TECHNICAL TRANSACTIONS | CZASOPISMO TECHNICZNE
FUNDAMENTAL SCIENCES | NAUKI PODSTAWOWE
2-NP/2014

CHRISTIANA DRAKE*, OSKAR KNAPIK**, JACEK LEŚKOW***

MISSING DATA ANALYSIS IN CYCLOSTATIONARY MODELS

ANALIZA BRAKUJĄCYCH OBSERWACJI W MODELU CYKLOSTACJONARNYM

Abstract

In recent years, there has been a growing interest in modeling cyclostationary time series. The survey of Gardner and others [5] is quoting over 1500 different recently published papers that are dedicated to this topic. Data that can be reasonable modeled with such time series is often incomplete. To our knowledge, no systematic research has been conducted on that problem. This paper attempts to fill this gap. In this paper we propose to use EM algorithms to extend estimation for situation when some observations are missing.

Keywords: (almost) periodically correlated time series, cyclostationary signals, EM algorithm, missing data

Streszczenie

W ostatnim czasie wzrasta zainteresowanie modelowaniem cyklostacjonarnych szeregów czasowych. W pracy Gardner i inni [5] cytowane jest ponad 1500 publikacji poświęconych temu zagadnieniu. Jednakże dane, dla których model cyklostacjonarny jest zasadny, są często niekompletne. Zgodnie z naszą wiedzą nie było do tej pory systematycznego omówienia tego problemu. Celem niniejszego artykułu jest uzupełnienie tej luki. W artykule proponujemy wykorzystanie algorytmu EM w celu estymacji parametrów modelu w sytuacji brakujących obserwacji.

Słowa kluczowe: (prawie) okresowe szeregi czasowe, sygnały cyklostacjonarne, algorytm EM, brakujące obserwacje

* Department of Statistics, University of California, Davis; cmdrake@ucdavis.edu
** Oskar Knapik, CREATES, Department of Economics and Business, Aarhus University; knapiko@uek.krakow.pl
*** Institute of Mathematics, Cracow University of Technology, Poland; jleskow@riad.pk.edu.pl

1. Cyclostationary time series

The starting point of this research is the analysis of nonstationary time series. Let us assume that a time series $\{y_t, t \in N\}$ where N represents the integers, is observed. The cyclostationarity of $\{y_t, t \in N\}$ means repeatable behavior of the first and second order characteristic of such a time series. Let us denote the mean $\mu_Y(t) = E(y_t)$ and the autocovariance function $B_Y(t, \tau) = Cov(y_t, y_{t+\tau})$. The time series will be called *cyclostationary* or *periodically correlated* if the mean $\mu_Y(t)$ and the autocovariance function $B_Y(t, \tau)$ are periodic functions of t (see [9]). A classical statistical inference for analysis of such time series was presented in paper [4] among others. In our research, we will focus on second-order properties of the time series $\{y_t, t \in N\}$. Therefore, we assume that the time series under study is zero-mean.

The aim of this paper is to provide statistical inference procedures in situations where complete observation of the cyclostationary time series is impossible. Missing data analysis of cyclostationary time series frequently arises in the context of economic time series, mechanical signals and also ocean signals [12]. So far no systematic research has been presented in the case of missing data for periodically correlated time series. This work attempts to fill this gap using likelihood based inference and EM algorithms. This paper is divided into six sections. Section 2 presents state of the art results of likelihood-based inference for cyclostationary time series. Section 3 presents inferential methods in the case of missing data. Section 4 describes the EM algorithm in our cyclostationary model with missing data. The original results of this paper are presented in Section 2 where the full likelihood approach in cyclostationary model is presented. In Section 4, the original results are concerned with conditional distributions of our model, thus enabling applications of EM algorithm. In Section 5 the original results of this paper are illustrated with the help of a simulation study. Finally, Section 6 describes further directions of our research.

2. Likelihood-based inference for cyclostationary time series

A special class of cyclostationary time series (amplitude-modulated time series) is studied. Such time series can be represented as:

$$y_t = x_t \cdot c_t, \tag{2.1}$$

where

$\{x_t\}$ $-$a stationary time series (e.g. Gaussian $AR(p)$) ,

c_t $-$a (deterministic) periodic function.

In model (2.1) it is assumed that:

(AS1) The deterministic sequence $c_t \neq 0$ $\forall t$ to prevent from deterministic zeros.

