IHS Economics Series
Working Paper 274
September 2011

Some Computational Aspects of Gaussian CARMA Modelling

Helgi Tómasson

IRIHS Institutional Repository at IHS
Vienna

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

Impressum

Author(s):
Helgi Tómasson

Title:
Some Computational Aspects of Gaussian CARMA Modelling

ISSN: Unspecified

2011 Institut für Höhere Studien - Institute for Advanced Studies (IHS)
Josefstädter Straße 39, A-1080 Wien
E-Mail: office@ihs.ac.at
Web: www.ihs.ac.at

All IHS Working Papers are available online:
http://irihs.ihs.ac.at/view/ihs_series/
This paper is available for download without charge at:
https://irihs.ihs.ac.at/id/eprint/2090/

274

Reihe Ökonomie

Economics Series

Some Computational Aspects of Gaussian CARMA Modelling

Helgi Tómasson

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

274

Reihe Ökonomie

Economics Series

Some Computational Aspects of Gaussian CARMA Modelling

Helgi Tómasson

September 2011

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

Helgi Tómasson
University of Iceland
Faculty of Economics,
IS-101, Reykjavík, ICELAND
☎: 354-552-6806,
email: helgito@hi.is

Founded in 1963 by two prominent Austrians living in exile – the sociologist Paul F. Lazarsfeld and the economist Oskar Morgenstern – with the financial support from the Ford Foundation, the Austrian Federal Ministry of Education and the City of Vienna, the Institute for Advanced Studies (IHS) is the first institution for postgraduate education and research in economics and the social sciences in Austria. The **Economics Series** presents research done at the Department of Economics and Finance and aims to share "work in progress" in a timely way before formal publication. As usual, authors bear full responsibility for the content of their contributions.

Das Institut für Höhere Studien (IHS) wurde im Jahr 1963 von zwei prominenten Exilösterreichern – dem Soziologen Paul F. Lazarsfeld und dem Ökonomen Oskar Morgenstern – mit Hilfe der Ford-Stiftung, des Österreichischen Bundesministeriums für Unterricht und der Stadt Wien gegründet und ist somit die erste nachuniversitäre Lehr- und Forschungsstätte für die Sozial- und Wirtschaftswissenschaften in Österreich. Die **Reihe Ökonomie** bietet Einblick in die Forschungsarbeit der Abteilung für Ökonomie und Finanzwirtschaft und verfolgt das Ziel, abteilungsinterne Diskussionsbeiträge einer breiteren fachinternen Öffentlichkeit zugänglich zu machen. Die inhaltliche Verantwortung für die veröffentlichten Beiträge liegt bei den Autoren und Autorinnen.

Abstract

Representation of continuous-time ARMA, CARMA, models is reviewed. Computational aspects of simulating and calculating the likelihood-function of CARMA are summarized. Some numerical properties are illustrated by simulations. Some real data applications are shown.

Keywords

CARMA, maximum-likelihood, spectrum, Kalman filter, computation

JEL Classification

C01, C10, C22, C53, C63

Contents

1 Introduction 1

2 Some properties of ARMA and CARMA 1

3 Simulation and estimation 3

4 Some technicalities 6

4.1 The time scale 6

4.2 Enforcing stationarity 6

4.3 Frequency domain approach 8

4.4 Some numerical considerations 9

4.5 Nested models and starting values 9

5 Some illustrative examples 10

5.1 The sampling of a simple CARMA(2,1) model 10

5.2 Simulation of a CARMA(4,3) 12

5.3 Sunspots data 14

5.4 Cycles in the Earth’s temperature 15

5.5 Analysis of IBM transaction data 17

6 Discussion 20

References 20

1 Introduction

The aim of this paper is to conceptualize a practical computational scheme for CARMA models. Time-dependency is a fundamental feature of many types of data analysis. Hence, time-series analysis is an important aspect of modelling many phenomena in business and science. In traditional time-series the main emphasis is on the case when a continuous variable is measured at discrete equi-spaced time-points. The famous book by Box & Jenkins (1970) has made the discrete time ARMA approach highly popular. As in classical time-series analysis analytical results are sparse and it is necessary to rely on numerical techniques. This paper will review various results that might be of use in applied continuous-time modelling. The paper will not deal with asymptotic properties of estimators nor with identification of parameters.

Many real phenomena consist of a continuous-time process and the discrete-time model is a technical approximation, due to the facts that data are only observed (sampled) at discrete time-points and the continuous-time models are sometimes analytically difficult. Irregular sampling is sometimes treated as missing data in the traditional discrete-time approach. A lot of missing data requires a lot of extra programming effort. If the model is defined as a continuous-time process from the beginning, then by design there is no such thing as a missing data problem. It is just discrete sampling of a continuous process.

Using a continuous-time model implies that irregularly spaced observations are treated in a natural and objective manner. Another feature is that high frequency variability is acknowledged. In discrete-time analysis high-frequency variability, i.e., the variability above the Nykvist frequency is mapped, *aliased*, into the frequency band defined by the sampling intensity. A classical textbook on the ARMA and CARMA models is by Priestley (1981).

The organization of this paper is as follows. In section 2 basic representations of discrete-time ARMA and continuous-time CARMA are reviewed. In section 3 key concepts of simulating and estimating CARMA models are reviewed. Numerical considerations are important for all practical work with time-series models. In section 4 some numerical aspects are discussed. Brief results of simulated models and textbook data are shown in section 5. Section 6 concludes.

2 Some properties of ARMA and CARMA

The traditional time-domain representation of an equispaced ARMA(p,q) process is:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q},$$

where Y_t is the observed process and ε_t is the unobserved innovation process. Among typical assumptions about the distribution, is that the ε_t's, are independent zero mean normal, or white-noise with finite variance. Relaxing the finite variance assumption is discussed in Mikosch, Gadrich, Kluppelberg & Adler (1995). Frequently the ARMA(p,q) process is stated in terms of polynomials of the backward operator B (or L), $BY_t = Y_{t-1}$.

$$\begin{aligned}
&\boldsymbol{\phi}(B)Y_t = \boldsymbol{\theta}(B)\varepsilon_t,\\
&\boldsymbol{\phi}(z) = 1 - \phi_1 z - \cdots - \phi_p z^p, \quad \boldsymbol{\theta}(z) = 1 + \theta_1 z + \cdots + \theta_q z^q.
\end{aligned}$$

Second order properties are described with the spectral density function,

$$f(\omega) = \frac{\sigma^2}{2\pi} \frac{\boldsymbol{\theta}(\exp(i\omega))\,\boldsymbol{\theta}(\exp(-i\omega))}{\boldsymbol{\phi}(\exp(i\omega))\,\boldsymbol{\phi}(\exp(-i\omega))}.$$

The process $Y(t)$ can also be represented as a stochastic integral,

$$\begin{aligned}
&Y_t = \int_{-\pi}^{\pi} \exp(i\omega\, t)\,\mathrm{d}Z(\omega),\\
&E(\mathrm{d}Z(\omega)) = 0, \quad E(\mathrm{d}Z(\omega)\,\overline{\mathrm{d}Z(\omega)}) = f(\omega)\,\mathrm{d}\omega, \quad E(\mathrm{d}Z(\omega)\,\overline{\mathrm{d}Z(\lambda)}) = 0, \quad \lambda \neq \omega.
\end{aligned}$$

A continuous-time ARMA, CARMA, process can be defined in terms of a continuous-time innovation process and a stochastic integral. A common choice of innovation process is the Wiener process, $W(t)$. A representation of a CARMA(p,q) process in terms of the differential operator D is:

$$\begin{aligned}
&Y^{(p)}(t) + \alpha_1 Y^{(p-1)}(t) + \cdots + \alpha_p Y(t) = \sigma\,\mathrm{d}(W(t) + \beta_1 W^1(t) + \cdots + \beta_q W^{(q))}(t)),\\
&\text{ or } \boldsymbol{\alpha}(D)\,Y(t) = \sigma\boldsymbol{\beta}(D)\ \mathrm{d}W(t),\\
&\boldsymbol{\alpha}(z) = z^p + a_1 z^{p-1} + \cdots + a_p, \quad \boldsymbol{\beta}(z) = 1 + \beta_1 z + \cdots + \beta_q z^q.
\end{aligned}$$

