1993

Mean Likelihood Estimation And Time Series Analysis

Joseph Gilles Quenneville

MEAN LIKELIHOOD ESTIMATION AND TIME SERIES ANALYSIS

by

Joseph Gilles Benoît Quenneville

Department of Statistical and Actuarial Sciences

Submitted in partial fulfilment

of the requirements of the degree of

Doctor of Philosophy

Faculty of Graduate Studies

The University of Western Ontario

London, Ontario

August 1993

LIST OF TABLES

ABSTRACT

In this thesis, I develop mean likelihood estimation (MeLE) and maximum likelihood estimation (MLE) for the parameters of a fractionally differenced autoregressive moving average (FARMA) model and its integrated form (IFARMA).

In chapter 1, I study the sampling distribution theory of MeLE by embedding the estimation problem into a Bayesian model.

In chapter 2, I apply MeLE and MLE to the estimation of the moving average parameter of a MA(1) model. The main result is that the MeLE has a smaller mean square error than the MLE when the moving average parameter is sufficiently inside the parameter space. I also demonstrate that the concentrated likelihood function of a time series with a MA(1) component has a local maximum or minimum at the boundary ± 1.

In chapter 3, I give the theory and methods needed to simulate FARMA series and to estimate the parameters of the FARMA model by MeLE and MLE. I explain how to calculate the exact value of the likelihood function for the parameters of a FARMA model. I show how to integrate functions over the stationary and invertible region of an ARMA process. Finally, I conduct simulation studies that compare the MLE and MeLE in combination with various algorithms to calculate them.

In the last chapter, I solve the problem of defining a likelihood function that unifies both the stationary FARMA and non-stationary IFARMA models, and I show how to select an appropriate model for a given time series by using a minimum information criteria strategy.

ACKNOWLEDGEMENTS

I would like to express my appreciation and gratitude to Dr. Estela Bee Dagum for her never ending support throughout my thesis. She motivated me to start a Ph.D. program, and provided me full support at work. Without her commitment, I would have never been able to finish this work. This thesis is dedicated to her.

It has been an honour and a privilege for me to have worked under the direction of Professor Angus Ian McLeod. I would like to thank him for his helpful guidance, stimulating discussions, and kind hospitality.

I am grateful to Statistics Canada, and especially to Gordon Brackstone, for providing me the environmental support. I thank all my colleagues: Zhao, Bernard, Helen, Guy, Marietta, Norma, Pierre, Paul, France, Kim and John for their encouragements. Special thanks to Bernard Lefrançois with whom I also had stimulating discussions, and who read one chapter of my thesis. I am indebted to Zhao-Guo Chen who read all the chapters of my thesis, and provided me valuable comments and advices. Both of them motivated me like no one else could have done.

I thank the members of the examining board: Dr. David Bellhouse, Dr. John Knight, Dr. Reginald Kulperger, and Dr. Rock Roy for their comments to improve the content of the thesis; and the thesis offic er, Renée Cordeiro, for her comments to improve the format of the thesis.

Last but not least, I thank Chantal, Danielle and Laurent who had to live with me through this intense emotional period where you strive for a Ph.D..

TABLE OF CONTENTS

CHAPTER 3

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

MEAN LIKELIHOOD ESTIMATION

In this chapter I study the sampling distribution theory of mean likelihood estimators by embedding the estimation problem into a Bayesian model. I show that the mean likelihood estimator has the following properties. It depends on a sufficient statistic if a sufficient statistic exists. When the parameter space is bounded, the mean likelihood estimator is usually biased, and it has the smallest average mean square error, where the average is over the parameter space. Finally, the mean likelihood estimator has the same asymptotic distribution as the maximum likelihood estimator.

1. Introduction.

In this chapter I study the sampling distribution theory of the mean likelihood estimator (MeLE). The method of mean likelihood estimation is an alternative to the method of maximum likelihood estimation (MLE), where a parameter is estimated by the average of the likelihood function instead of the maximum of the likelihood function. I show that the MeLE has the following properties. The MeLE depends on a sufficient statistic if a sufficient statistic exists. When the parameter space is bounded, the MeLE is usually biased, and it has the smallest average mean square error, where the average is over the parameter space. Finally, the MeLE has the same asymptotic distribution as the MLE.

To show these properties of the MeLE, I organize this chapter on mean likelihood estimation as follows. In section 2, I review the concepts and methods of the Bayesian inference that I find useful for MeLE. I put an emphasis on the sampling distribution theory of Bayes type estimators. In section 3, I review the likelihood approach to statistics, and define the mean likelihood estimator. Next, I study the sampling distribution theory of MeLE by embedding the estimation problem into a Bayesian model. I give the concluding remarks in section 4.

2. Bayesian methods.

This section reviews the Bayesian methodology that will be used later. First the likelihood principle is stated.

Definition 2.1: Let θ be an unknown quantity called the state of nature, x an observed vector of data from a random vector X whose probability distribution $f(x|\theta)$ depends on θ. The function $l(\theta|x) = f(x|\theta)$, considered as a function of θ, x fixed, is the likelihood function.

Remark 2.2: The likelihood principle says that in making inferences about θ after x is observed, all relevant experimental information is in the likelihood function. Also, if two likelihood functions are proportional to each other as a function of θ, then identical conclusions about θ should be drawn.

Definition 2.3: Let Φ be the parameter space in which θ takes values. Assume that θ is from the random variable Θ, where Θ has probability density $\pi(\theta)$, then $\pi(\theta)$ is the prior distribution of θ.

Definition 2.4: The posterior distribution of Θ given x, denoted by $\pi(\theta|x)$, is the conditional distribution of Θ given x. It is obtained by combining the prior distribution and the likelihood in the following way:

$$\pi(\theta|x) = \frac{\pi(\theta)\, l(\theta|x)}{\int_{\Phi} \pi(\theta)\, l(\theta|x)\, d\theta}. \tag{2.1}$$

Remark 2.5: The likelihood principle and the prior distribution are used to derive the posterior distribution. Berger (1985, p.32) points out:

> "If no prior information about θ is available it is natural to attempt to interpret the likelihood function as some kind of probability density for θ. The ambiguity arises in the need to then specify the "measure" with respect to which it is a density. There are often many plausible choices for this measure, and the choice can have a considerable effect on the conclusion reached. This problem is basically that of choosing a "noninformative" prior distribution. . . ."