(AS2) $\{x_t\}$ is a zero-mean stationary Gaussian sequence with a bounded and continuous spectral density.

We consider $\{x_t\}$to be a Gaussian autoregressive process of order p, $AR(p)$, given by

$$x_t - \sum_{i=1}^{p} \varphi_i x_{t-i} = \varepsilon_t \tag{2.2}$$

where

$\{\varepsilon_t\}$ $-$a sequence of independent Gaussian zero-mean random variables with finite variance σ^2.

The deterministic sequence c_t is assumed to be a known, periodic function of a finite dimensional unknown vector λ. The periodic function c_t with k periodicities can be expressed in the following way:

$$c_t = \exp\left(\sum_{j=1}^{k} (\lambda_{1j}\cos\omega_j t + \lambda_{2j}\sin\omega_j t)\right), \tag{2.3}$$

We assume that k is known. We assume furthermore that all frequencies are of the form $\omega_i \in (0, \pi], i = 1, \ldots, k : \omega_i = 2\pi r/P$ for some $r = 1, \ldots, (P-1)/2$, where P is a known period and also that $\lambda_{1j}^2 + \lambda_{2j}^2 > 0$ for $j = 1, \ldots, k$.

Following approach presented in [7] to ensure the identifiability of the model parameters, λ and the sequence $\{c_t\}_{t=1}^{T}$ are assumed to be linked via a one-to-one transformation, and it is assumed that there is no scale ambiguity in y_t.

Let $\theta = (\varphi, \lambda, \sigma^2)^T$ be the vector of all unknown parameters and suppose that we have $T = n + p$ observations from the model. The full likelihood for a vector of observations $y = (y_1, \ldots, y_T)^T$ corresponding to the model (2.1) is represented as:

$$L(y,\theta) = \frac{1}{(2\pi)^{T/2}} \cdot \det(R_Y)^{-1/2} \cdot \exp\left\{-\frac{1}{2} y^T R_Y^{-1} y\right\}, \tag{2.4}$$

where

$$R_Y = C_T R_X C_T^T,$$

R_X $-$the covariance matrix of $AR(p)$ process,

C_T $-$the $T \times T$diagonal matrix whose diagonal vector is $(c_1, \ldots, c_T)$ and $c_t = c_{i+mP} = c_i$.

Using [13] the inverse of the covariance matrix of an autoregressive process can be represented as follows

$$R_X^{-1} = \frac{1}{\sigma^2}\left(I_T + \sum_{i=1}^{p} \varphi_i^2 E_i - \sum_{i=1}^{p} \varphi_i F_i + \sum_{h=1}^{p-1}\sum_{i=1}^{p-h} \varphi_i \varphi_{i+h} G_{i,i+h}\right), \tag{2.5}$$

where

I_T $-$the identity matrix of the order T,

E_i $-$the identity with the first and last i ones set to zero,

F_i $-$the matrix which has ones along the upper and lower ith minor diagonals and zeros elsewhere,

$G_{i,i+h} = E_h F_i E_h$,thereby equaling F_i except the top and bottom h ones along the ith minor diagonals are replaced by zeros.

The sums in formula above are defined as zero if the upper limit of the summation is zero.

On the other hand, consider a vector of observations$y = (y_T, \ldots, y_1)^T$. From the model (2.1) it is obtained that

$$y = C \cdot x, \tag{2.6}$$

where

C $-$an $T \times T$ diagonal matrix with diagonal vector $c = (c_T, \ldots, c_1)$,

$x = (x_T, \ldots, x_1)^T$$-$a vector of observations from $AR(p)$ model.

Under the assumptions above, we have the following

Theorem 2.1.

Assume that (AS1) and (AS2) hold and that the time series $\{y_t\}$ follows model (2.1). Then, the log-likelihood function for the complete sample has the form

$$\begin{aligned} l\,(y;\theta) &= \log f_y\,(y;\theta) = -\log\left(|\det C|\right) + \log f_x\left(C^{-1}y;\theta\right) \\ &= -\log\left(|\det C|\right) - \tfrac{T}{2}\log\left(2\pi\right) - \tfrac{T}{2}\log\left(\sigma^2\right) + \tfrac{1}{2}\log\left|V_p^{-1}\right| - \tfrac{1}{2\sigma^2}\left(C_p\mathbf{y}_p\right)^T V_p^{-1}\left(C_p\mathbf{y}_p\right) \\ &- \sum_{t=p+1}^{T} \tfrac{1}{2\sigma^2}\left(\tfrac{y_t}{c_t} - \sum_{i=1}^{p}\tfrac{y_{t-i}}{c_{t-i}}\right)^2, \end{aligned} \tag{2.7}$$

where

$\theta = (\varphi, \lambda, \sigma^2)^T$ $-$the vector of all unknown parameters

Proof.