Here, $Y^{(p)} = D^p Y(t)$, denotes the p-th derivative of $Y(t)$. The path of a Wiener process is nowhere differentiable so the symbol $D\,W(t)$, and higher derivatives, is of a purely formal nature. The spectral density of $Y(t)$ is a rational function:

$$f(\omega) = \frac{\sigma^2}{2\pi} \frac{\boldsymbol{\beta}(i\omega)\boldsymbol{\beta}(-i\omega)}{\boldsymbol{\alpha}(i\omega)\boldsymbol{\alpha}(-i\omega)}.$$

The spectral representation of CARMA is:

$$\begin{aligned}
&Y(t) = \int_{-\infty}^{\infty} \exp(i\omega\, t)\,\mathrm{d}Z(\omega), \quad E(\mathrm{d}Z(\omega)) = 0,\\
&E(\mathrm{d}Z(\omega)\,\overline{\mathrm{d}Z(\omega)}) = f(\omega)\,\mathrm{d}\omega, \quad E(\mathrm{d}Z(\omega)\,\overline{\mathrm{d}Z(\lambda)}) = 0, \quad \lambda \neq \omega.
\end{aligned}$$

The stationarity condition of the ARMA requires that the roots of the polynomial $\boldsymbol{\phi}(z)$ are outside the unit circle. The stationarity condition of the CARMA requires the roots of the polynomial $\boldsymbol{\alpha}(z)$ to have negative real-parts and that $p > q$.

A regularly sampled CARMA(p,q) process is also ARMA process. For example the CAR(1) process,

$$\mathrm{d}Y(t) + aY(t) = \mathrm{d}W(t), \quad (\text{ The Ornstein-Uhlenbeck process}),$$

is if observed reglulary at time-points $t, t+\Delta, t+2\Delta, \ldots$, an AR(1) process:

$$Y(t+\Delta) = \exp(-a\Delta)Y(t) + \sigma \int_t^{t+\Delta} \mathrm{d}W(t),$$

i.e. $\phi = \exp(-a\Delta)$. Obviously this means that $\phi > 0$. Therefore an AR(1) process with a negative ϕ cannot be a CAR(1) process. In general a CARMA(p, q) process observed regularly can be written as an ARMA process. The converse is not true, i.e. there exist ARMA(p', q') processes which are not a discrete version of some CARMA(p, q). An intuitive explanation is seen by looking at the relation between the spectral density of a continuous-time process and a discretely equispaced sampled version of that process. If a continuous-time process with spectral density $f_c(\omega)$ is sampled at discrete time intervals, Δ, the discretely observed process has spectral density,

$$f_\Delta(\omega) = \sum_{k=-\infty}^{\infty} \frac{1}{\Delta} f_c((\omega + 2\,k\pi)/\Delta), \quad -\pi \leq 0 \leq \pi. \tag{1}$$

Finding a continuous-time process that, when observed at regular time intervals, Δ, has a particular spectral density f_Δ, by solving equation (1) is a non-trivial exercise. It is clear that different continuous-time processes can look the same when observed discretely. Equation (1) explains the concept of embedding a discrete time ARMA model in a continuous-time model. Chan & Tong (1987) give a description of a particular case. Further details on embedding of a CARMA within an ARMA and other aspects of CARMA processes are discussed by Brockwell (2009). Normal processes are completely defined by their second order properties, i.e. the spectral-density/auto-covariance function. For non-normal processes the question of embedding is more complicated because it addresses the whole distribution of the process, not just the first two moments.

A key feature of ARMA and CARMA models is that the spectral density has the form of a rational function. If the assumption of rational spectral density is abandoned, Priestley (1963) shows a method of finding a candidate for f_c based on f_Δ. Equation (1) also reflects the *aliasing* phenomenon. I.e. that variability due to higher frequencies is mapped into the interval defined by the sampling process. The estimation of a discrete-time model assigns all the variance to the interval of length $2\pi/\Delta$, i.e., the variance associated with higher frequencies is *aliased* with lower frequencies.

3 Simulation and estimation

The literature suggests several ways of simulating a CARMA process. A frequency domain approach is to use the fact that for a stationary process $Y(t)$,

$$V(Y(t)) = \int_{-\infty}^{\infty} f(\omega)\ \mathrm{d}\omega,$$

where $f(\omega)$ is the spectral density of $Y(t)$. Then, an interval $(-\omega_c, \omega_c)$, that represents a high proportion of the variability in $Y(t)$ is chosen. The interval $(0, \omega_c)$ is then divided into M subintervals with $\Delta_i = (\omega_i - \omega_{i-1})$. A classical approach is that of Rice (1954):

$$Y_{Rice}(t) = \sum_{i=1}^{M} 2\sqrt{f(\omega_i)\Delta_i}\,\cos(\omega_i t - U_i), \text{ with } U_i \text{ independent } U(-\pi, \pi). \tag{2}$$

Sun & Chaika (1997) give a modified version:

$$\begin{aligned} &Y_{SC}(t) = \sum_{i=1}^{M} R_i \cos(\omega_i t - U_i), \text{ with } U_i \text{ independent } U(-\pi, \pi), \text{ and} \\ &R_i \text{ independent Rayleigh with } E(R_i^2) = 4f(\omega_i)\Delta_i. \end{aligned} \tag{3}$$

The simulated processes $Y_{Rice}(t)$ and $Y_{SC}(t)$ have the same second order properties as a theoretical normal $Y(t)$ with spectral density $f(\omega)$. $Y_{SC}(t)$ is normally distributed, whereas $Y_{Rice}(t)$ is only approximately normal. The Kalman-filter algorithm offers an easy way of programming a time-domain approach. A traditional state-space representation of a normal CARMA process is:

$$\begin{aligned} &Y(t) = \boldsymbol{\beta'}\boldsymbol{X}(t), \quad t \geq 0, \\ &\mathrm{d}\boldsymbol{X}(t) = A\boldsymbol{X}(t) + \sigma\boldsymbol{R}\,\mathrm{d}W, \qquad (4) \\ &A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\alpha_p & -\alpha_{p-1} & \cdots & -\alpha_2 & -\alpha_1 \end{bmatrix}, \quad \boldsymbol{X}(t) = \begin{bmatrix} Y(t) \\ Y^{(1)}(t) \\ \vdots \\ Y^{(p-2)} \\ Y^{(p-1)} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 = 1 \\ \beta_1 \\ \vdots \\ \beta_q \\ \mathbf{0} \end{bmatrix}, \\ &\boldsymbol{R'} = [0 \cdots 0 \quad 1], \quad \mathbf{0} = p - q - 1 \text{ dimensional vector of zeroes,} \end{aligned}$$

with $p > q$, see, e.g., Tsai & Chan (2000) and Brockwell, Chadraa & Lindner (2006). Equation (4) is a multivariate linear stochastic differential equation. Given an initial value of the state vector $\boldsymbol{X}(t_0) = \boldsymbol{x}(t_0)$, the solution is, is given by

$$\boldsymbol{X}(t) = \exp(A(t - t_0))\boldsymbol{x}(t_0) + \sigma \int_{t_0}^{t} \exp(A(t - s))\boldsymbol{R}\,\mathrm{d}W(s).$$

Here exp, denotes the matrix exponential, $\exp(A) = I + A + A^2/2 + \cdots$. The conditional mean and covariance matrix of the state vector $\boldsymbol{X}(t)$ are given by,

$$E(\boldsymbol{X}(t)|\boldsymbol{X}(t_0) = \boldsymbol{x}(t_0)) = \exp(A(t - t_0))\boldsymbol{x}(t_0), \tag{5}$$

$$V_{t|t_0} = V(\boldsymbol{X}(t)|\boldsymbol{X}(t_0) = \boldsymbol{x}(t_0)) = \sigma^2 \int_{t_0}^{t} \exp(A(t - s))\boldsymbol{R}\boldsymbol{R'}\exp(A'(t - s))\,\mathrm{d}s. \tag{6}$$

The unconditional mean of $\boldsymbol{X}(t)$ is zero, and the relation between the unconditional variance, V_∞, and the innovation variance are related by:

$$V_\infty = \exp(A(t-t_0))V_\infty \exp(A(t-t_0))' + V_{t|t_0}. \tag{7}$$

Integration by parts shows that $V_{t|t_0}$ solves the equations system,

$$AV_{t|t_0} + V_{t|t_0}A' = \sigma^2 \left[-\exp(A(t-s))\boldsymbol{RR'}\exp(A'(t-s))(A')^{-1}\right]_{t_0}^{t}.$$

In particular, the stationary covariance matrix of the state vector, $V_\infty = \lim_{t_0 \to -\infty} V_{0|t_0}$ solves,

$$AV_\infty + V_\infty A' = -\sigma^2 \boldsymbol{RR}'.$$

Combining results in Shoji & Ozaki (1998) and Tsai & Chan (2000), equation (7), gives,

$$V_{t|t_0} = V_\infty - \exp(A(t-t_0))V_\infty \exp(A'(t-t_0)).$$

Applying a standard matrix algebra result on Kronecker products,

$$\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec}(B),$$

to

$$AV_\infty I \text{ and } IV_\infty A',$$

shows, that V_∞ solves:

$$\operatorname{vec}(I \otimes A)\operatorname{vec}(V_\infty) + \operatorname{vec}(A \otimes I)\operatorname{vec}(V_\infty) = -\sigma^2 \operatorname{vec}(\boldsymbol{RR}').$$

The number of equations in this system is p^2. However, the matrix $p \times p$ matrix V_∞ is a function of only p elements and has a particular structure. Tsai & Chan (2000) derive an explicit algorithm for calculating V_∞ by solving a system of p equations.