Definition 2.6: A prior distribution that contains no information about Θ is a noninformative prior. It refers to a prior distribution that does not favour any particular values of θ over others (see Berger (1985) for the problem of defining and determining noninformative priors rigorously).

Remark 2.7: The obvious choice of a noninformative prior distribution is to select $\pi(\theta) \propto c$, c is a constant, for all θ in Φ. This leads to a posterior distribution proportional to the likelihood function. This choice of prior distribution is not invariant under non-singular transformation of θ. Jeffreys (1961) has shown that selecting $\pi(\theta) \propto |I(\theta)|^{1/2}$, where $|I(\theta)|$ is the determinant of the Fisher information matrix $I(\theta)$, gives a noninformative prior distribution that is invariant under non-singular transformation. When Jeffreys prior distribution is selected, the posterior distribution is proportional to:

$$\pi(\theta|x) \propto l(\theta|x)\ |I(\theta)|^{1/2} \tag{2.2}$$

Remark 2.8: Bayesian estimation is based solely on the information contained in the posterior distribution $\pi(\theta|x)$. Two common Bayesian point estimates of θ are now given.

Definition 2.9: The generalized maximum likelihood estimate of θ is the value $\hat{\theta}^*$ that maximizes $\pi(\theta|x)$. In other words, $\hat{\theta}^*$ is the mode of the posterior distribution.

Definition 2.10: The posterior mean of Θ, denoted by $\mu^*(x)$, is the expectation of Θ with respect to its posterior distribution. That is:

$$\mu^*(x) = \int_{\bullet} \theta \pi(\theta|x) d\theta. \tag{2.3}$$

Definition 2.11: Let $\delta(X)$ be an estimator of θ.

a) The loss function $L(\theta,\delta(X))$ is a real-valued function that satisfies:

$$L(\theta,\delta(x)) \geq 0, \ \forall\theta \in \Phi, \ \delta(x)$$
$$L(\theta,\delta(x)) = 0, \ \delta(x) = \theta.$$

b) The posterior expected loss of an estimator $\delta(X)$ of θ is:

$$\rho(\pi(\theta|x), \ \delta(x)) = \int_{\bullet} L(\theta,\delta(x)) \pi(\theta|x) d\theta. \tag{2.4}$$

Definition 2.12: The estimate $\delta^*(x)$ that minimizes the posterior expected loss is the Bayes estimate.

Proposition 2.13: The posterior mean is the Bayes estimator under the square error loss, $L(\theta,\delta(X)) = (\delta(X)-\theta)^2$.

Definition 2.14: The usual Bayesian measure of the accuracy of the posterior mean, is the posterior covariance matrix,

$$V^{\pi}(x) = \mathcal{E}[\{\Theta - \mathcal{E}(\Theta|x)\}\{\Theta - \mathcal{E}(\Theta|x)\}'], \tag{2.5}$$

where the expectation is with respect to the posterior distribution.

Analysis of Bayes estimators.

Definition 2.15: The risk function of an estimator $\delta(X)$ is the average loss of using $\delta(X)$ as X varies over the sample space. The risk function is:

$$R(\theta,\delta(X)) = \mathcal{E}[L(\theta,\delta(X))] = \int L(\theta,\delta(x))f(x|\theta)dx. \tag{2.6}$$

Definition 2.16: The Bayes risk of an estimator $\delta(X)$ with respect to the loss function $L(\theta,\delta(X))$ and prior distribution $\pi(\theta)$ is:

$$r(\pi,\delta) = \int_{\Theta} R(\theta,\delta(X))\pi(\theta)d\theta. \tag{2.7}$$

Proposition 2.17: The Bayes estimator minimizes the Bayes risk.

Proof: Substitution of (2.6) for $R(\theta,\delta)$ in (2.7) gives:

$$r(\pi,\delta) = \int_{\Theta} \left\{ \int L(\theta,\delta(x))f(x|\theta)dx \right\} \pi(\theta)d\theta.$$

By Fubini's theorem, the order of integration can be interchanged:

$$r(\pi,\delta) = \int \left\{ \int_{\Theta} L(\theta,\delta(x))f(x|\theta)\pi(\theta)d\theta \right\} dx.$$

Whenever the density of X is greater than zero we can write $f(x|\theta)\pi(\theta)$ as $\pi(\theta|x)f(x)$ where

$$f(x) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta,$$

and the last integral becomes:

$$r(\pi,\delta) = \int\left\{\int_{\Theta} L(\theta,\delta(x))\pi(\theta|x)d\theta\right\} f(x)dx .$$

Since $f(x) \geq 0$, then the Bayes estimator minimizes $r(\pi, \delta)$.

Proposition 2.18: (Mood, Graybill and Boes, p. 343) Let $\delta^*(X)$ be the Bayes estimator of θ with respect to the square error loss function and prior distribution $\pi(\theta)$. If $\delta^*(X)$ has a finite variance with respect to the distribution of X and the variance of Θ is finite with respect to the prior distribution, then either $Var[\,\delta^*(X)|\theta\,] = 0$ with probability 1, or $\delta^*(X)$ is not an unbiased estimator of θ.

