Yule–Walker Estimates for Continuous Time Autoregressive Models

Rob J. Hyndman
Department of Statistics
The University of Melbourne

17 October 2003

Abstract

I consider continuous time autoregressive (CAR) processes of order p and develop estimators of the model parameters based on Yule–Walker type equations. For continuously recorded data, it is shown that these estimators are least squares estimators and have the same asymptotic distribution as maximum likelihood estimators.

In practice, though, data can only be observed discretely. For discrete data, I consider approximations to the continuous time estimators. It is shown that some of these discrete time estimators are asymptotically biased. Alternative estimators based on the autocovariance function are suggested. These are asymptotically unbiased and are a fast alternative to the maximum likelihood estimators described by Jones (1981). They may also be used as starting values for maximum likelihood estimation.

KEYWORDS: continuous-time autoregression, Yule–Walker estimates, continuously recorded time series, unequally spaced time series.

1 Introduction

A continuous time autoregressive (CAR) process of order p may be represented by the equation

$$X^{(p)}(t) + \alpha_{p-1}X^{(p-1)}(t) + \cdots + \alpha_0 X(t) = \sigma Z(t), \qquad -\infty < t < \infty, \quad (1.1)$$

where $X^{(j)}(t)$ denotes the jth derivative of $X(t)$ and $Z(t)$ denotes continuous time Gaussian white noise. Of course, $\{Z(t)\}$ does not exist as a stochastic process in the usual sense and so this is not a rigorous definition of a CAR process. Nevertheless, it is a useful description and is intuitively appealing. A rigorous definition of a CAR process is given in Section 2.

The problem of estimating the model parameters has been the subject of much statistical, econometric and engineering literature. For the estimation of the model from a continuous record, see Bartlett (1955), Dzhaparidze (1971) and Priestley (1981). In practice, a sample function is observed discretely at the points $t_0 < t_1 < \cdots < t_n$. For previous work in the case of equally spaced data ($t_{i+1} - t_i = \Delta$ for all i), see, for example, Bartlett (1946), Durbin (1961), Phillips (1972) and Bergstrom (1984). These authors exploit the well-known result that an equally spaced sample from a CAR(p) process is itself an ARMA(p,$p-1$) process. The case with unequally spaced observations where the times of observation are non-random has been considered by Jones (1981), Robinson (1977) and others. Robinson (1980) considers the case where the times of observation form a stationary point process.

Central to the development of Yule–Walker estimates for CAR models is the *derivative covariance function*. This is introduced in Section 3. Section 4 explores Yule–Walker equations for CAR models and estimators based on these equations are developed in Section 5. These are suitable for continuously recorded data and are generalisations of the estimators described by Bartlett (1955) and Priestley (1981). A quick method for estimating the residual variance is outlined in Section 6.

Section 7 discusses Yule–Walker estimators for discrete observations taken at non-random times. It is shown that discrete time approximations of the Yule–Walker estimators for continuously recorded data may be asymptotically biased. An alternative approach based on the autocovariance function is developed.

Finally, in Section 8, the discrete time estimators are applied to the Canadian lynx data and Wolf's sunspot numbers. The results are close to the maximum likelihood estimates reported in other studies of these data.

2 Definition of a CAR Process

In order to define a CAR(p) process rigorously, it is convenient to introduce the stationary p-variate process $\boldsymbol{S}(t) := \left[X(t), X^{(1)}(t), \cdots, X^{(p-1)}(t)\right]'$ satisfying the stochastic differential equation,

$$d\boldsymbol{S}(t) = A\boldsymbol{S}(t)dt + \sigma\boldsymbol{b}dW(t), \qquad t \geq 0, \tag{2.1}$$

where $\{W(t)\}$ denotes standard Brownian motion,

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -\alpha_0 & -\alpha_1 & \cdots & \cdots & -\alpha_{p-1} \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

and

$$\boldsymbol{S}(0) \stackrel{d}{=} N\left(\mathbf{0}\ ,\ \sigma^2 \int_0^\infty e^{Au}\boldsymbol{b}\boldsymbol{b}'e^{A'u}du\right) \tag{2.2}$$

is independent of $\{W(s), s \geq 0\}$. We also require (for stationarity) that all the eigenvalues of A have negative real parts (Arató, 1982, pp.118–119).

The CAR(p) process $\{X(t)\}$ is then defined to be the first component of the process $\{\boldsymbol{S}(t)\}$. That is,

$$X(t) := [1, 0, \cdots, 0]\, \boldsymbol{S}(t). \tag{2.3}$$

Equation (1.1) should be regarded as a formal representation of the well-defined Itô differential equation (2.1).

The solution of (2.1) is the Itô integral

$$\boldsymbol{S}(t) = e^{At}\boldsymbol{S}(0) + \sigma \int_0^t e^{A(t-y)}\boldsymbol{b}dW(y). \tag{2.4}$$

3 Derivative Covariance Function

For discrete time AR processes, the Yule–Walker equations are written in terms of the autocovariance function (ACVF) defined by $\gamma(h) := \mathrm{Cov}\,[X(t+h), X(t)]$. The analogous function for CAR processes is the derivative covariance function (DCVF) defined below.

Definition 1 *The derivative covariance function is defined as*

$$D_{j,k}(h) := Cov\left[X^{(j)}(t+h), X^{(k)}(t)\right], \qquad 0 \leq j,k \leq p-1. \tag{3.1}$$

The following theorem shows that the DCVF is closely related to the ACVF.

Theorem 3.1 *If $\{X(t)\}$ is a CAR(p) process then, for $h \geq 0$ and $0 \leq j,k \leq p-1$,*

$$\begin{aligned} D_{j,k}(h) &= (-1)^k\gamma^{(j+k)}(h) && (3.2)\\ \text{and}\quad D_{j,k}(-h) = (-1)^{j+k}D_{k,j}(h) &= (-1)^j\gamma^{(j+k)}(h) && (3.3) \end{aligned}$$

where $\gamma^{(j+k)}(h)$ denotes the $(j+k)$th derivative of the ACVF.

Proof: The first part of this theorem is proved by Ash and Gardner (1975, pp.263–264). The second equation follows using $D_{j,k}(h) = D_{k,j}(-h)$.

□

This result may be extended by defining $D_{j,p}(h)$ using the Itô integrals

$$D_{j,p}(h) := \lim_{T\to\infty} \frac{1}{T}\int_0^T X^{(j)}(t+h)\, d\Big[X^{(p-1)}(t)\Big] \tag{3.4}$$

where the limit exists in the sense of mean square convergence. Then $D_{j,p}(h) = -D_{j+1,p-1}(h) = (-1)^p\gamma^{(j+p)}(h)$, $j \leq p-2$. Similarly, for $h \neq 0$, $D_{p-1,p}(h) = (-1)^p\gamma^{(2p-1)}(h)$.

