Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 3793–3814
journal of statistical planning and inference
www.elsevier.com/locate/jspi

On multivariate nonlinear regression models with stationary correlated errors

Gy. Terdik[a], T. Subba Rao[b], S. Rao Jammalamadaka[c,*]

[a]*Department of Information Technology, Faculty of Informatics, University of Debrecen, 4010 Debrecen, Pf. 12, Hungary*
[b]*School of Mathematics, University of Manchester, P.O. Box 88, Manchester M60 1QD, UK*
[c]*Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, USA*

Available online 30 March 2007

Abstract

In this paper we consider the statistical analysis of multivariate multiple nonlinear regression models with correlated errors, using 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 the scaling factors are obtained explicitly for various types of nonlinear regression models including an interesting model which requires the estimation of unknown frequencies. The estimation of frequencies is a classical problem occurring in many areas like signal processing, environmental time series, astronomy and other areas of physical sciences. We illustrate our methodology using two real data sets taken from geophysics and environmental sciences. The data we consider from geophysics are 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 were first analyzed by Arato, Kolmogorov and Sinai who treat it as a bivariate time series satisfying a finite order time series model. They estimate the periodicities using the coefficients of the fitted models. Our analysis shows that the two dominant frequencies are 12 h and 410 days. The second example, we consider is the minimum/maximum monthly temperatures observed at the Antarctic Peninsula (Faraday/Vernadsky station). It is now widely believed that over the past 50 years there is a steady warming in this region, and if this is true, the warming has serious consequences on ecology, marine life, etc. as it can result in melting of ice shelves and glaciers. Our objective here is to estimate any existing temperature trend in the data, and we use the nonlinear regression methodology developed here to achieve that goal.

MSC: primary 62J02; 62M10; secondary 62H12

Keywords: Multivariate time series; Nonlinear regression estimation; Chandler Wobble; Global warming; Antarctic temperatures

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 $x_1, x_2, \ldots, x_p$ under suitable assumptions on the random errors. The usual assumption is that the errors are independent, identically distributed random variables. This has been 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. Asymptotic properties of the parameter estimates are now well known (see for example Jennrich, 1969).

* Corresponding author. Tel.: +1 805 893 3119; fax: +1 805 893 2334.
E-mail address: jammalam@pstat.ucsb.edu (S.R. Jammalamadaka).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2007.03.050

The results were later extended to the case of nonlinear multiple regression with correlated errors by Hannan (1971) when the stationary errors satisfy a certain linear representation. The methods used by Hannan are frequency domain oriented and depend on the properties of Finite Fourier Transforms (FFT). The methodology is dependent on making certain assumptions on regressor variables, now known as Grenander and Rosenblatt's conditions (see Grenander and Rosenblatt, 1957). The conditions depend on the nature of nonlinearity of the parameters in the regressor variables. The central limit theorems associated with the estimates obtained and the scaling factors required also depend on these functions. If the nonlinear parameters to be estimated are frequencies, even though the model looks linear, one has to impose a different set of 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 case, 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 for computational purposes. In this paper our objective is to consider both linear models and nonlinear models as well as a mixture of both. We state suitable scaling factors in each case for establishing asymptotic properties of the estimates. We refer to a recent paper by Subba Rao and Terdik (2006), for more supplementary details. A multivariate convolution model somewhat similar to our model was considered by Shumway et al. (1999). We propose a frequency domain approach which we believe is extremely useful in situations when the marginal distributions of the elements of the random vector have quite different distributions as in the case of minimum and maximum temperatures (which could have extreme value distributions with different sets of parameters) considered in this paper.

The minimum contrast estimates (Guyon, 1995; Heyde, 1997) of the unknown parameters are constructed in frequency domain. The mixed model containing the linear regression and linear combinations of the nonlinear regressions is also considered in detail.

To illustrate our methods we consider two important examples. One is the analysis of time series of Polar motion widely described in geophysics as the "Chandlers Wobble". One of the first papers dealing with the statistical analysis of these data was by Arato et al. (1962) who estimated the parameters of the polar motion, such as offset (trend), drift and periodicities. Using recent 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 h is also present. 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"—a children's toy. The second example we consider is the minimum and maximum monthly temperatures (and ozone levels) observed at a station in the Antarctic Peninsula and these data are considered to illustrate our methodology where the regression model is linear. For a different analysis of these temperature data, we refer to a recent paper by Hughes et al. (2006), referred to as HSS from now on.

1.1. The statistical model and assumptions

Consider a d-dimensional observational vector $\underline{Y}_t$, the random disturbances $\underline{Z}_t$ and the function $\underline{X}_t(\underline{\vartheta}_0)$ satisfying the usual model

$$\underline{Y}_t = \underline{X}_t(\underline{\vartheta}_0) + \underline{Z}_t. \tag{1}$$

The function $\underline{X}_t(\underline{\vartheta})$ can be a nonlinear function of both regressor variables and the p-dimensional parameter vector $\underline{\vartheta} \in \Theta \subset \mathbb{R}^p$, while $\underline{Z}_t$ is a d-dimensional stationary correlated time series. Here ϑ_0 is the true parameter to be estimated. The set Θ of admissible parameters $\underline{\vartheta}$ is defined by a number of possibly nonlinear equations, see Robinson (1972) for more details. We shall assume that the set Θ is chosen suitably in each case. For convenience we consider $\underline{X}_t(\underline{\vartheta})$ 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 $\underline{X}_t(\underline{\vartheta})$ can be written in the form

$$\underline{X}_t(\underline{\vartheta}) = \mathbf{B}_1\underline{X}_{1,t} + \mathbf{B}_2\underline{X}_{2,t}(\underline{\lambda}).$$

This can be considered as a mixed model since it is both linear as well as nonlinear in the parameters at the same time. The parameter vector $\underline{\vartheta}$ 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 $\underline{\lambda}$ are set of frequencies to be estimated we may have to put some constraints so that they lie within a compact set.

We assume that $\underline{Z}_t$ is a stationary linear processes, and has the moving average representation

$$\underline{Z}_t = \sum_{k=-\infty}^{\infty} \mathbf{A}_k \underline{W}_{t-k} \quad \text{where} \sum_{k=-\infty}^{\infty} \mathrm{Tr}(\mathbf{A}_k \mathbf{C}_W \mathbf{A}_k^*) < \infty.$$

Here we assume that $\underline{W}_t$ is a sequence of independent identically distributed random vectors (*i.i.d. vectors*) and $\mathbf{C}_W = \mathrm{Var}\, \underline{W}_t$, is non-singular We also assume that $\underline{Z}_t$ has a (element by element) piecewise *continuous* spectral density matrix $\mathbf{S}_Z(\omega)$, as in Hannan (1970) and Brillinger (2001). The model is *feedback free*, i.e. $\underline{Z}_t$ does not depend on $\underline{X}_t$.

1.2. The conditions on regressors and regression spectrum

Consider the regressor function $\underline{X}_t(\underline{\vartheta})$ 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 Grenander, 1954; Grenander and Rosenblatt, 1957) for the regressor $\underline{X}_t(\underline{\vartheta})$ are sufficient and in some situations (Wu, 1981) also necessary to establish the consistency of the least squares (LS) estimators. They are as follows. Let

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

denote the Euclidean norm of the kth regressor of the vector $X(\vartheta)$.

