Frequency Domain Identification of Continuous-Time ARMA Models: Interpolation and Non-uniform Sampling

Jonas Gillberg, Lennart Ljung

Division of Automatic Control
Department of Electrical Engineering
Linköpings universitet, SE-581 83 Linköping, Sweden
WWW: http://www.control.isy.liu.se
E-mail: gillberg@isy.liu.se, ljung@isy.liu.se

27th September 2004

Report no.: LiTH-ISY-R-2625
Submitted to

Technical reports from the Control & Communication group in Linköping are available at http://www.control.isy.liu.se/publications.
Abstract
In this paper is discussed how to estimate irregularly sampled continuous-time ARMA models in the frequency domain. In the process, the model output signal is assumed to be piecewise constant or piecewise linear, and an approximation of the continuous-time Fourier transform is calculated. ML-estimation in the frequency domain is then used to obtain parameter estimates.

Keywords:
Frequency Domain Identification of Continuous-Time ARMA Models: Interpolation and Non-uniform Sampling

Jonas Gillberg, Lennart Ljung

2004-09-27

Abstract

In this paper is discussed how to estimate irregularly sampled continuous-time ARMA models in the frequency domain. In the process, the model output signal is assumed to be piecewise constant or piecewise linear, and an approximation of the continuous-time Fourier transform is calculated. ML-estimation in the frequency domain is then used to obtain parameter estimates.

1 Introduction

Approaching parameter estimation from the discrete-time domain has been the dominating paradigm in system identification. Partly because the mathematics is straightforward and partly because methods are easily implemented on digital computers[8]. In the black-box discrete-time framework the identified model parameters often lack physical relevance. This is not a problem in some applications, for instance the synthesis of regulators. Modeling of physical systems on the other hand is often performed in continuous-time, and in many applications there is a genuine interest in the physical parameters.

Uniform sampling has also been a standard assumption. A single sensor delivered measurements at a constant rate. With the advent of networked asynchronous sensors this assumption has changed. In fields such as economics and finance uniform sampling might not be practically possible.

One route to obtain continuous-time parameters is via identification of a discrete time model. This, so called indirect approach, has proved to be sensitive to the size of the sampling interval. Numerical ill-conditioning can occur for very small sample intervals since the discrete-time poles gather around 1 in the complex plane[7].

1.1 Problem Formulation

In this report we take the so called direct approach, where the continuous-time parameters are identified directly from sampled data. This approach has been used in the time domain for non-uniformly sampled continuous-time ARMA (CARMA) models[3]. We discuss how to estimate an irregularly sampled continuous-time ARMA model in the frequency domain. In the process,
the model output is assumed to be piecewise constant or piecewise linear, and an approximation of the continuous-time Fourier transform is calculated. ML-estimation in the frequency domain is then used to obtain parameter estimates. Rules of thumb concerning the model bias and variance are derived and most important used in order to select the frequencies to be used in estimation.

1.2 Outline

The report will be structured as follows. In section 1 we will introduce the CARMA model together with basic properties for continuous-time stationary processes. We will then, in section 2, address the topic of spectral estimation and the use of interpolation in order to approximate the continuous time Fourier transform. We will also treat the effects of small sampling intervals on the spectral bias. After that, in section 4 and 5 we describe the maximum-likelihood approach in the time and most important in the frequency domain. In the following section 6 expression relating the spectral bias to parameter bias are derived. Asymptotic expressions for the variance can also be found. These expressions are used in section 7 in order to provide users with rules of thumb on which frequencies to choose. Finally, the preceding arguments are exemplified by numerical experiments in section 8. Comments on conclusions, further research and acknowledgments follow in section 9, 10 and 11.
2 Modeling and Simulation

In this section the basic model and data setup are introduced. First we present the continuous-time autoregressive moving average (CARMA) model in an informal manner. Then we more formally introduce the model in state space form. We also explain how this stochastic continuous-time dynamical system is simulated and how measurements are taken.

2.1 Continuous-Time ARMA Model

The continuous-time autoregressive moving average (CARMA) model can informally be described as

\[ y_t = \sigma^2 \frac{B(p)}{A(p)} e_t \]  \hspace{1cm} (1)

where \( e_t \) is continuous time white noise. This means that \( e_t dt = dw_t \) where \( w_t \) is a continuous-time Wiener process with

\[ E[ dw_t ] = 0 \]
\[ E[ dw_t dw_s ] = \begin{cases} dt & t = s \\ 0 & t \neq s \end{cases} \]

The operator \( p \) is here the differentiation operator while

\[ A(p) = p^n + a_1 p^{n-1} + a_2 p^{n-2} + \cdots + a_n \]
\[ B(p) = p^m + b_1 p^{m-1} + \cdots + b_m. \]

We denote the vector of parameters with \( \theta = [ a_1 \ a_2 \ \ldots \ a_n \ b_1 \ b_2 \ \ldots \ b_m \ \lambda]^T \) where \( \lambda = \sigma^2 \).

2.2 State-Space Representation

The model in (1) can be formally represented in a state-space controller canonical form

\[
\begin{align*}
    dx_t &= Ax_t dt + Bde_t \\
y_t &= Cx_t
\end{align*}
\]  \hspace{1cm} (2)

where
\[
A = \begin{bmatrix}
-a_1 & -a_2 & \cdots & -a_{n-1} & -a_n \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0
\end{bmatrix}
\]
\[
B = [\sigma \ 0 \ 0 \ \ldots \ 0]^T
\]
\[
C = [1 \ b_1 \ b_2 \ \ldots \ b_m].
\]

The exact solution to the state-space representation in (2) can be written as

\[ y_t = Ce^{At}x_0 + \int_0^t Ce^{A(t-s)}Bde_t. \]  \hspace{1cm} (3)
where the integral is in the sense of Ito [5]. If $y_t$ is to be a zero mean stationary Gaussian process the initial values $x_0$ must be Gaussian and distributed such that $E[x_0] = 0$ and $Q_0 = Cov[x_0]$ satisfies the Lyapunov equation

$$AQ_0 + Q_0A^T + \sigma^2 BB^T = 0.$$ 

For the rest of this report we will assume that the process $y_t$ is stationary Gaussian with zero mean value [2].

Finally we repeat the definition of the covariance function

$$r(\tau) =Ey_{t+\tau}y_t$$

and power spectrum

$$\Phi(i\omega) = \sigma^2B(i\omega)B(-i\omega)\over A(i\omega)A(-i\omega).$$

of the stationary process modeled in (1). Moving between spectrum $\Phi(i\omega)$ and covariance $r(\tau)$ representations of the process is facilitated by the well-known formulas

$$\Phi(i\omega) = \int_{-\infty}^{\infty} r(\tau)e^{-i\omega\tau}d\tau$$

and

$$r(\tau) = {1 \over 2\pi} \int_{-\infty}^{\infty} \Phi(i\omega)e^{i\omega\tau}d\omega.$$ (4)

Assume that we only observe the process $y_t$ at the discrete time instances $t = kT_s$. Then the covariance function of the discrete-time process $y_{kT_s}$ will be

$$r_d(k) = r(kT_s),$$

and the discrete-time power spectrum will be

$$\Phi_d(z) = T_s \sum_{k=-\infty}^{\infty} r_d(k)z^{-k}.$$ 

Finally the following relationship, similar to the Poisson summation formula, exists between the discrete-time and continuous-time spectrums [10]

$$\Phi_d(z) = \sum_{k=-\infty}^{\infty} \Phi(\log z + i2\pi k/T_s).$$

It can also be written as

$$\Phi_d(e^{i\omega T_s}) = \sum_{k=-\infty}^{\infty} \Phi(i\omega + i2\pi k/T_s).$$ (5)
3 Estimation of Power Spectrum

As mentioned earlier the parameter estimation method is divided into two stages. First the power spectrum is estimated and in a second stage the spectrum is used to estimate the model parameters. In the first stage we resort to interpolation in order to obtain an approximation of a continuous-time realization of the output. It turns out that this approach also interpolates the covariance function in two dimensions. The Fourier transform of the interpolated realization can then be computed and an estimate of the spectrum can be formed. The bias in this estimate can then be appreciated.

3.1 The Fourier Transform

Let us define Fourier transformation of the continuous time output \( \{ y_t : t \in [0, T] \} \) as

\[
Y_T(i\omega) = \frac{1}{\sqrt{T}} \int_0^T y_t e^{-i\omega t} dt. \tag{6}
\]

The values obtained for \( \omega = \frac{2\pi k}{T}, k = -\infty, \cdots -1, 0, 1, \cdots \infty \) are proportional to the familiar coefficients of the Fourier Series. Since \( y_t \) is a Gaussian process and \( Y(i\omega) \) is a linear combination of \( y_t \), it will be Gaussian too.