Difficulties arise with $D_{p-1,p}(0)$ since $\gamma^{(2p-2)}(h)$ is not differentiable at $h = 0$ (Doob, 1953, p.544). Nevertheless, $D_{p-1,p}(0)$ can be calculated using Itô's formula. Let $\boldsymbol{S}(t) = \left[X(t), X^{(1)}(t), \cdots, X^{(p-1)}(t)\right]'$ satisfy the Itô differential equation (2.1) and define $Y[\boldsymbol{S}(t), t] := \frac{1}{2}\left[\boldsymbol{b}'\boldsymbol{S}(t)\right]^2 = \frac{1}{2}\left[X^{(p-1)}(t)\right]^2$. Using the multi-dimensional Itô formula (Øksendal, 1989, p.33),

$$dY[\boldsymbol{S}(t), t] = X^{(p-1)}(t)\, d\left[X^{(p-1)}(t)\right] + \frac{\sigma^2}{2}dt.$$

So

$$\begin{aligned}\frac{1}{T}\int_0^T X^{(p-1)}(t)\, d\left[X^{(p-1)}(t)\right] &= \frac{1}{T}\left\{Y[\boldsymbol{S}(T), T] - Y[\boldsymbol{S}(0), 0]\right\} - \frac{\sigma^2}{2} \\ &= \frac{1}{2T}\left\{X^{(p-1)}(T)^2 - X^{(p-1)}(0)^2\right\} - \frac{\sigma^2}{2}.\end{aligned}$$

Thus, using (3.4),

$$D_{p-1,p}(0) = \frac{-\sigma^2}{2} \tag{3.5}$$

which is the right derivative of $-D_{p-1,p-1}(h)$ at $h = 0$ (Doob, 1953, p.544).

Lemma 3.2 *If $\{X(t)\}$ is a CAR(p) process then, for $0 \leq j, k \leq p-1$,*

$$\begin{aligned} D_{j,k}(0) &= 0 && \text{if } j+k \text{ is odd;} \\ \mathit{sign}(D_{j,k}(0)) &= (-1)^{(3k+j)/2} && \text{if } j+k \text{ is even.}\end{aligned}$$

Proof: Now $D_{j,k}(0) = D_{k,j}(0)$ by definition. Therefore, from Theorem 3.1, $(-1)^k\gamma^{(j+k)}(0) = (-1)^j\gamma^{(j+k)}(0)$. So, if $j+k$ is odd, then $(-1)^j \neq (-1)^k$ and $\gamma^{(j+k)}(0) = 0$. Hence, using Theorem 3.1 again, we obtain the first equation.

For even n where $n \leq 2p-2$, $\gamma^{(n)}(0) = (-1)^{n/2}\mathrm{Var}\left(X^{(n/2)}(t)\right)$, and so $\mathrm{sign}(\gamma^{(n)}(0)) = (-1)^{n/2}$. In particular, $\mathrm{sign}(\gamma^{(j+k)}(0)) = (-1)^{(j+k)/2}$ if $j+k$ is even. Hence, from Theorem 3.1, $\mathrm{sign}(D_{j,k}(0)) = (-1)^k(-1)^{(j+k)/2} = (-1)^{(3k+j)/2}$.

□

3.1 Estimating the DCVF

Let $\{x(t)\}$ denote a sample function of a CAR(p) process, $\{X(t)\}$, and suppose we are able to observe and record it continuously without error between the times 0 and T. I now consider the problem of estimating $D_{j,k}(0)$, $0 \leq j \leq p-1$, $0 \leq k \leq p$, based on $\{x(t)\}$.

For $j = 0, 1, \ldots, p-1$, $\{X^{(j)}(t)\}$ is equivalent to a process with almost all sample functions continuous (Jazwinski, 1970, p.111). Therefore, $\{X(t)\}$ is equivalent to a process with almost all sample functions possessing continuous derivatives up to order $p-1$. Hence, $x^{(j)}(t)$ $(0 \le j \le p-1)$ may be calculated from the data.

Then, $D_{j,k}(0)$, $0 \le j \le p-1$, $0 \le k \le p$ may be estimated by the sample covariances

$$\begin{aligned} \hat{D}_{j,k} &= \frac{1}{T}\int_0^T x^{(j)}(t)\, x^{(k)}(t)dt, \qquad k < p, && (3.6)\\ \text{and} \quad \hat{D}_{j,p} &= \frac{1}{T}\int_0^T x^{(j)}(t)\, dx^{(p-1)}(t). && (3.7) \end{aligned}$$

Then $\hat{D}_{j,k} \xrightarrow{p} D_{j,k}(0)$ as $T \to \infty$ (Yaglom, 1987, pp.231–233).

3.2 Covariance matrix

Let $\Gamma_p(h) = \operatorname{Cov}[\boldsymbol{S}(t), \boldsymbol{S}(t-h)]$ denote the autocovariance function of the stationary p-variate process $\boldsymbol{S}(t)$. Then the (j,k)th element of $\Gamma_p(h)$ is $D_{j-1,k-1}(h)$. Hence, for $h \ge 0$,

$$\Gamma_p(h) = \begin{pmatrix} \gamma^{(0)}(h) & -\gamma^{(1)}(h) & \cdots & (-1)^{p-1}\gamma^{(p-1)}(h) \\ \gamma^{(1)}(h) & -\gamma^{(2)}(h) & \cdots & (-1)^{p-1}\gamma^{(p)}(h) \\ \vdots & \vdots & & \vdots \\ \gamma^{(p-1)}(h) & -\gamma^{(p)}(h) & \cdots & (-1)^{p-1}\gamma^{(2p-2)}(h) \end{pmatrix}$$

and $\Gamma_p(-h) = \Gamma_p'(h)$ using Theorem 3.1. Note that in the case $h = 0$ every second element of this matrix is zero according to Lemma 3.2.

Multiplying both sides of (2.1) by $\boldsymbol{S}'(t-h)$ and taking expectations, we obtain

$$d\Gamma_p(h) = A\Gamma_p(h)dh. \tag{3.8}$$

Thus, using (2.2),

$$\Gamma_p(h) = e^{Ah}\Gamma_p(0) = \sigma^2 \int_0^\infty e^{A(u+h)}\boldsymbol{b}\boldsymbol{b}'e^{A'u}du. \tag{3.9}$$

$\boxed{\text{PROPOSITION 3.3}}$ *If $\{X(t)\}$ is a CAR(p) process with $\sigma > 0$, then $\Gamma_p(0)$ is strictly positive definite.*

PROOF: $\Gamma_p(0)$ is non-negative definite since it is a covariance matrix. It remains to show that $\boldsymbol{a}'\Gamma_p(0)\boldsymbol{a} \ne 0$ for any $\boldsymbol{a} \ne \boldsymbol{0}$.

Suppose there exists $\boldsymbol{a} \ne \boldsymbol{0}$ such that $\boldsymbol{a}'\Gamma_p(0)\boldsymbol{a} = 0$. Then, from (3.9),

$$\boldsymbol{a}'\Gamma_p(0)\boldsymbol{a} = \sigma^2 \int_0^\infty \boldsymbol{a}'e^{-Au}\boldsymbol{b}\boldsymbol{b}'e^{-A'u}\boldsymbol{a}\, du = 0$$

which means that $\boldsymbol{a}'e^{-Au}\boldsymbol{b} = 0$ for all $u \geq 0$. But, there exists a $u \geq 0$ such that $e^{-Au}\boldsymbol{b} = \boldsymbol{a}$, and for this u, $\boldsymbol{a}'e^{-Au}\boldsymbol{b} = \boldsymbol{a}'\boldsymbol{a} > 0$ since $\boldsymbol{a} \neq \boldsymbol{0}$. Hence, the assumption leads to a contradiction.