Proof: Assume $\delta^*(X)$ is an unbiased estimator of θ, i.e. $\mathcal{E}(\delta^*(X)|\theta) = \theta$. By definition $\delta^*(X) = \mathcal{E}(\Theta|X)$. Now:

$$\begin{aligned} Var(\delta^{\pi}(X)) &= \mathcal{E}[Var(\delta^{\pi}(X)|\Theta)] + Var[\mathcal{E}(\delta^{\pi}(X)|\Theta)] \\ &= \mathcal{E}[Var(\delta^{\pi}(X)|\Theta)] + Var(\Theta) \end{aligned} \tag{2.8}$$

and

$$\begin{aligned} Var(\Theta) &= \mathcal{E}[Var(\Theta|X)] + Var[\mathcal{E}(\Theta|X)] \\ &= \mathcal{E}[Var(\Theta|X)] + Var[\delta^{\pi}(X)]. \end{aligned} \tag{2.9}$$

Substitution of (2.9) into (2.8) and cancellation of the common factor $Var[\delta^{\pi}(X)]$ give:

$$0 = \mathcal{E}[Var(\delta^{\pi}(X)|\Theta)] + \mathcal{E}[Var(\Theta|X)].$$

Since both expectations are non-negative and their sum is zero, both are zero. In particular, $\mathcal{E}[\,Var\,(\,\delta^{\pi}(X)|\Theta\,)\,] = 0$, and since $Var\,(\,\delta^{\pi}(X)|\Theta\,)$ is non-negative and has zero expectation, $Var\,(\,\delta^{\pi}(X)|\Theta\,) = 0$.

Proposition 2.19: *A Bayesian version of the Central Limit Theorem (Berger 1985, p.224).* Suppose $X_1,\ldots,X_n$ are i.i.d. from the density $f(x_i|\theta)$, θ being an unknown vector of parameters of dimension p. Suppose $f(x_i|\theta)$ is twice differentiable near $\hat{\theta}$, the (assumed to exist) maximum likelihood estimate of θ. Then for large n and under commonly satisfied assumptions (explained below), the posterior density:

$$\pi_n(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta} = \frac{\prod_{i=1}^{n} f(x_i|\theta)\ \pi(\theta)}{\int_{\Theta} \prod_{i=1}^{n} f(x_i|\theta)\ \pi(\theta)d\theta}$$

can be approximated in the following four ways:

i) π_n is approximately $N_p(\ \mu^*(x),\ V^*(x)\ )$ where $\mu^*(x)$ and $V^*(x)$ are the posterior mean and covariance matrix.

ii) π_n is approximately $N_p(\ \hat{\theta}^*,\ [\ I^*(x)\ ]^{-1}\ )$ where $\hat{\theta}^*$ is the generalized maximum likelihood estimate and $I^*(x)$ the pxp matrix, having (i,j) element:

$$I_{ij}^*(x) = -\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log[\ f(x\,|\,\theta)\pi(\theta)\ ]\Big|_{\theta\,=\,\hat{\theta}^*}. \tag{2.10}$$

iii) π_n is approximately $N_p(\ \hat{\theta},\ I^{-1}(x)\ )$ where $\hat{\theta}$ is the maximum likelihood estimate and $I(x)$ the observed Fisher information matrix, having (i,j) element:

$$I_{ij}(x) = -\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log[\ f(x\,|\,\theta)\ ]\Big|_{\theta\,=\,\hat{\theta}}. \tag{2.11}$$

iv) π_n is approximately $N_p(\ \hat{\theta},\ I^{-1}(\hat{\theta})\ )$ where $\hat{\theta}$ is the maximum likelihood estimate and $I(\hat{\theta})$ the expected Fisher information matrix, having (i,j) element:

$$I_{ij}(\hat{\theta}) = -n\mathscr{E}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log[\ f(X_1\,|\,\theta)\ ]\Big|_{\theta\,=\,\hat{\theta}}\right], \tag{2.12}$$

where the expectation is with respect to the distribution of X_1.

The assumptions are the regularity conditions provided in Walker (1969). These conditions ensure that the MLE is consistent; the MLE has an asymptotic normal distribution; and the prior density is continuous at the true value θ_0 of the parameter, and $\pi(\theta_0) > 0$.

Bayesian calculation.

Remark 2.20: It is conceptually easy to calculate the posterior quantity, say $g(\theta)$, of interest. It suffices to calculate

$$\mathcal{E}[g(\theta)\,|x] = \int_{\Phi} g(\theta)\,\pi(\theta|x)\,d\theta = \frac{\int_{\Phi} g(\theta) f(x|\theta)\,\pi(\theta)\,d\theta}{\int_{\Phi} f(x|\theta)\,\pi(\theta)\,d\theta}. \tag{2.13}$$

In most cases there are no analytical expressions for the posterior density and a numerical integration method has to be used to calculate (2.13). Simpson's rule can be used when the parameter space is an interval. Monte-Carlo integration is preferable in higher dimensional space.

Remark 2.21: The method of Monte-Carlo integration is easily justified. Suppose $h(\theta)$ is a positive density on the parameter space Φ. Then with respect to this density:

$$\begin{aligned} \mathcal{E}\left[\frac{g(\theta) f(x|\theta)\,\pi(\theta)}{h(\theta)}\right] &= \int_{\Phi} \frac{g(\theta) f(x|\theta)\,\pi(\theta)}{h(\theta)}\, h(\theta)\,d\theta \\ &= \int_{\Phi} g(\theta) f(x|\theta)\,\pi(\theta)\,d\theta\,. \end{aligned} \tag{2.14}$$

If $\theta_1, \theta_2, \ldots, \theta_m$ is a sample from a random variable with density $h(\theta)$, it follows by the Weak Law of Large Number and (2.14) that

$$\lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} \frac{g(\theta_i) f(x|\theta_i)\,\pi(\theta_i)}{h(\theta_i)} \to \int_{\Phi} g(\theta) f(x|\theta)\,\pi(\theta)\,d\theta\,. \tag{2.15}$$

A suitable choice of the function g gives the desired quantity. For example the posterior mean is the fraction obtained with $g(\theta) = \theta$ and $g(\theta) = 1$ for the numerator and denominator respectively.

Definition 2.22: The function $h(\theta)$ is the importance function.