Condition 1 (*G1*). *For all* $k = 1, 2, \ldots, d$,

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

Condition 2 (*G2*). *For all* $k = 1, 2, \ldots, d$,

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

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

$$\begin{aligned}
\widehat{\mathbf{C}}_{\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2) &= \frac{1}{T} \sum_{t=1}^{T-h} \underline{X}_{t+h}(\underline{\vartheta}_1) \underline{X}_t^{\top}(\underline{\vartheta}_2), \\
\widehat{\mathbf{C}}_{\underline{X},T}(-h, \underline{\vartheta}_1, \underline{\vartheta}_2) &= \widehat{\mathbf{C}}_{\underline{X},T}^{\top}(h, \underline{\vartheta}_2, \underline{\vartheta}_1).
\end{aligned} \tag{2}$$

Whenever $\underline{\vartheta}_1 = \underline{\vartheta}_2 = \underline{\vartheta}$, throughout our paper we use the shorter notation $\widehat{\mathbf{C}}_{\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)|_{\underline{\vartheta}_1=\underline{\vartheta}_2=\underline{\vartheta}} = \widehat{\mathbf{C}}_{\underline{X},T}(h, \underline{\vartheta})$. The next condition we impose essentially means that the regressor $\underline{X}_t(\underline{\vartheta})$ is changing "slowly" in the following sense. For each integer h, $\|X_{k,t}(\underline{\vartheta})\|_{T+h}^2 \simeq T$.

Condition 3 (*G3*). *For each integer* h,

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

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

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

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

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 Brillinger, 2001 for details). We now introduce the discrete Fourier transform

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

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

$$\mathbf{I}_{\underline{X},T}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2) = \frac{1}{T} \underline{d}_{\underline{X},T}(\omega, \underline{\vartheta}_1) \underline{d}^*_{\underline{X},T}(\omega, \underline{\vartheta}_2),$$

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

$$\widehat{\mathbf{C}}_{\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2) = \int_{-1/2}^{1/2} \exp(\mathrm{i}\, 2\pi\lambda h) \mathbf{I}_{\underline{X},T}(\lambda, \underline{\vartheta}_1, \underline{\vartheta}_2)\, \mathrm{d}\lambda,$$
$$\mathbf{I}_{\underline{X},T}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2) = \sum_{|h|<T} \widehat{\mathbf{C}}_{\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)\, \mathrm{e}^{-\mathrm{i}\, 2\pi\omega h},$$
$$\mathbf{I}^{\top}_{\underline{X},T}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2) = \overline{\mathbf{I}_{\underline{X},T}(\omega, \underline{\vartheta}_2, \underline{\vartheta}_1)}$$

between $\widehat{\mathbf{C}}_{\underline{X},T}$ and the periodogram. Definition (2), which Jennrich (1969) calls tail product, reminds one 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.

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

For the univariate case, we refer to the classical books of Grenander and Rosenblatt (1957), Anderson (1971), and for vector-valued case, to Hannan (1970) and Brillinger (2001).

Now, define the empirical SDFR $\mathbf{F}_T$ as

$$\mathbf{F}_T(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2) = \int_0^{\omega} \mathbf{I}_{\underline{X},T}(\lambda, \underline{\vartheta}_1, \underline{\vartheta}_2)\, \mathrm{d}\lambda,$$

then it follows from the Grenander's conditions stated above that $\mathbf{F}$ is the weak limit of $\mathbf{F}_T$ and this is the condition we really need later. See also Ibragimov and Rozanov (1978, Chapter 7).

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*

$$\lim_{T\to\infty} \int_{-1/2}^{1/2} \varphi(\omega)\, \mathrm{d}\mathbf{F}_T(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2) = \int_{-1/2}^{1/2} \varphi(\omega)\, \mathrm{d}\mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2), \tag{3}$$

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, $\mathrm{d}\mathbf{F}(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2) = 0$. The matrix function F is Hermite symmetric metric since $\mathbf{F}_T$ satisfies following:

$$\mathbf{F}^{\top}_T(\omega, \underline{\vartheta}_1, \underline{\vartheta}_2) = \overline{\mathbf{F}_T(\omega, \underline{\vartheta}_2, \underline{\vartheta}_1)} = \mathbf{F}_T(-\omega, \underline{\vartheta}_2, \underline{\vartheta}_1).$$

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

1.3. The objective function

The frequency domain analysis has a number of advantages (see Brillinger, 2001). As we stated earlier our method depends on using FFT of model (1). The FFT is technically simple to use. We take the FFTs both sides of (1) at Finite Fourier frequencies $\omega_k = k/T \in [-\frac{1}{2}, \frac{1}{2}]$, $k = 0, \pm 1, \ldots, \pm T_1$, where $T_1 = \text{Int}[(T-1)/2]$, only). We obtain by taking FFT both sides

$$\underline{d}_{\underline{Y},T}(\omega) = \underline{d}_{\underline{X},T}(\omega, \underline{\vartheta}_0) + \underline{d}_{\underline{Z},T}(\omega).$$

The parameter vector $\underline{\vartheta}_0$ denotes the true unknown value and we would like to adjust the regressor $\underline{X}_t(\underline{\vartheta})$ in the model such that the distance

$$\underline{d}_{\underline{Y},T}(\omega) - \underline{d}_{\underline{X},T}(\omega, \underline{\vartheta}) = \underline{d}_{\underline{X},T}(\omega, \underline{\vartheta}_0) - \underline{d}_{\underline{X},T}(\omega, \underline{\vartheta}) + \underline{d}_{\underline{Z},T}(\omega) \tag{4}$$

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

$$\sum_{k=-T_1}^{T_1} \|\underline{d}_{\underline{Y},T}(\omega_k) - \underline{d}_{\underline{X},T}(\omega_k, \underline{\vartheta})\|^2 = \sum_{k=-T_1}^{T_1} \|\underline{d}_{\underline{X},T}(\omega_k, \underline{\vartheta}_0) - \underline{d}_{\underline{X},T}(\omega_k, \underline{\vartheta}) + \underline{d}_{\underline{Z},T}(\omega_k)\|^2,$$

which by Parseval's Theorem, actually corresponds to the sum of squares

$$\sum_{t=0}^{T-1} \|\underline{Y}_t - \underline{X}_t(\underline{\vartheta})\|^2 = \sum_{t=0}^{T-1} \|\underline{X}_t(\underline{\vartheta}_0) - \underline{X}_t(\underline{\vartheta}) + \underline{Z}_t\|^2.$$

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. sequence hence we are facing a generalized nonlinear regression problem with stationary residuals. The quadratic function we minimize, somewhat parallel to that suggested by Hannan (1971) for scalar valued case, is

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

where $\mathbf{\Phi}(\omega_k)$ is a series of matrix weights, originated from a continuous, Hermitian matrix function $\mathbf{\Phi}$, satisfying $\mathbf{\Phi}(\omega) \geqslant 0$. Eq. (4) can be rewritten as

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

The proof of $\mathbf{I}_{\underline{X},\underline{Z},T}(\omega_k, \underline{\vartheta}) \to 0$, a.s. and uniformly in $\underline{\vartheta}$ is given by Robinson (1972, Lemma 1). Now, suppose Conditions I-R, (or G1–G4) hold and we take the limit of the above to obtain

$$\begin{aligned}
Q(\underline{\vartheta}) &= \lim_{T\to\infty} Q_T(\underline{\vartheta}) \\
&= \int_{-1/2}^{1/2} \operatorname{Tr}(\mathbf{\Phi}(\omega)\, \mathrm{d}[\mathbf{F}(\omega, \underline{\vartheta}_0) + \mathbf{F}(\omega, \underline{\vartheta}) - \mathbf{F}(\omega, \underline{\vartheta}_0, \underline{\vartheta}) - \mathbf{F}(\omega, \underline{\vartheta}, \underline{\vartheta}_0)]) + \int_{-1/2}^{1/2} \operatorname{Tr}[\mathbf{S}_{\underline{Z}}(\omega)\mathbf{\Phi}(\omega)]\, \mathrm{d}\omega \\
&= R(\underline{\vartheta}, \underline{\vartheta}_0) + \int_{-1/2}^{1/2} \operatorname{Tr}[\mathbf{\Phi}(\omega)\mathbf{S}_{\underline{Z}}(\omega)]\, \mathrm{d}\omega.
\end{aligned} \tag{6}$$

The function

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

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 Robinson, 1972).