The traditional approaches of estimating the parameters, $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma)$, based on a set of observations $y(t_1), \ldots, y(t_n)$, are first, a frequency-domain approach, and second a time-domain least-squares/maximum-likelihood approach. For a frequency domain approach, a sample estimate, $\hat{f}(\omega)$, of the spectral density is needed, and then by minimizing (maximizing) some objective function, e.g.,

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma} \int_{-\infty}^{\infty} (\log(f(\omega) + \hat{f}(\omega)/f(\omega))\,\mathrm{d}\omega,$$

an estimator is obtained. Solving this optimization problem yields the Whittle estimator. A time-domain approach is to use the Kalman-filter to calculate the conditional log-likelihood, $l(y(t_i)|y(t_{i-1}), \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma)$ and solve,

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma} \sum_{i=1}^{n} l(y(t_i)|y(t_{i-1}), \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma).$$

When a value of the conditional expectation, (5), and variance, (6), are available it is straightforward to set up the Kalman-filter iterations and calculate the log-likelihood. The MLE (maximum-likelihood-estimates) are then obtained by some numerical optimization routine.

	α_1	α_2	β_1	σ
Original time-scale	2	40	0.15	8
Time multiplied by 10	0.2	0.4	1.5	0.253
Time multiplied by 0.1	20	4000	0.015	252.98

Table 1: Impact of scaling of time on CARMA parameters.

4 Some technicalities

4.1 The time scale

The continuous-time ARMA model has the property that the parameters are not a function of the sampling intensity. They are, however, a function of the definition of the time scale. The impact of transformations of the time scale are best understood by studying the spectral density. The spectral density of a particular CARMA process is:

$$f(\omega) = \frac{\sigma^2}{2\pi} \frac{\boldsymbol{\beta}(i\omega)\boldsymbol{\beta}(-i\omega)}{\boldsymbol{\alpha}(i\omega)\boldsymbol{\alpha}(-i\omega)} \, \mathrm{d}\omega. \tag{8}$$

The units of ω are radians per time unit. If the time scale is multiplied by a constant c, i.e., $\omega_* = c\,\omega$, then the spectral density of the time-transformed process will be,

$$f(\omega_*) = \frac{{\sigma_*}^2}{2\pi} \frac{\boldsymbol{\beta}_*(i\omega_*)\boldsymbol{\beta}_*(-i\omega_*)}{\boldsymbol{\alpha}_*(i\omega_*)\boldsymbol{\alpha}_*(-i\omega_*)} \, \mathrm{d}\omega_*. \tag{9}$$

The vectors $\boldsymbol{\alpha}_*$ and $\boldsymbol{\beta}_*$ in equation (9) are derived by solving for the corresponding powers of ω in (8). Solving for the $(\beta_j)_*$ is straightforward, $(\beta_j)_* = c^j\beta_j$. The $(\alpha_j)_*$ have to be scaled such that the coefficient of the highest power of the polynomial in the denominator is one, i.e., $(\alpha_j)_* = c^{-j}\alpha_j$. Then $\sigma* = c^{-(p-1/2)}\sigma$. The term $-1/2$ in the scaling transform of σ is due to the Jacobian of the transform. An example of the impact of scaling of a simple CARMA(2,1) model is shown in table 1. In numerical work a proper scaling of the time axis can be helpful.

4.2 Enforcing stationarity

The stability demand of a linear differential equation, $\mathrm{d}\boldsymbol{X}(t) = A\boldsymbol{X}(t)\,\mathrm{d}t$, restricts the real parts of the eigenvalues of the matrix A to be negative. For $p = 2$, this is equivalent to $\alpha_1 > 0$ and $\alpha_2 > 0$. Checking the condition of negative real parts of the eigenvalues of a matrix is in general a non-trivial exercise. Probably the best known approach for checking this condition is the Routh-Hurwitz theorem. Here two different, more computationally oriented approaches, that apply to the case where A is a companion matrix (like the matrix in equation (4)), are suggested.

The former approach is based on linking the stability condition of a continuous-time differential equation to the stability condition of a discrete-time AR process. A CAR(p),

$Y^{(p)} + \alpha_1 Y^{(p-1)} + \cdots + \alpha_p = \sigma \,\mathrm{d}W$, process is stationary if the roots of:

$$\boldsymbol{\alpha}(z) = z^p + \alpha_1 z^{p-1} + \cdots + \alpha_{p-1} z + \alpha_p, \tag{10}$$

have negative real parts. Analogously for the discrete-time AR(p) process, $Y_t + \phi_1 Y_{t-1} + \cdots \phi_p Y_{t-p} = \varepsilon_t$ is stationary if the roots of:

$$\boldsymbol{\phi}(z) = 1 + \phi_1 z + \cdots + \phi_p z^p, \tag{11}$$

lie outside the unit circle. The condition that the roots of $\boldsymbol{\phi}(z)$ lie outside the unit circle is equivalent to the roots of $z^p\boldsymbol{\phi}(1/z)$ lie inside the unit circle. Belcher, Hampton & Tunnicliffe Wilson (1994)(BHT) use the following transformation transforming stationary AR(p) parameter values to stationary CAR(p) parameter values. If z is a complex number within the unit circle then

$$s = -\kappa \frac{1-z}{1+z},$$

lies in the left half-plane. This fact is used to establish a connection between $\boldsymbol{\alpha}(z)$ and $\boldsymbol{\phi}(z)$ such that:

$$h(s) = h_0 s^p + h_1 s^{p-1} + h_{p-1} z + h_p =$$
$$\sum_{i=0}^{p} \phi_i (1 - s/\kappa)^i (1 + s/\kappa)^{p-i}, \quad (\phi_0 = 1).$$

Then define the coefficients of $\boldsymbol{\alpha}(z)$ are such that $\alpha_i = h_i/h_0$. If $w_1, \ldots, w_p$ are the roots of $z^p\boldsymbol{\phi}(1/z)$, then $\max(|w_i|) < 1$ is equivalent to that the roots of $\boldsymbol{\alpha}(z)$ are in the left-half plane. The κ coefficient in the above notation is taken from BHT. The κ reflects the impact of time-scaling on the CAR parameters. In this work κ is set to one, and time-scaling performed separately. A transformation described by Monahan (1984) gives a one-to-one relationship between the stationary parameter space of a stationary AR(p) and the p-dimensional cube $[-1,1]^p$. By combining the transformations described by Monahan (1984) and Belcher et al. (1994) on gets a transformation:

$$\boldsymbol{\gamma}_{MBHT} : [-1,1]^p \curvearrowright \text{the space of valid } (\alpha_1, \ldots, \alpha_p).$$

The transformation described in Monahan (1984) is essentially a recursive way of calculating the partial auto-correlation function for a set of discrete-time AR parameters $(\phi_1, \ldots \phi_p)$. The Durbin-Levinson algorithm is a well known algorithm for calculating the partial auto-correlation function.