3.2 Interpolating the Continuous-Time Realization

A complicating element in the estimation procedure is that we do not have access to the entire continuous time realization of the output. Instead we have, as we pointed out earlier, a finite number of samples of the continuous output \( y_t \) at time instances \( \{ t_1, t_2, \ldots, t_N \} \). Therefore it is in some way necessary to approximate or reconstruct the continuous time realization. In this report we reconstruct the output as

\[
\hat{y}_T^{k}(t) = \sum_{i=1}^{N} y_{t_i} \phi_{k}^{i}(t - t_i) \tag{7}
\]

where \( T \) is the time the process is observed, \( k \in \{-1, 0, 1\} \) is the order of interpolation and \( \phi \) is the interpolation kernels. Three types of kernels are used in this report. First we introduce

\[
\phi_{k}^{-1}(t) = (t_{i+1} - t_i) \delta(t - t_i) \tag{8}
\]

which we term “Riemann interpolation”. Next we introduce the piecewise-constant interpolation

\[
\phi_{0}^{i}(t) = \begin{cases} 
0 & t < t_i \\
1 & t_i \leq t \leq t_{i+1} \\
0 & t_{i+1} < t
\end{cases}
\]
which will often go under the name Zero-Order Hold (ZOH). Finally we have piecewise linear interpolation

\[
\phi^i_t(t) = \begin{cases} 
0 & t < t_{i-1} \\
\frac{t - t_{i-1}}{t_i - t_{i-1}} & t_{i-1} \leq t < t_i \\
\frac{t_i - t}{t_{i+1} - t} & t_i \leq t < t_{i+1} \\
0 & t_{i+1} \leq t
\end{cases}
\]

which is usually termed First-Order Hold (FOH).

### 3.3 Transforming the Interpolated Output

From the interpolated output it is possible to compute the Fourier transform

\[
\hat{Y}_T = \frac{1}{\sqrt{T}} \int_0^T \hat{y}^{T,k}(t)u(t)e^{-i\omega t}dt
\]

where \( u(t) \) will be called the sampling function. If we sample in the discrete points \( t_k \) then

\[
u(t) = \sum_{k=-\infty}^{\infty} \delta(t - t_k).
\]

If we sample uniformly at \( t_k = kT_s \) we get

\[
u(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT_s)
\]

and if we sample continuously

\[
u(t) = 1.
\]

Since multiplication in the time-domain is equivalent to convolving in the frequency domain the transform will be

\[
\hat{Y}_T(i\omega) = \int_{-\infty}^{\infty} \hat{Y}(i(\omega - v))K(iu)du
\]

where

\[
K(i\omega) = \int_{-\infty}^{\infty} u(t)e^{-i\omega t}dt
\]

and

\[
\hat{Y}(i\omega) = \int_{-\infty}^{\infty} \hat{y}(t)e^{-i\omega t}dt.
\]

If we sample the interpolated signal equidistantly we will get

\[
K(i\omega) = \sum_{k=-\infty}^{\infty} \delta(w - \frac{2\pi}{T_s}k)
\]
and

\[ \hat{Y}_T(i\omega) = \sum_{k=-\infty}^{\infty} \hat{Y}(i(w - \frac{2\pi}{T_s}k)) \]

From this we see that the only effect of re-sampling is that we fold the transform around the sampling frequency.

3.4 Fourier Transforms for Continuous Sampling of Interpolated Output

If we sample the interpolated output \( \hat{y}(t) \) continuously the “Riemann interpolation” case will yield a transform

\[ \hat{Y}_{(-1)}(i\omega) = \frac{1}{\sqrt{T}} \sum_{k=1}^{N-1} (t_{k+1} - t_k) y(t_k) e^{i\omega t_k} \]

In the piecewise constant case it will be

\[ \hat{Y}_{(0)}(i\omega) = \frac{1}{\sqrt{T}} \sum_{k=1}^{N-1} y(t_k) \frac{e^{i\omega t_{k-1}} - e^{i\omega t_k}}{i\omega} \]

while in the piecewise linear case we have

\[ \hat{Y}_{(1)}(i\omega) = \frac{1}{\sqrt{T}} \frac{1}{i\omega} (y(t_1) e^{i\omega t_1} - y(t_N) e^{i\omega t_N}) + \frac{1}{\sqrt{T} (i\omega)^2} \sum_{k=1}^{N-1} y(t_{k+1}) - y(t_k) \frac{e^{i\omega t_{k+1}} - e^{i\omega t_k}}{t_{k+1} - t_k} \]

This might seem as an awkward and expensive way of computing the Fourier transform. If we however do this on-line, we can restrict us to a time window of size \( T \) and just remove old data as new samples arrive.

3.5 Spectrum Estimates for Uniform Sampling of Output

In the case of uniform sampling \( t_k = kT_s \) the Fourier transform of the interpolated data will be

\[ \hat{Y}_{(-1)}(i\omega) = \frac{1}{\sqrt{T}} \int_0^T \sum_{i=1}^N y_{kT_s} \phi^i(t) e^{-i\omega t} dt \]  

\[ = F_k(i\omega) \frac{1}{\sqrt{N_t}} \sum_{k=1}^{N_t} y_{kT_s} e^{-i\omega kT_s} \]

where

\[ F_k(i\omega) = \frac{1}{\sqrt{T_s}} \int_{-\infty}^{\infty} \phi^k(t) e^{-i\omega t} dt. \]
The ‘Riemann interpolation’ will yield the ordinary discrete-time Fourier transform

\[ \hat{Y}_T^{(-1)}(i\omega) = \frac{\sqrt{T_s}}{\sqrt{N_t}} \sum_{k=1}^{N_t-1} y(kT_s)e^{i\omega kT_s} \]

which is related to the discrete-time power spectrum

\[ E\hat{\Phi}(i\omega) = \lim_{T \to \infty} E|\hat{Y}_T^{(-1)}(i\omega)|^2 = \Phi_d(e^{i\omega T_s}) = \sum_{k=-\infty}^{\infty} \Phi(i\omega + i\frac{2\pi}{T_s}k). \]

Zero-order hold and first order hold sampling on the other hand will provide filtered versions of the discrete-time Fourier transform since

\[ \hat{F}_0(i\omega) = T_s \text{sinc}\left(\frac{\omega T}{2}\right) \]
\[ \hat{F}_1(i\omega) = T_s \text{sinc}^2\left(\frac{\omega T}{2}\right) \]

and hence

\[ E\hat{\Phi}_{ZOH}(i\omega) = \text{sinc}^2\left(\frac{\omega T}{2}\right) \sum_{k=-\infty}^{\infty} \Phi(i\omega + i\frac{2\pi}{T_s}k) \]
\[ E\hat{\Phi}_{FOH}(i\omega) = \text{sinc}^4\left(\frac{\omega T}{2}\right) \sum_{k=-\infty}^{\infty} \Phi(i\omega + i\frac{2\pi}{T_s}k). \]

The objective of the interpolation is to make the estimate consistent with the continuous-time power spectrum. Therefore we would like to remove the effects of folding. In view of what has been said earlier with respect to interpolation and filtering we realize that the optimal interpolation kernel \( \phi^{OPT} \) should have a power spectrum which is

\[ |F_{OPT}(i\omega)|^2 = \frac{\Phi(i\omega)}{\sum_{k=-\infty}^{\infty} \Phi(i\omega + i\frac{2\pi}{T_s}k)} \] (12)

where \( \Phi \) is the power spectrum for the true parameters. In Figure 1 we have compared the spectrum of the optimal kernel with ZOH and FOH. From this perspective we notice that FOH as an approximation of the optimal kernel seems to be worse than ZOH. This will also be confirmed by experiments later on.

### 3.6 Approximate Covariance Function

The approximation \( \hat{y}_{T,k}^{\tau}(t) \) of a realization of the stationary process \( y_t \) will have the following second order properties

\[ \hat{n}_T^{\tau,k}(t) \triangleq E[\hat{y}(t)] = 0 \]
\[ \hat{r}_T^{\tau,k}(t, s) \triangleq E[(\hat{y}(t) - \hat{n}(t))(\hat{y}(s) - \hat{n}(s))] \]
\[ = \sum_{i=1}^{N} \sum_{j=1}^{N} r(t_i - t_j) \phi_i^k(t) \phi_j^k(s) \] (13)
Figure 1: Kernel Spectrums for $T_s = 0.1$. The optimal spectrum is for a second order CAR model where $A(p) = p^2 + 2p + 1$, $B(p) = 1$ and $\sigma = 1$.

Where $r$ is the covariance function defined in (4).

An interesting question is how well $\hat{r}_{T,k}$ approximates $R$. From the representation in (14) we see that piecewise constant interpolation of $y(t)$ will result in a piecewise constant interpolation of $r(t - s)$. Piecewise linear interpolation of $y(t)$ on the other hand, will result in piecewise bilinear interpolation of $r(t - s)$.

### 3.6.1 Interpolation in Two Dimensions

In the following lemma we explain how we can bound the interpolation error.