The starting point is the observation that

$$f_y\,(y) = |\det C|^{-1}\, f_x\left(C^{-1}y\right). \tag{2.8}$$

Using that fact and properties of $AR(p)$ process (see [8]) the joint density for the complete data set can be written as

$$f_y\,(y;\theta) = |\det C|^{-1} \cdot f_{x_p,\ldots,x_1}\left(\frac{y_p}{c_p},\ldots,\frac{y_1}{c_1}\right) \cdot \prod_{t=p+1}^{T} f_{x_t|x_{t-1},\ldots,x_{t-p}}\left(\frac{y_t}{c_t}\,\middle|\,\frac{y_{t-1}}{c_{t-1}},\ldots,\frac{y_{t-p}}{c_{t-p}}\right). \tag{2.9}$$

The density of the first p observations $f_{x_p,\dots,x_1}(\cdot)$ is of a $N(0,\sigma^2 V_p)$ variable:

$$f_{x_p,\dots,x_1}\left(\frac{y_p}{c_p},\dots,\frac{y_1}{c_1};\theta\right) = (2\pi)^{-p/2}\left(\sigma^{-2}\right)^{p/2}\left|V_p^{-1}\right|^{1/2}\exp\left[-\frac{1}{2\sigma^2}\left(C_p^{-1}\mathbf{y}_p\right)^T V_p^{-1}\left(C_p^{-1}\mathbf{y}_p\right)\right], \tag{2.10}$$

where $\mathbf{y}_p = (y_p,\dots,y_1)^T$ and C_p is a $p\times p$ diagonal matrix with diagonal vector $(c_p,\dots,c_1)$ and V_p^{-1} is the inverse covariance matrix given by [3] .

$$v^{ij}(p) = v^{ji}(p) = \left[\sum_{k=0}^{i-1}\varphi_k\varphi_{k+j-i} - \sum_{k=p+1-j}^{p+i-j}\varphi_k\varphi_{k+j-i}\right] \tag{2.11}$$

for $1,,i,,j,,p$, where $\varphi_0 = -1$.

For the remaining observations in the sample the prediction-error decomposition can be used as in $AR(p)$ case. Conditional on the first $t-1$ observations, the density of tth observation $f_{x_t|x_{t-1},\dots,x_{t-p}}(\cdot)$ is

$$f_{x_t|x_{t-1},\dots,x_{t-p}}\left(\frac{y_t}{c_t}\middle|\frac{y_{t-1}}{c_{t-1}},\dots,\frac{y_{t-p}}{c_{t-p}}\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}\left(\frac{y_t}{c_t}-\sum_{i=1}^{p}\varphi_i\frac{y_{t-i}}{c_{t-i}}\right)^2\right] \tag{2.12}$$

The log-likelihood function for the sample has the form

$$\begin{aligned} l(y;\theta) &= \log f_y(y;\theta) == -\log(|\det C|) + \log f_x\left(C^{-1}y;\theta\right) = \\ &= -\log(|\det C|) - \tfrac{T}{2}\log(2\pi) - \tfrac{T}{2}\log\left(\sigma^2\right) + \tfrac{1}{2}\log\left|V_p^{-1}\right| - \tfrac{1}{2\sigma^2}\left(C_p\mathbf{y}_p\right)^T V_p^{-1}\left(C_p\mathbf{y}_p\right) \\ &- \sum_{t=p+1}^{T}\tfrac{1}{2\sigma^2}\left(\tfrac{y_t}{c_t}-\sum_{i=1}^{p}\tfrac{y_{t-i}}{c_{t-i}}\right)^2, \end{aligned} \tag{2.13}$$

This completes the proof of Theorem 2.1.

We have the following result

Corrollary 2.2. Let us consider the model $y_t = c_t x_t$, where $\{x_t\}$ is $AR(1)$ process. For the $AR(1)$ process V_p^{-1} is a scalar whose value is found by taking $i=j=p=1$:

$$V_1^{-1} = \left(1-\varphi^2\right). \tag{2.14}$$

Thus $\sigma^2 V_1 = \sigma^2\left(1-\varphi^2\right)$ which indeed reproduces the formula for the unconditional variance of the $AR(1)$ process.

