terms of the parameters

$$\begin{aligned} Q_T(\mathbf{B}_1, \mathbf{B}_2, \lambda) &= \sum_{k=-T_1}^{T_1} \text{Tr}(\mathbf{I}_{Y,T}(\omega_k) \Phi(\omega_k)) + \text{Tr} \mathbf{B}_1 \mathbf{I}_{X_1,T}(\omega_k) \mathbf{B}_1^\top \Phi(\omega_k) \\ &+ \text{Tr} \mathbf{B}_1 \mathbf{I}_{X_1, X_2, T}(\omega_k, \lambda) \mathbf{B}_2^\top \Phi(\omega_k) + \text{Tr} \mathbf{B}_2 \mathbf{I}_{X_2, X_1, T}(\omega_k, \lambda) \mathbf{B}_1^\top \Phi(\omega_k) \\ &+ \text{Tr} \mathbf{B}_2 \mathbf{I}_{X_2, T}(\omega_k, \lambda) \mathbf{B}_2^\top \Phi(\omega_k) \\ &- \text{Tr} \mathbf{I}_{Y, X_1, T}(\omega_k) \mathbf{B}_1^\top \Phi(\omega_k) - \text{Tr} \mathbf{B}_1 \mathbf{I}_{X_1, Y, T}(\omega_k) \Phi(\omega_k) \\ &- \text{Tr} \mathbf{I}_{Y, X_2, T}(\omega_k, \lambda) \mathbf{B}_2^\top \Phi(\omega_k) - \text{Tr} \mathbf{B}_2 \mathbf{I}_{X_2, Y, T}(\omega_k, \lambda) \Phi(\omega_k) \quad . \end{aligned} \quad (31)$$

Consider now the normal equations

$$\begin{aligned} \frac{\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \lambda)}{\partial \mathbf{B}_1} &= \mathbf{0}, \\ \frac{\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \lambda)}{\partial \mathbf{B}_2} &= \mathbf{0}, \\ \text{Vec} \frac{\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \lambda)}{\partial \lambda^\top} &= \mathbf{0}. \end{aligned}$$

They can be written as

$$\begin{aligned} \frac{\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \underline{\lambda})}{\partial \mathbf{B}_1} &= \sum_{k=-T_1}^{T_1} \Phi^\top(\omega_k) \mathbf{B}_1 \mathbf{I}_{X_1, T}^\top(\omega_k) + \Phi(\omega_k) \mathbf{B}_1 \mathbf{I}_{X_1, T}(\omega_k) \\ &+ \Phi^\top(\omega_k) \mathbf{B}_2 \mathbf{I}_{X_1, X_2, T}^\top(\omega_k, \underline{\lambda}) + \Phi(\omega_k) \mathbf{B}_2 \mathbf{I}_{X_2, X_1, T}(\omega_k, \underline{\lambda}) \\ &- \Phi(\omega_k) \mathbf{I}_{Y, X_1, T}(\omega_k) - \Phi^\top(\omega_k) \mathbf{I}_{X_1, Y, T}^\top(\omega_k). \end{aligned}$$

Similarly

$$\begin{aligned} \frac{\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \underline{\lambda})}{\partial \mathbf{B}_2} &= \sum_{k=-T_1}^{T_1} \Phi^\top(\omega_k) \mathbf{B}_2 \mathbf{I}_{X_2, T}^\top(\omega_k, \underline{\lambda}) + \Phi(\omega_k) \mathbf{B}_2 \mathbf{I}_{X_2, T}(\omega_k, \underline{\lambda}) \\ &+ \Phi^\top(\omega_k) \mathbf{B}_1 \mathbf{I}_{X_1, X_2, T}^\top(\omega_k, \underline{\lambda}) + \Phi(\omega_k) \mathbf{B}_1 \mathbf{I}_{X_2, X_1, T}(\omega_k, \underline{\lambda}) \\ &- \Phi(\omega_k) \mathbf{I}_{Y, X_2, T}(\omega_k, \underline{\lambda}) - \Phi^\top(\omega_k) \mathbf{I}_{X_2, Y, T}^\top(\omega_k, \underline{\lambda}), \end{aligned}$$

and finally

$$\begin{aligned} \text{Vec} \frac{\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \underline{\lambda})}{\partial \underline{\lambda}^\top} &= \sum_{k=-T_1}^{T_1} \frac{\partial \mathbf{B}_1 \mathbf{I}_{X_1, X_2, T}(\omega_k, \underline{\lambda}) \mathbf{B}_2^\top}{\partial \underline{\lambda}^\top} + \frac{\partial \mathbf{B}_2 \mathbf{I}_{X_2, X_1, T}(\omega_k, \underline{\lambda}) \mathbf{B}_1^\top}{\partial \underline{\lambda}^\top} \\ &+ \frac{\partial \mathbf{B}_2 \mathbf{I}_{X_2, T}(\omega_k, \underline{\lambda}) \mathbf{B}_2^\top}{\partial \underline{\lambda}^\top} - \frac{\partial \mathbf{I}_{Y, X_2, T}(\omega_k, \underline{\lambda}) \mathbf{B}_2^\top}{\partial \underline{\lambda}^\top} - \frac{\partial \mathbf{B}_2 \mathbf{I}_{X_2, Y, T}(\omega_k, \underline{\lambda})}{\partial \underline{\lambda}^\top} \text{Vec}[\Phi^\top(\omega_k)]. \end{aligned}$$