**Lemma 3.1** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a continuous once differentiable function with bounded first derivatives. Assume that the function is interpolated by a constant function $\hat{f} : \mathbb{R}^2 \to \mathbb{R}$ in a rectangle $\Omega_{ij} = \{x, y : 0 \leq x \leq h_i, 0 \leq y \leq h_j\}$. Then we have for some constant $C > 0$

$$\max_{(x,y) \in \Omega_{ij}} |f(x, y) - \hat{f}(x, y)| \leq Ch_{\max}$$  \hspace{1cm} (15)

If we interpolate $f$ with a bilinear polynomial and assume it is twice differentiable with bounded second derivatives we get for some $C > 0$

$$\max_{(x,y) \in \Omega_{ij}} |f(x, y) - \hat{f}(x, y)| \leq Ch_{\max}^2$$ \hspace{1cm} (16)

**Proof:** Both results come from Taylor's theorem. In the case of constant interpolation we have

$$f(x, y) - f(0, 0) = f_x(0, 0)x + f_y(0, 0)y + O(x^2 + y^3)$$ \hspace{1cm} (17)

In the bilinear case we have the interpolant

$$\hat{f}(x, y) = f(0, 0) + \frac{f(h_1, 0) - f(0, 0)}{h_1} x + \frac{f(0, h_2) - f(0, 0)}{h_2} y$$ \hspace{1cm} (18)

$$+ \frac{f(h_1, h_2) - f(h_1, 0) - f(0, h_2) + f(0, 0)}{h_1 h_2} xy$$ \hspace{1cm} (19)
Define $\tilde{f}(x, y) = f(x, y) - \check{f}(x, y)$. Then
\[
\tilde{f}(x, y) = f(0, 0) + f_x(0, 0)h_1 + f_y(0, 0)h_2 + \frac{1}{2} \left( x \begin{pmatrix} f_{xx}(0, 0) & f_{xy}(0, 0) \\ f_{yx}(0, 0) & f_{yy}(0, 0) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \mathcal{O}(\sqrt{x^2 + y^2}) \right)
\]
\[
+ \frac{1}{2} \left( x \begin{pmatrix} \tilde{f}_{xx}(0, 0) & \tilde{f}_{xy}(0, 0) \\ \tilde{f}_{yx}(0, 0) & \tilde{f}_{yy}(0, 0) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \mathcal{O}(\sqrt{x^2 + y^2}) \right)
\]
Now
\[
\tilde{f}(0, 0) = 0
\]
\[
\tilde{f}_x(0, 0) = f_x(0, 0) - f(0, 0) = \mathcal{O}(h_1^2)
\]
\[
\tilde{f}_y(0, 0) = \mathcal{O}(h_2^2)
\]
Similarly we have
\[
\tilde{f}_{xx}(0, 0) = f_{xx}(0, 0)
\]
\[
\tilde{f}_{yy}(0, 0) = f_{yy}(0, 0)
\]
\[
\tilde{f}_{xy}(0, 0) = f_{xy}(0, 0) - f(h_1, h_2) - f(h_1, 0) - f(0, h_2) + f(0, 0)
\]
\[
= \frac{\mathcal{O}(\sqrt{h_1^2 + h_2^2})}{h_1 h_2}
\]
Thus we have
\[
\tilde{f}(x, y) = \mathcal{O}(h_1^2) x + \mathcal{O}(h_2^2) y + \frac{1}{2} \left( x \begin{pmatrix} f_{xx}(0, 0) & f_{xy}(0, 0) \\ f_{yx}(0, 0) & f_{yy}(0, 0) \end{pmatrix} \frac{\mathcal{O}(\sqrt{h_1^2 + h_2^2})}{h_1 h_2} \begin{pmatrix} x \\ y \end{pmatrix} + \mathcal{O}(\sqrt{x^2 + y^2}) \right)
\]
\[
+ \mathcal{O}(\sqrt{x^2 + y^2})
\]
From this we get the inequality
\[
|\tilde{f}(x, y)| \leq \frac{1}{2} |f_{xx}(0, 0)| h_1^2 + \frac{1}{2} |f_{yy}(0, 0)| h_2^2 + \mathcal{O}(\sqrt{h_1^2 + h_2^2})
\]
\[
\leq \frac{|f_{xx}(0, 0)| + |f_{yy}(0, 0)|}{2} h_{\text{max}}^2 + \mathcal{O}(h_{\text{max}}^3)
\]
This result can be used to bound the covariance function approximation error and its contribution to the bias in the periodogram.

3.6.2 Covariance Function Interpolation Error

From Lemma 3.1 we can now prove that the interpolation error in the whole area $\Omega = \{ t, s : 0 \leq t \leq T, 0 \leq s \leq T \}$ is bounded. Let
\[
h_{\text{max}} = \max_{1 \leq i < N_1-1} (t_{i+1} - t_i).
\]
Then we have the following result.
Corollary 3.1 Assume that \( r \) is bounded, continuous and has bounded first derivatives. Let \( \hat{r}^{T,0} \) be defined as in (13) then

\[
\max_{t,s \in [0,T]} |r(t-s) - \hat{r}^{T,0}(t,s)| \leq C h_{\text{max}}
\] (34)

Proof: Follows directly from Lemma 3.1.

In the case of piecewise linear interpolation of \( y_t \) there will be a bilinear interpolation of \( r(t-s) \) where the approximation error is given by the following lemma

Lemma 3.2 Assume that \( r_y(t,s) \) is bounded, continuous and has bounded second derivatives. Let \( \hat{r}^{T,1} \) be defined as in (13) then

\[
\max_{t,s \in [0,T]} |r(t-s) - \hat{r}^{T,1}(t,s)| \leq C h_{\text{max}}^2
\] (35)

Proof: Follows directly from Lemma 3.1.

We will now continue to treat the bias in the periodogram.

3.7 Periodogram Bias

From (6) and (7) we can compute the approximate periodogram which we denote as

\[
\hat{\Phi}^{T,k}(i\omega) = \left| \hat{Y}_k^T(i\omega) \right|^2
\] (36)

The periodogram which is an estimate of the power spectrum will be biased due to interpolation and leakage effects. This bias as we will see later on translates into the bias in the parameter estimate.

3.7.1 Interpolation Contribution to the Periodogram Bias

The following expression relates the bias in the approximate periodogram to the error from the interpolation of the covariance function.

Lemma 3.3 Let \( y(t) \) be a stationary stochastic process. Let \( \hat{\Phi}^T_y(i\omega) \) be defined by (51) and let \( \hat{\Phi}^{T,k}_y(i\omega) \) be defined by (36). Then, for

\[
\Delta \hat{\Phi}_1 = \left| E \left[ \hat{\Phi}^{T,k}(i\omega) - \hat{\Phi}^{T}(i\omega) \right] \right|
\] (37)

we have

\[
\Delta \hat{\Phi}_1 \leq C_1 \max_{t,s \in [0,T]} |r(t-s) - \hat{r}^{T,k}(t,s)|
\] (38)

Proof: From the definitions we get

\[
\Delta \hat{\Phi}_1 = \left| E[|Y_k^T(i\omega)|^2 - E|Y_T(i\omega)|^2] \right|
\]

\[
= \left| \frac{1}{T} \int_0^T \int_0^T \left( \hat{r}^{T,k}(t,s) - r(t-s) \right) e^{i\omega_\alpha(t-s)} dt ds \right|
\]

\[
\leq \frac{1}{T} \int_0^T \int_0^T |\hat{r}^{T,k}(t,s) - r(t-s)| dt ds.
\]
Let $\Delta r(t, s) = |\hat{r}^T(t, s) - r(t - s)|$ and make a change of variables. Then we get

$$
\frac{1}{T} \int_0^T \int_0^T \Delta r(t, s) dt ds = \frac{1}{T} \int_0^T \int_{-t}^{T-t} \Delta r(t, t + \tau) d\tau dt
$$

$$
= \frac{1}{T} \int_0^T \int_{\max\{-a, -t\}}^{\min\{a, T-t\}} \Delta r(t, t + \tau) d\tau dt
$$

$$
+ \frac{1}{T} \int_0^T \int_{-t}^{-a} \Delta r(t, t + \tau) d\tau dt
$$

$$
+ \frac{1}{T} \int_0^T \int_{a}^{T-t} \Delta r(t, t + \tau) d\tau dt
$$

These terms can be separately bounded. For the first term we have

$$
\frac{1}{T} \int_0^T \int_{\max\{-a, -t\}}^{\min\{a, T-t\}} \Delta r(t, t + \tau) d\tau dt \leq \frac{1}{T} \int_0^T \int_{\max\{-a, -t\}}^{\min\{a, T-t\}} C h_{max} d\tau dt
$$

$$
\leq \alpha C h_{max}
$$

For some $\alpha > 0$ and $\lambda > 0$ we have

$$
\Delta r(\tau) \leq |r(\tau)| + |\hat{r}(t, t + \tau)| < e^{-\lambda|\tau|}.
$$

The second term can then be bounded the following way

$$
\frac{1}{T} \int_{\alpha}^{T} \int_{-t}^{-a} \Delta r(t, t + \tau) d\tau dt \leq \frac{1}{T} \int_{\alpha}^{T} \int_{-t}^{-a} e^{-\lambda|\tau|} d\tau dt
$$

$$
\leq \int_{-\infty}^{-a} e^{-\lambda|\tau|} d\tau.
$$

We then choose $\alpha > 0$ such that

$$
\alpha C h_{max} \geq \int_{-\infty}^{-a} e^{-\lambda|\tau|} d\tau
$$

and we have our result. This is the first part of the bias. The second part comes from the leakage.