3. Likelihood methods.

Remark 3.1: *The likelihood approach to statistics.* It is well known that the likelihood function and likelihood inference was introduced into modern statistics by R.A. Fisher. The likelihood approach to statistical inference was further developed by Barnard (1949) and Barnard, Jenkins and Winsten (1962). It is summed up in the statement that probability distributions are useful in describing the data before data collection, whereas likelihoods are useful in describing parameters after data collection. The likelihood approach asserts that the properties of the estimate must depend on the data available, not on the data that might have been obtained. This contrasts with the sampling distribution approach to statistical inference and agrees with the Bayesian approach. The main difference between the likelihood approach and the Bayesian approach is that the parameters have a prior distribution in the Bayesian approach but the parameters are fixed quantities in the likelihood approach. The likelihood approach is then straightforward and is summarized as follows. 1) The sampling probability distribution function for the observations is assumed to be known and to depend on a set of unknown parameters. 2) After data collection, the likelihood function is calculated by substituting the actual values of the observations into the probability density function (p.d.f.). 3) Finally, suitable means of summarizing the likelihood functions are found.

Example 3.2: Assume $X = (X_1,\ldots,X_n)$ is a random sample from the Normal density with an unknown mean θ and variance 1. The joint p.d.f. of the vector X is:

$$f_X(x_1,\ldots,x_n \mid \theta) = (2\pi)^{-\frac{n}{2}} \, Exp\left[-\frac{1}{2}\left\{ \sum_{i=1}^{n} (x_i-\bar{x})^2 + n(\bar{x}-\theta)^2 \right\} \right] \tag{3.1}$$

and the likelihood function is:

$$l(\theta \mid x) \propto Exp\left[-\frac{n}{2}(\theta-\bar{x})^2 \right]. \tag{3.2}$$

That is, $l(\theta \mid x)$ is proportional to a Normal p.d.f. with mean $\bar{x}$ and variance $1/n$.

Example 3.3: Assume $X = (X_1,\ldots,X_n)$ is a random sample from the binomial density with parameter θ. The joint p.d.f. of the vector X is:

$$f_X(x_1,\ldots,x_n \mid \theta) = \theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}} \tag{3.3}$$

and the likelihood function for θ is proportional to a $Beta(n\bar{x}+1, n-n\bar{x}+1)$ distribution.

Recall: The density of a random variable from a $Beta(a, b)$ distribution is:

$$f_Y(y) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \, y^{a-1} \, (1-y)^{b-1} \, I_{(0,\,1)}(y) \;,\; a > 0,\; b > 0.$$

Methods of summarizing the likelihood function.

Remark 3.4: Likelihood methods are essentially methods that summarize the likelihood function. Common likelihood types of estimates are:

Definition 3.5: The maximum likelihood estimate (MLE) $\hat{\theta} = (\hat{\theta}_1,...,\hat{\theta}_p)$ of a vector of parameter $\theta = (\theta_1,...,\theta_p)$ is a point in the parameter space where the likelihood function has its maximum value.

Definition 3.6: The mean likelihood estimate (MeLE) $\bar{\theta} = (\bar{\theta}_1,...,\bar{\theta}_p)$ of a vector of parameters $\theta = (\theta_1,...,\theta_p)$ has i^{th} element:

$$\bar{\theta}_i = \frac{\int \theta_i l(\theta \mid x) d\theta}{\int l(\theta \mid x) d\theta}. \tag{3.4}$$

Definition 3.7: The likelihood covariance matrix V_L has (i,j) element:

$$V_{L_{ij}} = \frac{\int \theta_i \theta_j l(\theta \mid x) d\theta}{\int l(\theta \mid x) d\theta} - \bar{\theta}_i \bar{\theta}_j. \tag{3.5}$$

Remark 3.8: Note that the likelihood covariance matrix depends on the data, and that it is not the covariance matrix of any random variable since θ is a fixed vector of parameters. V_L could be called the likelihood curvature matrix, but the term likelihood covariance matrix will be kept for reasons that will be given later in the thesis.

Remark 3.9: When the likelihood function is proportional to a multivariate normal distribution then the MeLE, the MLE and V_L are sufficient to describe the likelihood. Also the matrix of second derivatives of the log-likelihood, i.e. the observed Fisher information matrix with $(i,j)^{th}$ element,

$$I_{i,j}(x) = -\frac{\partial^2 \log l(\theta \mid x)}{\partial \theta_i \partial \theta_j}\bigg|_{\theta = \hat{\theta}}, \tag{3.6}$$

is the inverse of the covariance matrix associated with this multivariate Normal distribution.

Remark 3.10: When the likelihood function is not proportional to a multivariate Normal distribution, an approach, often utilized, is to find a transformation of the parameters, such that the likelihood function for the transformed parameters is proportional to a multivariate Normal distribution.

Analysis of the likelihood method.

I now study the sampling distribution of the MeLE.

Proposition 3.11: Let $\bar{\theta} = \bar{\theta}(X)$ be the MeLE of θ. If the MeLE has a finite variance and the parameter space is in a bounded interval then, either $Var(\bar{\theta}) = 0$ or the MeLE is a biased estimator.

Proof: Embedding the estimation problem into a Bayesian model by assuming Θ uniformly distributed over a bounded interval, it follows by proposition 2.13 that the MeLE is the Bayes estimator because

$$\pi(\theta|x) = \frac{l(\theta|x)}{\int l(\theta|x)d\theta}$$

is the posterior distribution. The result now follows from proposition 2.18.

Example 3.12: In example 3.2 the MeLE of θ is $\bar{X}$. This is an unbiased estimator that has a positive variance for every n. However, the conditions of proposition 3.11 are not satisfied because $-\infty < \theta < \infty$ is not a bounded interval.

Proposition 3.13: Suppose that the parameter space is bounded, then the MeLE has the smallest average mean square error, where the average is over all values of θ.

Proof: Consider the corresponding Bayesian model by assuming a uniform prior, $\pi(\theta) \propto 1$, and the square error loss. Let Φ be the parameter space and $\mathbb{R}^n$ be the sample space. Then the MeLE is the Bayes estimator and by proposition 2.17 minimizes the Bayes risk:

$$r(\pi,\delta) = \int_{\Phi}\int_{\mathbb{R}^n} (\bar{\theta}(x) - \theta)^2 f(x|\theta)\, dx\, d\theta,$$

which is the average mean square error.

Proposition 3.14: If a sufficient statistic S exists, then the MeLE is a function of it.