Applying a similar idea in the continuous-time setting gives another approach of enforcing the stationarity restriction. Pham & Breton (1991) give a continuous-time version of the Durbin-Levinson algorithm. Their approach results in a one-to-one transformation that maps a p-dimensional vector, $(\gamma_1, \ldots, \gamma_p)$ of positive real numbers into to the parameter space of the stationary CAR.

$$\boldsymbol{\gamma}_{PB} : \mathbb{R}_+^p \curvearrowright \text{the space of valid } (\alpha_1, \ldots, \alpha_p).$$

These two transformations offer two ways of enforcing the stationarity restriction of the parameter $(\alpha_1, \ldots, \alpha_p)$. They make programming of restricted numerical maximization of the likelihood function straightforward. The results shown in this paper are based on numerically maximizing the log-likelihood enforcing the stationary restriction by performing unconstrained optimization of $\log(\boldsymbol{\gamma}_{PB})$ over $\mathbb{R}^p$,

$$\max_{\log(\gamma_{PB})} \sum_{i=1}^{n} l(y(t_i)|y(t_{i-1}).$$

A similar type of transformation, $\boldsymbol{\gamma}_{MBHT}$ or $\boldsymbol{\gamma}_{PB}$, of the MA part of the model ensures a unique stationary parameterization of the CARMA(p,q) model.

4.3 Frequency domain approach

The Whittle estimator requires an estimate of the spectral density. Obtaining an estimate of the spectral density is a demanding numerical tasks. For the case of equally spaced observations the fast-Fourier-transform (FFT) is an efficient way of getting an estimate, $\hat{f}(\omega)$, of the spectral density, $f(\omega)$. For irregularly spaced observations the case is not so simple. One way is to use the approach described by Masry (1978a,b,c). The idea is to use a bias-correction term in the Fourier transform.

$$\frac{1}{2\pi\bar{\Delta}n}|\sum_{j=1}^{[n/2]} e^{i\omega_k t_j} y(t_j)|^2 - \frac{1}{2\pi\bar{\Delta}n}\sum_{j=1}^{n} y(t_j)^2, \quad \bar{\Delta} = \frac{1}{n}\sum_{j=2}^{n}(t_j - t_{j-1}).$$

This can be calculated for a set of frequencies, $\omega_1 < \omega_2, \ldots$, and used for calculating the Whittle-objective function. For large n this will be a computationally demanding task. Another approach is to interpolate the observations, e.g., linearly, and then sample equally spaced observations from the interpolated time-series. Then one can use the FFT to get an estimate of the spectral density and use a discrete-time model spectrum as an approximation. Greengard & Lee (2004) derive an acceleration of the non-uniform Fourier transform and call the result NUFFT. Using such an algorithm gives a computationally fast way of computing a frequency-domain based estimate of the spectrum. Typically, using bias correction methods such as those above, results in negative values of the estimated spectrum for a range of ω_k's. If the variance of the process is estimated with:

$$2\sum_{k=1}^{K} \hat{f}(\omega_k)(\omega_k - \omega_{k-1}),$$

there is a possibility of a negative estimate of the variance. In general, as pointed out in Greengard & Lee (2004) the non-uniform Fourier transform is sensitive to the choice of frequencies, ω_k, and it is not clear how to choose them. Therefore some tweaking of sets of frequencies is inevitable for getting a good candidate for the empirical spectrum in the irregular sampling case. An alternative might be to approach the empirical spectrum by a positive function.

4.4 Some numerical considerations

The Kalman-filter algorithm offers an analytical way of calculating the normal likelihood. For a given set of parameter values all terms in the likelihood function are straightforward, except for the stationary covariance matrix of the state vector, V_∞. Tsai & Chan (2000) give an analytical iterative method of calculating V_∞. Here the parameter values are restricted in such a manner that stationarity is enforced. This restricts the eigenvalues of the matrix A to have negative real parts. The estimates shown in this paper are based on combining the algorithms of Tsai & Chan (2000) and Pham & Breton (1991) mentioned earlier. This yields an analytical way of calculating the likelihood function. A $\boldsymbol{\gamma}_{PB}$-type transformation is also used for MA parameters to ensure a unique MA representation corresponding to the numerator of the spectral density.

The matrix-exponent that appears in the likelihood function is a numerically challenging object. Moler & Van Loan (2003) review the progress of the last 25 years of several methods to calculate the matrix exponent. The results shown in this paper are based on the EXPOKIT FORTRAN subroutines, Sidje (1998). Many numerical optimization packages demand the derivative of the log-likelihood. Tsai & Chan (2003) give analytical methods for calculating the derivative of the matrix-exponent with respect to the matrix A. Their results show that the calculation of the analytical derivative will be quite computationally demanding so here numerical methods are used to calculate the derivative of the log-likelihood. The scoring algorithm is a convenient way of numerically maximizing the likelihood. The parameter space can easily be transformed in such a way that unrestricted optimization can be performed. Then the application of standard optimization packages, e.g., in R (R Development Core Team, 2011), is straightforward.

A standard way of calculating an estimate of the covariance matrix of the estimated is to numerically calculate the information matrix, by either:

$$I_{\boldsymbol{\theta}} = \frac{-\partial^2 \log L(\boldsymbol{\theta}|\boldsymbol{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \text{ or } I_{\boldsymbol{\theta}} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log L(\boldsymbol{\theta}|y(t_i))}{\partial\boldsymbol{\theta}} \frac{\partial \log L(\boldsymbol{\theta}|y(t_i))}{\partial\boldsymbol{\theta}'}$$

Precision of function of the parameters, such as e.g., the logged-spectrum, $g(\boldsymbol{\theta}, w) = \log(f(\omega|\boldsymbol{\theta}))$ can then be approximated by the delta-method, i.e.,:

$$\text{var}(g(\hat{\boldsymbol{\theta}})) \simeq \frac{\partial g}{\partial\boldsymbol{\theta}'} I^{-1}(\boldsymbol{\theta}) \frac{\partial g}{\partial\boldsymbol{\theta}}$$

The derivative of the log-spectral density $\log(f(\omega))$, with respect to the parameters, can be calculated analytically. The confidence bands shown in graphs in this paper are based on this method. BHT give a different method suitable for their parameterization.

4.5 Nested models and starting values

In the discrete-time ARMA all AR(1) models are a subset of any ARMA(p,q) models with p>1. The AR(1) model is nested within an AR(2) model with $\phi_2 = 0$. In the continuous-time case the CAR(1) is not a subset of any CAR(2). A feature shared with the discrete, and continuous-time ARMA is that if a common root is added to the AR and MA components of the model, then the dynamic structure is the same.

$\boldsymbol{\alpha}(D)Y(t) = \sigma\boldsymbol{\beta}(D)\,\mathrm{d}W(t)$ is the same as
$(k + D)\boldsymbol{\alpha}(D)Y(t) = \sigma(k + D)\boldsymbol{\beta}(D)\,\mathrm{d}W(t)$.

If fit of a CARMA(p,q) is available it is always possible to find infinitely many exactly equivalent CARMA(p+1,q+1)'s. It is therefore to be expected that the correlations between the CARMA parameter estimates are very close to one in an overparameterized CARMA(p,q). In general it is difficult to find suitable starting values for the numerical maximization of the log-likelihood function of a CARMA model. Even if transformations such as $\boldsymbol{\gamma}_{PB}$ and γ_{MBHT} ensure a mathematically valid parameter value, the log-likelihood value can be so flat that numerical maximization is difficult. Sometimes a good guess can be obtained by a Whittle estimate. In the case of irregular sampling it may take some tweaking of which frequencies to use in deriving a useful spectral estimate. BHT give another way of nesting models by designing a special structure on the MA component. Their idea is that if $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ is a valid set of parameters for a stationary AR(p), then if the MA coefficients of CARMA are defined in a particular manner,

$$\beta_k = \binom{(p-1)}{(k-1)}, \quad k = 1, \ldots, p-1, \quad \beta_0 = 1, \tag{12}$$

then calculating the AR coefficients using the function $\boldsymbol{\gamma}_{MBHT}(r_1, \ldots, r_p)$, $-1 \leq r_k \leq 1$, will give a stationary CARMA(p,p-1) which will be nested in a stationary CARMA(p+1,p) with AR coefficients $\boldsymbol{\gamma}_{MBHT}(r_1, \ldots, r_p, 0)$ and $\boldsymbol{\beta}$ calculated by equation (12). The BHT method offers a way of numerically estimating a CARMA(p,p-1) directly by imposing these restriction on the MA part. Both methods, adding a common root to the AR and MA components, and using the BHT transforms offer a way of getting a sequence of nested models.