The variance of the derivatives

Let us denote the limit of the Hessian matrix of the estimates $\text{Vec} \widehat{\mathbf{B}}_1, \text{Vec} \widehat{\mathbf{B}}_2, \widehat{\underline{\lambda}}$ by

$$\Sigma = 2 \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{1\lambda} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{2\lambda} \\ \Sigma_{\lambda 1} & \Sigma_{\lambda 2} & \Sigma_{\lambda\lambda} \end{bmatrix}.$$

The second derivative of $\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \underline{\lambda}) / \partial \mathbf{B}_1$ by \mathbf{B}_1 does not depend either on \mathbf{B}_2 nor $\underline{\lambda}$, therefore according to (12)

$$\Sigma_{11} = \int_{-1/2}^{1/2} d\mathbf{F}_{11}^\top(\omega) \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega).$$

By setting $\Phi = \mathbf{S}_{\underline{Z}}^{-1}$. The matrix Σ_{12} between $\text{Vec} \widehat{\mathbf{B}}_1$ and $\text{Vec} \widehat{\mathbf{B}}_2$ follows from

$$\frac{\partial^2 Q_T(\mathbf{B}_1, \mathbf{B}_2, \underline{\lambda})}{\partial \mathbf{B}_2 \partial \mathbf{B}_1} = \sum_{k=-T_1}^{T_1} \mathbf{I}_{X_1, X_2, T}(\omega_k, \underline{\lambda}) \otimes \Phi^\top(\omega_k) + \mathbf{I}_{X_2, X_1, T}^\top(\omega_k, \underline{\lambda}) \otimes \Phi(\omega_k),$$

it is

$$\Sigma_{12} = \int_{-1/2}^{1/2} d\mathbf{F}_{1,2}^\top(\omega, \underline{\lambda}_0) \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega).$$

The second derivative of $\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \underline{\lambda}) / \partial \mathbf{B}_2$ by \mathbf{B}_2 is similar except the SDFR depends on $\underline{\lambda}$

$$\Sigma_{22} = \int_{-1/2}^{1/2} d\mathbf{F}_2(\omega, \underline{\lambda}_0) \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega)^\top.$$

Now for the matrix $\Sigma_{1\lambda}$ consider

$$\begin{aligned}
\frac{\partial^2 Q_T(\mathbf{B}_1, \mathbf{B}_2, \lambda)}{\partial \mathbf{B}_1 \partial \lambda^\top} &= \sum_{k=-T_1}^{T_1} \frac{\partial \text{Vec } \Phi^\top(\omega_k) \mathbf{B}_2 \mathbf{I}_{\underline{X}_1, \underline{X}_2, T}^\top(\omega_k, \lambda)}{\partial \lambda^\top} + \frac{\partial \text{Vec } \Phi(\omega_k) \mathbf{B}_2 \mathbf{I}_{\underline{X}_2, \underline{X}_1, T}(\omega_k, \lambda)}{\partial \lambda^\top} \\
&= \frac{1}{T} \sum_{k=-T_1}^{T_1} \left[\frac{\partial \text{Vec } \Phi^\top(\omega_k) \mathbf{B}_2 \overline{d_{\underline{X}_2, T}(\omega_k, \lambda)} d_{\underline{X}_1, T}^\top(\omega_k)}{\partial \lambda^\top} \right. \\
&\quad \left. + \frac{\partial \text{Vec } \Phi(\omega_k) \mathbf{B}_2 \overline{d_{\underline{X}_2, T}(\omega_k, \lambda)} d_{\underline{X}_1, T}^*(\omega_k)}{\partial \lambda^\top} \right] \\
&= \frac{1}{T} \sum_{k=-T_1}^{T_1} (\mathbf{U}_p \otimes \Phi^\top(\omega_k) \mathbf{B}_2) \left[\overline{d_{\underline{X}_1, T}(\omega_k)} \otimes \frac{\partial \overline{d_{\underline{X}_2, T}(\omega_k, \lambda)}}{\partial \lambda^\top} \right] \\
&\quad + (\mathbf{U}_p \otimes \Phi(\omega_k) \mathbf{B}_2) \left[\overline{d_{\underline{X}_1, T}(\omega_k)} \otimes \frac{\partial \overline{d_{\underline{X}_2, T}(\omega_k, \lambda)}}{\partial \lambda^\top} \right],
\end{aligned}$$