□

4 Yule–Walker Equations for CAR processes

Let $\{X(t)\}$ be the CAR(p) process defined by (2.1), (2.2) and (2.3). Then, multiplying equation (1.1) by $X^{(j)}(t+h)$ and taking expectations gives

$$\alpha_0 D_{j,0}(h) + \alpha_1 D_{j,1}(h) + \cdots + \alpha_p D_{j,p}(h) = 0, \qquad j = 0, 1, \ldots, p-1. \quad (4.1)$$

These are the Yule–Walker equations for CAR processes. They are analogous to the traditional discrete time Yule–Walker equations which can be obtained in a similar way replacing $X^{(j)}(t)$ by $X(t-j)$ and letting $h = 0$.

Let $h \geq 0$. Then substituting $(-1)^k\gamma^{(j+k)}(h)$ for $D_{j,k}(h)$ into (4.1) gives

$$(-1)^p\gamma^{(j+p)}(h) + (-1)^{p-1}\alpha_{p-1}\gamma^{(j+p-1)}(h) + \cdots + \alpha_0\gamma^{(j)}(h) = 0 \qquad h \geq 0 \quad (4.2)$$

where $j = 0, 1, \cdots, p-1$ and $\gamma^{(2p-1)}(0)$ is to be interpreted as the right derivative of $\gamma^{(2p-2)}(h)$ at $h = 0$.

Similarly, for $h \leq 0$, substitute $(-1)^j\gamma^{(j+k)}(-h)$ for $D_{j,k}(h)$ into (4.1) to obtain

$$\gamma^{(j+p)}(h) + \alpha_{p-1}\gamma^{(j+p-1)}(h) + \cdots + \alpha_0\gamma^{(j)}(h) = 0 \qquad h \geq 0. \quad (4.3)$$

Note that these equations are also obtained from the last row of (3.8).

It will be convenient to write (4.2) in the matrix form

$$\Gamma_p(h)\boldsymbol{\alpha} + \boldsymbol{\gamma}_p(h) = 0 \quad (4.4)$$

where $\boldsymbol{\gamma}_p(h) = (-1)^p\left[\gamma^{(p)}(h), \gamma^{(p+1)}(h), \cdots, \gamma^{(2p-1)}(h)\right]'$ and $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \cdots, \alpha_{p-1}]'$.

Note that $\Gamma_p(0)$ is non-singular by Theorem 3.3. So, for $h = 0$ we may write

$$\boldsymbol{\alpha} = -\Gamma_p^{-1}(0)\boldsymbol{\gamma}_p(0). \quad (4.5)$$

5 Estimates Based on the Yule–Walker Equations

In this section, I consider a continuous time analogue of the Yule–Walker method for estimating discrete time autoregressive models. Following the Yule–Walker approach, I shall use the Yule–Walker equations developed in the previous section to estimate the model coefficients by replacing the covariances by sample estimates based on continuously recorded data. The Yule–Walker estimators are shown to be least squares estimators which converge in distribution to maximum likelihood estimators, thus continuing the close analogy between the continuous time Yule–Walker estimators and the discrete time Yule–Walker estimators.

Consider equation (4.5) and let $\Gamma_{j,k}$ denote the (j,k)th element of $\Gamma_p(0)$ and γ_j denote the jth element of $\boldsymbol{\gamma}_p(0)$ $(j,k=1,\ldots,p)$. Then, replacing $\Gamma_{j,k}$ by the sample estimate $\hat{D}_{j-1,k-1}$ and γ_j by the sample estimate $\hat{D}_{j-1,p}$, we obtain the Yule–Walker estimator

$$\hat{\boldsymbol{\alpha}} = -\hat{\Gamma}_p^{-1}(0)\hat{\boldsymbol{\gamma}}_p(0). \tag{5.1}$$

Bartlett (1955) and Priestley (1981) derive these estimators in the cases $p=1$ and $p=2$ using a least squares approach. The following theorem shows that the Yule–Walker estimates satisfy least squares criteria for all p.

$\boxed{\text{Theorem 5.1}}$ *Let $\{x(t)\}$ be a sample function from a CAR(p) process observed at times $0 \le t \le T$. Then the least squares estimators of $\boldsymbol{\alpha}$, obtained by minimising*

$$Q(\boldsymbol{\alpha}) = \int_0^T [\sigma dW(t)]^2 = \int_0^T \left[dx^{(p-1)}(t) + \alpha_{p-1}x^{(p-1)}(t)dt + \cdots + \alpha_0 x(t)dt\right]^2 \tag{5.2}$$

are the Yule–Walker estimators given by (5.1).

Proof: Differentiating (5.2) with respect to α_j gives

$$\int_0^T x^{(j)}(t)\left[dx^{(p-1)}(t) + \breve{\alpha}_{p-1}x^{(p-1)}(t)dt + \cdots + \breve{\alpha}_0 x(t)dt\right] = 0, \qquad j=0,\ldots,p-1$$

where $\breve{\alpha}_j$ is the least squares estimator of α_j. Equivalently,

$$\breve{\alpha}_0\hat{D}_{j,0} + \breve{\alpha}_1\hat{D}_{j,1} + \cdots + \breve{\alpha}_{p-1}\hat{D}_{j,p-1} = -\hat{D}_{j,p} \qquad j=0,1,\ldots,p-1,$$

which are the Yule–Walker equations with $D_{j,k}$ replaced by $\hat{D}_{j,k}$. Hence $\breve{\alpha}_j \equiv \hat{\alpha}_j$ for all j.

□

The asymptotic distribution of the Yule–Walker estimators is given in the following theorem, the proof of which is given in the appendix.

$\boxed{\text{Theorem 5.2}}$ *Let $\hat{\boldsymbol{\alpha}}$ denote the estimator of $\boldsymbol{\alpha}$ defined by (5.1). Then*

$$\sqrt{T}\,(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{d} N\left(\mathbf{0}\,,\ \sigma^2\Gamma_p^{-1}(0)\right) \qquad \text{as } T \to \infty.$$

Arató (1982, p.261) shows that the maximum likelihood estimator of $\boldsymbol{\alpha}$ for a continuously recorded process between 0 and T is asymptotically normally distributed with mean $\boldsymbol{\alpha}$ and covariance matrix $\frac{1}{T}\sigma^2\Gamma_p^{-1}(0)$. Thus, the Yule–Walker estimator has the same asymptotic distribution as the maximum likelihood estimator.

The results of Section 3 suggest some improvements to the Yule–Walker estimates. I shall use a bar ($\bar{\ }$) to distinguish the improved estimates from those above denoted by a hat ($\hat{\ }$).