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*

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

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

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

One can easily see (using Magnus and Neudecker, 1999, Theorem 7, Chapter 7) under some additional assumptions given below, that $Q_T(\underline{\vartheta})$ is convex since the Hessian $\mathsf{H}Q(\underline{\vartheta}_0)$ is nonnegative definite. This shows that the next Theorem, due to Robinson (1972), is indeed valid not only for a compact set Θ, but also for more general classes of parameters, for example convex parameter sets 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 Anh et al. (2004).

Theorem 7. *Under the assumptions I-R* (*or G*1–*G*4), *and R*, *the minimum contrast estimator* $\widehat{\underline{\vartheta}}_T$ *converges a.s. to* $\underline{\vartheta}_0$.

2. Approximate sampling distributions of the estimates

For obtaining the sampling distribution it is necessary to consider the second derivatives of the SDFR and their limits for the objective function, see Robinson (1972). The matrix of the second derivatives of $\widehat{\mathbf{C}}_{\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)$ can be calculated, by using the matrix differential calculus (Magnus and Neudecker, 1999)

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

Here the differentiation of the RHS can be carried out directly. Notice, the order of the variables $\underline{\vartheta}_1, \underline{\vartheta}_2$ in $\widehat{\mathbf{C}}_{\underline{X},T}$, is opposite to the order of the partial derivatives: $\partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top$. The later means that one differentiates first by $\underline{\vartheta}_2$ and then by $\underline{\vartheta}_1$ which operates on the right-hand side. Starting the differentiating by $\underline{\vartheta}_1$, and then followed by $\underline{\vartheta}_2$, it can be written as 'direct' one

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

where we used the commutation matrix $\mathbf{K}_{p\cdot d}$, see Magnus and Neudecker (1999), $\otimes$ denotes the Kronecker product. Here and elsewhere $^\top$ stands for transpose. Here $\mathbf{U}_d$ is the $d \times d$ identity matrix. Following Hannan (1971) we assume the following condition:

Condition 8 (H). *All the second partial derivatives of the regressor* $\underline{X}_t(\underline{\vartheta})$ *exist and* $\partial^2 \widehat{\mathbf{C}}_{\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)/\partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top$ *converges to a limit*, *and let it be denoted by* $\partial^2 \mathbf{C}_{\underline{X}}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)/\partial \underline{\vartheta}_2^\top \partial \underline{\vartheta}_1^\top$.

It is necessary to emphasize that Condition H is

$$\frac{\partial^2 \mathbf{C}_{\underline{X}}(h,\underline{\vartheta}_1,\underline{\vartheta}_2)}{\partial\underline{\vartheta}_2^{\top}\partial\underline{\vartheta}_1^{\top}} \doteq \lim_{T\to\infty} \frac{\partial^2 \widehat{\mathbf{C}}_{\underline{X},T}(h,\underline{\vartheta}_1,\underline{\vartheta}_2)}{\partial\underline{\vartheta}_2^{\top}\partial\underline{\vartheta}_1^{\top}},$$

where the left-hand side is defined by the limit (which is the derivative of $\mathbf{C}_{\underline{X}}$). Here and elsewhere we use the symbol $\doteq$ 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* $\partial^2\mathbf{F}_T(\omega,\underline{\vartheta}_1,\underline{\vartheta}_2)/\partial\underline{\vartheta}_1^{\top}\partial\underline{\vartheta}_2^{\top}$, *of the matrix function* $\mathbf{F}_T$ *converges weakly to* $\partial^2\mathbf{F}(\omega,\underline{\vartheta}_1,\underline{\vartheta}_2)/\partial\underline{\vartheta}_2^{\top}\partial\underline{\vartheta}_1^{\top}$.

Again

$$\frac{\partial^2 \mathbf{F}(h,\underline{\vartheta}_1,\underline{\vartheta}_2)}{\partial\underline{\vartheta}_2^{\top}\partial\underline{\vartheta}_1^{\top}} \doteq \lim_{T\to\infty} \frac{\partial^2 \mathbf{F}_T(\omega,\underline{\vartheta}_1,\underline{\vartheta}_2)}{\partial\underline{\vartheta}_1^{\top}\partial\underline{\vartheta}_2^{\top}},$$

by definition. Using the above formulae for the derivatives we calculate the Hessian $\mathsf{H}\mathbf{F}$ for the SDFR $\mathbf{F}$ as well.

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

$$\mathsf{H}\mathbf{F}(\omega,\underline{\vartheta}) = \left.\left[\mathsf{H}_{\underline{\vartheta}_1}\mathbf{F}(\omega,\underline{\vartheta}_1,\underline{\vartheta}_2) + \frac{\partial^2\mathbf{F}(\omega,\underline{\vartheta}_1,\underline{\vartheta}_2)}{\partial\underline{\vartheta}_1^{\top}\partial\underline{\vartheta}_2^{\top}} + \mathsf{H}_{\underline{\vartheta}_2}\mathbf{F}(\omega,\underline{\vartheta}_1,\underline{\vartheta}_2) + \frac{\partial^2\mathbf{F}(\omega,\underline{\vartheta}_1,\underline{\vartheta}_2)}{\partial\underline{\vartheta}_2^{\top}\partial\underline{\vartheta}_1^{\top}}\right]\right|_{\underline{\vartheta}_1=\underline{\vartheta}_2=\underline{\vartheta}}, \tag{7}$$

where the indirect derivative satisfies

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

2.1. An expression for the asymptotic variance

For the variance of $\operatorname{Vec}\partial Q_T(\underline{\vartheta}_0)/\partial\underline{\vartheta}^{\top}$ consider the expression

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

Let $\mathbf{\Psi}$ be some matrix function of appropriate dimension, and introduce the following notation, which will be frequently used:

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

where $\mathbf{U}_p$ denotes the identity matrix of order p. The proof of the following lemma can be found in Subba Rao and Terdik (2006).

Lemma 11.

$$\lim_{T\to\infty}\operatorname{Var}\left[\sqrt{T}\operatorname{Vec}\frac{\partial Q_T(\underline{\vartheta}_0)}{\partial\underline{\vartheta}^{\top}}\right] = 4\mathbf{J}(\mathbf{\Phi}\mathbf{S}_{\underline{Z}}\mathbf{\Phi},\mathbf{F}).$$

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

Lemma 12.

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

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

$$\mathbf{J}_T=\operatorname{Var}\left[\sqrt{T}\operatorname{Vec}\frac{\partial Q_T(\underline{\vartheta}_0)}{\partial\underline{\vartheta}^{\mathsf{T}}}\right],$$

and suppose the following condition holds.

Condition 13 (*R*). *The limiting form of the variance matrix given by* $\mathbf{J}(\mathbf{\Phi}\mathbf{S}_{\underline{Z}}\mathbf{\Phi},\mathbf{F})=\operatorname{Var}[\sqrt{T}\operatorname{Vec}\frac{\partial Q_T(\underline{\vartheta}_0)}{\partial\underline{\vartheta}^{\mathsf{T}}}]$, *be positive definite, for all admissible spectral density functions.* $\mathbf{S}_{\underline{Z}}$ *and SDFR* $\mathbf{F}$. *Further suppose that* $\mathbf{J}(\mathbf{\Phi},\mathbf{F})>0$.