Proof: The p.d.f of the observation can be written as

$$f(x|\theta) = h(x)g(S(x),\theta).$$

It follows by the factorization theorem that:

$$\bar{\theta}(x) = \frac{\int_{\Phi}\theta f(x|\theta)\,d\theta}{\int_{\Phi} f(x|\theta)\,d\theta} = \frac{\int_{\Phi}\theta\, g(S(x),\theta)\,d\theta}{\int_{\Phi} g(S(x),\theta)\,d\theta} = \bar{\theta}(S(x)).$$

Remark 3.15: The next proposition links together the MeLE, the MLE, the likelihood covariance matrix, and the observed and expected Fisher information matrices.

Proposition 3.16: *A central limit theorem for the likelihood.* Suppose $X_1,\ldots,X_n$ are i.i.d. from the density $f(x_i|\theta)$, θ being an unknown vector of parameters. Suppose $f(x_i|\theta)$ is twice

differentiable near $\hat{\theta}$, the (assumed to exist) maximum likelihood estimate of θ. Then for large n and under commonly satisfied assumptions, the normalized likelihood function :

$$l_n(\theta|x) = \frac{f(x|\theta)}{\int_\Theta f(x|\theta)\,d\theta} = \frac{\prod_{i=1}^{n} f(x_i|\theta)}{\int_\Theta \prod_{i=1}^{n} f(x_i|\theta)\,d\theta}$$

can be approximated as follows:

i) $l_n(\theta|x)$ is approximately $N_p(\bar{\theta}, V_L)$ where $\bar{\theta}$ is the mean likelihood estimate and V_L is the likelihood covariance matrix.

ii) $l_n(\theta|x)$ is approximately $N_p(\hat{\theta}, I^{-1}(x))$ where $\hat{\theta}$ is the maximum likelihood estimate and $I(x)$ is the observed Fisher information matrix, having (i,j) element:

$$I_{ij}(x) = -\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log[f(x|\theta)]\Big|_{\theta=\hat{\theta}}.$$

iii) $l_n(\theta|x)$ is approximately $N_p(\hat{\theta}, I^{-1}(\hat{\theta}))$ where $\hat{\theta}$ is the maximum likelihood estimate and $I(\hat{\theta})$ the expected Fisher information matrix, having (i,j) element:

$$I_{ij}(\hat{\theta}) = -n\mathcal{E}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log[f(X_1|\theta)]\Big|_{\theta=\hat{\theta}}\right]$$

where the expectation is with respect to the distribution of X_1.

Proof: This follows from proposition 2.19 by assuming a uniform prior distribution for θ, i.e.: $\pi(\theta) \propto 1$.

Remark 3.17: Important implications of this proposition are that the MeLE and the MLE are asymptotically equivalent, and that the likelihood covariance matrix is asymptotically equivalent to the inverse of the observed and expected Fisher information matrices.

Example 3.18: (Example 3.2 contd.): $\pi(\theta|x)$ is $N(\bar{x}, 1/n)$ for every n. Therefore: $\bar{\theta} = \hat{\theta} = \bar{x}$. The information matrices are $V_L = I^1(x) = I^1(\hat{\theta}) = 1/n$. Consequently i), ii) and iii) of proposition 3.16 give the same asymptotic distribution.

Example 3.19: (Example 3.3 contd.). If the prior distribution for Θ is $U[0,1]$ then the posterior distribution of Θ, $\pi(\theta|x)$, is $Beta(n\bar{x}+1, n-n\bar{x}+1)$. It follows that the MLE and the MeLE are:

$$\hat{\theta} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

$$\bar{\theta} = \frac{n\bar{x} + 1}{n + 2},$$

with:

$$\mathcal{E}(\hat{\theta}) = \theta,$$

$$\mathcal{E}(\bar{\theta}) = \frac{n\theta + 1}{n + 2},$$

$$V(\hat{\theta}) = \frac{\theta(1-\theta)}{n},$$

$$V(\bar{\theta}) = \frac{n\theta(1-\theta)}{(n+2)^2}.$$

The mean square error of the MLE and MeLE are:

$$MSE(\hat{\theta}) = \frac{\theta(1-\theta)}{n},$$

$$MSE(\bar{\theta}) = \frac{n\theta(1-\theta) + (1-2\theta)^2}{(n+2)^2},$$

and Average MSE (AMSE):

$$AMSE(\hat{\theta}) = \frac{1}{6n},$$

$$AMSE(\bar{\theta}) = \frac{1}{12 + 6n}.$$

The likelihood variance is:

$$V_L(\Theta|x) = \frac{(n\bar{x} + 1)(n - n\bar{x} + 1)}{(n + 3)(n + 2)^2}$$

$$\approx \frac{\bar{x}(1 - \bar{x})}{n}.$$

The inverse of the Fisher information is:

$$I^{-1}(\bar{x}) = \frac{\bar{x}(1-\bar{x})}{n}.$$

From the above discussions, it follows that i) the MeLE is not unbiased and has a non-zero variance (proposition 3.11), ii) AMSE($\bar{\theta}$) < AMSE($\hat{\theta}$) for all sample size n (proposition 3.13), iii) the MeLE depends on $\bar{X}$, which is a sufficient statistic (proposition 3.14) and, iv) the asymptotic distributions of the MeLE and MLE are the same (proposition 3.16). Also, the MeLE has a smaller MSE than the MLE when

$$\frac{1}{2} - \frac{1}{2}\sqrt{\frac{1+n}{1+2n}} < \theta < \frac{1}{2} + \frac{1}{2}\sqrt{\frac{1+n}{1+2n}}.$$

The interval is $[0.1127, 0.8873)$ when $n = 2$ and is $[0.1464, 0.8536)$ when n increases to ∞.

Example 3.20: *An earlier application of likelihood methods to time series analysis.* Barnard, Jenkins and Winsten (1962) have discussed likelihood inference for the first order autoregressive model. Copas (1966) has calculated the mean square errors of the MeLE and other estimators of the autoregressive parameter that were available then. He conducted a small simulation experiment with ***100*** replicates and sample size ***10*** and ***20***. The experiment had been previously proposed by Barnard et al. (1962) in his reply to the discussion of their original paper. Copas showed that the MeLE gives the smallest average mean square error. He points out that:

> "on theoretical grounds it is not surprising, since we have already noted that b_1 is a Bayes estimate with respect to a uniform prior weighting in $(-1,1)$."