### 3.7.2 Leakage Contribution to Bias

Since we only observe our process during a finite time interval $[0, T]$ the expected periodogram and the spectrum will be slightly different. The following lemma by Ljung [4] quantifies this difference.

**Lemma 3.4** Let $y(t)$ be a stationary stochastic process with spectrum $\Phi_y$ and let $\hat{\Phi}_y(i\omega)$ be defined by (51), then

$$
\Delta \hat{\Phi}_2 = \left|E \hat{\Phi}^T(i\omega) - \Phi(i\omega)\right| \leq \frac{C^2}{T} \tag{39}
$$
Proof: Lemma 6.1 in [4].

\[ EY_t(i\omega)Y_T(-i\xi) = \frac{1}{T} \int_0^T \int_0^T E_y(r)y(s)e^{i\omega(r-s)}dsdr \]
\[ = \frac{1}{T} \int_0^T \int_0^T r(t-s)e^{-i\omega(t-s)}dsdt \]
\[ = \frac{1}{T} \int_0^T e^{-i(\omega-\xi)t} \int_{t-T}^t r(\tau)e^{-i\xi\tau}d\tau d\tau \]

Now

\[ \int_{t-T}^t r(\tau)e^{-i\xi\tau}d\tau = \Phi_y(i\xi) - \int_{-\infty}^{t-T} r(\tau)e^{-i\xi\tau}d\tau \]
\[ = \int_{t}^{T} R(\tau)e^{-i\xi\tau}d\tau \]

and

\[ \frac{1}{T} \int_0^T e^{-i(\omega-\xi)t}dt = \begin{cases} 1, & \text{if } \omega = \xi \\ 0, & \text{if } (\omega - \xi) = \frac{2\pi}{T} k, \quad k = \pm 1, \pm 2, \ldots, \pm \infty \end{cases} \]

Consider

\[ \left| \frac{1}{T} \int_0^T e^{-i(\omega-\xi)t} \int_{-\infty}^{t-T} r(\tau)e^{-i\xi\tau}d\tau d\tau \right| \leq \frac{1}{T} \int_0^T \int_{-\infty}^{t-T} |r(\tau)| d\tau d\tau \]
\[ \leq \frac{1}{T} \int_{-\infty}^{0} |\tau| |r(\tau)| d\tau \leq \frac{C}{T} \]

provided

\[ \int_{-\infty}^{\infty} |\tau r_y(\tau)| d\tau \]

Similarly

\[ \left| \frac{1}{T} \int_0^T e^{-i(\omega-\xi)t} \int_{-\infty}^{t-T} r(\tau)e^{-i\xi\tau}d\tau d\tau \right| \leq \frac{1}{T} \int_0^{\infty} |\tau| |r(\tau)| d\tau \leq \frac{C}{T} \]

\[ \quad \] \halmos

3.7.3 Power Spectrum Bias Expression

The previous two lemmas can now be used to estimate the difference between the expected approximate periodogram and the spectrum

**Theorem 3.1** If we define

\[ \Delta \hat{\Phi} = \left| E\hat{\Phi}^{T,k} - \Phi \right| \]

then we have

\[ \Delta \hat{\Phi} \leq C_1 h_{\max}^{k+1} + \frac{C_2}{T} \]

(40)
Proof: Application of the triangle inequality, Lemma 3.4 and Lemma 3.3 yields

\[ \Delta \hat{\Phi} = \left| E \hat{\Phi}^{T,k} - \Phi \right| \]
\[ \leq \left| E \hat{\Phi}^{T,k} - E \hat{\Phi}^T + E \hat{\Phi}^T - \Phi \right| \]
\[ \leq \left| E \left[ \hat{\Phi}^{T,k} - \hat{\Phi}^T \right] \right| + \left| E \left[ \hat{\Phi}^T - \Phi \right] \right| \]
\[ \leq C_1 \max_{r,s \in [0,T]} \left| \hat{R}^{T,k}(r,s) - R(r-s) \right| + \frac{C_2}{T} \]

We will now turn to the procedure of estimating parameters from the approximate periodogram.
4 Maximum Likelihood Estimation in the Time Domain

A set of samples $y(t_k), k = 1 \ldots N_t$ of the stationary output of the CARMA process in (1) will be distributed as

$$Y = \begin{pmatrix} y(t_1) \\ \vdots \\ y(t_{N_t}) \end{pmatrix} \in N(0, R(\theta))$$

where

$$R(\theta) = \begin{pmatrix} r(t_1 - t_1, \theta) & \ldots & r(t_1 - t_{N_t}, \theta) \\ r(t_2 - t_1, \theta) & \ldots & r(t_2 - t_{N_t}, \theta) \\ \vdots & \vdots & \vdots \\ r(t_{N_t} - t_1, \theta) & \ldots & r(t_{N_t} - t_{N_t}, \theta) \end{pmatrix}.$$

For the stationary case [3]

$$r(\tau, \theta) = Ce^{A\tau}PC^T$$

where

$$AP + PA + \sigma^2BB^T = 0. \quad (41)$$

This paves the way for straightforward Maximum-Likelihood Estimation by the criteria

$$\hat{\theta} = \arg \min_{\theta} Y^T R(\theta)^{-1} Y + \log \det R(\theta) \quad (42)$$

The negative log-likelihood function of this distribution will be

$$- \log p(Y|\theta) = \frac{N_t}{2} \log 2\pi + \frac{1}{2} \log \det R(\theta) + \frac{1}{2} Y^T R(\theta)^{-1} Y.$$

The Maximum Likelihood (ML) method for estimating the parameters would hence be

$$\hat{\theta} = \arg \min_{\theta} Y^T R(\theta)^{-1} Y + \log \det R(\theta) \quad (43)$$

A big obstacle is that we have to compute $R(\theta)^{-1}Y$ and retrieve the eigenvalues of $R(\theta)$ in order to get $\log \det R(\theta)$. 

15
4.1 Efficient Computation

The matrix $R(\theta)$ will have an almost banded structure since $r(\tau) \to 0$ as $e^{-\lambda \tau}$ for some $\lambda$. Hence we can effectively set all $r(\tau) = 0$ for $\tau > B$ for some bandwidth $B$. This will yield an $R(\theta)$ which is exactly banded and algorithms optimized for such matrices can be used.

Since the new matrix $R(\theta)$ is symmetric, positive definite and banded it is possible to perform a banded $LDL^T$ factorization which is a special formulation of the Cholesky factorization

$$R(\theta) = L(\theta)D(\theta)L^T(\theta).$$ (44)

The objective function in the optimization could then be rewritten as

$$\hat{\theta} = \arg\min_{\theta} \|L(\theta)^{-1}Y\|_{D(\theta)}^2 + \sum_{k=1}^{N_t} \log D_{kk}(\theta)$$

where the $D_{kk}(\theta)$ are the eigenvalues $\lambda_k(R(\theta))$. This approach tends to work only if we have a medium or small number of measurements $N_t$ since the size of $R(\theta)$ grows as $N_t \times N_t$.

4.2 Cramer-Rao Lower Bound

A property of the maximum likelihood estimator is that it under certain conditions is consistent i.e.

$$\lim_{N_t \to \infty} \hat{\theta} \to \theta^*$$

and the covariance matrix of the estimated parameters is bounded as

$$P = E(\hat{\theta} - \theta^*)(\hat{\theta} - \theta^*)^T \geq M^{-1}.$$ Here

$$M = -E \frac{d^2}{d\theta^2} \log p(Y|\theta^*)$$

is known as the Fisher information matrix and the bound is known as the Cramer-Rao Lower Bound (CRLB)[4].

In the case that the distribution is Gaussian it is possible to give a more explicit expression for the CRLB known as the Slepian-Bang formula [9].

$$M_{ij} = \frac{1}{2} \text{Tr} R(\theta)^{-1} \frac{\partial R(\theta)}{\partial \theta_i} R(\theta)^{-1} \frac{\partial R(\theta)}{\partial \theta_j}$$

This expression can and will be used later on in order to evaluate the efficiency of different parameter estimators for a fairly large number of samples. We will now move on and consider a ML method in the frequency domain.
5 Maximum Likelihood Estimation in the Frequency Domain

When a continuous time measurement \( y(t), \ t \in [0, T] \) of the output is available the parameters of CARMA models can be estimated by solving the following minimization problem

\[
\hat{\theta} \triangleq \arg \min_{\theta} \sum_{k=1}^{N_\omega} \frac{\hat{\Phi}(i\omega_k)}{\Phi(i\omega_k, \theta)} + \log \Phi(i\omega_k, \theta) \tag{45}
\]

where the frequencies \( \omega_k, k = 1, \ldots, N_\omega \) are chosen such that

\[
\omega_k = \frac{2\pi}{T} l l \in \mathbb{Z} \tag{46}
\]

This is a maximum-likelihood procedure where the model has been transformed into the frequency domain. The frequencies have been deliberately selected such that the Fourier transforms of the output at different frequencies are uncorrelated. If we compare the objective function with the time-domain expression we see that the quadratic form in the time-domain criteria have been replaced with as summation in the frequency domain. Hence the, complexity has been greatly reduced.