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

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

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

$$\lim_{T\to\infty}\operatorname{Var}[\sqrt{T}(\widehat{\underline{\vartheta}}_T-\underline{\vartheta}_0)]=\mathbf{J}^{-1}(\mathbf{\Phi},\mathbf{F})\mathbf{J}(\mathbf{\Phi}\mathbf{S}_{\underline{Z}}\mathbf{\Phi},\mathbf{F})\mathbf{J}^{-1}(\mathbf{\Phi},\mathbf{F})|_{\underline{\vartheta}=\underline{\vartheta}_0}. \tag{9}$$

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

$$\lim_{T\to\infty}\operatorname{Var}[\sqrt{T}(\widehat{\underline{\vartheta}}_T-\underline{\vartheta}_0)]=\mathbf{J}^{-1}(\mathbf{S}_{\underline{Z}}^{-1},\mathbf{F}), \tag{10}$$

where

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

3. Scaling

Now we consider various regression models and obtain scaling factors for each. As our first special case, consider the linear multivariate multiple regression model

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

In this case $\underline{\vartheta}=\operatorname{Vec}\mathbf{B}$, 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 $\|X_{k,t}\|_T\simeq D_k(T)$ which tends to infinity by Grenander's condition G1, then the

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

might not converge unless each $D_k(T) \simeq \sqrt{T}$. This condition is not satisfied for polynomial regression models. Grenander's condition can be interpreted in the following way. Define the diagonal matrix $\mathbf{D}_T = \operatorname{diag}(D_1, D_2, \ldots, D_p)$, where $D_k = D_k(T) \simeq \|X_{k,t}\|_T$. Now, consider the new scaled regression model

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

where $\underline{V}_t = \sqrt{T}\mathbf{D}_T^{-1}\underline{X}_t$. We note

$$\widetilde{\mathbf{B}}\underline{V}_t = \left(\frac{1}{\sqrt{T}}\mathbf{B}\mathbf{D}_T\right)(\sqrt{T}\mathbf{D}_T^{-1}\underline{X}_t).$$

Therefore the asymptotic variance of the estimate of the parameter matrix $\mathbf{B}$ can now be written as

$$\lim_{T\to\infty} \operatorname{Var}\sqrt{T}(\widehat{\widetilde{\mathbf{B}}} - \widetilde{\mathbf{B}}_0) = \lim_{T\to\infty} \operatorname{Var}[(\widehat{\mathbf{B}} - \mathbf{B}_0)\mathbf{D}_T].$$

We call this type of transformation as "primary" scaling. We observe that the procedure of scaling opens the possibility of considering random regressors which are not necessarily weakly stationary because their second-order moments may not exist (see Klüppelberg and Mikosch, 1996) 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 $T \to \infty$, for each $k = 1, 2, \ldots, d$.

In general, let $D_k(T) \simeq \|X_{k,t}\|_T$, for each k and define $\mathbf{D}_T = \operatorname{diag}(D_1, D_2, \ldots, D_d)$, 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 $\underline{X}_t(\underline{\vartheta})$ is possible if $D_k(T)$ does not depend on the unknown parameter $\underline{\vartheta}$. Even if the regressors $\underline{X}_t(\underline{\vartheta})$ are properly scaled, we may have a new set of problems 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 $\underline{X}_t(\underline{\vartheta})$.

Consider first, the diagonal matrix $\mathbf{D}_T = \operatorname{diag}(D_{X,k}(T), k = 1, 2, \ldots, d)$ and apply the scaling which results in $\sqrt{T}\mathbf{D}_T^{-1}\underline{X}_t(\underline{\vartheta})$. Another type of scaling can be obtained by the process of differentiation. We define the *scaled* partial derivative $\partial_{s,T}(\underline{\vartheta})$ of the diagonal matrix $\mathbf{D}_{1,T} = \operatorname{diag}(D_k^{(1)}(T), k = 1, 2, \ldots, p)$ by $\partial(\mathbf{D}_{1,T}^{-1}\underline{\vartheta})$, resulting

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

which gives

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

Notice, the entries of the scaled partial derivatives are $[D_{X,j}(T)D_k^{(1)}(T)]^{-1}\partial X_{j,t}(\underline{\vartheta})/\partial\vartheta_k$. The scaled second derivatives of $\widehat{\mathbf{C}}_{\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)$ are given by

$$\begin{aligned}\frac{\partial_{s,T}^2 \widehat{\mathbf{C}}_{\mathbf{D}_T\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial_{s,T}\underline{\vartheta}_2^\top \partial_{s,T}\underline{\vartheta}_1^\top} &= (\mathbf{D}_{1,T}^{-1} \otimes \mathbf{U}_{d^2})\frac{\partial^2 \widehat{\mathbf{C}}_{\sqrt{T}\mathbf{D}_T^{-1}\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial\underline{\vartheta}_2^\top \partial\underline{\vartheta}_1^\top}\mathbf{D}_{1,T}^{-1} \\ &= T(\mathbf{D}_{1,T}^{-1} \otimes \mathbf{D}_T^{-1} \otimes \mathbf{D}_T^{-1})\frac{\partial^2 \widehat{\mathbf{C}}_{\underline{X},T}(h, \underline{\vartheta}_1, \underline{\vartheta}_2)}{\partial\underline{\vartheta}_2^\top \partial\underline{\vartheta}_1^\top}\mathbf{D}_{1,T}^{-1}.\end{aligned}$$

Note that $1/T$ in the expression of $\widehat{\mathbf{C}}_{\underline{X},T}$ is canceled and the scaling factor has been absorbed completely by the scaling matrices.

Condition 16 (H'). *All the second partial derivatives of the regressor vector $\underline{X}_t(\underline{\vartheta})$ exist. There exist diagonal matrices $\mathbf{D}_T$ and $\mathbf{D}_{1,T}$ such that, uniformly in $\underline{\vartheta}$, the scaled derivative $\partial^2_{s,T}\widehat{\mathbf{C}}_{\sqrt{T}\mathbf{D}_T^{-1}\underline{X},T}(h,\underline{\vartheta}_1,\underline{\vartheta}_2)/\partial_{s,T}\underline{\vartheta}_2^\top\partial_{s,T}\underline{\vartheta}_1^\top$ converges to a limit, which we denote it by $\partial^2_s\mathbf{C}_{\underline{X}}(h,\underline{\vartheta}_1,\underline{\vartheta}_2)/\partial_s\underline{\vartheta}_1\partial_s\underline{\vartheta}_2$.*

The above condition H implies that

$$\frac{\partial^2_s\mathbf{C}_{\underline{X}}(h,\underline{\vartheta}_1,\underline{\vartheta}_2)}{\partial_s\underline{\vartheta}_2^\top\partial_s\underline{\vartheta}_1^\top} \doteq \lim_{T\to\infty}\frac{\partial^2_{s,T}\widehat{\mathbf{C}}_{\mathbf{D}_T\underline{X},T}(h,\underline{\vartheta}_1,\underline{\vartheta}_2)}{\partial_{s,T}\underline{\vartheta}_2^\top\partial_{s,T}\underline{\vartheta}_1^\top}.$$

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

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

then $\mathbf{D}_T = \operatorname{diag}(B_{X,j}(T), k=1,2,\ldots,d)$, and $\mathbf{D}_{1,T} = \operatorname{diag}(B_k^{(1)}(T), k=1,2,\ldots,p)$, say. Note here that the matrix $\mathbf{D}_T$ contains the factors of primary scaling. There may be regressors $\underline{X}_t(\underline{\vartheta})$, which may require other forms of scaling.

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

Condition 17 (*I-R-H'*). *The scaled derivative $\partial^2_{s,T}\mathbf{F}_T(\omega,\underline{\vartheta}_1,\underline{\vartheta}_2)/\partial_{s,T}\underline{\vartheta}_1^\top\partial_{s,T}\underline{\vartheta}_2^\top$, of matrix function $\mathbf{F}_T$ converges weakly to a function which we denote by $\partial^2_s\mathbf{F}(h,\underline{\vartheta}_1,\underline{\vartheta}_2)/\partial_s\underline{\vartheta}_2^\top\partial_s\underline{\vartheta}_1^\top$.*

Define

$$\mathbf{J}_T(\mathbf{D}_T,\mathbf{\Psi},\mathbf{F}) = \int_{-1/2}^{1/2}(\mathbf{U}_p\otimes[\operatorname{Vec}(\mathbf{D}_T\mathbf{\Psi}^\top(\omega_k)\mathbf{D}_T)]^\top)\,\mathrm{d}\left(\left.\frac{\partial^2_s\mathbf{F}(\omega,\underline{\vartheta}_1,\underline{\vartheta}_2)}{\partial_s\underline{\vartheta}_2^\top\partial_s\underline{\vartheta}_1^\top}\right|_{\underline{\vartheta}_1=\underline{\vartheta}_2=\underline{\vartheta}_0}\right).$$