hence the limit

$$\begin{aligned}
\Sigma_{1\lambda} &= \int_{-1/2}^{1/2} (\mathbf{U}_p \otimes \Phi(\omega) \mathbf{B}_{2,0}) d \frac{\partial \mathbf{F}_{1,2}(\omega, \lambda_0)}{\partial \lambda^\top} \\
&= \int_{-1/2}^{1/2} \mathbf{U}_p \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega) \mathbf{B}_{2,0} d \frac{\partial \mathbf{F}_{1,2}(\omega, \lambda_0)}{\partial \lambda^\top}.
\end{aligned}$$

The matrix $\Sigma_{2\lambda}$ based on

$$\begin{aligned}
\frac{\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \lambda)}{\partial \mathbf{B}_2} &= \sum_{k=-T_1}^{T_1} \Phi^\top(\omega_k) \mathbf{B}_2 \mathbf{I}_{\underline{X}_2, T}^\top(\omega_k, \lambda) + \Phi(\omega_k) \mathbf{B}_2 \mathbf{I}_{\underline{X}_2, T}(\omega_k, \lambda) \\
&\quad + \Phi^\top(\omega_k) \mathbf{B}_1 \mathbf{I}_{\underline{X}_1, \underline{X}_2, T}^\top(\omega_k, \lambda) + \Phi(\omega_k) \mathbf{B}_1 \mathbf{I}_{\underline{X}_2, \underline{X}_1, T}(\omega_k, \lambda) \\
&\quad - \Phi(\omega_k) \mathbf{I}_{\underline{Y}, \underline{X}_2, T}(\omega_k, \lambda) - \Phi^\top(\omega_k) \mathbf{I}_{\underline{X}_2, \underline{Y}, T}^\top(\omega_k, \lambda), \\
\frac{\partial^2 Q_T(\mathbf{B}_1, \mathbf{B}_2, \lambda)}{\partial \mathbf{B}_2 \partial \lambda^\top} &= \sum_{k=-T_1}^{T_1} \frac{\partial \text{Vec } \Phi^\top(\omega_k) \mathbf{B}_2 \mathbf{I}_{\underline{X}_2, T}^\top(\omega_k, \lambda)}{\partial \lambda^\top} + \frac{\partial \text{Vec } \Phi(\omega_k) \mathbf{B}_2 \mathbf{I}_{\underline{X}_2, T}(\omega_k, \lambda)}{\partial \lambda^\top} \\
&\quad + \frac{\partial \text{Vec } \Phi^\top(\omega_k) \mathbf{B}_1 \mathbf{I}_{\underline{X}_1, \underline{X}_2, T}^\top(\omega_k, \lambda)}{\partial \lambda^\top} + \frac{\partial \text{Vec } \Phi(\omega_k) \mathbf{B}_1 \mathbf{I}_{\underline{X}_2, \underline{X}_1, T}(\omega_k, \lambda)}{\partial \lambda^\top} \\
&\quad - \frac{\partial \text{Vec } \Phi(\omega_k) \mathbf{I}_{\underline{Y}, \underline{X}_2, T}(\omega_k, \lambda)}{\partial \lambda^\top} - \frac{\partial \text{Vec } \Phi^\top(\omega_k) \mathbf{I}_{\underline{X}_2, \underline{Y}, T}^\top(\omega_k, \lambda)}{\partial \lambda^\top} \\
&= 2 \sum_{k=-T_1}^{T_1} \frac{\partial \text{Vec } \Phi^\top(\omega_k) \mathbf{B}_2 \mathbf{I}_{\underline{X}_2, T}^\top(\omega_k, \lambda_1, \lambda_2)}{\partial \lambda_1^\top}, \quad \lambda_1 = \lambda_2 = \lambda
\end{aligned}$$

Use the equation (13) then the limit of the derivative at $\underline{\vartheta} = \underline{\vartheta}_0$ is zero,

$$\Sigma_{2\lambda} = \int_{-1/2}^{1/2} (\mathbf{U}_q \otimes \Phi(\omega) \mathbf{B}_{2,0}) \frac{\partial \mathbf{F}_{2,2}(\omega, \lambda_1, \lambda_2)}{\partial \lambda_1^\top} \Big|_{\lambda_1 = \lambda_2 = \lambda_0}.$$