The criteria here is the negative log-likelihood function for the real and imaginary parts of the periodogram given the parameters \( \theta \). Most of the material in this section can be found in classical textbooks such as the one by Brillinger [1] but we provide a different perspective on the problem.

5.1 Frequency Domain Model

In this subsection we show what the result will be if we apply the Fourier transform

\[
Y_T(i\omega) = \frac{1}{\sqrt{T}} \int_0^T y_t e^{-i\omega t} dt
\]

to the stochastic process \( y_t \) defined in (1). This will produce a set of Gaussian complex-valued stochastic variables \( Y(i\omega) \).

**Theorem 5.1** Assume that we have stochastic process \( y_t \) generated by the CARMA model in (1). Then the Fourier transform of the process from time \( t = 0 \) to \( t = T \) will be

\[
Y(i\omega) = \sigma \frac{B(i\omega)}{A(i\omega)} \frac{1}{\sqrt{T}} \int_0^T e^{-i\omega t} dt + \frac{1}{\sqrt{T}} C(i\omega I - A)^{-1}(x_0 - e^{-i\omega T} x_T) \tag{47}
\]

where \( x_0 \) and \( x_T \) are the states at the initial point and the end point.

**Proof:** Assume that we have the following system

\[
y_t = Ce^{At} x_0 + \int_0^t Ce^{A(t-\tau)} B \sigma d\tau
\]
Assume that this signal is observed through a window

\[ W(t) = \frac{1}{\sqrt{T}} I_{[0,T]} \]

where \( I \) is the indicator function. Then we have the transform

\[ Y(i\omega) = \int_{-\infty}^{\infty} W(t) y_t e^{-i\omega t} dt \]

which is a stochastic variable now. The integral of the first term will be

\[
\frac{1}{\sqrt{T}} \int_{0}^{T} C e^{(A - i\omega I)t} x_0 dt \\
= \frac{1}{\sqrt{T}} C(i\omega I - A)^{-1}(I - e^{(A - i\omega I)T}) x_0
\]

were \( x_0 \) can be a stochastic variable. Integration by parts

\[
\int_{0}^{T} f(t) B_t dt = F(T) B_T - \int_{0}^{T} F(t) dB_t
\]

\[
f(t) = C e^{(A - i\omega I)t} \\
F(t) = C(A - i\omega I)^{-1} e^{(A - i\omega I)t}
\]

\[
B_t = \int_{0}^{t} e^{-At} B d\tau
\]

will give

\[
\frac{1}{\sqrt{T}} \int_{0}^{T} \int_{0}^{t} C e^{A(t-\tau)} B \sigma d\tau e^{-i\omega \tau} dt =
\]

\[
\frac{1}{\sqrt{T}} C(A - i\omega I)^{-1} e^{(A - i\omega I)T} \int_{0}^{T} e^{-At} B \sigma d\tau - \frac{1}{\sqrt{T}} \int_{0}^{T} C(A - i\omega I)^{-1} e^{(A - i\omega I)t} e^{-At} B \sigma d\tau
\]

\[
\frac{1}{\sqrt{T}} C(A - i\omega I)^{-1} e^{-i\omega T} \int_{0}^{T} e^{A(T-\tau)} B \sigma d\tau - \frac{1}{\sqrt{T}} \int_{0}^{T} C(A - i\omega I)^{-1} e^{(A - i\omega I)t} e^{-At} B \sigma d\tau
\]

\[
\frac{1}{\sqrt{T}} C(A - i\omega I)^{-1} e^{-i\omega T} \int_{0}^{T} e^{A(T-\tau)} B \sigma d\tau + \frac{1}{\sqrt{T}} \sigma B(i\omega) \int_{0}^{T} e^{-i\omega \tau} d\tau
\]

One effect of the initial and end point conditions is that

\[
\int_{0}^{T} e^{A(T-\tau)} B \sigma d\tau + e^{AT} x_0 = x_T
\]

and by this the transform will become

\[
X(i\omega) = \frac{B(i\omega)}{A(i\omega)} \frac{1}{\sqrt{T}} \int_{0}^{T} e^{-i\omega t} dt + \frac{1}{\sqrt{T}} C(i\omega I - A)^{-1}(x_0 - e^{-i\omega T} x_T)
\]
We will now show that for a specific set of frequencies $\omega_k$ the random variables $E_T(w_k)$ will be uncorrelated with each other.

**Theorem 5.2** If

$$E_T(i\omega_k) = \frac{1}{\sqrt{T}} \int_0^T e^{-i\omega t} dt$$

and

$$\omega_k = \frac{2\pi}{T} l \quad l \in \mathbb{Z}.$$

then for $k, l \in \mathbb{Z}$

$$EE(i\omega_k)E(-i\omega_l) = \delta_{k-l}$$

**Proof:** Study the following vector of random variables

$$
\begin{pmatrix}
\Re E_T(i\omega_k) \\
\Im E_T(i\omega_k)
\end{pmatrix}
= \begin{pmatrix}
\frac{1}{\sqrt{T}} \int_0^T \cos \omega t dw_t \\
\frac{1}{\sqrt{T}} \int_0^T -\sin \omega t dw_t
\end{pmatrix}
$$

Assume that we take two different frequencies $\omega_1$ and $\omega_2$ and check the correlation

$$E \left( \begin{pmatrix}
\Re E_T(i\omega_1) \\
\Im E_T(i\omega_1)
\end{pmatrix} \left( \begin{pmatrix}
\Re E_T(i\omega_2) \\
\Im E_T(i\omega_2)
\end{pmatrix} \right)^T
\right) =
\begin{pmatrix}
\frac{1}{T} \int_0^T \cos \omega_1 t \cos \omega_2 t dt & -\frac{1}{T} \int_0^T \cos \omega_1 t \sin \omega_2 t dt \\
-\frac{1}{T} \int_0^T \sin \omega_1 t \cos \omega_2 t dt & \frac{1}{T} \int_0^T \sin \omega_1 t \sin \omega_2 t dt
\end{pmatrix}
$$

The term in the upper left corner becomes

$$\frac{1}{T} \int_0^T \cos \omega_1 t \cos \omega_2 t dt = \frac{1}{2T} \int_0^T \cos(\omega_1 - \omega_2) t + \cos(\omega_1 + \omega_2) t dt$$

$$= \frac{1}{2} \left( \frac{\sin(\omega_1 - \omega_2) T}{(\omega_1 - \omega_2) T} + \frac{\sin(\omega_1 + \omega_2) T}{(\omega_1 + \omega_2) T} \right)$$

and the term in the lower right will become

$$\frac{1}{T} \int_0^T \sin \omega_1 t \sin \omega_2 t dt = \frac{1}{2T} \int_0^T \cos(\omega_1 - \omega_2) t - \cos(\omega_1 + \omega_2) t dt$$

$$= \frac{1}{2} \left( \frac{\sin(\omega_1 - \omega_2) T}{(\omega_1 - \omega_2) T} - \frac{\sin(\omega_1 + \omega_2) T}{(\omega_1 + \omega_2) T} \right).$$

The term in the upper right will be

$$-\frac{1}{T} \int_0^T \cos \omega_1 t \sin \omega_2 t dt = -\frac{1}{2T} \int_0^T \sin(\omega_2 - \omega_1) t + \sin(\omega_2 + \omega_1) t dt$$

$$= \frac{1}{2} \left( \frac{\cos(\omega_2 - \omega_1) T - 1}{(\omega_2 - \omega_1) T} + \frac{\cos(\omega_2 + \omega_1) T - 1}{(\omega_2 + \omega_1) T} \right).$$
and the lower left is
\[ -\frac{1}{T} \int_{0}^{T} \sin \omega t \cos \omega t \, dt \]
\[ = -\frac{1}{2T} \int_{0}^{T} \sin(\omega_1 - \omega_2) t + \sin(\omega_1 + \omega_2) t \, dt \]
\[ = \frac{1}{2} \left( \frac{\cos(\omega_1 - \omega_2) T - 1}{(\omega_1 - \omega_2)} + \frac{\cos(\omega_2 + \omega_1) T - 1}{(\omega_1 + \omega_2)} \right). \]

If we want the random variables at different frequencies to be uncorrelated, then following must be satisfied.

\[
\begin{align*}
\sin(\omega_1 - \omega_2) T &+ \sin(\omega_1 + \omega_2) T = 0 \\
\frac{(\omega_1 - \omega_2) T}{\sin(\omega_1 - \omega_2) T} &+ \frac{1}{\sin(\omega_1 + \omega_2) T} = 0 \\
\frac{(\omega_1 - \omega_2) T - 1}{\cos(\omega_1 - \omega_2) T} &+ \frac{1}{\cos(\omega_1 + \omega_2) T} = 0.
\end{align*}
\]