Here b_1 refers to the MeLE. Moreover, his simulation shows that the MeLE gives the least MSE for the approximate range $(-0.3,0.6)$.

4. Conclusions.

In this chapter I have investigated the sampling distribution theory of mean likelihood estimation by embedding the estimation problem into a Bayesian model. I have derived various propositions on mean likelihood estimation, and I have given various examples to illustrate the applications and benefits of mean likelihood estimation.

5. References.

Barnard,G.O. (1949), "Statistical Inference", **JRSS-B 11**, p. 116.

Barnard,G.O., Jenkins,G.M. and Winsten,G.B. (1962), "Likelihood Inference in Time Series", **JRSS-A 125**, p. 321-372.

Berger,J.O. (1985), **Statistical Decision Theory and Bayesian Analysis**, Second edition, Springer-Verlag, New-York.

Copas,J.B. (1966), "Monte Carlo Results for Estimation in Stable Markov Time Series", **JRSS-A 129**, p. 110-116.

Jeffreys,H. (1961), **Theory of Probability**, Third edition, Oxford University Press, Oxford,

Jenkins,G.M. and Watts,D.G. (1968), **Spectral Analysis and its Applications**, Holden Day, San Francisco.

Mood,A.M., Graybill,F.A. and Boes,D.C. (1974), **Introduction to the Theory of Statistics**, Third edition, McGraw-Hill, New-York.

Walker, A.M. (1969), "On the Asymptotic Behaviour of Posterior Distributions". **JRSS-B 31**, p. 80-88.

CHAPTER 2

MEAN LIKELIHOOD ESTIMATION OF THE MA(1) PROCESS

In this chapter I study the mean likelihood estimator (MeLE) and the maximum likelihood estimator (MLE) of the parameter θ of a moving average model of order 1, MA(1). I first show that the likelihood function of a MA(1) series has always a local maximum or minimum at the boundary $\theta = \pm 1$ of the parameter space, and consequently the MLE has a positive probability to be on the boundary. On the other hand, I show that the MeLE is never on the boundary. I calculate the theoretical mean square error (MSE) of the MLE and MeLE when the sample sizes are $n=1$ and $n=2$, and show that the MSE of the MeLE is uniformly smaller than that of the MLE. For larger sample size, I conduct a simulation study to calculate the MSE of the MeLE and the MLE, and show that the MSE of the MeLE is smaller than that of the MLE when the true value of the moving average parameter is sufficiently inside the parameter space (the interval $[-0.8, 0.8]$, for example). Finally, I show that there are other time series models, with a MA(1) component, that have a likelihood function with a local maximum or minimum at the boundary $\theta = \pm 1$ of the parameter space.

Keywords: Autoregressive integrated moving average model; maximum likelihood estimation; mean likelihood estimation.

Foreword: The tables and figures are at the end of the current chapter. A reference to an equation is in the current chapter unless otherwise specified.

1. Introduction.

Cryer and Ledolter (1981) found that the Gaussian maximum likelihood estimator (MLE) of the parameter θ in a moving average process of order 1, MA(1), can take the values ± 1 with a positive probability, whatever is the true value of θ in the open interval $(-1,1)$. Their research was motivated by an earlier study of Ansley and Newbold (1980) who conducted a simulation to investigate the small sample properties of various estimators in autoregressive moving average (ARMA) time series models. They report a concentration of the MLE to take its value on the boundary ± 1 whenever the true value of θ is close to it. Anderson and Takemura (1986) give the limiting probabilities that the likelihood function attains a local maximum at ± 1 whatever is the value of θ in $[-1,1]$. They also explain why for a general ARMA process there is a positive probability that the MLE of the MA parameters fall on the boundary of the MA parameter space.

In this chapter, I propose to estimate the parameter θ of a MA(1) model by mean likelihood estimation (MeLE) to avoid the estimate of θ to be on the boundary. Barnard, Jenkins and Winsten (1962) have already suggested MeLE for the autoregressive model of order 1, AR(1), and Copas (1966) has compared the MeLE of the AR(1) parameter by simulation techniques. Copas (1966) showed that the MeLE gives the smallest average mean square (ref: chapter 1, example 3.20).

In chapter 1, I have derived the sampling distribution properties of the MeLE. It follows that the MeLE of the parameters of an ARMA model are biased (chapter 1, proposition 3.11), and have the smallest average mean square error, where the average is with respect to a uniform prior

weighting of the parameter space (chapter 1, proposition 3.13). Also, the MeLE has the same asymptotic distribution as the MLE (chapter 1, remark 3.17).

In this chapter I investigate in more detail mean likelihood estimation of the parameter θ of a MA(1) process. I organize this chapter as follow. First, I give a simpler proof than that of Cryer and Ledolter (1981) to show that the likelihood function of a MA(1) process has a local maximum or minimum at the boundary $\theta = \pm 1$ of the parameter space. Next, I show that the MeLE of θ is always in the open interval $(-1,1)$, i.e. it is never on the boundary. I calculate the theoretical moments of the MeLE and MLE when the sample sizes are $n=1$ and $n=2$, and show that the MeLE of θ has a smaller mean square error (MSE) than the MLE for all values of θ in the interval $[-1,1]$. For larger sample sizes ($n = 2, 5, 10, 25, 50, 100$), I conduct a simulation study to calculate the MSE of the MeLE and the MLE, and show that the MeLE of θ has a smaller MSE than the MLE when the true value of the parameter θ is in the interval $[-0.8, 0.8]$. The simulation also shows that the MSE of the MeLE and the MLE become close to one another when the sample size increases. This is because the two estimators have the same asymptotic distribution. I conclude this chapter by showing that the likelihood functions of the ARMA(1,1) and of the seasonal ARMA$(0,1)(0,1)_s$ models have a local maximum or minimum at the boundaries of the MA parameter space.