5 Some illustrative examples

5.1 The sampling of a simple CARMA(2,1) model

A CARMA process with spectral density:

$$f(\omega) = \frac{{\sigma_*}^2}{4\pi}\left(\frac{1}{(\omega_c + \omega)^2 + a^2} + \frac{1}{(\omega_c - \omega)^2 + a^2}\right), \tag{13}$$

has a peak in the spectrum at ω_0 and an overall variance of $\sigma_*^2/(2a)$. If $\omega_0 = 0$ it is an Ornstein-Uhlenbeck, CAR(1). If $a = \sigma_* = 1$ and $\omega_0 = 2\pi$, it has a CARMA(2,1) representation with $\boldsymbol{\alpha} \simeq (2, 40.478)'$, $\boldsymbol{\beta} \simeq (1, 0.1572)'$ and $\sigma \simeq 6.362$. Regular sampling of one observation per time unit will obviously not be informative as the process has a cycle with frequency one cycle per time unit ($\omega_0 = 2\pi$). The term for this well known phenomenon in time-series analysis is *aliasing*. In the continuous-time case the question how much data is needed has two aspects. Observations are dependent, and the dependency decreases with the increase of time between observation. It is therefore not only a question of how many data points are obtained, but also the timing of the observations has an impact. In tables 2 and 3 estimates of a particular replication of this process are shown. The observation periods are of length $T = 100$ and $T = 1000$ units of time, respectively. The sampling frequencies are 10, 20, and 100 observations per unit of time. The estimated standard errors of the estimates are shown in tables 4 and 5. The pattern is clear. Increasing the number of datapoints by increasing the length of the observation period increases the precision of the estimates. However, increasing the number of datapoints by sampling more observations per unit of time has only marginal impact on the precision of the parameters describing the cyclical properties. Increasing the number of observations increases the precision of the overall variability, σ, of the process. This is natural because the main feature of this process is its cyclical structure and for getting a precise information about its cyclical nature it is necessary to observe many cycles. I.e., what is needed is, a reasonable number of datapoints within each cycle and then a large T, i.e. many cycles. For this particular process over 99% of the variation is due to frequencies below 5π, two and a half cycle per time unit. Therefore it is understandable that not much information is gained by sampling more frequently than 10 observations per unit of time.

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\sigma}$
Δ=0.1	1.733	41.269	0.174	5.678
Δ=0.05	1.727	41.477	0.167	5.784
$\Delta = 0.01$	1.741	40.890	0.179	5.573

Table 2: Parameter estimates for $T = 100$.

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\sigma}$
Δ=0.1	1.980	39.516	0.168	6.012
Δ=0.5	1.985	39.662	0.166	6.060
Δ=0.01	1.995	39.830	0.163	6.141

Table 3: Parameter estimates for T=1000.

	s.e.($\hat{\alpha}_1$)	s.e.($\hat{\alpha}_2$)	s.e.($\hat{\beta}_1$)	s.e.($\hat{\sigma}$)
Δ=0.1	0.284	1.888	0.013	0.762
Δ=0.05	0.286	1.802	0.012	0.689
Δ=0.01	0.213	1.422	0.015	0.444

Table 4: Standard errors of parameter estimates for T=100.

	s.e.($\hat{\alpha}_1$)	s.e.($\hat{\alpha}_2$)	s.e.($\hat{\beta}_1$)	s.e.($\hat{\sigma}$)
Δ=0.1	0.090	0.632	0.004	0.229
Δ=0.05	0.088	0.574	0.003	0.201
Δ=0.01	0.075	0.482	0.004	0.154

Table 5: Standard errors of parameter estimates for T=1000.

Another virtue of defining the statistical model in continuous time is that the parameterization of the model is not a function of the sampling frequency. Comparing the parameter estimates of discrete-time ARMA model of the data generated by the above reveals this difference between the discrete-time and the continuous-time case. Compare table 6 and table 3.

Regular sampling cannot give information about cycles with frequency above the Nykvist frequency. If there is substantial variation in the process above the Nykvist frequency it will be aliased in to a low frequency band of the spectrum. Random sampling is in principle alias-free, i.e., all frequencies have a possibility of being represented by the data. It is not clear how to define a ,,Nykvist" frequency in irregular finite sample cases. In finite samples it is obvious that there is a bound on which part of the spectrum can be reasonably estimated. It is clear that sometimes it is possible to measure a cycle that has higher frequency that than the average sampling frequency. In table 7 results of 5000 observations of two simulated cases of the above model are shown. In one case the time between observations are exponentially distributed with mean $\bar{\Delta} = 1$, i.e., on average one observation per unit of time, in the other case the time interval are exponentially distributed with $\bar{\Delta} = 4$. In both cases reasonable estimates of the parameters are obtained. The precision is worse in the sparser sampling case. This suggests that there exists some kind of optimal sampling rate. Here on average one measurement per time unit is better than on average one observation every four time units. A long period is needed to observe many cycles and sufficiently dense observations are needed to get information about the nature of the cycles. For the case of table 7 sufficiently many fragments of cycles are available to get reasonable parameter estimates.

	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}$	$\hat{\sigma}$
Δ=0.1	1.485	-0.840	-0.490	0.358
Δ=0.05	1.819	-0.914	-0.714	0.242
Δ=0.01	1.978	-0.982	-0.939	0.102

Table 6: Parameter estimates in an unevenly sampled ARMA.

5.2 Simulation of a CARMA(4,3)

Fifty replications of a particular CARMA(4,3) models were simulated. The length of the simulated series is T=1000 time units and the interval between observations are exponentially distributed with an average of 10 observation per day. Table 8 shows summary

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\sigma}$
estimate - $\bar{\Delta} = 1$	1.963	40.439	0.146	6.521
estimate - $\bar{\Delta} = 4$	1.851	39.178	0.145	6.348
s.e. - $\bar{\Delta} = 1$	0.085	0.673	0.013	0.359
s.e. - $\bar{\Delta} = 4$	0.121	1.032	0.031	0.726

Table 7: Parameter estimates and standard errors of an irregularly observed CARMA(2,1).

statistics of the average value of the maximum-likelihood estimates of the parameters, the standard deviation of the estimates within the simulation and the average estimated standard errors. The estimated standard errors are calculated by inverting the observed estimated information matrix in each replication. For the case of about 10000 observations averaging 10 observations per time unit the level of the MLE estimates and the estimated standard errors are of a correct order of magnitude. This particular CARMA was chosen because it contains two cycles of similar amplitude at frequencies π and 2π (a half cycle and a full cycle per unit of time, respectively). The simulation method was a frequency domain method based on adding two spectral function of the type in equation (13), $\omega_c = \pi$ and $\omega_c = 2\pi$, respectively. The true theoretical spectrum is shown in figure 5.2. In this example the magnitude of the parameters α_4 and β_3 is, 10^2 and 10^{-2}, respectively. Therefore an observation period of 1000 units of time results in about 500 cycles of the longer type and about 1000 oft the shorter one. This is an easy example. A reasonable number of both types of cycle was observed and both cycles were of comparable size and frequency. In all fifty replications the correct type of model would have been chosen if say a BIC-type, or comparable model selection method was used.

If the process would consist of two very different cycles, say 2π and 2000π, this difference in size between α_4 and β_3, would be more dramatic, i.e. 10^9 and 10^{-9}, respectively. This shows that the numerical treatment of a CARMA model with very different frequencies is difficult. If indeed, one believes that there are two important frequencies, one cycle per time unit, and one thousand cycles per time unit, a practical approach could be to take some subsamples within one period and try to correct for the long cycle with some deterministic function. Similarly, then one could try to filter out the short cycle and get a mean value within the short cycle and then estimate the dynamics of the long cycle.

	α_1	α_2	α_3	α_4	β_1	β_2	β_3	σ
True value	4.000	55.348	102.696	439.984	0.398	0.075	0.009	212.567
Average MLE	3.968	55.462	102.538	442.105	0.401	0.075	0.009	212.562
Std. sim.	0.247	1.730	10.864	30.894	0.044	0.007	0.001	20.339
Average Std.est	0.258	1.580	10.092	30.523	0.063	0.009	0.002	30.076

Table 8: Summary statistics of 50 replication of a CARMA(4,3) model. T=1000, Δ independent exponentially distributed, $\bar{\Delta}$=0.1.