By subtracting the second equation from the first and the fourth equation from the third we and then eliminating we will get

\[
\begin{align*}
\sin(\omega_1 - \omega_2) T &+ \sin(\omega_1 + \omega_2) T = 0 \\
\frac{(\omega_1 - \omega_2) T}{\sin(\omega_1 - \omega_2) T} &+ \frac{1}{\sin(\omega_1 + \omega_2) T} = 0 \\
\frac{(\omega_1 - \omega_2) T - 1}{\cos(\omega_1 - \omega_2) T} &+ \frac{1}{\cos(\omega_1 + \omega_2) T} = 0.
\end{align*}
\]

This yields that

\[
\begin{align*}
\omega_1 - \omega_2 &= \frac{2\pi}{\omega_1} k & k \in \mathbb{Z} \\
\omega_1 + \omega_2 &= \frac{2\pi}{\omega_2} l & l \in \mathbb{Z}
\end{align*}
\]

This motivates choosing the frequencies \( \omega_k, k = 0, \ldots, N \omega \) such that

\[ \omega_k = \frac{2\pi}{\omega} l \quad l \in \mathbb{Z}. \quad (48) \]

If we make the above choice then

\[ \text{if } \omega_k \in N_c \left( K(i\omega_k), \sigma^2 \frac{|B(i\omega_k)|^2}{|A(i\omega_k)|^2} \right) \quad (49) \]

where \( N_c \) denotes the complex normal distribution. That is

\[ \begin{pmatrix} \text{Re} Y_T(i\omega_k) \\ \text{Im} Y_T(i\omega_k) \end{pmatrix} \in N \begin{pmatrix} \text{Re} K(i\omega_k) \\ \text{Im} K(i\omega_k) \end{pmatrix}, \frac{1}{2} \begin{pmatrix} \sigma^2 \frac{|B(i\omega_k)|^2}{|A(i\omega_k)|^2} & 0 \\ 0 & \sigma^2 \frac{|B(i\omega_k)|^2}{|A(i\omega_k)|^2} \end{pmatrix} \quad (50) \]

for this particular choice of frequencies.
5.2 Objective Function

The likelihood function for the estimated values of the Fourier transform 
\{Y_T(i\omega_1), Y_T(i\omega_2), \ldots, Y_T(i\omega_N)\}
will according to the expression in (50) be

\[ p(Y_T(i\omega_1), Y_T(i\omega_2), \ldots, Y_T(i\omega_N)|\theta) = \prod_{k=1}^{N_\omega} \frac{1}{2\pi\sigma^2} e^{-\frac{|Y_T(i\omega_k) - K(i\omega_k)|^2}{\sigma^2|B(i\omega_k)|^2}}. \]

The negative log likelihood function will be

\[ L(\theta) = -\log p(Y_T(i\omega_1), Y_T(i\omega_2), \ldots, Y_T(i\omega_N)|\theta) = N_\omega \log 2\pi + \sum_{k=1}^{N_\omega} \frac{|Y_T(i\omega_k) - K(i\omega_k)|^2}{\sigma^2|B(i\omega_k)|^2} + \log \sigma^2|B(i\omega_k)|^2. \]

Suppose now that a whole continuous time realization \{y(t) : t \in [0, T]\} of the output of a CAR model is known. Define the periodogram of this output as

\[ \hat{\Phi}_T(i\omega, \hat{\theta}) = |Y_T(i\omega) - K(i\omega)|^2 \]

since \(K(i\omega)\) will depend on the parameters, endpoint and initial values. The ML-procedure for estimating the parameters and the covariance is then

\[ \hat{\theta} \triangleq \arg \min_\theta V(\theta, \hat{\Phi}_T) \]

where

\[ V(\theta, \hat{\Phi}_T) \triangleq \sum_{k=1}^{N_\omega} \frac{\hat{\Phi}_T(i\omega_k, \theta)}{\Phi(i\omega_k, \theta)} + \log \Phi(i\omega_k, \theta) \]

and

\[ \Phi(i\omega, \theta) = \sigma^2|B(i\omega_k)|^2|A(i\omega_k)|^2. \]

In the remaining theory in this report we will disregard the effects of initial and endpoint conditions. In practice, while estimating models, we will always estimate initial and endpoint values. We refer the reader to the book by Pintelon and Schoukens[6] which treats the topic of initial values more thoroughly.
6 Properties of Bias and Variance

In this section we show how the bias and variance of the parameter estimates will be related to the bias and variance of the periodogram. In the case of bias we will have

\[ E(\hat{\theta}_T^k - \theta^*) \approx (V''_{\theta\theta}(\theta^*, \Phi))^{-1} \sum_{k=1}^{N_\omega} V''_{\theta\Phi_k}(\theta^*, \Phi) \left( \hat{\Phi}^{T,k} \omega_k - \Phi(\omega_k, \theta^*) \right) \]

where

\[ V''_{\theta\theta}(\theta^*, \Phi) \approx \sum_{k=1}^{N_\omega} \frac{\Phi'_\theta(\omega_k, \theta^*)}{\Phi(\omega_k, \theta^*)^2} \left( 1 - \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta^*)} \right) \]

if \( T \) is large and \( h_{\text{max}} \) is small, and

\[ V''_{\theta\Phi_k}(\theta^*, \hat{\Phi}) = \frac{\Phi'_\theta(\omega_k, \theta^*)}{\Phi(\omega_k, \theta^*)} \left( \frac{\Phi'_\theta(\omega_k, \theta^*)}{\Phi(\omega_k, \theta^*)} \right)^T \]

In the case of variance we resort to asymptotics where

\[ E(\hat{\theta}_T^k - \theta^*)(\hat{\theta}_T^k - \theta^*)^T \rightarrow (V''_{\theta\theta}(\theta^*, \Phi))^{-1} \]

when \( T \rightarrow \infty \) and \( h_{\text{max}} \rightarrow 0 \). In the subsections below we derive these expressions for the derivatives, bias and variance.

6.1 Derivatives

In both the expressions for the bias and the asymptotic variance second order derivatives with respect to the parameters are needed.

Lemma 6.1 Let \( V \) be defined as in (52). Then

\[ V''_{\theta\Phi_k}(\theta^*, \Phi) = \sum_{k=1}^{N_\omega} \frac{\Phi'_\theta(\omega_k, \theta^*)}{\Phi(\omega_k, \theta^*)^2} \left( 1 - \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta^*)} \right) \]

\[ + \sum_{k=1}^{N_\omega} \frac{\Phi'_\theta(\omega_k, \theta^*)}{\Phi(\omega_k, \theta^*)^3} \]

Proof: From the definition in (52) we have

\[ V'_\theta(\theta, \Phi) = \sum_{k=1}^{N_\omega} \frac{\Phi'_\theta(\omega_k, \theta)}{\Phi(\omega_k, \theta)} \left( 1 - \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta)} \right) \]

and therefore

\[ V''_{\theta\Phi_k}(\theta, \Phi) = \sum_{k=1}^{N_\omega} \frac{\Phi'_\theta(\omega_k, \theta)}{\Phi(\omega_k, \theta)^2} \left( 1 - \frac{\hat{\Phi}(\omega_k)}{\Phi(\omega_k, \theta)} \right) \]

\[ + \sum_{k=1}^{N_\omega} \frac{\Phi'_\theta(\omega_k, \theta)^2}{\Phi(\omega_k, \theta)^3} \]

\[ \square \]
If $T$ is large and $h_{\text{max}}$ is small then then $\hat{\Phi} \approx \Phi$ and

$$V''_{\theta \hat{\Phi}_k} (\theta^*, \hat{\Phi}) = \frac{\Phi'_k(i\omega_k, \theta^*)}{\Phi(i\omega_k, \theta^*)^2}. $$

We also take a look at the derivatives of the parameters with respect to the parameters and the information from the estimated spectrum. This is summarized in the following lemma

**Lemma 6.2** Let $V$ be defined as in (52). Then

$$V''_{\theta \hat{\Phi}_k} (\theta^*, \hat{\Phi}) = -\frac{\Phi'_k(i\omega_k, \theta^*)}{\Phi(i\omega_k, \theta^*)^2} \Phi(i\omega_k, \theta^*)^2. $$

**Proof:** As in the previous lemma we have

$$V'_\theta(\hat{\theta}_k, \hat{\Phi}) = -\sum_{k=1}^{N_\omega} \frac{\Phi'_k(i\omega_k, \theta)}{\Phi(i\omega_k, \theta)} \left( 1 - \frac{\hat{\Phi}(i\omega_k)}{\Phi(i\omega_k, \theta)} \right)$$

we now let $\hat{\Phi}_k = \hat{\Phi}(i\omega_k)$ and then

$$V''_{\theta \hat{\Phi}_k} (\theta^*, \hat{\Phi}) = -\frac{\Phi'_k(i\omega_k, \theta^*)}{\Phi(i\omega_k, \theta^*)^2} \Phi(i\omega_k, \theta^*)^2. $$


\[\blacksquare\]

### 6.2 Bias Expression

If we define the exact parameters as

$$\hat{\theta}_k^T \triangleq \arg \min_{\theta} V(\theta, \hat{\Phi}_T^T)$$

and the estimated parameters as

$$\theta^* \triangleq \arg \min_{\hat{\theta}} V(\theta, \hat{\Phi}_y)$$

we get the following result for the bias.