2. The moving average process of order 1.

Definition 2.1: The process $\{z_t, t = 0, \pm 1, \pm 2, \ldots\}$ is an autoregressive moving average process of order (p,q), ARMA(p,q), if it satisfies $\phi(B)z_t = \theta(B)a_t$, $\phi(B) = 1-\phi_1 B-\ldots-\phi_p B^p$ and $\theta(B) = 1-\theta_1 B-\ldots-\theta_q B^q$. B is the backshift operator $(B^i z_t = z_{t-i})$. The sequence $\{a_t, t = 0, \pm 1, \pm 2, \ldots\}$ is independently and Normally distributed with mean zero and variance σ^2.

Remark 2.2: The roots of $\phi(B)$ and $\theta(B)$ must lie outside the unit circle to ensure the stationarity and the invertibility of the process. These conditions define restrictions on the parameter values that are admissible for a particular ARMA process.

Definition 2.3: The MA(1) process is an ARMA(0,1) process. It satisfies $z_t = a_t - \theta a_{t-1}$.

Proposition 2.4: The MA(1) process has the following properties:

i) The variance of the process is:

$$\gamma_0(\theta) = (1+\theta^2)\sigma^2.$$

ii) The autocovariances are:

$$\gamma_1(\theta) = -\theta\sigma^2$$

$$\gamma_k(\theta) = 0 \quad k \geq 2.$$

iii) The covariance matrix $\sigma^2 \Omega(\theta)$ of $Z = (Z_1,\ldots,Z_n)$ from a MA(1) process is tridiagonal with:

$$\omega_{i,i}(\theta) = 1+\theta^2$$
$$\omega_{i,i+1}(\theta) = \omega_{i-1,i}(\theta) = -\theta,$$

where $\omega_{i,j}(\theta)$ is the $(i,j)^{th}$ element of $\Omega(\theta)$.

iv) The determinant of the covariance matrix is:

$$|\Omega(\theta)| = \frac{1-\theta^{2(n+1)}}{1-\theta^2}.$$

v) The likelihood is:

$$L(\theta,\sigma^2|z) = (2\pi\sigma^2)^{-\frac{n}{2}}\left\{\frac{1-\theta^2}{1-\theta^{2(n+1)}}\right\}^{\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}z'\Omega(\theta)^{-1}z\right\}.$$

vi) Given θ, the maximum likelihood estimate of σ^2 is:

$$\hat{\sigma}^2 = \frac{z'\Omega(\theta)^{-1}z}{n}.$$

vii) The concentrated likelihood function is:

$$L(\theta|z) \propto |\Omega(\theta)|^{-\frac{1}{2}}\{z'\Omega(\theta)^{-1}z\}^{-\frac{n}{2}}.$$

Remark 2.5: The concentrated likelihood function for θ is also obtained if the likelihood function is integrated with respect to $(1/\sigma)\, d\sigma$. The Bayesian motivation is to assume a uniform prior, $\pi(\theta) \propto 1$, for the parameter θ, and Jeffrey's non-informative prior, $\pi(\sigma) \propto 1/\sigma$, for the scale parameter. Box-Jenkins (1976 p. 252) have called this the exact posterior distribution. The likelihood motivation is to transform the scale parameter σ to $\eta = \ln \sigma$ with jacobian $1 / \sigma$. This transformation of the scale parameter ensures that, when plotted against η, the likelihood function "looks" more like a Normal distribution. The integration of the likelihood function with respect to $1/\sigma$ has been suggested as a natural way to get rid of σ by Barnard et al. (1962)

Proposition 2.6: For all $\theta \in (-1, 1)$, the concentrated likelihood function of the MA(1) process has the following property:

$$L(\theta \,|\, z) = L(\theta^{-1} \,|\, z).$$

Remark 2.7: This result means that the concentrated likelihood function has either a local minimum, a local maximum, or a point of inflexion at $\theta = \pm 1$ (see remark 3.4 for the case $n = 2$). When $n = 1$, the concentrated likelihood is constant (Anderson and Mentz, 1980) since:

$$\Omega(\theta) = 1 + \theta^2 ,$$

$$\hat{\sigma}^2 = \frac{z_1^2}{1 + \theta^2} ,$$

$$L(\theta \,|\, z_1) = \frac{1}{z_1} .$$

Cryer and Ledolter (1981) have derived the result in proposition 2.6 by evaluating the determinant and the inverse of the covariance matrix. Here I give a simpler proof.

Proof of proposition 2.6: First note that:

$$\Omega(\theta^{-1}) = \theta^{-2}\Omega(\theta),$$

and

$$|\Omega(\theta^{-1})| = \theta^{-2n}|\Omega(\theta)|.$$

Consequently:

$$\begin{aligned} L(\theta^{-1}|z) &= |\Omega(\theta^{-1})|^{-\frac{1}{2}} \{z'\Omega(\theta^{-1})z\}^{-\frac{n}{2}} \\ &= \{\theta^{-2n}|\Omega(\theta)|\}^{-\frac{1}{2}} \{\theta^2 z'\Omega(\theta)z\}^{-\frac{n}{2}} \\ &= L(\theta|z). \end{aligned}$$

Remark 2.8: This feature of the likelihood function causes the MLE to have a positive probability to be on the boundary ± 1 whatever is the true value of θ (Cryer and Ledolter, 1981, and Anderson and Takemura, 1986). Consider the following estimator to avoid the boundary problem associated with the MLE:

Definition 2.9: The mean likelihood estimator (MeLE) $\bar{\theta}(z)$ of θ and its likelihood variance V_L are (ref.: chapter 1, definitions 3.6 and 3.7):

$$\bar{\theta}(z) = \frac{\int_{-1}^{1} \theta L(\theta|z) d\theta}{\int_{-1}^{1} L(\theta|z) d\theta},$$

$$V_L = \frac{\int_{-1}^{1} \theta^2 L(\theta|z) d\theta}{\int_{-1}^{1} L(\theta|z) d\theta} - \bar{\theta}(z)^2 .$$

Proposition 2.10: The MeLE $\bar{\theta}(z)$ of the parameter θ of and MA(1) process is a function from the sample space into the open interval $(-1,1)$.