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

$$\sqrt{T}\mathbf{J}_T^{-1/2}\mathsf{H}_{s,T}Q_T(\underline{\vartheta}_0)\mathbf{D}_{1,T}(\widehat{\underline{\vartheta}}_T-\underline{\vartheta}_0)\xrightarrow{\mathscr{D}}N(0,\mathbf{U}_p).$$

In other words the variance of $(\widehat{\underline{\vartheta}}_T-\underline{\vartheta}_0)$ can be approximated by

$$\mathbf{D}_{1,T}^{-1}\mathbf{J}_T^{-1}(\mathbf{D}_T,\mathbf{\Phi},\mathbf{F})\mathbf{J}_T(\mathbf{D}_T,\mathbf{\Phi}\mathbf{S}_{\underline{Z}}\mathbf{\Phi},\mathbf{F})\mathbf{J}_T^{-1}(\mathbf{D}_T,\mathbf{\Phi},\mathbf{F})\mathbf{D}_{1,T}^{-1}|_{\underline{\vartheta}=\underline{\vartheta}_0}.$$

Moreover if $\mathbf{\Phi}=\mathbf{S}_{\underline{Z}}^{-1}$, one obtains the asymptotic variance in the concise form

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

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

Remark 19. The spectrum $\mathbf{S}_{\underline{Z}}$ in general is not known, in which case it becomes 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 (1971) apply and both the estimator of the unknown parameter $\underline{\vartheta}$ and the estimator for the parameters of the spectrum are consistent.

4. Examples of regression models and their estimation

We now consider various interesting special models and discuss the asymptotic properties of the estimates.

4.1. Multiple linear regression model with stationary errors

Consider once again the linear model

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

Taking FFT both sides we obtain

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

Now the periodogram matrices we defined earlier for the above will now reduce to

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

The normal equations are obtained by solving

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

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

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

In cases when the inverse does not exist, one can use the generalized inverse. This estimate is *linear and unbiased* since

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

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

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

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

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

We caution that we use the superscript $\top$ for transpose only, and not the sample size T which we use as a subscript. In particular we have the expressions in both situations when the estimates are the ordinary LS estimates (when the errors are independent) and also in the case of weighted least squares LS estimates. In the first case $\mathbf{\Phi}(\omega) = \mathbf{U}_d$ and in the later case $\mathbf{\Phi}(\omega) = \mathbf{S}_{\underline{Z}}^{-1}(\omega)$ which leads to the best linear unbiased estimates, the BLUEs. Grenander (1954) has shown that under certain assumptions asymptotically the LS and BLUE are equivalent. When such conditions are satisfied, we have $\mathrm{Var}[\sqrt{T}\,\mathrm{Vec}(\widehat{\mathbf{B}})] = [\int_{-1/2}^{1/2} \mathrm{d}\mathbf{F}^\top(\omega) \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega)]^{-1}$. This limit does not depend on $\mathbf{B}_0$. This result can also be obtained from the general expression (10) for the variance.

Remark 20. The estimation of the transpose of the matrix $\mathbf{B}$ is used more often in time domain (see Hannan, 1970). Then the variance of Vec $(\widehat{\mathbf{B}}^\top)$ follows from (12) easily

$$\begin{aligned}\operatorname{Var}[\sqrt{T}\operatorname{Vec}(\widehat{\mathbf{B}}^\top)] &= \operatorname{Var}[\mathbf{K}_{p\cdot d}\operatorname{Vec}(\widehat{\mathbf{B}})]\\ &= \mathbf{K}_{p\cdot d}\operatorname{Var}\operatorname{Vec}(\widehat{\mathbf{B}})\mathbf{K}_{d\cdot p},\end{aligned}$$

hence

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

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 $(1/\sqrt{T})\mathbf{B}\mathbf{D}_T$, writing

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

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

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

(see also Hannan, 1970, 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 $\mathbf{D}_T=\operatorname{diag}(T_1,T_2,\dots T_d)$. In this case the SDFR $\mathbf{F}$ is concentrated at zero with values $\mathrm{d}\mathbf{F}_{j,k}(0)=\sqrt{(2k-1)(2j-1)}/(k+j-1)$, so the asymptotic variance of Vec $\widehat{\mathbf{B}}$ is

$$[\mathbf{D}_T^{-1}\,\mathrm{d}\mathbf{F}^{-1}(0)\mathbf{D}_T^{-1}]\otimes\mathbf{S}_{\underline{Z}}(0), \tag{13}$$

(see Grenander and Rosenblatt, 1957, p. 247, for scalar valued case).

4.2. A mixed model involving parameters both linearly and nonlinearly

An interesting model which one comes across very often, is when the regression function is linear and also contains a nonlinear part of the type

$$\underline{Y}_t=\underline{X}_t(\underline{\vartheta}_0)+\underline{Z}_t,$$

where

$$\begin{aligned}\underline{X}_t(\underline{\vartheta}) &= \mathbf{B}_1\underline{X}_{1,t}+\mathbf{B}_2\underline{X}_{2,t}(\underline{\lambda})\\ &= [\mathbf{B}_1,\mathbf{B}_2]\begin{bmatrix}\underline{X}_{1,t}\\ \underline{X}_{2,t}(\underline{\lambda})\end{bmatrix}\\ &= \mathbf{B}\underline{X}_{3,t}(\underline{\lambda}).\end{aligned} \tag{14}$$

Here the unknown parameter $\underline{\vartheta}=\operatorname{Vec}(\operatorname{Vec}\mathbf{B}_1,\ \operatorname{Vec}\mathbf{B}_2,\underline{\lambda})$, where $\mathbf{B}_1$ 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, $\mathbf{B}=[\mathbf{B}_1,\mathbf{B}_2]$ and $\underline{X}_{3,t}(\underline{\lambda})=\begin{bmatrix}\underline{X}_{1,t}\\ \underline{X}_{2,t}(\underline{\lambda})\end{bmatrix}$. First we consider the problem of estimation, by minimizing the objective function

$$\begin{aligned}Q_T(\mathbf{B},\underline{\lambda}) = \frac{1}{T}\sum_{k=-T_1}^{T_1}&[\operatorname{Tr}(\mathbf{I}_{\underline{Y},T}(\omega_k)\mathbf{\Phi}(\omega_k))+\operatorname{Tr}(\mathbf{B}\mathbf{I}_{\underline{X}_3,T}(\omega_k,\underline{\lambda})\mathbf{B}^\top\mathbf{\Phi}(\omega_k))\\ &-\operatorname{Tr}(\mathbf{I}_{\underline{Y},\underline{X}_3,T}(\omega_k,\underline{\lambda})\mathbf{B}^\top\mathbf{\Phi}(\omega_k))-\operatorname{Tr}(\mathbf{B}\mathbf{I}_{\underline{X}_3,\underline{Y},T}(\omega_k,\underline{\lambda})\mathbf{\Phi}(\omega_k))].\end{aligned} \tag{15}$$

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

$$\frac{\partial Q_T(\mathbf{B}, \underline{\lambda})}{\partial \mathbf{B}} = \mathbf{0},$$
$$\operatorname{Vec}\frac{\partial Q_T(\underline{\lambda})}{\partial \underline{\lambda}^{\mathsf{T}}} = \underline{0}.$$

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 $\underline{\lambda} = \widetilde{\underline{\lambda}}$ (chosen), the normal equations result in

$$\operatorname{Vec}(\widehat{\mathbf{B}}) = \left(\sum_{k=-T_1}^{T_1} \mathbf{I}^{\mathsf{T}}_{\underline{X}_3,T}(\omega_k, \widetilde{\underline{\lambda}}) \otimes \mathbf{\Phi}(\omega_k) \right)^{-1} \operatorname{Vec} \sum_{k=-T_1}^{T_1} \mathbf{\Phi}(\omega_k) \mathbf{I}_{\underline{Y},\underline{X}_3,T}(\omega_k, \widetilde{\underline{\lambda}}).$$