Using Theorem 3.1, we define $\bar{\gamma}^{(2n)}(0) := (-1)^n \hat{D}_{n,n}$ and $\bar{\Gamma}_{j,k} := (-1)^{k-1}\bar{\gamma}^{(j+k-2)}(0)$ for $j+k$ even. For $j+k$ odd, we define $\bar{\Gamma}_{j,k} := 0$ according to Lemma 3.2.

Similarly, for $j+p-1$ even, define $\bar{\gamma}_j := (-1)^p \bar{\gamma}^{(j+p-1)}(0)$ and for $j+p-1 < 2p$ odd, define $\bar{\gamma}_j := 0$. Finally, $\bar{\gamma}_p$ is estimated by $(-1)^p \bar{\gamma}^{(2p-1)}(0) := \hat{D}_{p-1,p}$.

Then (5.1) may be replaced by

$$\bar{\boldsymbol{\alpha}} = -\bar{\Gamma}_p^{-1}(0)\bar{\boldsymbol{\gamma}}_p(0). \tag{5.3}$$

Consider the structure of $\bar{\Gamma}_p(0)$. Using the above definitions,

$$\bar{\Gamma}_p(0) = \begin{pmatrix} \bar{\gamma}^{(0)}(0) & 0 & \bar{\gamma}^{(2)}(0) & 0 & \bar{\gamma}^{(4)}(0) & \cdots \\ 0 & -\bar{\gamma}^{(2)}(0) & 0 & -\bar{\gamma}^{(4)}(0) & \cdots & \\ \bar{\gamma}^{(2)}(0) & 0 & \bar{\gamma}^{(4)}(0) & \ddots & & \\ 0 & -\bar{\gamma}^{(4)}(0) & \ddots & & & \\ \bar{\gamma}^{(4)}(0) & \vdots & & & & \\ \vdots & & & & & \end{pmatrix}. \tag{5.4}$$

Separating the odd and even rows of $\bar{\Gamma}_p(0)$, we can write equation (5.3) as

$$\begin{aligned} \bar{\boldsymbol{\alpha}}_0 &= -F_p^{-1}\boldsymbol{f}_p & (5.5)\\ \text{and} \quad \bar{\boldsymbol{\alpha}}_1 &= -G_p^{-1}\boldsymbol{g}_p & (5.6) \end{aligned}$$

where

$$F_p = \begin{pmatrix} \bar{\gamma}^{(0)}(0) & \bar{\gamma}^{(2)}(0) & \cdots & \bar{\gamma}^{(f)}(0) \\ \bar{\gamma}^{(2)}(0) & \bar{\gamma}^{(4)}(0) & \cdots & \bar{\gamma}^{(f+2)}(0) \\ \vdots & \vdots & & \vdots \\ \bar{\gamma}^{(f)}(0) & \bar{\gamma}^{(f+2)}(0) & \cdots & \bar{\gamma}^{(2f)}(0) \end{pmatrix}, \qquad \boldsymbol{f}_p = \begin{pmatrix} \bar{\gamma}^{(p)}(0) \\ \bar{\gamma}^{(p+2)}(0) \\ \vdots \\ \bar{\gamma}^{(p+f)}(0) \end{pmatrix},$$

$$G_p = \begin{pmatrix} \bar{\gamma}^{(2)}(0) & \bar{\gamma}^{(4)}(0) & \cdots & \bar{\gamma}^{(g)}(0) \\ \bar{\gamma}^{(4)}(0) & \bar{\gamma}^{(6)}(0) & \cdots & \bar{\gamma}^{(g+2)}(0) \\ \vdots & \vdots & & \vdots \\ \bar{\gamma}^{(g)}(0) & \bar{\gamma}^{(g+2)}(0) & \cdots & \bar{\gamma}^{(2g-2)}(0) \end{pmatrix}, \qquad \boldsymbol{g}_p = \begin{pmatrix} \bar{\gamma}^{(p+1)}(0) \\ \bar{\gamma}^{(p+3)}(0) \\ \vdots \\ \bar{\gamma}^{(p+g-1)}(0) \end{pmatrix},$$

$\bar{\boldsymbol{\alpha}}_0 = [\bar{\alpha}_0, \bar{\alpha}_2, \cdots, \bar{\alpha}_f]'$, $\quad \bar{\boldsymbol{\alpha}}_1 = [\bar{\alpha}_1, \bar{\alpha}_3, \cdots, \bar{\alpha}_{g-1}]'$,

$$f = \begin{cases} p-2 & p \text{ even;} \\ p-1 & p \text{ odd;} \end{cases} \qquad \text{and} \quad g = \begin{cases} p & p \text{ even;} \\ p-1 & p \text{ odd.} \end{cases}$$

For p odd, $\boldsymbol{f}_p = [0, \cdots, 0, \bar{\gamma}^{(2p-1)}(0)]'$ and for p even, $\boldsymbol{g}_p = [0, \cdots, 0, \bar{\gamma}^{(2p-1)}(0)]'$.

So instead of inverting a matrix of order $p \times p$ as in equation (5.1), one need only invert two matrices about half that size.

Clearly, $\bar{\boldsymbol{\alpha}} \xrightarrow{p} \hat{\boldsymbol{\alpha}}$ and so these have the same asymptotic distribution as $\hat{\boldsymbol{\alpha}}$.

One useful property of the improved Yule–Walker estimates is that they allow more rapid computation when estimating models of successively higher orders. In particular, note that if p is even, $F_p = F_{p-1}$ and if p is odd, $G_p = G_{p-1}$.

6 Estimating the Instantaneous Variance

In this section, I consider how to estimate the parameter, σ^2.

For CAR models there is no direct analogue of the Yule–Walker estimate of residual variance. Nevertheless, it is possible to derive a quick and simple estimate based on the process variance $\gamma(0)$ and coefficients $\alpha_0, \ldots, \alpha_{p-1}$.

Let $\mathcal{C}(z) = \sum_{j=0}^{p} \alpha_p z^j = 0$, $\alpha_p = 1$, have q distinct roots $\lambda_1, \ldots, \lambda_q$ where λ_i has multiplicity m_i. Using contour integration, Doob (1953, p.543) showed that

$$\gamma(h) = \sigma^2 \sum_{i=1}^{q} c_i(h) e^{\lambda_i |h|} \tag{6.1}$$

where $c_i(h)$ is a polynomial in h of order m_i. Where all the roots are distinct ($m_i = 1, \forall i$), Jones (1981) gives

$$c_i(h) = \left[-2\mathrm{Re}(\lambda_i) \prod_{\substack{l=1 \\ l \neq i}}^{p} (\lambda_l - \lambda_i)(\bar{\lambda}_l + \lambda_i) \right]^{-1}$$

where $\bar{\lambda}_l$ denotes the complex conjugate of λ_l.

The roots λ_i may be estimated from the coefficient estimates and $\gamma(0)$ may be estimated from the data. Then (6.1) can be used to estimate σ^2.