The matrix $\Sigma_{\lambda\lambda}$ can be obtained from the general result (10)

$$\begin{aligned}
\Sigma_{\lambda\lambda} &= \int_{-1/2}^{1/2} \mathbf{U}_r \otimes \left[\text{Vec } \mathbf{S}_{\underline{Z}}^{-1}(\omega) \text{ }^\top \text{ }^\top (\mathbf{B}_{2,0} \otimes \mathbf{B}_{2,0}) \right] d \frac{\partial^2 \mathbf{F}_{2,2}(\omega, \lambda_1, \lambda_2)}{\partial \lambda_2^\top \partial \lambda_1^\top} \Big|_{\lambda_1 = \lambda_2 = \lambda_0} \\
&= \int_{-1/2}^{1/2} \mathbf{U}_r \otimes \text{Vec } \mathbf{B}_{2,0}^\top \mathbf{S}_{\underline{Z}}^{-1}(\omega) \mathbf{B}_{2,0} \text{ }^\top \text{ }^\top d \frac{\partial^2 \mathbf{F}_{2,2}(\omega, \lambda_1, \lambda_2)}{\partial \lambda_2^\top \partial \lambda_1^\top} \Big|_{\lambda_1 = \lambda_2 = \lambda_0}.
\end{aligned}$$

Finally $\text{Vec } \widehat{\mathbf{B}}_1, \text{Vec } \widehat{\mathbf{B}}_2, \widehat{\lambda}$

$$\lim_{T \rightarrow \infty} \text{Var} \left[\text{Vec } \widehat{\mathbf{B}}_1, \text{Vec } \widehat{\mathbf{B}}_2, \widehat{\lambda} \right] = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{1\lambda} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{2\lambda} \\ \Sigma_{\lambda 1} & \Sigma_{\lambda 2} & \Sigma_{\lambda\lambda} \end{bmatrix}^{-1}.$$

Use Theorem 2, [25] p.16 for the inverse. We use the general formula for the variance of the estimate of $\underline{\vartheta} = \text{Vec}(\text{Vec } \mathbf{B}_1, \text{Vec } \mathbf{B}_2, \underline{\lambda})$. Here \mathbf{B}_1 is $d \times p$, \mathbf{B}_2 is $d \times q$, $\underline{\lambda}$ is $r \times 1$, $\underline{X}_{1,t}$ is of dimension p , $\underline{X}_{2,t}(\underline{\lambda})$ is of dimension q . For the mixed derivative

$$\frac{\partial^2 \mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top} \quad (32)$$

we use

$$\begin{aligned} \mathbf{I}_{\underline{X}, T}(\omega_k, \underline{\vartheta}) &= \mathbf{B} \mathbf{I}_{\underline{X}_3, T}(\omega_k, \underline{\lambda}) \mathbf{B}^\top \\ &= \left[\mathbf{B}_1 \underline{d}_{\underline{X}_1, T}(\omega_k) + \mathbf{B}_2 \underline{d}_{\underline{X}_2, T}(\omega_k, \underline{\lambda}) \right] \left[\mathbf{B}_1 \underline{d}_{\underline{X}_1, T}(\omega_k) + \mathbf{B}_2 \underline{d}_{\underline{X}_2, T}(\omega_k, \underline{\lambda}) \right]_{\mathbf{B}_1 = \mathbf{B}_2}^* \end{aligned}$$

For the parameters $\underline{\vartheta}_1, \underline{\vartheta}_2$ we have $\underline{\vartheta}_i = \text{Vec}(\text{Vec } \mathbf{B}_{1i}, \text{Vec } \mathbf{B}_{2i}, \underline{\lambda}_i)$, $i = 1, 2$. Write

$$\begin{aligned} \mathbf{I}_{\underline{X}, T}(\omega_k, \underline{\vartheta}_1, \underline{\vartheta}_2) &= \left[\mathbf{B}_{11} \underline{d}_{\underline{X}_1, T}(\omega_k) + \mathbf{B}_{21} \underline{d}_{\underline{X}_2, T}(\omega_k, \underline{\lambda}_1) \right] \left[\mathbf{B}_{12} \underline{d}_{\underline{X}_1, T}(\omega_k) + \mathbf{B}_{22} \underline{d}_{\underline{X}_2, T}(\omega_k, \underline{\lambda}_2) \right]^* \\ &= \mathbf{B}_{11} \underline{d}_{\underline{X}_1, T}(\omega_k) \underline{d}_{\underline{X}_1, T}^*(\omega_k) \mathbf{B}_{12}^\top + \mathbf{B}_{21} \underline{d}_{\underline{X}_2, T}(\omega_k, \underline{\lambda}_1) \underline{d}_{\underline{X}_1, T}^*(\omega_k) \mathbf{B}_{12}^\top \\ &\quad + \mathbf{B}_{11} \underline{d}_{\underline{X}_1, T}(\omega_k) \underline{d}_{\underline{X}_2, T}^*(\omega_k, \underline{\lambda}_2) \mathbf{B}_{22}^\top + \mathbf{B}_{21} \underline{d}_{\underline{X}_2, T}(\omega_k, \underline{\lambda}_1) \underline{d}_{\underline{X}_2, T}^*(\omega_k, \underline{\lambda}_2) \mathbf{B}_{22}^\top. \end{aligned}$$