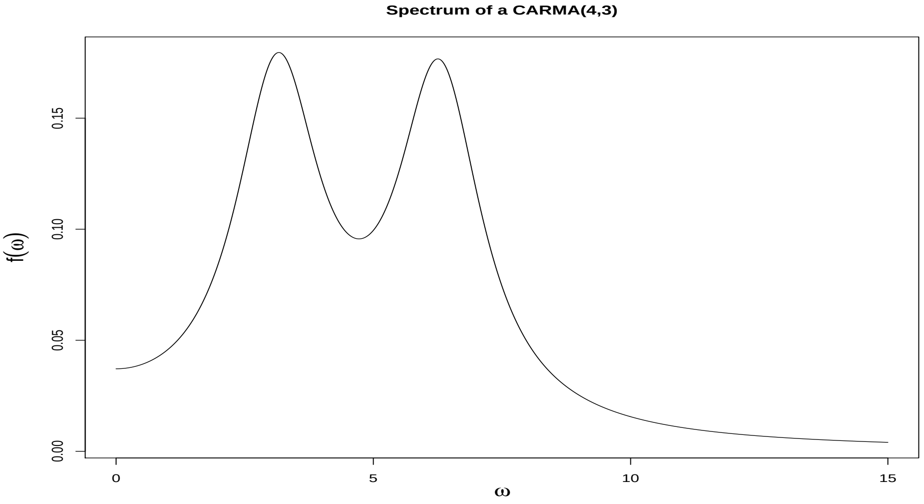

Figure 1: The spectum of the CARMA(4,3) of table 8.

5.3 Sunspots data

The Wolfer sunspots data is well known from time-series textbooks. As many other authors Phadke & Wu (1974) use that series to illustrate methodology. Phadke and Wu give a method for transforming discrete-time ARMA estimates to continuous-time CARMA. Their data sets is the average monthly number of sunspots from 1749 to 1924. The dataset is available for R-datasets (R Development Core Team, 2011). It seems to differ slightly from that used by Phadke and Wu, i.e., their average is 44.75, whereas in R-datasets the average is 44.78. Graphical inspection, figure 2, of the 176 datapoints suggests that this is the same series. Table 9 compares the discrete-time results of Phadke and Wu with the

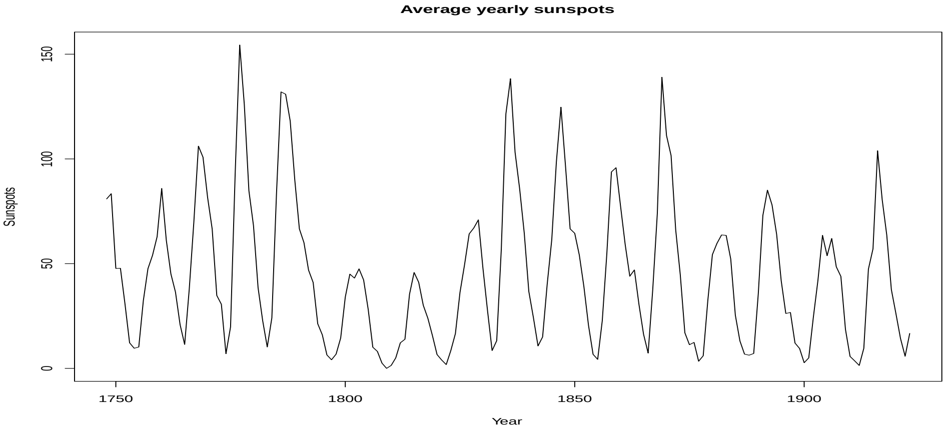

Figure 2: Average montly number of sunspots in the period 1749 to 1924 (data from R-datasets).

	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	$\hat{\sigma}$
Phadke & Wu (1974)	1.424	-0.721	-0.151	15.51
Results of arima in R	1.426	-0.721	-0.159	15.30

Table 9: Result of a discrete ARMA(2,1) modelling of sunspots 1749 to 1924.

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\sigma}$
Phadke & Wu (1974)	0.327	0.359	0.633	15.51
Author's R-program	0.327	0.357	0.645	15.52

Table 10: Results of a continuous time CARMA(2,1) modelling of sunspots 1749 to 1924.

results of a standard discrete-time ARMA program, arima from the R-package. Table 10 compares the derived CARMA(2,1) estimates of Phadke and Wu with the author's implementations of direct CARMA estimation. The cycle of this model is about 10 years. The author also tested the sunspots series up to 1983. Direct estimation. based on removing the mean, suggests a CARMA(4,3), but if trend also is removed, a CARMA(2,1) seems usable. Still the cycle is roughly 10 years.

5.4 Cycles in the Earth's temperature

Jouzel & et al. (2007) show data describing the evolution of the climate on Earth for the past 800 Kyears. One of their data series is used for describing the evolution of the Earth's average temperature. A variable, deltaT, is used as an indicator temperature. The 800 Kyear past is shown in figure 3. There are 5.788 observation points and thereof 4.921 in the past 400 Kyears. A quick look a the series and an estimated spectrum suggests that the main action in this variable is due to a low frequency component. In figure 4 an empirical estimate of the spectrum is shown. The spectrum is calculated by use of the Masry (1978c) bias correction and the NUFFT, the non-uniform fast Fourier transform of Greengard & Lee (2004). Table 11 shows the maximized log-likelihood value of some CARMA(p,p-1) models. Usual model selection critera, AIC, BIC, etc., suggest that a value of p, between 3 and 6 is a good choice. A typical form of the logged spectrum and the respective confidence interval, of these CARMA models is shown in figure 5. The CARMA(6,5) shown in figure 5 suggests that the most important cycle is of length 80.5 Kyears. The other CARMA(p,p-1)($p \geq 2$) models also suggest a similar cycle. The estimated CARMA models also agree on allocating about 50% of the variance to cycles longer than 50 Kyears. Substantial variance is also allocated to the high frequencies, 1% to frequencies above 360 radians per Kyear (about a cycle of 15 years). Splitting the data in two equally long periods suggests that the dynamics are similar in both periods.

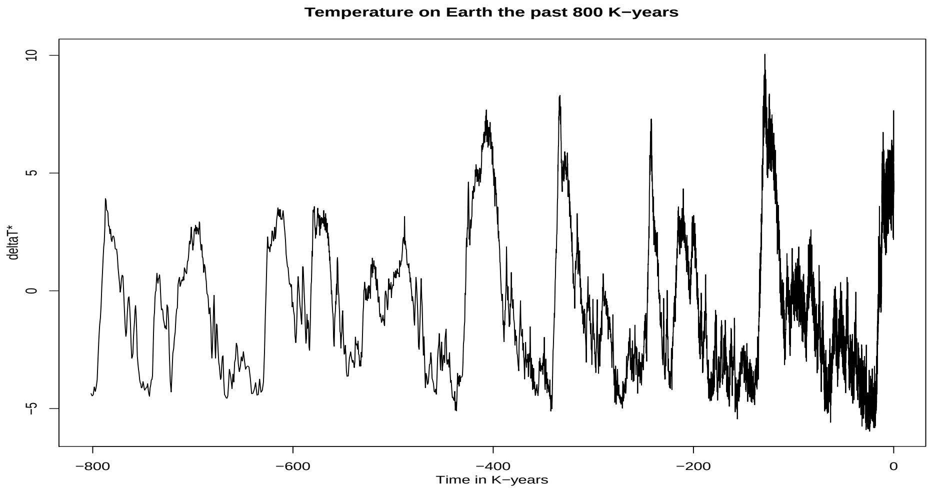

Figure 3: Evolution in climate on Earth for the past 800.000 years.