For $p = 1$, this yields the simple estimate

$$\hat{\sigma}^2 = 2\hat{\alpha}_0 \hat{\gamma}(0)$$

and for $p = 2$,

$$\hat{\sigma}^2 = 2\hat{\alpha}_0 \hat{\alpha}_1 \hat{\gamma}(0).$$

In these cases, if the Yule–Walker estimates of the previous section are used, then $\hat{\sigma}^2 = -2\hat{D}_{p-1,p}$; a result which could have been derived from (3.5).

7 Yule–Walker Estimates for Discrete Observations

In practice, observations are recorded at discrete times. So to use the above estimators, it is necessary to derive discrete time approximations to them.

Suppose the data consists of observations taken at times $0 = t_1 < t_2 < \cdots < t_n = T$ and let $\Delta_i = t_{i+1} - t_i$ and $\Delta = \sup_i \Delta_i$. Then, for small Δ, it seems reasonable to estimate $x^{(j)}(t_i)$ by numerical derivatives $\hat{x}^{(j)}(t_i)$ and $\hat{D}_{j,k}$ by the numerical integrals

$$\begin{aligned} \tilde{D}_{j,k} &= \frac{1}{T}\sum_{i=1}^{n} \hat{x}^{(j)}(t_i)\,\hat{x}^{(k)}(t_i)\Delta_i, \qquad k < p, && (7.1)\\ \text{and} \quad \tilde{D}_{j,p} &= \frac{1}{T}\sum_{i=1}^{n} \hat{x}^{(j)}(t_i)\,\left[\hat{x}^{(p-1)}(t_{i+1}) - \hat{x}^{(p-1)}(t_i)\right] && (7.2)\end{aligned}$$

Define a discrete form of the Yule–Walker estimators by replacing $\hat{D}_{j,k}$ by $\tilde{D}_{j,k}$ in (5.3). Let $\tilde{\boldsymbol{\alpha}}$ denote the estimators which are obtained in this way.

Now suppose the data are equally spaced ($\Delta_i = \Delta$ for all i). Then for $p = 1$, $\tilde{\alpha}_0 = -\frac{1}{\Delta}\left[\frac{\hat{\gamma}(\Delta)}{\hat{\gamma}(0)} - 1\right]$ where $\hat{\gamma}(\Delta) = \frac{1}{n}\sum_{i=1}^{n-1} x(t_i)x(t_{i+1})$ and

$$E(\tilde{\alpha}_0) = -\frac{1}{\Delta}\left[e^{-\alpha_0\Delta} - 1]\right] + O(n^{-1}).$$

Thus the discrete time approximation to the Yule–Walker estimator is asymptotically unbiased (as $n \to \infty$ and $\Delta \to 0$).

However, for $p > 1$, (7.2) gives an asymptotically biased estimate of $D_{j,p}$. This is most easily seen by considering the case $p = 2$ with $\hat{x}^{(1)}(t_i) := [x(t_{i+1}) - x(t_i)]/\Delta$. Then, by the mean value theorem, $\hat{x}^{(1)}(t_i) = x^{(1)}(\xi)$ for some $\xi \in (t_i, t_{i+1})$. Thus, the integrand in $\tilde{D}_{1,2}$ is not evaluated at t_i and so $\tilde{D}_{1,2}$ does not converge to the required Itô integral. Specifically,

$$\begin{aligned} E(\tilde{D}_{1,2}) &= E\left\{\frac{1}{T}\sum_i \hat{x}^{(1)}(t_i)\left[\hat{x}^{(1)}(t_{i+1}) - \hat{x}^{(1)}(t_i)\right]\right\}\\ &= E\Big\{\frac{1}{n\Delta^3}\sum_i [x(t_{i+1})x(t_{i+2}) - x(t_i)x(t_{i+2}) - 2x^2(t_{i+1}) + 2x(t_i)x(t_{i+1})\\ &\qquad\qquad + x(t_{i+1})x(t_i) - x^2(t_i)]\Big\}\\ &\approx \frac{1}{\Delta^3}\left\{-3\gamma(0) + 4\gamma(\Delta) - \gamma(2\Delta)\right\} \quad \text{for large } n. && (7.3)\end{aligned}$$

Now $\gamma(h)$ is the top left element of $\Gamma_p(h) = e^{Ah}\Gamma_p(0)$ using (3.9). For small h, $e^{Ah} \approx I + Ah + A^2h^2/2 + A^3h^3/6$. So the top left element of e^{Ah} is approximately $1 - \alpha_0 h^2/2 + \alpha_0\alpha_1 h^3/6$ and

$$\gamma(h) \approx \gamma(0)\left[1 - \tfrac{\alpha_0 h^2}{2} + \tfrac{\alpha_0\alpha_1 h^3}{6}\right]. \qquad (7.4)$$

Substituting this expression into (7.3) gives $E(\tilde{D}_{1,2}) \approx -2\alpha_0\alpha_1\gamma(0)/3$ for small Δ and large n. Similarly, $E(\tilde{D}_{1,1}) \approx \gamma(0)\alpha_0(1 - \alpha_1\Delta/3)$. But using (3.5) and the results of Section 6, $D_{1,2} = -\alpha_0\alpha_1\gamma(0) \approx \frac{3}{2}E(\tilde{D}_{1,2})$. Hence $E(\tilde{\alpha}_1) = E(-\tilde{D}_{1,2}/\tilde{D}_{1,1}) \approx 2\alpha_1/(3 - \alpha_1\Delta)$. The estimate of α_0 is not asymptotically biased in this way since it doesn't rely on $\tilde{D}_{1,2}$. The above approximations give $E(\tilde{\alpha}_0) = E(\tilde{D}_{1,1}/\tilde{D}_{0,0}) \approx \alpha_0 - \alpha_0\alpha_1\Delta/3$.

Clearly, for discrete time data, another approach is required. If the data are equally spaced, one approach is to equate the ACVF with its sample estimate rather than equating the DCVF with its sample estimate. This continues the flavour of Yule–Walker estimation for discrete autoregressive models although the resulting estimates can no longer be calculated by solving a set of p linear equations. In fact, we obtain a set of p non-linear equations which can be solved with only a little more computation.

A modification of this approach suitable for unequally spaced data will be given at the end of this section.

7.1 Equally spaced data

First, consider the simple CAR(1) model observed at equally spaced intervals of length Δ. Then

$$\gamma(\Delta) = e^{-\alpha_0 \Delta}\gamma(0).$$

This suggests the estimator

$$\tilde{\tilde{\alpha}}_0 := -\frac{1}{\Delta}\log\left[\frac{\hat{\gamma}(\Delta)}{\hat{\gamma}(0)}\right] = -\frac{1}{\Delta}\log\left[1-\Delta\tilde{\alpha}_0\right].$$

Now the restriction of small Δ can be dropped. Note that $e^{-\tilde{\tilde{\alpha}}_0\Delta}$ is the discrete time Yule–Walker estimator for the discrete time AR(1) process formed by the observations. Of course, $\tilde{\tilde{\alpha}}_0$ is also biased because of the ratio estimate of $\gamma(\Delta)/\gamma(0)$. The estimate may be further improved by replacing $\hat{\gamma}(\Delta)/\hat{\gamma}(0)$ with a jackknife estimator.