Now, to obtain the estimates for $\underline{\lambda}$, we keep $\mathbf{B} = \widehat{\mathbf{B}}$ fixed and then minimize (15), i.e. find the solution to the equation

$$\sum_{k=-T_1}^{T_1} \left[\frac{\partial \mathbf{B} \mathbf{I}_{\underline{X}_3,T}(\omega_k, \underline{\lambda}) \mathbf{B}^{\mathsf{T}}}{\partial \underline{\lambda}^{\mathsf{T}}} - \frac{\partial \mathbf{I}_{\underline{Y},\underline{X}_3,T}(\omega_k, \underline{\lambda}) \mathbf{B}^{\mathsf{T}}}{\partial \underline{\lambda}^{\mathsf{T}}} - \frac{\partial \mathbf{B} \mathbf{I}_{\underline{X}_3,\underline{Y},T}(\omega_k, \underline{\lambda})}{\partial \underline{\lambda}^{\mathsf{T}}} \right]^{\mathsf{T}}_{\underline{\lambda} = \widehat{\underline{\lambda}}} \operatorname{Vec} \mathbf{\Phi}^{\mathsf{T}}(\omega_k) = \underline{0}.$$

The primary scaling of $\underline{X}_{3,t}(\underline{\lambda}) = \begin{bmatrix} X_{1,t} \\ \underline{X}_{2,t}(\underline{\lambda}) \end{bmatrix}$, is given by $\mathbf{D}_T = \operatorname{diag}(\mathbf{D}_{X_1,T}, \mathbf{D}_{X_2,T})$ where $\mathbf{D}_{X_1,T} = \operatorname{diag}(D_{X_1,k}(T), k = 1, 2, \ldots, p)$, and $\mathbf{D}_{X_2,T} = \operatorname{diag}(D_{X_2,k}(T), k = 1, 2, \ldots, q)$. The secondary scaling of the regressors are $\mathbf{D}_{1,T} = \operatorname{diag}(\mathbf{U}_{\mathrm{d}p+\mathrm{d}q}, \mathbf{D}_{3,T})$. Denoting the limit of the variance–covariance of the derivatives

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

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 $\mathbf{\Phi} = \mathbf{S}_{\underline{Z}}^{-1}$). The part that is linear in parameters results in

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

In the mixed model context we get

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

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

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

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

Finally the variance matrix of the estimates $\mathrm{Vec}(\mathrm{Vec}\,\widehat{\mathbf{B}}_1, \mathrm{Vec}\,\widehat{\mathbf{B}}_2, \widehat{\underline{\lambda}})$ is

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

4.3. Some special examples of interest: a model with linear trend and with harmonic components

Here we consider a special case of the mixed model mentioned above. We give some detailed calculations to illustrate the methodology. As this model will be used later to demonstrate our methods on real data (here $d=2$). Let

$$\underline{Y}_t = \underline{X}_t(\underline{\vartheta}_0) + \underline{Z}_t,$$

where

$$\underline{X}_t(\underline{\vartheta}) = \mathbf{B}\begin{bmatrix} 1 \\ t \end{bmatrix} + \mathbf{A}\begin{bmatrix} \cos(2\pi t\lambda_1) \\ \sin(2\pi t\lambda_1) \\ \cos(2\pi t\lambda_2) \\ \sin(2\pi t\lambda_2) \end{bmatrix}. \tag{16}$$

The parameter is $\underline{\vartheta}^{\top} = ([\mathrm{Vec}\,\mathbf{B}_1]^{\top}, [\mathrm{Vec}\,\mathbf{B}_2]^{\top}, [\lambda_1,\lambda_2])$, $|\lambda_i| \leqslant \pi$, $\lambda_1 \neq \lambda_2$, $\lambda_i \neq 0, \pm 1/2$. 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 $\mathbf{D}_{\underline{X}_1,T} = \mathrm{diag}(T^{1/2}, T^{3/2}/\sqrt{3})$, and for $\underline{X}_{2,t}(\underline{\lambda})$ is $\mathbf{D}_{\underline{X}_2,T} = T^{1/2}/\sqrt{2}\mathbf{U}_4$, since $\underline{X}_{1,t} = [1,t]^{\top}$ and $\underline{X}_{2,t}(\underline{\lambda}) = [\cos(2\pi t\lambda_1), \sin(2\pi t\lambda_1), \cos 2\pi t\lambda_2, \sin(2\pi t\lambda_2)]$. 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 $\underline{\lambda}$ 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 $[\underline{X}_{1,t}^{\top}, \underline{X}_{2,t}^{\top}(\underline{\lambda})]^{\top}$ is $\mathbf{D}_T = \mathrm{diag}(\mathbf{D}_{\underline{X}_1,T}, \mathbf{D}_{\underline{X}_2,T})$, and $\mathbf{D}_{1,T} = \mathrm{diag}(\mathbf{U}_{12}, \mathbf{D}_{3,T})$. 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 $\widehat{\mathbf{C}}_{\underline{X},T}$ 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

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

Using obvious notation, we partition the second derivative of SDFR according to the parameters

$$\frac{\partial^2 \mathbf{F}(\omega, \underline{\lambda}_1, \underline{\lambda}_2)}{\partial \underline{\lambda}_2^{\top} \partial \underline{\lambda}_1^{\top}} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} & \mathbf{F}_{1\lambda} \\ \mathbf{F}_{21} & \mathbf{F}_{22} & \mathbf{F}_{2\lambda} \\ \mathbf{F}_{\lambda 1} & \mathbf{F}_{\lambda 2} & \mathbf{F}_{\lambda\lambda} \end{bmatrix}.$$

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

(1) The regression spectrum of the linear part

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

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

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

(2) It is seen here that the mixed model parameters has no effect, $\mathbf{F}_{12}(\omega, \underline{\lambda}_0) = \mathbf{0}$, $\Sigma_{12} = \mathbf{0}$, and $\mathbf{F}_{1\lambda}(\omega, \underline{\lambda}_0) = \mathbf{0}$, $\Sigma_{1\lambda} = \mathbf{0}$.

(3) The $\mathbf{F}_{22}(\omega, \underline{\lambda}_0)$ 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

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

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

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

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

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

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

$$\begin{aligned} \Sigma_{22} &= \int_{-1/2}^{1/2} (\mathbf{D}_{X_2, T} \mathrm{d}\mathbf{F}_{22}^{\top}(\omega, \underline{\lambda}_0) \mathbf{D}_{X_2, T}) \otimes \mathbf{S}_{\underline{Z}}^{-1}(\omega) \\ &= T/2 \begin{bmatrix} \begin{bmatrix} \mathrm{Re}\, \mathbf{S}_{\underline{Z}}^{-1}(\lambda_1) & \mathrm{Im}\, \mathbf{S}_{\underline{Z}}^{-1}(\lambda_1) \\ -\mathrm{Im}\, \mathbf{S}_{\underline{Z}}^{-1}(\lambda_1) & \mathrm{Re}\, \mathbf{S}_{\underline{Z}}^{-1}(\lambda_1) \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \begin{bmatrix} \mathrm{Re}\, \mathbf{S}_{\underline{Z}}^{-1}(\lambda_2) & \mathrm{Im}\, \mathbf{S}_{\underline{Z}}^{-1}(\lambda_2) \\ -\mathrm{Im}\, \mathbf{S}_{\underline{Z}}^{-1}(\lambda_2) & \mathrm{Re}\, \mathbf{S}_{\underline{Z}}^{-1}(\lambda_2) \end{bmatrix} \end{bmatrix}. \end{aligned}$$

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

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

then we have

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

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

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

Suppose we have now three frequencies $\underline{\lambda} = [\lambda_1, \lambda_2, \lambda_3]$ then 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 jth entry in the diagonal which is 1.