The exact likelihood for the vector of observations y from the above model is given as

$$l\left(y;\theta\right)=-\log\left(\left|\det C\right|\right)-\frac{T}{2}\log\left(2\pi\right)-\frac{T}{2}\log\left(\sigma^2\right)+\frac{1}{2}\log\left|\left(1-\varphi^2\right)\right|$$
$$-\frac{1}{2\sigma^2}\frac{y_1^2}{c_1^2}\left(1-\varphi^2\right)-\sum_{t=2}^{T}\frac{1}{2\sigma^2}\left(\frac{y_t}{c_t}-\varphi\frac{y_{t-1}}{c_{t-1}}\right)^2. \tag{2.15}$$

It is possible to give an exact formula for the ML estimate of σ^2 conditional on φ and λ. There is no analytical form of the ML estimates of φ and λ even conditionally on the other parameters. Because of this the maximization of the exact log likelihood function must be accomplished numerically. It is a high dimensional nonlinear optimization problem. It is worth mentionning that parameters of the deterministic and stochastic parts of this model are linked via a nonlinear relationship.

3. Missing Data Mechanisms and Inference

Data that may reasonably be modeled by cyclostationary time series are often incomplete for a number of reasons, e.g. the interruption of measurements due to instrument failure or extreme natural phenomena, the accidental loss of data or erroneous measurements, among others ([12]). Similar problems are encountered in vibroacoustic and economic time series.

Missing data present some potentially serious problems in drawing inferences from time series. The degree to which conclusions under such circumstances are affected, depends on the mechanism by which the data is missing. Missing data mechanisms can be divided into roughly three categories (see [10] and [11]): *missing completely at random* (MCAR), *missing at random* (MAR) and *missing not at random* (MNAR).

For the purpose of formal description of missingness mechanism let us define the complete data $y=(y_t)$, where observation y_t comes from the data generating process P_θ parameterized by the unknown vector of parameters θ for $t=1,\ldots,T$. Let us also define the missing-data indicator vector $m=(m_t)$, where a random variable m_t

$$m_t=\begin{cases}1, & \text{if } y_t \text{ is missing}\\ 0, & \text{if } y_t \text{ is present}\end{cases}, \tag{3.1}$$

has a distribution P_ψ. Let us also assume that θ and Ψ are distinct. Denote the joint distribution of (y,m) by $P(y,m;\theta,\psi)$, where y consists of two parts. The observed part is denoted by y_{obs} and the missing part by y_{mis}. The missing-data mechanism is defined by means of conditional distribution of m given y, which is$P(m|y)$.

If data is *missing completely at random* (MCAR) we have

$$P\left(m_t=1|y;\theta,\psi\right)=P\left(m_t=1|\psi\right). \tag{3.2}$$

When data is *missing at random* (MAR), we have

$$P\left(m_t=1|y;\theta,\psi\right)=P\left(m_t=1|y_{obs};\theta,\psi\right). \tag{3.3}$$

When neither MCAR nor MAR hold, we say the data are missing not at random, abbreviated MNAR. Missingness depends on the unobserved data.

Missing completely at random is easiest to deal with. The missing observations constitute a random sub-sample of all observations. Estimates of population parameters are unbiased. Precision of estimates will be affected, particularly if the original sample size was modest and the numbers of missing observations constitute a substantial fraction of all observations.

MAR allows likelihood based and Bayesian inference and does not require modeling of the missingness mechanism. Inference can be based on the observed data likelihood. When data is MNAR unbiased inference is not possible without further assumptions and additional information. Finally, MAR and/or MCAR cannot be established on the basis of observed data alone and require additional information.

Often in cyclostationary data, values are missing, because of instrument failure. This instrumental failure may or may not depend on the missing values.

Instruments sometimes have a limited range for signal detection. This can cause data to be missing, because missing values are below or above the detection limits of the instrument. This is a special type of informative missingness.

The type of missingness mechanism considered in this paper will be MCAR or MAR. The first missingness mechanism (MCAR) assumes that failure to observe the data does not depend on the data. The second missingness mechanism (MAR) is more general than the first: here it is possible that the missingness mechanism depends on observed data.