In general, we have a set of p non-linear equations

$$f_k(\boldsymbol{\alpha}) = \gamma(k\Delta) - \hat{\gamma}(k\Delta), \qquad k = 1, 2, \ldots, p.$$

We wish to find $\boldsymbol{\alpha}$ such that $f_k(\boldsymbol{\alpha}) = 0$ for all k. For $p > 1$ a closed form solution is not possible although the equations may be solved using the Newton–Raphson method as follows. Expanding f_k in a Taylor series yields

$$f_k(\boldsymbol{\alpha} + \delta\boldsymbol{\alpha}) = f_k(\boldsymbol{\alpha}) + \sum_{j=0}^{p-1}\frac{\partial f_k}{\partial \alpha_j}\delta\alpha_j + O(\delta\boldsymbol{\alpha}^2). \tag{7.5}$$

Ignoring the terms of order $\delta\boldsymbol{\alpha}^2$ and higher, we can calculate the corrections $\delta\boldsymbol{\alpha}$ from the set of linear equations

$$\sum_{j=0}^{p-1} f_{k,j}\delta\alpha_j = \beta_k \tag{7.6}$$

where $f_{k,j} = \frac{\partial f_k}{\partial \alpha_j}$ and $\beta_k = -f_k$. Given initial values, $\boldsymbol{\alpha}$, equation (7.5) can be iterated to convergence.

To calculate f_k and $f_{k,j}$ we need to find $\gamma(k\Delta)$ given $\boldsymbol{\alpha}$. This may be determined from the top left element of $e^{Ak\Delta}\Gamma_p(0)$. Elements of e^{Ah} may be calculated using Taylor series approximations and elements of $\Gamma_p(0)$ may be estimated using $\tilde{D}_{j,k}$.

Thus, we obtain

$$\gamma(h) \approx \sum_{k=0}^{p-1} \hat{\gamma}^{(k)}(0) \sum_{m=0}^{M} \frac{c_{m,k} h^m}{m!} \tag{7.7}$$

for some sufficiently large M where $\hat{\gamma}^{(2j)}(0) = \tilde{D}_{j,j}$, $\hat{\gamma}^{(2j-1)}(0) = 0$, $2j \leq p$, and $c_{m,k}$ is a function of the coefficients $\alpha_0, \alpha_1, \ldots, \alpha_{p-1}$.

For example, if $p = 2$, further terms may be added to (7.4) to give

$$\gamma(h) \approx \hat{\gamma}(0) \left[1 - \tfrac{\alpha_0 h^2}{2} + \tfrac{\alpha_0 \alpha_1 h^3}{6} + \tfrac{\alpha_0(-\alpha_1^2+\alpha_0)h^4}{24} + \tfrac{\alpha_0(\alpha_1^3-2\alpha_0\alpha_1)h^5}{120} + \tfrac{\alpha_0(-\alpha_1^4+3\alpha_0\alpha_1^2-\alpha_0^2)h^6}{720}\right]. \tag{7.8}$$

Then f_k and $f_{k,j}$ may be easily calculated for all k and j. In particular,

$$\begin{aligned} f_{k,0}(\boldsymbol{\alpha}) &\approx \gamma(0)\left[-\tfrac{h^2}{2} + \tfrac{\alpha_1 h^3}{6} + \tfrac{(-\alpha_1^2+2\alpha_0)h^4}{24} + \tfrac{(\alpha_1^3-4\alpha_0\alpha_1)h^5}{120} + \tfrac{(-\alpha_1^4+6\alpha_0\alpha_1^2-3\alpha_0^2)h^6}{720}\right] \\ \text{and} \quad f_{k,1}(\boldsymbol{\alpha}) &\approx \gamma(0)\left[\tfrac{\alpha_0 h^3}{6} - \tfrac{2\alpha_0\alpha_1 h^4}{24} + \tfrac{\alpha_0(3\alpha_1^2-2\alpha_0)h^5}{120} + \tfrac{\alpha_0(-4\alpha_1^3+6\alpha_0\alpha_1)h^6}{720}\right] \end{aligned}$$

where $h = k\Delta$. Again, we do not need Δ small for this approximation, although for $p > 2$ this restriction is necessary to calculate $\hat{\gamma}^{(2j)}(0)$, $j \geq 1$.

For $p = 3$ and $p = 4$, values of $c_{m,k}$ are given in Table 1.

	$p=1$	$p=2$	$p=3$	$p=4$
$c_{0,0}$	1	1	1	1
$c_{1,0}$	$-\alpha_0$	0	0	0
$c_{2,0}$	α_0^2	$-\alpha_0$	0	0
$c_{3,0}$	$-\alpha_0^3$	$\alpha_0\alpha_1$	$-\alpha_0$	0
$c_{4,0}$	α_0^4	$\alpha_0(-\alpha_1^2+\alpha_0)$	$\alpha_0\alpha_2$	$-\alpha_0$
$c_{5,0}$	$-\alpha_0^5$	$\alpha_0(\alpha_1^3-2\alpha_0\alpha_1)$	$\alpha_0(-\alpha_2^2+\alpha_1)$	$\alpha_0\alpha_3$
$c_{6,0}$	α_0^6	$\alpha_0(-\alpha_1^4+3\alpha_0\alpha_1^2-\alpha_0^2)$	$\alpha_0(\alpha_2^3-2\alpha_1\alpha_2+\alpha_0)$	$\alpha_0(-\alpha_3^2+\alpha_2)$
$c_{7,0}$	$-\alpha_0^7$	$\alpha_0(\alpha_1^5-4\alpha_0\alpha_1^3+3\alpha_0^2\alpha_1)$	$\alpha_0(-\alpha_2^4+3\alpha_1\alpha_2^2-2\alpha_0\alpha_2+\alpha_1^2)$	$\alpha_0(\alpha_3^3-2\alpha_2\alpha_3+\alpha_1)$
$c_{0,2}$			0	0
$c_{1,2}$			0	0
$c_{2,2}$			1	1
$c_{3,2}$			$-\alpha_2$	0
$c_{4,2}$			$\alpha_2^2-\alpha_1$	$-\alpha_2$
$c_{5,2}$			$-\alpha_2^3+2\alpha_1\alpha_2-\alpha_0$	$\alpha_2\alpha_3-\alpha_1$
$c_{6,2}$			$\alpha_2^4-3\alpha_1\alpha_2^2+2\alpha_0\alpha_2+\alpha_1^2$	$-\alpha_2\alpha_3^2+\alpha_1\alpha_3-\alpha_0+\alpha_2^2$
$c_{7,2}$			$-\alpha_2^5+4\alpha_1\alpha_2^3-3\alpha_0\alpha_2^2-3\alpha_1^2\alpha_2+2\alpha_0\alpha_1$	$\alpha_2\alpha_3^3-\alpha_1\alpha_3^2-2\alpha_2^2\alpha_3+2\alpha_1\alpha_2+\alpha_0\alpha_3$

Table 1: *Values of $c_{m,k}$ for use with equation (7.7).*

7.2 Unequally spaced data

Let $\bar{\gamma}_k = \frac{1}{n}\sum_i X(t_{i+k})X(t_i)$ so that $E[\bar{\gamma}_k] = \frac{1}{n}\sum_i \gamma(\Delta_{i,k})$ where $\Delta_{i,k} = t_{i+k} - t_i$. Note that for equally spaced data, $\bar{\gamma}_k = \hat{\gamma}(k\Delta)$. Using (7.7) and interchanging the order of summation, we obtain

$$E[\bar{\gamma}_k] \approx \sum_{k=0}^{p-1} \hat{\gamma}^{(k)}(0) \sum_{m=0}^{M} \frac{c_{m,k}}{m!} \frac{1}{n} \sum_{i=1}^{n-k} \Delta_{i,k}^{m}.$$

Hence, the estimates derived above may be modified for unequally spaced data by replacing $\hat{\gamma}(k\Delta)$ with $\bar{\gamma}_k$, $\gamma(k\Delta)$ with $E[\bar{\gamma}_k]$ and $(k\Delta)^m$ with $\frac{1}{n}\sum_i \Delta_{i,k}^m$.