**Theorem 6.1**

$$E(\hat{\theta}_T^T - \theta^*) \approx (V''_{\theta \Phi}(\theta^*, \Phi))^{-1} \sum_{k=1}^{N_\omega} V''_{\theta \Phi}(\theta^*, \Phi) \left( \Phi(i\omega_k, \theta^*) - E\hat{\Phi}_T^T(i\omega_k) \right) \tag{55} $$

**Proof:** From the definition in (45) we have

$$V'_\theta(\hat{\theta}_T^T, \hat{\Phi}_T^T) = 0$$

$$V'_\theta(\theta^*, \Phi) = 0$$

A Taylor expansion of $V'_\theta(\theta, \Phi)$ around $\theta^*$ and $\Phi_k = \Phi(i\omega_k)$ yields

$$V'_\theta(\hat{\theta}_T^T, \hat{\Phi}_T^T) \approx V'_\theta(\theta^*, \Phi) + V''_{\theta \Phi}(\theta^*, \Phi)(\hat{\theta}_T^T - \theta^*) $$

$$+ \sum_{k=1}^{N_\omega} V''_{\theta \Phi}(\theta^*, \Phi) \left( \hat{\Phi}_T^T(i\omega_k) - \Phi(i\omega_k) \right). $$
From this we get
\[
\hat{\theta}_k^T - \theta^* \approx (V''_{\theta\theta}(\theta^*, \Phi))^{-1} \sum_{k=1}^{N_x} V''_{\phi_k}(\theta^*, \Phi) \left( \hat{\Phi}^{T,k}(i\omega_k) - \Phi(i\omega_k, \theta^*) \right).
\]

By taking expectations on both sides we get the conclusion.

### 6.3 Variance Expressions

An expression for the bias can be computed in a similar manner.

**Theorem 6.2** *The asymptotic variance will be*

\[
E(\hat{\theta}_k^T - \theta^*)(\hat{\theta}_k^T - \theta^*)^T \approx \sum_{k=1}^{N_x} \sum_{l=1}^{N_x} \Psi(i\omega_k, \theta^*) \Psi(i\omega_l, \theta^*)^T E[\Delta\Phi(i\omega_k, \theta^*) \Delta\Phi(i\omega_l, \theta^*)]
\]

*where*

\[
\Psi(i\omega_k, \theta^*) = V''_{\theta\theta}(\theta^*, \Phi)^{-1} V''_{\phi_k}(\theta^*, \Phi)
\]

*and*

\[
\Delta\Phi(i\omega_k, \theta^*) = \hat{\Phi}^{T,k}(i\omega_k) - \Phi(i\omega_k, \theta^*)
\]

**Proof:** This is a simple application of the same Taylor expansion as in the case of bias expression

\[
\hat{\theta}_k^T - \theta^* \approx (V''_{\theta\theta}(\theta^*, \Phi))^{-1} \sum_{k=1}^{N_x} V''_{\phi_k}(\theta^*, \Phi) \left( \hat{\Phi}^{T,k}(i\omega_k) - \Phi(i\omega_k, \theta^*) \right).
\]

Multiply this expression by the transpose of itself and take expectation, and we are done.

As you can see this is a translation of the covariance of the spectrum to the covariance of the parameter estimates. A bound on the spectrum covariance is therefore needed. The expression above is quite elaborate to calculate and therefore we settle with the asymptotic properties. When \(T \to \infty\) and \(h_{\text{max}} \to 0\) we get

\[
E\Delta\Phi(i\omega_k, \theta^*) \Delta\Phi(i\omega_l, \theta^*) \to \begin{cases} \Phi(i\omega_k, \theta^*)^2 & \text{if } i\omega_k = i\omega_l \\ 0 & \text{if } i\omega_k \neq i\omega_l. \end{cases}
\]

Then by simple calculation we will get the asymptotic result

\[
E(\hat{\theta}_k^T - \theta^*)(\hat{\theta}_k^T - \theta^*)^T \approx V''_{\theta\theta}(\theta^*, \Phi)^{-1}.
\]

when \(T \to \infty\) and \(h_{\text{max}} \to 0\). 

24
7 Practical Considerations for Frequency Selection

So far we have said very little about which frequencies to use in the criteria (52) in order to make a tradeoff between bias and variance. We have only pointed out earlier that we are restricted to the frequencies

\[ \omega_k = \frac{2\pi}{T} l, \quad l \in \mathbb{Z} \]

where \( k = 1, \ldots, N_{\omega} \). As will be explained below it is important to avoid high frequencies, because they will produce a considerable parameter bias if the estimated spectrum is biased. We will also explain why some frequencies have a significant effect on the bias while others have very little effect.

7.1 Minimizing the Variance

According the the asymptotic expression in (56) the variance is roughly inversely proportional to the quantity

\[ \sum_{k=1}^{N_{\omega}} \Psi(i\omega_k, \theta^*) \Psi(i\omega_k, \theta^*)^T \]

where

\[ \Psi(i\omega, \theta) = \frac{\Phi'(i\omega, \theta)}{\Phi(i\omega, \theta)} \]

For a CAR model where \( \theta = [a^T \sigma^2]^T, \quad a = [a_1 \ldots a_n]^T \) and

\[ \Phi(i\omega, \theta) = \frac{\sigma^2}{|A(i\omega, \theta)|^2} \]

we have

\[ \Psi(i\omega, \theta) = \left[ -\frac{(|A(i\omega)|^2)^{\frac{1}{2}}}{|A(i\omega)|^2} \right] \]

In Figure 2 we show the first two elements of \( \Psi(i\omega, \theta^*) \) for a CAR model. From the figure we see that these quantities are only significantly different from zero on a small frequency interval. Therefore a rule of thumb would be to only include frequencies where the magnitude of the elements of \( \Psi(i\omega, \theta) \) are large in order to reduce the variance of the parameter estimates. The circles represent the discrete frequencies we are allowed to use when \( T = 25 \).

7.2 Minimizing the Bias

From expression (55) we know that the contribution to the bias from each individual frequency component of the power spectrum is roughly proportional to

\[ V_{\theta \tilde{\theta}}'(\theta^*, \tilde{\Phi})^{-1} \left( \frac{\Phi'_{\theta_k}(i\omega_k, \theta^*)}{\Phi(i\omega_k, \theta^*)} \right)^T \left( \frac{\Phi'_{\theta}(i\omega_k, \theta^*)}{\Phi(i\omega_k, \theta^*)} \right) E\tilde{\Phi}(i\omega_k) - \frac{\Phi(i\omega_k, \theta)}{\Phi(i\omega_k, \theta^*)} \]

25
Figure 2: Relative sensitivity for $a_1$ (upper) and $a_2$ (lower) for a CAR model. Here $A(p) = p^2 + a_1 p + a_2$, $B(p) = 1$, $\sigma = 1$, $a_1 = 2$ and $a_2 = 1$.

In Figure 3 the term
\[
\frac{E\hat{\Phi}(i\omega_k) - \Phi(i\omega_k, \theta)}{\Phi(i\omega_k, \theta^*)}
\]
is plotted as a function of the frequency $\omega$. The system is the same as in the previous figure and the output is uniformly sampled with sampling interval $T_s = 0.5$. The solid line is the relative spectral bias for the Discrete-Time Fourier Transform. The dotted line indicates the same bias for ZOH interpolation and the dash-dotted for FOH. The contribution from an individual frequency to the

Figure 3: Relative spectral bias for an uniformly sampled CAR model. Here $A(p) = p^2 + 2p + 1$, $B(p) = 1$, $\sigma = 1$ and $T_s = 0.5$.

total bias is proportional to
\[
\frac{\Phi'(i\omega_k, \theta^*) \cdot E\hat{\Phi}(i\omega_k) - \Phi(i\omega_k, \theta)}{\Phi'(i\omega_k, \theta^*) \cdot \Phi(i\omega_k, \theta^*)}
\]

In Figure 4 we illustrate the contribution to the bias as a function of $\omega$ for the same system and circumstances as above.
Figure 4: Bias contribution for each frequency for $a_1$ (left) and $a_2$ (right). The model is uniformly sampled and of CAR type. Here $A(p) = p^2 + a_1 p + a_2$, $B(p) = 1$, $\sigma = 1$, $a_1 = 2$, $a_2 = 1$ and $T_s = 0.5$.

7.3 Re-parameterization

Consider the second order CAR model

$$y_t = \frac{\sigma}{p^2 + a_1 p + a_2} e_t.$$  \hfill (57)

We can also re-parameterize this model as

$$y_t = \frac{\omega_0^2}{p^2 + 2\zeta \omega_0 p + \omega_0^2} \sigma e_t.$$  \hfill (58)

where $\zeta$ and $\omega_0$ are named damping ratio and the undamped natural frequency.

The vector of parameters for this model is

$$\theta = [\zeta \omega_0 \lambda]^T$$  \hfill (59)

where $\lambda = \sigma^2$.