Within this research we consider only the case of missing at random (MAR)

4. EM algorithm in likelihood-based inference for cyclostationary time series with missing observations

The expectation-maximization (EM) algorithm is an iterative procedure for computing the maximum likelihood estimator for data set which is not complete. [2] showed the wide applicability of the EM algorithm in statistics. The convergence and performance of the EM algorithm was proved by [14].

Let y be a complete data vector which consists of y_{mis} missing data and y_{obs} "observed" data. The EM algorithm is an iterative procedure for computing the maximum likelihood estimator only on the basis of the observed datay_{obs}. Each iteration of EM algorithm consists of two steps. If $\theta^{(i)}$ denotes the estimated value of the parameter θ after i iterations, then the two steps in the $(i+1)$th iteration are

E-step: Calculate $Q\left(\theta|\theta^{(i)}\right) = E_{\theta^{(i)}}\left[l\left(\theta; y_{obs}, y_{mis}\right)|y_{obs}\right]$

M-step: Maximize $Q\left(\theta|\theta^{(i)}\right)$ with respect toθ.

Then $\theta^{(i+1)}$ is set to the maximizer of Q in the M-step.

For our model, we have two possible realizations of EM algorithm: one on the basis of conditional likelihood function which leads to modified EM algorithm for normal linear regression settings and on the basis of full likelihood function which is well-known problem of inference for incomplete data within multivariate normal distribution.

Before description of EM algorithm, let us take a look at properties of considered model.

Many properties of the cyclostationary model (2.1) follow from the autoregressive structure of $\{x_t\}$. It is clear that if $\{x_t\}$ follows the zero-mean Gaussian $AR(p)$ process then

$$x_t|x_{t-1},\ldots,x_{t-p} \sim N\left(\sum_{j=1}^{p}\varphi_i x_{t-i},\sigma^2\right). \tag{4.1}$$

When $y_t = c_t \cdot x_t$, where $\{x_t\}$ is $AR(p)$ process and c_t is a deterministic periodic function then assuming that $c_t \neq 0$ for all t we can write $x_t = y_t/c_t$.

The following result establishes the form of the conditional distribution, the conditional expectation and the conditional variance given the past.

Theorem 4.1. Under the assumptions (AS1) and (AS2) and the model equation (2.1) one obtains

$$y_t|y_{t-1},\ldots,y_{t-p} \sim N\left(c_t\left(\sum_{i=1}^{p}\varphi_i\frac{y_{t-i}}{c_{t-i}}\right),c_t^2\sigma^2\right), \tag{4.2}$$

$$E\left(y_t|y_{t-1},\ldots,y_{t-p}\right) = c_t\left(\sum_{i=1}^{p}\varphi_i\frac{y_{t-i}}{c_{t-i}}\right), \tag{4.3}$$

and

$$Var\left(y_t|y_{t-1},\ldots,y_{t-p}\right) = c_t^2\sigma^2. \tag{4.4}$$

Proof is straightforward and will be omitted.

In missing data analysis one frequently confronts the situation of 'filling the gaps' that is calculating the conditional expectation given the past and future. For that purpose consider the distribution $p(y_{mis}|y_{obs})$. Due to $AR(p)$ structure of $\{x_t\}$ one obtains

$$x_t = \varphi_1 x_{t-1} + \ldots + \varphi_p x_{t-p} + \varepsilon_t,\ \{\varepsilon_t\} \sim N\left(0,\sigma^2\right) \tag{4.5}$$

taking $x_t = y_t/c_t$, we have

$$\frac{y_t}{c_t} = \varphi_1\frac{y_{t-1}}{c_{t-1}} + \ldots + \varphi_p\frac{y_{t-p}}{c_{t-p}} + \varepsilon_t. \tag{4.6}$$

Denote observed values $y = (y_{i_1}, \ldots, y_{i_r})^T$, with $1?i_1 < \ldots < i_r?T$. If there are no missing observations in the first p observations, then the best estimates of the missing values are found by minimizing

$$\sum_{t=p+1}^{T} \left(\frac{y_t}{c_t} - \varphi_1 \frac{y_{t-1}}{c_{t-1}} - \ldots - \varphi_p \frac{y_{t-p}}{c_{t-p}} \right)^2 \tag{4.7}$$

with respect to the missing values. The minimization of the sum above gives the form of the conditional expectation.