8 Applications

In this section, the modified Yule–Walker estimators derived in Section 7 are applied to two data sets which have been subjected to previous analyses.

8.1 Wolf's Sunspot Numbers

Brockwell and Hyndman (1991) fitted a CAR(2) model to Wolf's sunspots numbers for 1770–1869 (Brockwell and Davis, 1991, p.6) using maximum likelihood estimation. They give the model

$$X^{(2)}(t) + 0.495X^{(1)}(t) + 0.435X(t) = 24.72W^{(1)}(t)$$

where $X(t)$ denotes the mean-corrected data. The discrete time approximation of the Yule–Walker estimators gives $\tilde{\alpha}_0 = 0.3630$ and $\tilde{\alpha}_1 = 0.4255$. Starting with these values, (7.5) was iterated to obtain the model

$$X^{(2)}(t) + 0.5496X^{(1)}(t) + 0.4427X(t) = 25.996W^{(1)}(t).$$

8.2 Canadian Lynx Data

The logged Canadian lynx data have been analysed by Tong and Yeung (1991) who

fit a CAR(2) model using maximum likelihood estimation. They give the model

$$X^{(2)}(t) + 0.491X^{(1)}(t) + 0.433X(t) = 0.736W^{(1)}(t)$$

where, again, $X(t)$ denotes the mean-corrected data. The Yule–Walker estimates give $\tilde{\alpha}_0 = 0.4169$ and $\tilde{\alpha}_1 = 0.4380$. Iterating (7.5) gives

$$X^{(2)}(t) + 0.4207X^{(1)}(t) + 0.4969X(t) = 0.3590W^{(1)}(t).$$

Acknowledgments

I am greatly indebted to Professor Peter Brockwell and Dr Gary Grunwald for many valuable discussions in connection with this work and to BHP Melbourne Research Laboratories, Australia, for supporting this work with a BHP Research Grant.

References

ARATÓ, M. (1982) *Linear stochastic systems with constant coefficients: a statistical approach*, Lecture notes in control and information sciences, **45**, (Springer-Verlag: Berlin).

ASH, R.B. GARDNER, M.F. (1975) *Topics in stochastic processes*, (Academic Press: New York).

BARTLETT, M.S. (1946) On the theoretical specification of sampling properties of autocorrelated time-series, *J. Roy. Statist. Soc. Ser. B*, **8**, 27–41.

BARTLETT, M.S. (1955) *An introduction to stochastic processes with special reference to methods and applications*, (Cambridge University Press: London).

BERGSTROM, A.R. (1984) Continuous time stochastic models and issues of aggregation over time, In *Handbook of econometrics*, Vol. 2, ed. Griliches, Z. and Intriligator, M.D., (Elsevier Science Publishers: BV), 1146–1212.

BROCKWELL, P.J. DAVIS, R.A. (1991) *Time series: theory and methods,* second edition, (Springer-Verlag: New York).

BROCKWELL, P.J. HYNDMAN, R.J. (1991) On continuous-time threshold autoregression, *J. Forecasting*, To appear, 1992.

DOOB, J.L. (1953) *Stochastic processes*, (John Wiley and Sons: New York).

DURBIN, J. (1961) Efficient fitting of linear models for continuous stationary time series from discrete data, *Bull. Int. Statist. Inst.*, **38**, 273–281.

DZHAPARIDZE, K.O. (1971) On methods for obtaining asymptotically efficient spectral parameter estimates for a stationary Gaussian process with rational spectral density, *Th. Prob. Appl.*, **16**, 550–554.

JAZWINSKI, A.H. (1970) *Stochastic processes and filtering theory*, (Academic Press: New York).

JONES, R.H. (1981) Fitting a continuous time autoregression to discrete data, In *Applied Time Series Analysis II*, ed. D.F. Findlay, (Academic Press: New York), 651–682.

ØKSENDAL, B. (1989) *Stochastic differential equations: an introduction with applications*, second edition, (Springer–Verlag: Berlin).

PHILLIPS, P.C.B. (1972) The structural estimation of a stochastic differential equation system, *Econometrica*, **40**, 1021–1041.

PRIESTLEY, M.B. (1981) *Spectral analysis and time series, Vol. 1: univariate series*, (Academic Press: London).

ROBINSON, P.M. (1977) Estimation of a time series model from unequally spaced data, *Stoch. Proc. Appl.*, **6**, 9–24.

ROBINSON, P.M. (1980) Continuous model fitting from discrete data, In *Directions in time series*, ed. Brillinger, D. and Tiao, G., 263–278.

TONG, H. YEUNG, I. (1991) Continuous time threshold autoregressive models, *Statistica Sinica*, **1**, 411–430.

YAGLOM, A.M. (1987) *Correlation theory of stationary and related random functions, Vol. 1: basic results*, (Springer-Verlag: New York).

Appendix

Proof of Theorem 5.2

Let $\{x(t)\}$, $0 < t < T$ denote a sample function of a CAR(p) process. Also let n be some positive integer and define $\Delta = T/n$ and $t_i = i\Delta$. Then, for large n, $\hat{D}_{j,k}$ may be approximated by

$$\begin{aligned} \hat{D}_{j,k,n} &= \frac{1}{n}\sum_{i=0}^{n} x^{(j)}(t_i)\, x^{(k)}(t_i), \qquad k < p, \\ \text{and} \quad \hat{D}_{j,p,n} &= \frac{1}{n\Delta}\sum_{i=0}^{n-1} x^{(j)}(t_i)\, \left[x^{(p-1)}(t_{i+1}) - x^{(p-1)}(t_i)\right] \end{aligned}$$

so that $\hat{D}_{j,k,n} \xrightarrow{p} \hat{D}_{j,k}$, $0 \le k \le p$ as $n \to \infty$ (Yaglom, 1987).