For this special parameterization the relative sensitivity functions with respect to respective parameters are shown in Figure 5. Here we have chosen $a_1 = 2$ and $a_2 = 2$. This means that $w_0 = 1/\sqrt{2}$ and $\zeta = \sqrt{2}$. From this figure, we conclude that the relative damping $\zeta$ is sensitive near the natural resonance frequency $\omega_0$ of the system. The natural frequency $\omega_0$ on the other hand is sensitive at high frequencies.
Figure 5: Relative sensitivities for $\zeta$ (upper) and $\omega_0$ (lower). Here $A(p) = p^2 + 2\zeta\omega_0 p + \omega_0^2$ and $B(p) = \omega_0^2$, $\sigma = 1$, $\zeta = 1/\sqrt{2}$ and $\omega_0 = \sqrt{2}$.

8 Numerical Experiments

In this section we will illustrate the previous theory with a few examples. First the method is applied to a first order continuous-time autoregressive model. Then we proceed with a second order model. The examples are used to show the effects of interpolation and time of observations on the parameter estimates. Finally we illustrate how the choice of frequencies will affect the parameter estimates.

8.1 Measurement & Simulation

In this report we set out to estimate $\theta$ from uniformly and non-uniformly distributed discrete-time samples of the continuous-time output process $\{y_t : t \in [0, T]\}$. We denote the instances of these samples by $\{t_1, t_2, \ldots, t_N\}$ and the observations themselves by $\{y_{t_1}, y_{t_2}, \ldots, y_{t_N}\}$. We use samples that are additively randomly distributed around a nominal sampling time $T_s$ according to the following model

$$t_k = kT_s + \sum_{l=1}^{k} \delta_l$$

where $\delta_l \in U(-\delta_0, \delta_0)$.

In order to simulate the discrete-time samples we resort to discretizing the model at $T_{is} = T_s/100$ where $T_{is}$ stands for the “inter-sample” sample time. The discretized system will become

$$\begin{cases}
x_{kT_{is}} = e^{AT_{is}} y_{(k-1)T_{is}} + \int_{(k-1)T_{is}}^{kT_{is}} e^{A(tT_{is}-\tau)} B dW_t, \\
y_{kT_{is}} = Cx_{kT_{is}}
\end{cases}$$

The last term in this expression is approximated as

$$\int_{(k-1)T_{is}}^{kT_{is}} e^{A(tT_{is}-\tau)} B dW_t \approx \frac{1}{\sqrt{T_{is}}} \int_{(k-1)T_{is}}^{kT_{is}} e^{A(tT_{is}-\tau)} B d\tau v$$
where \( v \in N(0,1) \). This approach has proved to work well in practice and samples have been picked randomly according to a discrete version of the scheme in (60).

### 8.2 First Order CAR Model

In this section we will apply the method described earlier to a simple first order continuous-time autoregressive model. The first system is

\[
y_t = \frac{\sigma}{p + a} e_t^t \quad (60)
\]

where \( a = 1 \) and \( \sigma = 1 \). We will observe the effect of: sampling time, the amount of non-uniformity in the sampling and in the time of observation. We will notice that the amount of non-uniformity have little effect on the amount of bias and variance. The variance and bias decreases as the time of observation increases. Finally, the bias decreases as the sampling interval decreases.

#### 8.2.1 Observation Time

In Figure 6 we have simulated and estimated the system for different values of \( T \). The dotted lines represent the Cramer-Rao lower bound computed by

![Figure 6: Monte-Carlo Simulations of estimates of \( a \) for different observation times \( T \). The model is \( A(p) = p + a, B(p) = 1, \sigma = 1 \) and \( a = 1 \). Here \( N_{MC} = 100, T_s = 0.1 \) and \( \delta_0 = T_s/5 \). The dotted line represents a CRLB estimate. Vertical bars indicate the standard deviation of estimates.](image)

Monte-Carlo Simulations using the Slepian-Bang formula [9]

\[
E[(\hat{a} - a)(\hat{a} - a)] \leq \frac{2}{T^r R^{-1} R_0^{t} R^{-1} R_0^t} \quad (61)
\]

where \( R \) is the covariance matrix for the measurements and \( R_0^t \) is the derivative of the covariance matrix with respect to \( a \). The bias and standard deviation decrease with the time of observation. For moderately large \( T \) the standard deviation is approximately same as the Cramer-Rao bound.
8.2.2 Sampling Interval

In Figure 7 we have varied the sample time $T_s$ in order to illustrate the effect on the bias. From the figure we see that the jitter have very little effect on the quality of the estimates.

Figure 7: Monte-Carlo Simulations of estimates of $a$ for different sample times $T_s$. The model is $A(p) = p + a$, $B(p) = 1$, $\sigma = 1$ and $a = 1$. Here $N_{MC} = 250$, $T = 200$ and $\delta_0 = T_s/5$. The vertical bars indicate the standard deviation of estimates.

8.2.3 Jitter

In Figure 8 we have plotted the effect of different amount of jitter $\delta_0$ on the sampled data. Here the magnitude of the $\delta_0$ has very little effect on the sampling interval.

8.2.4 Effect of Linear Interpolation

An interesting observation is that the FOH interpolation does not improve the performance of the method. This is illustrated in the Figure 9 where there is significant bias compared to ZOH. The reason for this as we have seen earlier is that the spectrum is more biased for FOH than for ZOH.
Figure 8: Monte-Carlo Simulations of estimates of $a$ for different $\delta_0$. The model is $A(p) = p + a$, $B(p) = 1$, $\sigma = 1$ and $a = 1$. Here $N_{MC} = 250$, $T = 200$ and $T_s = 0.1$. The vertical bars indicate the standard deviations.

8.3 Numerical Example for a second order CAR model

We have also estimated the parameters of a second order continuous-time autoregressive model where

$$y_t = \frac{\sigma}{p^2 + a_1 p + a_2} e_t$$  \hspace{1cm} (62)

First we will illustrate how different kind of interpolation will effect the parameter estimates. Then we show how the choice of frequency interval affects the quality of estimates.

8.3.1 Interpolation

The effects of “Riemann”, ZOH interpolation and FOH interpolation are illustrated below in Figure 10. As in the first order CAR case FOH is no better than ZOH. The performance of the “Riemann” case is even worse. The reasons for this can be attributed to the spectral bias described earlier.

8.4 Choice of Frequencies

In Figure 11 we have estimated the parameters in the model in (62). First we used the frequency span $\{2\pi \frac{1}{200}, \ldots, 2\pi \frac{200}{200}\}$ while in the second case we have used $\{2\pi \frac{8}{200}, \ldots, 2\pi \frac{208}{200}\}$. As can be seen from the picture this doubled the variance of the $a_2$ parameter estimate. Standard deviation for $a_2$ is 0.0201 in the first case which is indicated by dots. In the second case which is indicated by plus signs we have 0.0491. The reason is as we can see from Figure 2 that the parameter is especially sensitive at low frequencies.

While there are negative effects of the exclusion of low frequencies the use of high frequencies can be equally detrimental. This is illustrated in Figure 12 where we have used a frequency window up to the the frequency indicated by the x-axis. The reason for the large RMSE
Figure 9: Monte-Carlo Simulations of estimates of $a$ for different $T_s$ and ZOH (solid) and FOH (dot-dashed) assumptions. The model is $A(p) = p + a$, $B(p) = 1$, $\sigma = 1$ and $a = 1$. Here $N_{MC} = 250$, $T = 250$ and $\delta_0 = T_s/5$.

$$RMSE = \sqrt{\frac{1}{N_{MC}} \sum_{j=1}^{N_{MC}} (63)}$$

at high frequencies is that there is a very large relative bias at these frequencies.
9 Conclusion

In this paper we identify a continuous-time ARMA process from non-uniformly distributed samples of the output. The continuous-time realization is approximated by piecewise constant and piecewise linear interpolation. The continuous-time Fourier transform is then computed for the interpolated output. Parameters are then estimated from the power spectrum obtained from the Fourier transform. A bound on the parameter bias in terms of interpolation and leakage is derived. These results focus on the effect of the sampling interval. By a few numerical examples we show that piecewise linear and “Riemann”. interpolation is not necessarily better than piecewise constant interpolation. Guidelines on frequency selection in order to reduce the parameter bias and variance are also provided.

10 Further Research

One way to proceed would be to consider the Output Error Model Structure together with the Box-Jenkins Case. It would also be interesting to see if tapering and faders could reduce the bias due to leakage.

References


Figure 11: Parameter estimate dependence on frequency choice. Here $N_{MC} = 200$, $T = 200$, $T_s = 0.1$ and $\delta_0 = T_s/5$.


Figure 12: RMSE for a second order CAR model with respect to maximum frequency. Here $Nmc = 200$, $T = 200$, $T_s = 0.1$ and $\delta_0 = T_s/5$. 
In this paper is discussed how to estimate irregularly sampled continuous-time ARMA models in the frequency domain. In the process, the model output signal is assumed to be piecewise constant or piecewise linear, and an approximation of the continuous-time Fourier transform is calculated. ML-estimation in the frequency domain is then used to obtain parameter estimates.