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

where

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

Applying the general formula for $\Sigma_{2\lambda}$, see Section 4.2,

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

Put

$$\Gamma_2 = \begin{bmatrix} i & -1 \\ 1 & i\end{bmatrix},$$
$$\Lambda_2(\omega) = \mathbf{U}_4\otimes\mathbf{S}_{\underline{Z}}^{-1}(\omega)\mathbf{A}_0,$$

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

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

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

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

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

where

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

The SDFR is given by

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

where

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

The corresponding variance matrix is

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

For computational purposes, set

$$\begin{aligned} &\Gamma_3 = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}, \\ &\Lambda(\omega) = \mathbf{U}_2 \otimes [\operatorname{Vec}(\mathbf{A}^\top [\mathbf{S}_{\underline{Z}}^{-1}(\omega)]^\top \mathbf{A})]^\top, \end{aligned}$$

then the variance matrix has the form

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

After further simplification

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

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

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

Now we return to the asymptotic variance matrix of the estimated parameters. Collecting 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 $\mathbf{D}_{1,T} = \operatorname{diag}(\mathbf{U}_{12}, \mathbf{D}_{2,T})$. The variance matrix of the coefficient $\widehat{\mathbf{A}}$

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

and of $\widehat{\underline{\lambda}}$

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

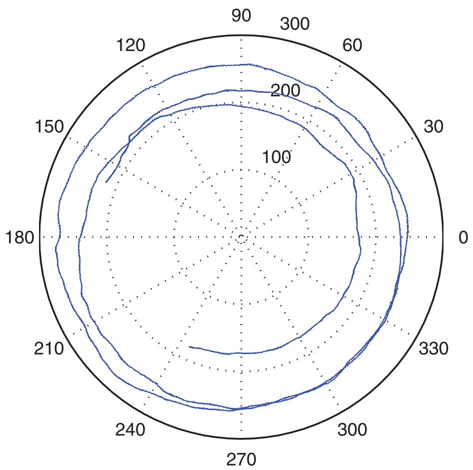
where $\mathbf{D}_{3,T} = 2\pi T/\sqrt{3}\mathbf{U}_{12}$, and Σ' denotes the covariance matrix without scaling. The speed of convergence of the variance matrix of the coefficient $\widehat{\mathbf{A}}$ is $T/2$ and the frequency $\widehat{\underline{\lambda}}$ is $(2\pi)^2T^3/6$.

5. Applications to real data analysis: I. Chandler Wobble

5.1. 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. It has been estimated by several researchers using different models and methods (see, for instance Brillinger, 1973; Arató et al., 1962) and the period of this wobble is known to be approximately between 430 and 435 days. Some properties of the monthly data have been described in Iglói and Terdik (1997).

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



Centralized wobbling motion in polar coordinates

Rotational variations of polar motion are due to the superposition of the influences of six partial tides. Different techniques suggest that these are real oscillations of polar motion. Rapid oscillations with periods of 12 h has already been considered, see IVS 2004 General Metting Proceedings (2005), Gross (2000).

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

One model that has been considered (see Schuh et al., 2001), has both a linear trend term as well as some harmonic components (usually termed as drift, offset and elliptical periodic motions parameters),

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

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, $(|\lambda_i| \leqslant \pi)$ are unknown and are to be estimated. This model is a special case of the nonlinear model we considered in this paper. We start 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

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

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

and

$$\widehat{\underline{\lambda}} = \begin{bmatrix}0.0001\\ 0.0833\end{bmatrix}.$$

The estimated frequencies correspond to the periods 410.5626 days and 11.9999 h which are close to the estimates obtained by many geophysicists.

5.2. II. Minimum and maximum monthly temperatures in the Antarctic Peninsula

It is now widely believed that there is a Global warming, and especially a major warming has been happening over the last 50 years in Antarctica. Though there is an increase in temperature at a majority of the stations in the Antarctica, Turner et al. (2005), have found a very significant increase in the monthly average surface temperatures at Faraday/Vernadsky station in the Antarctic Peninsula. Using linear LS methodology (under Gaussianity assumption) they conclude that the increase is of the order of 0.56 °C per decade historically and 1.09 °C per decade over the winter during the last 50 years. There is evidence of strong temporal dependence in the observations and also the data are not Gaussian. In view of this, it is obvious that their methodology does not lead to good estimates. Recently Hughes et al. (2006), have considered the monthly minimum/maximum temperatures observed during January 1951–December 2004 and they have analyzed the data using linear time series models where the innovations have extreme value distributions. They conclude that there is a significant increase in minimum temperatures, but the increase is not statistically significant in the time-series on maximum temperatures. This conclusion is very important, as any increase in temperatures in winter months can result in melting of ice shelves, glaciers and increase in sea levels. They have analyzed the two data sets separately, even though there is dependence between the minimum and maximum temperatures in a given month. This suggests that we should perform a joint analysis using multivariate multiple regression methodology. However, we note that minimum and maximum have two distinct distributions, and as such, the classical time domain methodology may not be that appropriate (Figs. 1 and 2). Therefore the methodology we developed in this paper is very appropriate and we use the model and analysis suggested in Section 4.3. Since it is widely believed that the ozone levels in the Antarctica have an affect on temperatures (see Hughes et al., 2006), we include this as our third explanatory variable (Fig. 3). The data we analyze here are monthly minimum, maximum temperatures and also ozone levels observed at Faraday/Vernadsky station during the period Jan 1958–Dec 2004. The ozone levels are measured in Dobson units. The plots of the data are given below.

The model we fitted is a linear trend model for the three variables with three known periodicities, and they are 12 months, 6 months and 4 months). We found these periods using the classical periodogram methodology. In the present

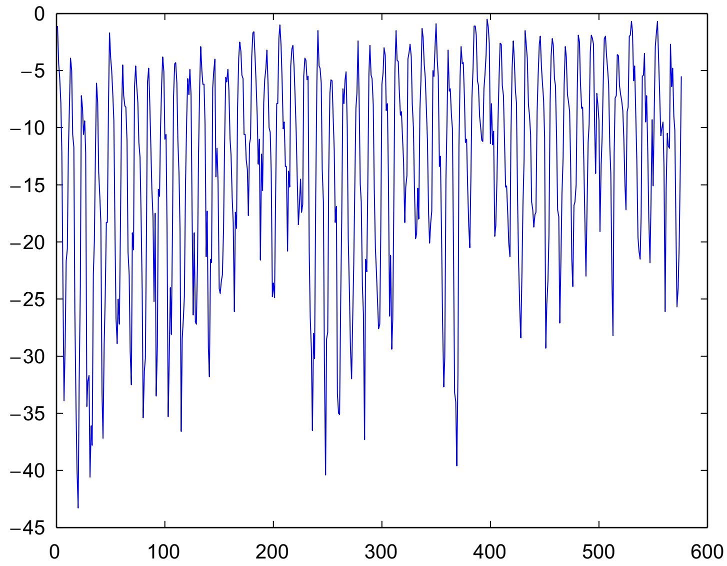


Fig. 1. Monthly minimum temperatures.

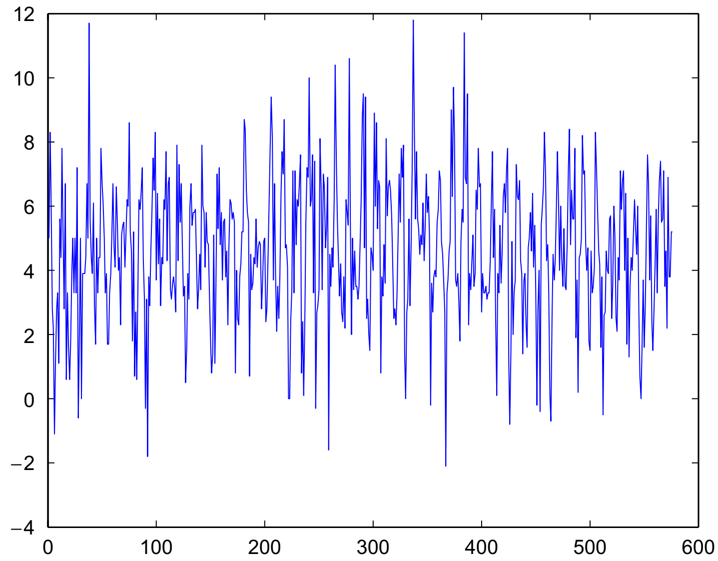


Fig. 2. Monthly maximum temperatures.

context, the matrix **B** is of order 3×2, the matrix **A** is of order 3×6, (16). The estimates of these coefficients and their standard errors (in brackets) calculated using the formula (13), are given in Table 1.