Define $\hat{\boldsymbol{\alpha}}_n$ to be the estimator obtained by replacing $\hat{D}_{j,k}$ by $\hat{D}_{j,k,n}$ in (5.1). Then $\hat{\boldsymbol{\alpha}}_n$ is the standard least squares estimator for the regression problem

$$Y(t_i) = -\sqrt{\Delta}\left(\alpha_0 x(t_i) + \cdots + \alpha_{p-1}x^{(p-1)}(t_i)\right) + e(t_i)$$

$$\begin{aligned} \text{where} \quad Y(t_i) &= \left[x^{(p-1)}(t_{i+1}) - x^{(p-1)}(t_i)\right]/\sqrt{\Delta} \\ \text{and} \quad e(t_i) &= \sigma\left[W(t_{i+1}) - W(t_i)\right]/\sqrt{\Delta}. \end{aligned}$$

That is

$$\begin{aligned} \boldsymbol{Y} &= -X\boldsymbol{\alpha} + \boldsymbol{e} \\ \text{and} \quad \hat{\boldsymbol{\alpha}}_n &= -(X'X)^{-1}(X'\boldsymbol{Y}) \end{aligned}$$

$$\begin{aligned} \text{where} \quad \boldsymbol{Y} &= \left[Y(t_0), \ldots, Y(t_{n-1})\right]', \\ \boldsymbol{e} &= \left[e(t_0), \ldots, e(t_{n-1})\right]' \\ \text{and} \quad X &= \sqrt{\Delta}\begin{bmatrix} x(t_0) & x^{(1)}(t_0) & \cdots & x^{(p-1)}(t_0) \\ x(t_1) & x^{(1)}(t_1) & \cdots & x^{(p-1)}(t_1) \\ \vdots & \vdots & & \vdots \\ x(t_{n-1}) & x^{(1)}(t_{n-1}) & \cdots & x^{(p-1)}(t_{n-1}) \end{bmatrix}. \end{aligned}$$

Thus, $e(t_i)$ is a sequence of independent, normally distributed random variables with zero mean and variance σ^2.

Consider

$$T^{\frac{1}{2}}(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}) \;=\; T^{\frac{1}{2}}\left[(X'X)^{-1}X'(X\boldsymbol{\alpha} - \boldsymbol{e}) - \boldsymbol{\alpha}\right] \;=\; -\,T(X'X)^{-1}\left[T^{-\frac{1}{2}}X'\boldsymbol{e}\right].$$

By setting $\boldsymbol{U}(t_i) = \left[x(t_i), x^{(1)}(t_i), \ldots, x^{(p-1)}(t_i)\right]' e(t_i)$ we have

$$T^{-\frac{1}{2}}X'\boldsymbol{e} \;=\; T^{-\frac{1}{2}}\sum_{i=0}^{n-1}\sqrt{\Delta}\boldsymbol{U}(t_i) \;\xrightarrow{p}\; n^{-\frac{1}{2}}\sum_{i=0}^{n}\boldsymbol{U}(t_i) \qquad \text{as } n \to \infty.$$

Also, $E\left[\boldsymbol{U}(t_i)\right] = \mathbf{0}$, $\operatorname{Cov}\left[\boldsymbol{U}(t_i), \boldsymbol{U}(t_j)\right] = O_{p\times p}$, $i \neq j$ and $\operatorname{Cov}\left[\boldsymbol{U}(t_i), \boldsymbol{U}(t_i)\right] = \sigma^2\Gamma_p(0)$ since $W(t_i)$ and $W(t_{i+1})$ are independent of $x(t_i), \ldots, x^{(p-1)}(t_i)$.

Now $x^{(k)}(t)$ may be denoted by $\sigma\int_0^\infty \psi_k(y)\, dW(t-y)$ (Arató, 1982, p.119). Hence, we may define the sequence of approximations $x_m^{(k)}(t_i) = \sigma\int_0^{t_m} \psi_k(y)\, dW(t_i - y)$ where m is a fixed positive integer. Let $\boldsymbol{S}_m(t_i) = \left[x_m(t_i), x_m^{(1)}(t_i), \ldots, x_m^{(p-1)}(t_i)\right]'$ and $\boldsymbol{U}_m(t_i) = \boldsymbol{S}_m(t_i)e(t_i)$ and let $\boldsymbol{\lambda}$ be a fixed but arbitrary vector in $\mathbb{R}^p$. Then $Z_i = \boldsymbol{\lambda}'\boldsymbol{U}_m(t_i)$ is a strictly stationary sequence with zero mean and variance $\sigma^2\boldsymbol{\lambda}'\Gamma_{m,p}(0)\boldsymbol{\lambda}$ where $\Gamma_{m,p}(0)$ is the covariance matrix of $\boldsymbol{S}_m(t)$. The sequence Z_i also has the property $E\left[Z_i Z_{i+h}\right] = 0$ when $h > m+1$. That is, it is $(m+1)$-dependent. Hence, we may apply the central limit theorem for dependent stationary sequences (see, e.g., Brockwell and Davis, 1991, p.213) to obtain

$$n^{-\frac{1}{2}}\sum_{i=0}^{n-1}\boldsymbol{\lambda}'\boldsymbol{U}_m(t_i) \xrightarrow{d} \boldsymbol{\lambda}'\boldsymbol{V}_m \quad \text{where} \quad \boldsymbol{V}_m \stackrel{d}{=} N\left(\mathbf{0}\ ,\ \sigma^2\Gamma_{m,p}(0)\right).$$

Since $\sigma^2\Gamma_{m,p}(0) \to \sigma^2\Gamma_p(0)$ as $m\to\infty$, we have $\boldsymbol{\lambda}'\boldsymbol{V}_m \xrightarrow{d} \boldsymbol{\lambda}'\boldsymbol{V}$ where $\boldsymbol{V} \stackrel{d}{=} N\left(\mathbf{0}\ ,\ \sigma^2\Gamma_p(0)\right)$. Also, $n^{-1}\operatorname{Var}\left(\boldsymbol{\lambda}'\sum_{i=0}^{n-1}\left[\boldsymbol{U}_m(t_i) - \boldsymbol{U}(t_i)\right]\right) \to 0$ as $m\to\infty$.

Since $x_m(t) \xrightarrow{ms} x(t)$, application of Brockwell and Davis's Proposition 6.3.9 (1991, p.207–208) and the Cramer–Wold device gives

$$T^{-\frac{1}{2}}X'\boldsymbol{e} \xrightarrow{d} N\left(\mathbf{0}\ ,\ \sigma^2\Gamma_p(0)\right).$$

It follows from (3.6) that $T(X'X)^{-1} \xrightarrow{p} \Gamma_p^{-1}(0)$. Then, using Brockwell and Davis's Proposition 6.4.2 (1991, p.211),

$$\hat{\boldsymbol{\alpha}}_n \stackrel{d}{=} AN\left(\boldsymbol{\alpha}\ ,\ T^{-1}\sigma^2\Gamma_p^{-1}(0)\right).$$

Now $\hat{\boldsymbol{\alpha}}_n \xrightarrow{p} \hat{\boldsymbol{\alpha}}$ as $n\to\infty$ since $D_{j,k,n} \xrightarrow{p} D_{j,k}$. So

$$\hat{\boldsymbol{\alpha}} \stackrel{d}{=} AN\left(\boldsymbol{\alpha}\ ,\ T^{-1}\sigma^2\Gamma_p^{-1}(0)\right).$$