Theorem 4.2.
Consider the following sequence (y_j, y_{j+1}, y_{j+2}) from the process $y_t = c_t \cdot x_t$, where c_t is (deterministic) periodic function and $\{x_t\}$ is $AR(1)$ process defined by

$$x_t = \varphi x_{t-1} + \varepsilon_t,\, \varepsilon_t \sim N\left(0, \sigma^2\right). \tag{4.8}$$

The conditional expectation of y_{j+1} given y_j (past) and y_{j+2} (future) has the following form:

$$E\left(y_{j+1} | y_j, y_{j+2}\right) = c_{j+1}\varphi \left(\frac{y_j}{c_j} + \frac{y_{j+2}}{c_{j+2}} \right) \Big/ (1 + \varphi^2). \tag{4.9}$$

Proof.
The form of the conditional expectation can be obtained as the minimization of the above sum

$$\left(\frac{y_{j+1}}{c_{j+1}} - \varphi \frac{y_j}{c_j} \right)^2 + \left(\frac{y_{j+2}}{c_{j+2}} - \varphi \frac{y_{j+1}}{c_{j+1}} \right)^2, \tag{4.10}$$

with respect to y_{j+1}. Setting the derivative of this expression with respect to y_{j+1} equal to 0 and solving for y_{j+1} and using properties of conditional expectation.

Suppose $y = (y_1, \ldots, y_T)^T$ be the "complete" data vector of which r are observed and $T - r$ are missing. Denote the "observed" data by $y_{obs} = (y_{i_1}, \ldots, y_{i_r})$ (called "incomplete" data) and missing data by $y_{mis} = (y_{j_1}, \ldots, y_{j_{T-r}})$.

If we work with full likelihood function, we have that $y = (y_{obs}, y_{mis})$ has a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix R_Y which depends on the parameter θ, the log-likelihood of the complete data is given by

$$l\left(\theta, y\right) = -\frac{T}{2} \ln\left(2\pi\right) - \frac{1}{2} \ln \det\left(R_Y\right) - \frac{1}{2} y^T R_Y^{-1} y \tag{4.11}$$

The E step requires that we compute the expectation of $l(\theta, y)$ with respect to the conditional distribution of y given y_{obs} with $\theta = \theta^i$. Following the approach presented in [1] let us consider $R_Y(\theta)$ as the block matrix

$$R_Y = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{4.12}$$

which is conformable with y_{mis} and y_{obs}, the conditional distribution of y given y_{obs} is multivariate normal with mean
and covariance matrix $\begin{pmatrix} \Sigma_{11|2}(\theta) & 0 \\ 0 & 0 \end{pmatrix}$, where $\hat{y}_{mis} = E_\theta(y_{mis}|y_{obs}) = \Sigma_{12}\Sigma_{22}^{-1}y_{obs}$ and $\Sigma_{11|2}(\theta) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. It can be shown that

$$E_{\theta^{(i)}}\left(\left(y_{mis}^T, y_{obs}^T\right)^T R_Y^{-1}(\theta)\left(y_{mis}^T, y_{obs}^T\right)|y_{obs}\right) = trace\left(\Sigma_{11|2}\left(\theta^{(i)}\right)\Sigma_{11|2}^{-1}(\theta)\right) + \hat{y}^T R_Y^{-1}(\theta)\hat{y}, \tag{4.13}$$

where $\hat{y} = (\hat{y}_{mis}, y_{obs})$.

As a consequence

$$Q\left(\theta|\theta^{(i)}\right) = l(\theta;\hat{y}) - \frac{1}{2}trace\left(\Sigma_{11|2}\left(\theta^{(i)}\right)\Sigma_{11|2}^{-1}(\theta)\right) \tag{4.14}$$

The first term on the right is the log-likelihood based on the complete data, but with y_{mis} replaced by its "best estimate" $\hat{y}_{mis}$ calculated from the previous iteration. If the increments $\theta^{(i+1)} - \theta^{(i)}$ are small, then the second term on the right is nearly constant ($\approx T - r$) and can be ignored. To make computation easier we can use the modified version

$$\tilde{Q}\left(\theta|\theta^{(i)}\right) = l(\theta;\hat{y}). \tag{4.15}$$

With this modification, the steps in the EM algorithm are as follows:

E-step: Calculate $E_{\theta^{(i)}}(y_{mis}|y_{obs})$ and form $\tilde{Q}\left(\theta|\theta^{(i)}\right)$.

M-step: Find the maximum likelihood estimator for "complete" data problem, i.e. maximize $l(\theta, \hat{y})$.