From these results, we see that the increase in minimum temperatures (trend coefficient $= 0.0139$) is highly significant, and also the decrease in ozone levels, but there is no significant increase in the maximum temperatures. The results are somewhat similar to those reported by Hughes et al. (2006), where they assumed that the errors are correlated and Gaussian. However, there is some difference in the estimates (and their standard errors) when they fitted a linear model with converse extreme value distribution. The estimates obtained by them using maximum likelihood methodology, have smaller standard errors, as one would expect. By comparing the results obtained here with those of

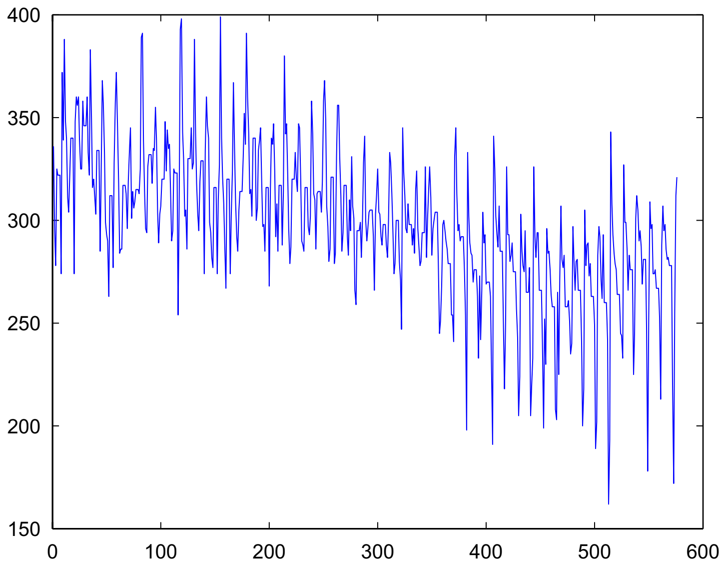


Fig. 3. Monthly ozone levels.

Table 1
Estimates of the parameters

	Minimum	Maximum	Ozone levels
Constant	−16.8014	4.5228	342.6448
Linear trend	0.0139 (0.0051)	0.0006 (0.0009)	−0.158 (0.0201)
$\sin(2\pi t/12)$	9.1013 (0.4512)	1.3126 (0.1325)	−0.2592 (0.1618)
$\cos(2\pi t/12)$	5.0523 (0.4512)	1.5558 (0.1325)	9.4127 (0.1618)
$\sin(2\pi t/6)$	0.0399 (0.1072)	0.2334 (0.0735)	−5.1302 (0.0913)
$\cos(2\pi t/6)$	1.87 (0.1072)	−0.22 (0.0735)	15.7086 (0.0913)
$\sin(2\pi t/4)$	0.0401 (3.9718)	−0.0747 (1.9769)	−5.8617 (1.5507)
$\cos(2\pi t/4)$	0.3352 (3.9718)	0.2137 (1.9769)	2.1319 (1.5507)

Hughes et al. (2006), we observe that the estimates obtained here are very close to Gaussian estimates, although our estimates are obtained without any restrictive assumption on the errors.

References

Anderson, T.W., 1971. The Statistical Analysis of Time Series. Wiley, New York.

Anh, V.V., Leonenko, N.N., Sakhno, L.M., 2004. Quasilikelihood-based higher-order spectral estimation of random processes and fields with possible long-range dependence. J. Appl. Probab. 41A, 35–54.

Arató, M., Kolmogorov, A.N., Sinay, J.G., 1962. Evaluation of the parameters of a complex stationary Gauss–Markov process. Dokl. Akad. Nauk SSSR 146, 747–750.

Brillinger, D.R., 1973. An empirical investigation of the Chandler Wobble and two proposed excitation processes. Bull. Int. Statist. Inst. 45 (3), 413–434.

Brillinger, D.R., 2001. Time Series; Data Analysis and Theory. Society for Industrial and Applied Mathematics (SIAM). Philadelphia, PA. (Reprint of the 1981 edition).

Grenander, U., 1954. On the estimation of regression coefficients in the case of an auto-correlated disturbance. Ann. Math. Statist. 25, 252–272.

Grenander, U., Rosenblatt, M., 1957. Statistical Analysis of Stationary Time Series. Wiley, New York.

Gross, R.S., 2000. The excitation of the Chandler Wobble. Geophys. Res. Lett. 27 (15), 2329.

Guyon, X., 1995. Random Fields on a Network Modeling, Statistics, and Applications. Springer, New York, Berlin, Heidelberg.

Hannan, E.J., 1970. Multiple Time Series. Springer, New York.

Hannan, E.J., 1971. Non-linear time series regression. Appl. Probab. Ser. Trust, 767–780.

Heyde, C.C., 1997. Quasi-likelihood and its application: A general approach to optimal parameter estimation. Springer Series in Statistics. Springer, New York.

Hughes, G.L., Subba Rao, S., Subba Rao, T., 2006. Statistical analysis and time series models for minimum/maximum temperatures in the Antarctic Peninsula. Proc. Roy. Soc. London Ser. A.

Ibragimov, I.A., Rozanov, Yu.A., 1978. Gaussian Random Processes. Springer, New York. (Translated from Russian by A.B. Aries).

Iglói, E., Terdik, Gy., 1997. Bilinear modelling of Chandler Wobble. Theory Probab. Appl. 44 (2), 398–400.

IVS 2004 General Meeting Proceedings, 2005. Proceedings of the workshop: Forcing of polar motion in the Chandler frequency band: a contribution to understanding interannual climate variations. Chandler Wobble Workshop, April 21-23, 2004, Luxembourg-city. Cahiers du Centre Europeen de Geodynamique et de Seismologie, vol. 24.

Jennrich, R.I., 1969. Asymptotic properties of non-linear least squares estimators. Ann. Math. Statist. 40, 633–643.

Klüppelberg, C., Mikosch, T., 1996. The integrated periodogram for stable processes. Ann. Statist. 24 (5), 1855–1879.

Magnus, J.R., Neudecker, H., 1999. Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, Chichester. (Revised reprint of the 1988 original).

Robinson, P.M., 1972. Non-linear regression for multiple time-series. J. Appl. Probab. 9, 758–768.

Schuh, H., Nagel, T., Seitz, T., 2001. Linear drift and periodic variations observed in long time series of polar motion. J. Geodesy 74, 701–710.

Shumway, R.H., Kim, S.-E., Blandford, R.R., 1999. Nonlinear estimation for time series observed on arrays. Asymptotics, Nonparametrics, and Time Series. Statistics: Textbooks and Monography, vol. 158. Dekker, New York, pp. 227–258.

Subba Rao, T., Terdik, Gy., 2006. Multivariate non-linear regression with applications. In: Bertail, P., Doukhan, P., Soulier, P. (Eds.), Dependence in Probability and Statistics. Springer, New York, pp. 431–470.

Turner, J., Colwell, S.R., Marshall, G.J., Lachlan-Cope, T.A., Carleton, A.M., Jones, P.D., Lagin, V., Reid, P.A., Igovkina, S., 2005. Antarctic climate change during the past 50 years. J. Climatol. 25, 279–294.

Wu, C.-F., 1981. Asymptotic theory of nonlinear least squares estimation. Ann. Statist. 9 (3), 501–513.