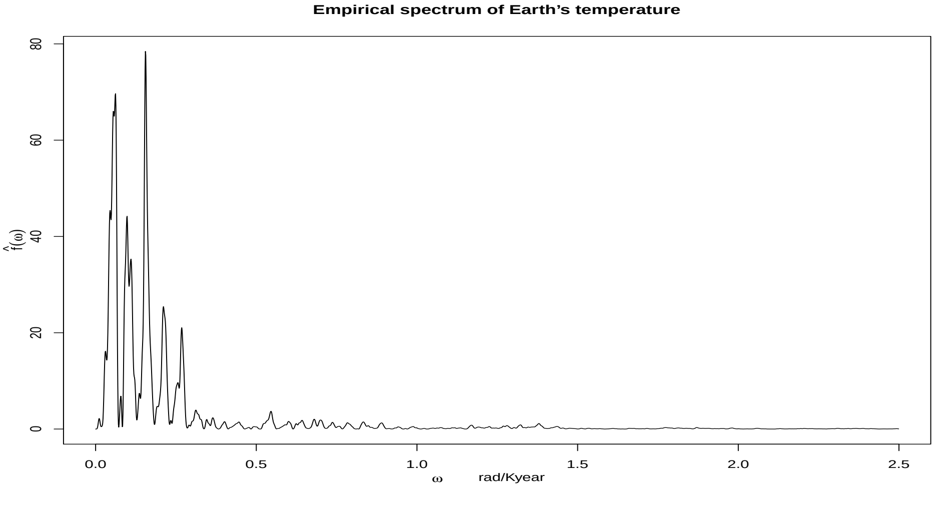

Figure 4: Empirical spectrum of the climate on Earth for the past 800.000 years.

	p=1	p=2	p=3	p=4	p=5	p=6	p=8
Log-likelihood	-8580.4	-5696.5	-5655.1	-5648.8	-5645.7	-5642.3	-5637.9

Table 11: The maximized log-likelihood of some CARMA(p,p-1) models for climate data.

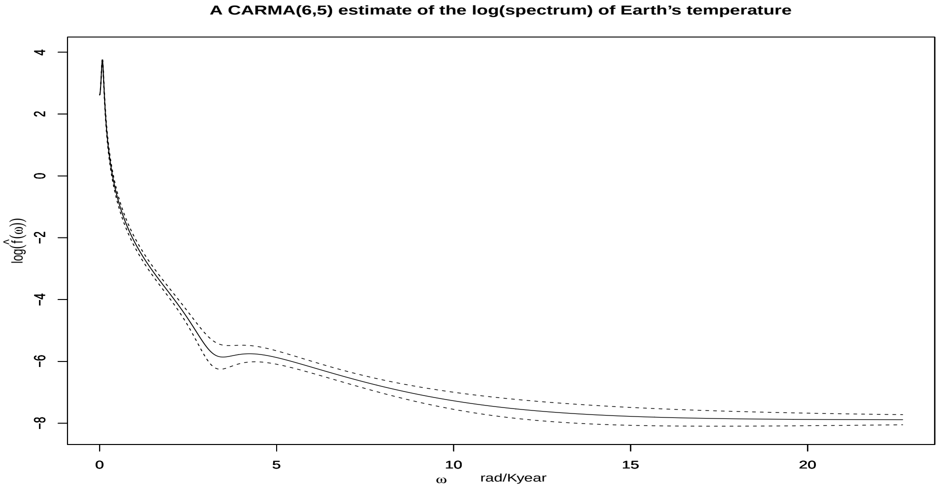

Figure 5: Log-spectrum and 95% confidence band of an estimated CARMA(6,5) model of the earth's temperature.

5.5 Analysis of IBM transaction data

Data on IBM transactions at the New York Stock Exchange from first of November 1990 to 31st of January 1991 have been used as an example in the high frequency financial literature, (Engle & Russel, 1998; Tsay, 2010). The data contain both transaction time measured in seconds and transaction prices. In the 91 day period, there are 63 days of trading and roughly 60.000 transactions, whereof about 53.000 have distinct transaction times. This makes these data a candidate for continuous-time modelling. The transactions per day range from 304 a day up to 1844, with an average of 854. As to be expected with financial market price data the long-term linear dynamic structure is weak. Figure 6 shows an empirical estimate of the spectrum of the logged returns. It is clear that the variability is not concentrated to the lower frequencies.

One can also analyse the dynamics within each day, e.g., measuring the time in minutes rather than days. An empirical spectral estimate was calculated for each of the 63 days and the average of the 63 spectral curves is shown in figure 7. The figure does not show a decrease in the spectrum. CAR(1) and a CARMA(2,1) models were compared for each of the 63 days. The average value of twice the log-likelihood-ratio was 26, suggesting that frequently the CARMA(2,1) was giving a better fit than the CAR(1). The estimated CARMA(2,1) models have a peak in the spectrum corresponding to a very high frequency.

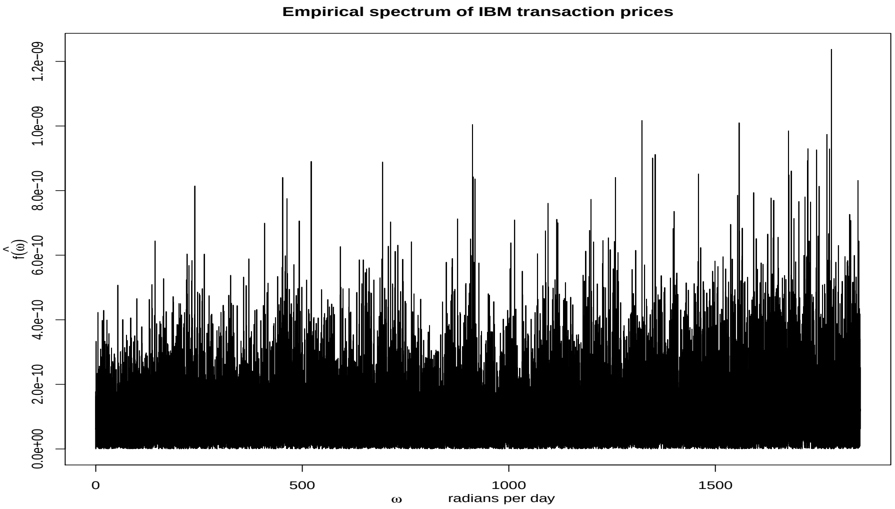

Figure 6: Empirical spectrum of 91 days of IBM transaction prices.

This peak represented a much higher frequency than the average trading frequency. The average trading frequency corresponds to about two transactions per minute on average, but the peak in the spectrum was in the range from 4 to 20 seconds. That means 3-15 cycles a minute. Some restricted models were estimated, demanding that all 63 days have exactly the same CARMA(p,p-1) dynamic structure. Inspection of log-likelihood values, suggests that the main features are already captured by a CARMA(5,4). These main features are a two-peaked spectral curve. The estimated spectrum of a common CARMA(5,4) for the 63 days is shown in figure 8. The first peak suggest a cycle of about 10 seconds and the second a cycle of about 4 seconds.

From the simulation example presented earlier it is clear that irregular sampling of a CARMA can give information on cycles that are of higher frequency than the average sampling frequency. It is however, not clear how to interpret the result for the returns in the IBM transaction prices. Even if one believes in efficient market hypothesis of no linear dynamics of prices, the observed prices might show some dynamics due to market micro-structure. A plausible explanation for the high frequency variance is that prices mostly bounce between the bid and ask quotes. A transaction at the ask price is often followed by a transaction at the bid price. Perhaps the New York Stock Exchange market specialist is balancing his portfolio according to some rule. In this example the transaction times are treated as exogenous. The discussion of whether that is realistic is outside the scope of this paper. Engle & Russel (1998) use the same data to illustrate the ACD (Autoregressive-Conditional-Duration) model. They show that the transaction times have a dynamic structure. Simple examination of mean, variance, and quantiles of the durations also reveal that many thousand durations are between 1 and 3 seconds and that the standard deviation of durations divided by the mean duration within a day is about 1.9, suggesting more clustering of transaction than in a Poisson process.

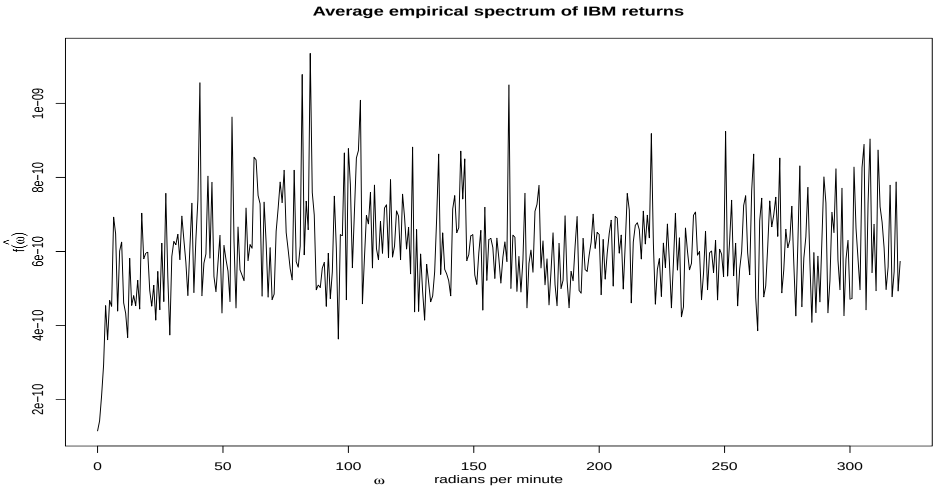

Figure 7: Average of 63 intraday spectral estimates for the returns of IBM.