The best linear predictor of a missing observation y_{j_k} from the vector y_{mis} is $E(y_{j_k}|y_{obs})$, so within E-step we reconstruct "complete" observations in the following way:

$$\hat{y}_t^{(i)} = \begin{cases} y_t, \text{ if } y_t \text{ is present} \\ E\left(y_t|y_{obs};\theta^{(i)}\right), \text{ if } y_t \text{ is missing} \end{cases} \tag{4.16}$$

In M-step, we maximize likelihood function of vector $\hat{y} = (\hat{y}_1, \ldots, \hat{y}_T)$ with respect to θ.

Tab. 1: Results of ML and EM algorithm-based estimation

Method	$\hat{\varphi}$	$\hat{\sigma}^2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$
ML estimation	0,4976	1,2946	0,4077	0,5387
EM algorithm	0,5599	1,3329	0,3877	0,5523

5. Simulation study

To see how approximation of the EM algorithm works in practice, we restrict our attention to the class of PC time series of the form

$$y_t = c_t x_t, \tag{5.1}$$

where $x_t = \varphi x_{t-1} + ?_t$, $\{?_t\} \sim N(0, \sigma^2)$ and $c_t = \exp\left(\lambda_1 \cos\left(\frac{2\pi}{20}t\right) + \lambda_2 \sin\left(\frac{2\pi}{20}t\right)\right)$. The following values of parameters were chosen $\varphi = 0.5$, $\sigma^2 = 1$, $\lambda_1 = 0.4$ and $\lambda_2 = 0.5$.

Firstly we simulate $T = 100$ observations from our model and estimate unknown vector of parameters θ on the basis of complete sample. Then we randomly choose 10% of the data to be missing.

The results are presented in the Table 1.

It can be seen that the approximation of the EM algorithm gives reasonable estimates of unknown parameters θ in situation when some observations are missing. The question of convergence, however, needs to be further explored.

Oskar Knapik gratefully acknowledges the Kosciuszko Foundation for financial support for this research

References

[1] Brockwell P.J., Davis R.A., *Introduction to time series and forecasting*, Springer, 2002.

[2] Dempster A.P., Laird N.M. and Rubin D.B., *Maximum Likelihood from Incomplete Data via The EM Algorithm (with discussion)*, J. Roy. Statist. Soc. B, Vol. 39, 1977, 1—38.

[3] Galbraith R.F., Galbraith J.I., *On the inverses of some patterned matrices arising in the theory of stationary time series*, Journal of Applied Probability, Vol. 11, 1974, 63—71.

[4] Gardner W.A., *Representation and estimation of cyclostationary processes*, IEEE Transactions on Information Theory, Vol. 19, No. 3, 1973, 375—376.

[5] Gardner W.A., Napolitano A., Paura L., *Cyclostationarity: Half a century of research*, Signal Processing, Vol. 86, No. 4, 2006, 639—697.

[6] Giannakis,G.B, Dandawate, A.V., *Consistent Kth-order time-frequency representations for (almost) cyclostationary signals*, in: IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, Victoria, BC, Canada, 79 October 1992, 123—126.

[7] Ghogho, M., Garel, B., *Maximum likelihood estimation of amplitude-modulated time series*, Signal Processing, Vol. 75, 1999 99—116.

[8] Hamilton J.D., *Time Series Analysis*, Princeton University Press, 1994.

[9] Hurd H. L., Miamee A., *Periodically Correlated Random Sequences: Spectral Theory and Practice (Wiley Series in Probability and Statistics),* Wiley-Interscience, 2007.

[10] Little R.J.A., Rubin D.B., *Statistical Analysis with Missing Data.* John Wiley & Sons, New York, 2nd edn., 2002.

[11] Schaffer J.L., *Analysis of Incomplete Multivariate Data.* Chapman & Hall/CRC, 1997.

[12] Stefanakos Ch.N., Athanassoulis G.A., *A unified methodology for the analysis, completion and simulation of nonstationary time series with missing values, with application to wave data*, Applied Ocean Research, Vol. 23, Issue 4, 2001, 207—220.

[13] Verbyla A.P., *A note on the inverse covariance matrix of the autoregressive process*, Australian Journal of Statistics, Vol. 27, Issue 2, 1985, 221—224.

[14] Wu C.F.J., *On the convergence properties of the EM algorithm*, Annals of Statistics, Vol. 11, 1983, 95—103.