The variance of

$$\begin{bmatrix} \left[\frac{\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \underline{\lambda})}{\partial \text{Vec } \mathbf{B}_1^\top} \right]^\top \\ \left[\frac{\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \underline{\lambda})}{\partial \text{Vec } \mathbf{B}_2^\top} \right]^\top \\ \left[\frac{\partial Q_T(\mathbf{B}_1, \mathbf{B}_2, \underline{\lambda})}{\partial \underline{\lambda}^\top} \right]^\top \end{bmatrix}$$

according to the mixed derivative(32) contains nine nonzero terms which are discussed below:

1. We have already considered the case

$$\begin{aligned} &\frac{\partial^2 \mathbf{B}_{11} \underline{d}_{\underline{X}_1, T}(\omega_k) \underline{d}_{\underline{X}_1, T}^*(\omega_k) \mathbf{B}_{12}^\top}{\partial \mathbf{B}_{12} \partial \mathbf{B}_{11}} \\ &= \overline{\underline{d}_{\underline{X}_1, T}(\omega_k)} \otimes \text{Vec } \mathbf{U}_d \otimes \underline{d}_{\underline{X}_1, T}^\top(\omega_k) \otimes \mathbf{U}_d \end{aligned}$$

(see (30)). For the linear model we have

$$\begin{aligned} &(\mathbf{U}_{pd} \otimes [\text{Vec } \Phi^\top(\omega_k)]^\top) \overline{\underline{d}_{\underline{X}_1, T}(\omega_k)} \otimes \text{Vec } \mathbf{U}_d \otimes \underline{d}_{\underline{X}_1, T}^\top(\omega_k) \otimes \mathbf{U}_d \\ &= \left[\underline{d}_{\underline{X}_1, T}(\omega_k) \underline{d}_{\underline{X}_1, T}^*(\omega_k) \right]^\top \otimes \Phi(\omega_k). \end{aligned}$$

The cases

$$\begin{aligned} &\frac{\partial^2 \mathbf{B}_{21} \underline{d}_{\underline{X}_2, T}(\omega_k, \underline{\lambda}_1) \underline{d}_{\underline{X}_1, T}^*(\omega_k) \mathbf{B}_{12}^\top}{\partial \mathbf{B}_{12} \partial \mathbf{B}_{21}} \\ &= \overline{\underline{d}_{\underline{X}_1, T}(\omega_k)} \otimes \text{Vec } \mathbf{U}_d \otimes \underline{d}_{\underline{X}_2, T}^\top(\omega_k, \underline{\lambda}_1) \otimes \mathbf{U}_d, \end{aligned}$$

and

$$\begin{aligned} &\frac{\partial^2 \mathbf{B}_{11} \underline{d}_{\underline{X}_1, T}(\omega_k) \underline{d}_{\underline{X}_2, T}^*(\omega_k, \underline{\lambda}_2) \mathbf{B}_{22}^\top}{\partial \mathbf{B}_{22} \partial \mathbf{B}_{11}} \\ &= \overline{\underline{d}_{\underline{X}_2, T}(\omega_k, \underline{\lambda}_2)} \otimes \text{Vec } \mathbf{U}_d \otimes \underline{d}_{\underline{X}_1, T}^\top(\omega_k) \otimes \mathbf{U}_d, \\ &\frac{\partial^2 \mathbf{B}_{21} \underline{d}_{\underline{X}_2, T}(\omega_k, \underline{\lambda}_1) \underline{d}_{\underline{X}_2, T}^*(\omega_k, \underline{\lambda}_2) \mathbf{B}_{22}^\top}{\partial \mathbf{B}_{22} \partial \mathbf{B}_{21}} \\ &= \overline{\underline{d}_{\underline{X}_2, T}(\omega_k, \underline{\lambda}_2)} \otimes \text{Vec } \mathbf{U}_d \otimes \underline{d}_{\underline{X}_2, T}^\top(\omega_k, \underline{\lambda}_1) \otimes \mathbf{U}_d. \end{aligned}$$

are similar since the parameters $\underline{\lambda}_i$ are fixed here. Also

2.