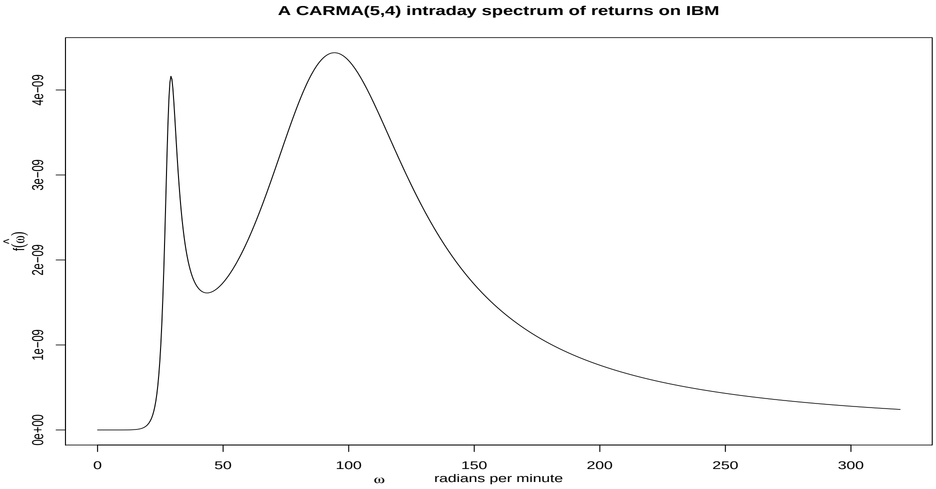

Figure 8: Spectrum of an estimated CARMA(5,4) of IBM returns.

6 Discussion

With access to modern computer and software application of CARMA models is merely a technical implementation. The tools have been scattered in the literature for years. Many of the usual textbook examples of ARMA can easily be analysed with the CARMA tools described in this paper. Application of CARMA models to scientific problems, as any statistical model, requires intuition and understanding of the underlying scientific process. If the process in question is by construction a continuous-time process, issues such as stationarity, and deterministic components have to be addressed. The path of a stationary CARMA(p,q) process is p-differentiable, so at very dense sampling it is a virtually a constant. In the case of the returns in the IBM transaction it seems plausible that the variance does not fade away with increasing sampling frequency, i.e., there is an inherit variation, perhaps due to the bid/ask structure of the data generating process. In the case of the Earth's temperature it is natural to assume that this is a slowly evolving process, i.e., if there was a possibility of high density sampling, a near constant pattern would be measured. If variation of a process is due to cycles of very different frequencies, special measures are necessary due to the fact that many short cycles are observed, but relatively, only a few long cycles are observed. The CARMA representation of very heterogeneous cycles can also cause numerical problems due to very big range of the parameter values. There exist several ways of enforcing the stationarity restriction of the AR parameters so that application of standard numerical software for optimization of the likelihood function can be applied.

Acknowledgements

Part of this work was done while the author visited the Institute for Advanced Studies(IAS/IHS) in Vienna, Austria on an ERASMUS scientist exchange program. The author thanks the IAS/IHS for their hospitality, the participants in their seminars, as well as participants in seminars and conferences in Iceland and Sweden for comments on this work. The computations in this paper are done using standard packages in R, R Development Core Team (2011), the authors FORTRAN progams together with slightly modified versions of EXPOKIT, Sidje (1998), and NUFFT, Greengard & Lee (2004) FORTRAN subroutines. The programs used in this paper are available from the author upon request.

References

Belcher, J., Hampton, J., & Tunnicliffe Wilson, G. (1994). Parameriztion of cont autoregressive models for irregularly sampled time series data. *Journal of the Royal Statistical Association, series B*, *56*(1), 141–155.

Box, G. E. P. & Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control.* Holden Day, San Fransisco.

Brockwell, P., Chadraa, E., & Lindner, A. (2006). Continuous-time GARCH processes. *The Annals of Probability*, *16*(2), 790–826.

Brockwell, P. J. (2009). Levi-driven continuous-time ARMA processes. In T. G. Andersen, R. A. Davis, J.-P. Kreiss, & T. Mikosch (Eds.), *Handbook of Financial Time Series* (pp. 457–480). Springer, Berlin Heidelberg.

Chan, K. & Tong, H. (1987). A note on embedding a discrete time ARMA model in a continuous parameter ARMA model. *Journal of Time Series Analysis*, *8*, 277–281.

Engle, R. F. & Russel, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced data. *Econometrica*, *66*(5), 1127–1162.

Greengard, L. & Lee, J.-Y. (2004). Accelerating the nonuniform fast fourier transform. *SIAM Review*, *46*(3), 443–454.

Jouzel, J. & et al. (2007). Epica dome c ice core 800kyr deuterium data and temperature estimates. IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series # 2007-091. NOAA/NCDC Paleoclimatology Program.

Masry, E. (1978a). Alias-free sampling: An alternative conceptualization and its applications. *IEEE Transactions on Information Theory*, *IT-24*(3), 317–324.

Masry, E. (1978b). Poisson sampling and spectral estimation of continous-time processes. *[IEEE] Transactions on Information Theory*, *IT-24*(2).

Masry, E. (1978c). Poisson sampling and spectral estimation of continuous-time processes. *IEEE Transactions on Information Theory*, *IT-24*(2), 173–183.

Mikosch, T., Gadrich, T., Kluppelberg, C., & Adler, R. J. (1995). Parameter estimation for arma models with infinite variance innovations. *The Annals of Statistics*, *23*(1), pp. 305–326.

Moler, C. & Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *[SIAM] Review*, *45*(1), 1–46.

Monahan, J. F. (1984). A note on enforcing stationarity in autoregressive-moving average models. *Biometrika*, *71*(2), pp. 403–404.

Phadke, M. & Wu, S. (1974). Modeling of continuous stochastic processes from discrete observations with application to sunspots data. *Journal of the American Statistical Association*, *69*(346), 325–329.

Pham, D. T. & Breton, A. L. (1991). Levinson-Durbin-type algorithms for continuous-time autoregressive models and applications. *Mathematics of Control, Signals, and Systems*, *4*(1), 69–79.

Priestley, M. (1963). The spectrum of a continuous process derived from a discrete process. *Biometrika*, *50*, 517–520.

Priestley, M. (1981). *Spectral analysis and time series.* Academic Press.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rice, S. (1954). *Mathematical analysis of random noise. Monograph B-1589.* Bell Telephone Labs Inc., New York.

Shoji, I. & Ozaki, T. (1998). A statistical method of estimation and simulation for systems of stochastic differential equations. *Biometrika*, *85*, 240–243.

Sidje, R. B. (1998). EXPOKIT. A software package for computing matrix exponentials. *ACM Trans. Math. Softw.*, *24*(1), 130–156.

Sun, T. & Chaika, M. (1997). On simulation of a Gaussian stationary process. *Journal of Time Series Analysis*, *18*(1), 79–93.

Tsai, H. & Chan, K. (2000). A note on the covariance structure of a continuous-time ARMA process. *Statistica Sinica*, *10*, 989–998.

Tsai, H. & Chan, K. (2003). A note on parameter differentiation of matrix exponentials, with application to continuous time modelling. *Bernoulli*, *9*(5), 895–919.

Tsay, R. S. (2010). *Analysis of Financial Time Series* (Third ed.). John Wiley & Sons.

Author: Helgi Tómasson

Title: Some Computational Aspects of Gaussian CARMA Modelling

Reihe Ökonomie / Economics Series 274

Editor: Robert M. Kunst (Econometrics)
Associate Editors: Walter Fisher (Macroeconomics), Klaus Ritzberger (Microeconomics)

ISSN: 1605-7996
© 2011 by the Department of Economics and Finance, Institute for Advanced Studies (IHS),
Stumpergasse 56, A-1060 Vienna • ☎ +43 1 59991-0 • Fax +43 1 59991-555 • http://www.ihs.ac.at

ISSN: 1605-7996