$$\begin{aligned}
& (\mathbf{U}_{pd} \otimes [\text{Vec } \Phi^\top(\omega_k)]^\top) \overline{d_{X_1, T}(\omega_k)} \otimes \text{Vec } \mathbf{U}_d \otimes d_{X_2, T}^\top(\omega_k, \lambda_1) \otimes \mathbf{U}_d \\
&= \left[d_{X_2, T}(\omega_k, \lambda_1) d_{X_1, T}^*(\omega_k) \right]^\top \otimes \Phi(\omega_k), \\
& (\mathbf{U}_{pd} \otimes [\text{Vec } \Phi^\top(\omega_k)]^\top) \overline{d_{X_1, T}(\omega_k)} \otimes \text{Vec } \mathbf{U}_d \otimes d_{X_2, T}^\top(\omega_k, \lambda_2) \otimes \mathbf{U}_d \\
&= \left[d_{X_1, T}(\omega_k) d_{X_2, T}^*(\omega_k, \lambda_2) \right]^\top \otimes \Phi(\omega_k), \\
& (\mathbf{U}_{qd} \otimes [\text{Vec } \Phi^\top(\omega_k)]^\top) \overline{d_{X_2, T}(\omega_k, \lambda_2)} \otimes \text{Vec } \mathbf{U}_d \otimes d_{X_2, T}^\top(\omega_k, \lambda_1) \otimes \mathbf{U}_d \\
&= \left[d_{X_2, T}(\omega_k, \lambda_1) d_{X_2, T}^*(\omega_k, \lambda_2) \right]^\top \otimes \Phi(\omega_k),
\end{aligned}$$

3. Taking the derivative by λ_2 and λ_1

$$\begin{aligned}
\frac{\partial^2 \mathbf{B}_{21} d_{X_2, T}(\omega_k, \lambda_1) d_{X_2, T}^*(\omega_k, \lambda_2) \mathbf{B}_{22}^\top}{\partial \lambda_2^\top \partial \lambda_1^\top} &= \text{Vec } \mathbf{B}_{22} \frac{\partial \overline{d_{X_2, T}(\omega_k, \lambda_2)}}{\partial \lambda_2^\top} \otimes \mathbf{B}_{21} \frac{\partial d_{X_2, T}(\omega_k, \lambda_1)}{\partial \lambda_1^\top} \\
&= \left[(\mathbf{U}_r \otimes \mathbf{B}_{22}) \text{Vec } \frac{\partial d_{X_2, T}(\omega_k, \lambda_2)}{\partial \lambda_2^\top} \right] \otimes \mathbf{B}_{21} \frac{\partial d_{X_2, T}(\omega_k, \lambda_1)}{\partial \lambda_1^\top} \\
&= (\mathbf{U}_r \otimes \mathbf{B}_{22} \otimes \mathbf{B}_{21}) \frac{\partial^2 d_{X_2, T}(\omega_k, \lambda_1) d_{X_2, T}^*(\omega_k, \lambda_2)}{\partial \lambda_2^\top \partial \lambda_1^\top},
\end{aligned}$$

therefore we can apply the earlier results.

4. Consider now

$$\begin{aligned}
& (\mathbf{U}_{pd} \otimes [\text{Vec } \Phi^\top(\omega_k)]^\top) \frac{\partial^2 \mathbf{B}_{21} d_{X_2, T}(\omega_k, \lambda_1) d_{X_1, T}^*(\omega_k) \mathbf{B}_{12}^\top}{\partial \mathbf{B}_{12} \partial \lambda_1^\top} \\
&= (\mathbf{U}_{pd} \otimes [\text{Vec } \Phi^\top(\omega_k)]^\top) \frac{\partial}{\partial \lambda_1^\top} \left(\text{Vec } \frac{\partial \text{Vec } [\mathbf{B}_{21} d_{X_2, T}(\omega_k, \lambda_1) d_{X_1, T}^*(\omega_k) \mathbf{B}_{12}^\top]}{\partial (\text{Vec } \mathbf{B}_{12})^\top} \right) \\
&= (\mathbf{U}_{pd} \otimes [\text{Vec } \Phi^\top(\omega_k)]^\top) \frac{\partial}{\partial \lambda_1^\top} (\mathbf{U}_{dp} \otimes \mathbf{K}_{d,d}) \text{Vec } \mathbf{B}_{21} d_{X_2, T}(\omega_k, \lambda_1) d_{X_1, T}^*(\omega_k) \otimes \mathbf{U}_d \\
&= (\mathbf{U}_{pd} \otimes [\text{Vec } \Phi(\omega_k)]^\top) \frac{\partial}{\partial \lambda_1^\top} \text{Vec } \mathbf{B}_{21} d_{X_2, T}(\omega_k, \lambda_1) d_{X_1, T}^*(\omega_k) \otimes \mathbf{U}_d \\
&= \frac{\partial}{\partial \lambda_1^\top} \text{Vec } \left[[\text{Vec } \Phi(\omega_k)]^\top \mathbf{B}_{21} d_{X_2, T}(\omega_k, \lambda_1) d_{X_1, T}^*(\omega_k) \otimes \mathbf{U}_d \right] \\
&= \frac{\partial}{\partial \lambda_1^\top} \text{Vec } \left[\mathbf{B}_{21} d_{X_2, T}(\omega_k, \lambda_1) d_{X_1, T}^*(\omega_k) \otimes \mathbf{U}_d \text{Vec } \Phi(\omega_k) \right] \\
&= \frac{\partial}{\partial \lambda_1^\top} \text{Vec } \left[\Phi(\omega_k) \mathbf{B}_{21} d_{X_2, T}(\omega_k, \lambda_1) d_{X_1, T}^*(\omega_k) \right] \\
&= (\mathbf{U}_p \otimes \Phi(\omega_k) \mathbf{B}_{21}) \left[\overline{d_{X_1, T}(\omega_k)} \otimes \frac{\partial d_{X_2, T}(\omega_k, \lambda)}{\partial \lambda^\top} \right].
\end{aligned}$$

5. The case

$$\begin{aligned}
& (\mathbf{U}_{qd} \otimes [\text{Vec } \Phi(\omega_k)]^\top) \frac{\partial^2 \mathbf{B}_{21} d_{X_2, T}(\omega_k, \lambda_1) d_{X_2, T}^*(\omega_k, \lambda_2) \mathbf{B}_{22}^\top}{\partial \mathbf{B}_{22} \partial \lambda_1^\top} \\
&= (\mathbf{U}_q \otimes \Phi(\omega_k) \mathbf{B}_{21}) \left[\overline{d_{X_2, T}(\omega_k, \lambda_2)} \otimes \frac{\partial d_{X_2, T}(\omega_k, \lambda_1)}{\partial \lambda_1^\top} \right].
\end{aligned}$$

is similar to the previous one.

6.

$$\begin{aligned}
& \frac{\partial^2 \mathbf{B}_{21} \underline{d}_{X_2, T}(\omega_k, \lambda_1) \underline{d}_{X_2, T}^*(\omega_k, \lambda_2) \mathbf{B}_{22}^\top}{\partial \lambda_2^\top \partial \mathbf{B}_{21}} \\
&= \frac{\partial}{\partial (\text{Vec } \mathbf{B}_{21})^\top} \text{Vec} \left[(\mathbf{B}_{22} \otimes \mathbf{B}_{21}) \frac{\overline{d_{X_2, T}(\omega_k, \lambda_2)}}{\partial \lambda_2^\top} \otimes \underline{d}_{X_2, T}(\omega_k, \lambda_1) \right] \\
&= \frac{\partial}{\partial (\text{Vec } \mathbf{B}_{21})^\top} \left[\frac{\overline{d_{X_2, T}(\omega_k, \lambda_2)}}{\partial \lambda_2^\top} \otimes \underline{d}_{X_2, T}(\omega_k, \lambda_1) \right]^\top \otimes \mathbf{U}_{d^2} \text{Vec}(\mathbf{B}_{22} \otimes \mathbf{B}_{21}) \\
&= \left[\frac{\overline{d_{X_2, T}(\omega_k, \lambda_2)}}{\partial \lambda_2^\top} \otimes \underline{d}_{X_2, T}(\omega_k, \lambda_1) \right]^\top \otimes \mathbf{U}_{d^2} (\mathbf{U}_q \otimes \mathbf{K}_{q \cdot d} \otimes \mathbf{U}_d) (\text{Vec } \mathbf{B}_{22} \otimes \mathbf{U}_{dq}).
\end{aligned}$$

7.

$$\begin{aligned}
& \frac{\partial^2 \mathbf{B}_{11} \underline{d}_{X_1, T}(\omega_k) \underline{d}_{X_2, T}^*(\omega_k, \lambda_2) \mathbf{B}_{22}^\top}{\partial \lambda_2^\top \partial \mathbf{B}_{11}} \\
&= \left[\frac{\overline{d_{X_2, T}(\omega_k, \lambda_2)}}{\partial \lambda_2^\top} \otimes \underline{d}_{X_1, T}(\omega_k) \right]^\top \otimes \mathbf{U}_{d^2} (\mathbf{U}_q \otimes \mathbf{K}_{p \cdot d} \otimes \mathbf{U}_d) (\text{Vec } \mathbf{B}_{22} \otimes \mathbf{U}_{dp}).
\end{aligned}$$

Also we can obtain

$$\int_{-1/2}^{1/2} \mathbf{U}_r \otimes ([\text{Vec}(\boldsymbol{\Phi}^\top(\omega_k))]^\top [\mathbf{B}_{22} \otimes \mathbf{B}_{21}]) d \frac{\partial^2 \mathbf{F}_2(\omega, \lambda_1, \lambda_2)}{\partial \lambda_2^\top \partial \lambda_1^\top}.$$

8 REFERENCES

- [1] M. Arató, A. N. Kolmogorov, and J. G. Sinay. Evaluation of the parameters of a complex stationary gauss-markov process. *Dokl. Akad. Nauk SSSR*, 146:747–750, 1962.
- [2] V. V. Anh, N. N. Leonenko, and L. M. Sakhno. Quasilikelihood-based higher-order spectral estimation of random processes and fields with possible long-range dependence. *Journal of Applied Probability*, 41A:35–54, 2004.
- [3] T. W. Anderson. *The statistical analysis of time series*. John Wiley & Sons Inc., New York, 1971.
- [4] D. R. Brillinger. An empirical investigation of the chandler wobble and two proposed excitation processes. *Bulletin of the International Statistical Institute*, 45(3):413–434, 1973.
- [5] D. R. Brillinger. *Time Series; Data Analysis and Theory*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. Reprint of the 1981 edition.
- [6] E. N. Brown. A note on the asymptotic distribution of the parameter estimates for the harmonic regression model. *Biometrika*, 77(3):653–656, 1990.
- [7] K. Choy and M. Taniguchi. Stochastic regression model with dependent disturbances. *J. Time Ser. Anal.*, 22(2):175–196, 2001.
- [8] R. Dahlhaus. Efficient location and regression estimation for long range dependent regression models. *Ann. Statist.*, 23(3):1029–1047, 1995.
- [9] U. Grenander and M. Rosenblatt. *Statistical Analysis of Stationary Time Series*. Wiley, New York, 1957.
- [10] U. Grenander. On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Ann. Math. Statistics*, 25:252–272, 1954.
- [11] R. S. Gross. The excitation of the chandler wobble. *Geophys. Res. Lett.*, 27(15):2329, 2000.
- [12] X. Guyon. *Random Fields on a Network Modeling, Statistics, and Applications*. Springer-Verlag, New York Berlin Heidelberg, 1995.
- [13] E. J. Hannan. *Multiple Time Series*. Springer-Verlag, New York, 1970.
- [14] E. J. Hannan. Non-linear time series regression. *Applied Probability Trust*, pages 767–780, 1971.
- [15] E. J. Hannan. The asymptotic theory of linear time series models. *J. Applied Probability*, 10:130–145, 1973.
- [16] C. C. Heyde. *Quasi-likelihood and its application*. Springer Series in Statistics. Springer-Verlag, New York, 1997. A general approach to optimal parameter estimation.
- [17] Y. Hosoya. A limit theory for long-range dependence and statistical inference on related models. *Ann. Statist.*, 25(1):105–137, 1997.
- [18] I. A. Ibragimov and Yu. A. Rozanov. *Gaussian random processes*. Springer-Verlag, New York, 1978. Translated from the Russian by A. B. Aries.
- [19] E. Iglói and Gy. Terdik. Bilinear modelling of chandler wobble. *Theory of Probability and its Applications*, 44(2):398–400, 1997.
- [20] E. Iglói and Gy. Terdik. Superposition of diffusions with linear generator and its multifractal limit process. *ESAIM Probab. Stat.*, 7:23–88 (electronic), 2003.
- [21] R. I. Jennrich. Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.*, 40:633–643, 1969.
- [22] C. Klüppelberg and T. Mikosch. The integrated periodogram for stable processes. *Ann. Statist.*, 24(5):1855–1879, 1996.
- [23] I. Lobato. Consistency of the averaged cross-periodogram in long memory series. *Journal of time series analysis*, 18(2):137–155, 1997.
- [24] I. N. Lobato. A semiparametric two-step estimator in a multivariate long memory model. *J. Econometrics*, 90(1):129–153, 1999.
- [25] J. R. Magnus and H. Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons Ltd., Chichester, 1999. Revised reprint of the 1988 original.
- [26] P. M. Robinson and F. J. Hidalgo. Time series regression with long-range dependence. *Ann. Statist.*, 25(1):77–104, 1997.

- [27] P. M. Robinson. Non-linear regression for multiple time-series. *J. Appl. Probability*, 9:758–768, 1972.
- [28] H. Schuh, T. Nagel, and T. Seitz. Linear drift and periodic variations observed in long time series of polar motion. *Journal of Geodesy*, 74:701–710, 2001.
- [29] Gy. Terdik. Higher order statistics and multivariate vector Hermite polynomials for nonlinear analysis of multidimensional time series. *Teor. Īmovĳr. Mat. Stat.*, 66:147–168, 2002.
- [30] J. Wang. Asymptotics of least-squares estimators for constrained nonlinear regression. *Ann. Statist.*, 24(3):1316–1326, 1996.
- [31] L. Wang. Asymptotics of estimates in constrained nonlinear regression with long-range dependent innovations. *Ann. Inst. Statist. Math.*, 56(2):251–264, 2004.
- [32] Chien-Fu Wu. Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.*, 9(3):501–513, 1981.
- [33] Y. Yajima. On estimation of a regression model with long-memory stationary errors. *Ann. Statist.*, 16(2):791–807, 1988.
- [34] Y. Yajima. Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Ann. Statist.*, 19(1):158–177, 1991.