Proof: It suffices to show that $|\bar{\theta}| < 1$. The function $L(\theta|z)$ is non-negative and so:

$$|\bar{\theta}| < 1 \quad \leftrightarrow \quad \left| \int_{-1}^{1} \theta L(\theta|z) d\theta \right| < \int_{-1}^{1} L(\theta|z) d\theta.$$

Also

$$\left| \int_{-1}^{1} \theta L(\theta|z) d\theta \right| \leq \int_{-1}^{1} |\theta| L(\theta|z) d\theta.$$

Assume there exist an open interval $O = (a,b)$ in $[-1,1]$ where $L(\theta|z) > 0$. Select $0 < \delta < b-a$, and let $A = (a+\delta, b-\delta)$ and A^c be the complement of A in $[-1, 1]$. Then:

$$\int_{-1}^{1} |\theta| L(\theta|z) d\theta = \int_{A} |\theta| L(\theta|z) d\theta + \int_{A^c} |\theta| L(\theta|z) d\theta$$

$$< \int_{A} L(\theta|z) d\theta + \int_{A^c} L(\theta|z) d\theta = \int_{-1}^{1} L(\theta|z) d\theta,$$

because

$$\int_{A} |\theta| L(\theta|z) d\theta < \int_{A} L(\theta|z) d\theta,$$

and

$$\int_{A^c} |\theta| L(\theta|z) d\theta \leq \int_{A^c} L(\theta|z) d\theta.$$

If $L(\theta|z) = 0$ almost everywhere in $[-1, 1]$, then its average value is zero and define $\bar{\theta}(z) = 0$. Unless $z = 0$, this situation is impossible because the function is continuous since (Brockwell and Davis 1987, example 5.2.1 and equation 8.6.8):

$$L(\theta|z) = \left[\frac{1-\theta^{2(n+1)}}{1-\theta^2}\right]^{-\frac{1}{2}} \left[\sum_{t=1}^{n} \frac{(z_t - \hat{z}_t)^2}{r_{t-1}}\right]^{-\frac{n}{2}}$$

$$\hat{z}_t = \theta \frac{(z_{t-1} - \hat{z}_{t-1})}{r_{t-2}}$$

$$\hat{z}_1 = 0$$

$$r_0 = 1 + \theta^2$$

$$r_{t+1} = 1 + \theta^2 - \frac{\theta^2}{r_t}.$$

Corollary 2.11: The probability that the MeLE of θ gives a value equal to ± 1 is zero.

Example 2.12: *The MeLE and MLE of the parameter θ for a MA(1) process with sample size $n = 1$.* It is shown in remark 2.7 that the concentrated likelihood is constant, in which case the MLE has a uniform distribution in the interval $[-1, 1]$, and the MeLE takes the value 0 with probability 1. Consequently:

$$\mathcal{E}(\hat{\theta}) = \mathcal{E}(\bar{\theta}) = 0,$$

$$Var(\hat{\theta}) = \frac{1}{3},$$

$$Var(\bar{\theta}) = 0,$$

$$MSE(\hat{\theta}) = \theta^2 + \frac{1}{3},$$

$$MSE(\bar{\theta}) = \theta^2.$$

It follows that the MSE of the MeLE is uniformly smaller than that of the MLE for all the values of θ in the interval $[-1, 1]$.

3. The moments of the MLE and MeLE of the parameter θ for a MA(1) process with sample size $n = 2$.

Proposition 3.1: i) The likelihood of the MA(1) process with sample size $n = 2$ is:

$$L(\theta,\sigma^2|z) \propto \sigma^2(1+\theta^2+\theta^4)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}\,\frac{(z_1^2+z_2^2)(1+\theta^2)+2\theta z_1 z_2}{1+\theta^2+\theta^4}\right\}, \tag{3.1}$$

where $z = (z_1, z_2)$.

ii) The concentrated likelihood function for θ is:

$$L(\theta|z) \propto \frac{\sqrt{1+\theta^2+\theta^4}}{(1+\theta^2)(z_1^2+z_2^2)+2\theta z_1 z_2}.$$

iii) If $w = -z_1 z_2 / (z_1^2 + z_2^2)$ then the concentrated likelihood for θ is:

$$L(\theta|w) \propto \frac{\sqrt{1+\theta^2+\theta^4}}{1+\theta^2-2\theta w}. \tag{3.2}$$

Remark 3.2: It has to be noticed that $|w| \leq 0.5$. This follows from $(z_1-z_2)^2 \geq 0$ for every z_1 and z_2. Also, w is the negative value of the estimated sample autocorrelation at lag one with the mean assumed to be equal to zero.

Proposition 3.3: Cryer and Ledolter(1981) have derived the MLE of θ for a MA(1) process with sample size $n = 2$. Let $W = -Z_1 Z_2 / (Z_1^2 + Z_2^2)$, then the MLE is:

$$\hat{\theta}(W) = \begin{cases} -1 & \text{if } -0.5 \le W \le -0.25 \\ \dfrac{1-\sqrt{1-16W^2}}{4W} & \text{if } -0.25 \le W \le 0.25 \\ 1 & \text{if } 0.25 \le W \le 0.5 \ . \end{cases} \tag{3.3}$$

Proof: The local minima/maxima of the concentrated likelihood (3.2) are provided by the roots of the equation:

$$\frac{\partial L(\theta|w)}{\partial \theta} \propto 2w\theta^4 - \theta^3 + \theta - 2w = 0,$$

which are ± 1 and $(1 \pm (1 - 16w^2)^{1/2}) / 4w$.

Remark 3.4: Figure 1 (located at the end of the current chapter) shows the ensemble of concentrated likelihood functions (3.2). For a given w, the MLE is the maximum of the curve along the θ-axis. The maximum occurs at the value provided by equation (3.3). When $w \le -0.25$, the concentrated likelihood has a maximum at $\theta = -1$ and a minimum at $\theta = 1$. When $w \ge 0.25$, the concentrated likelihood has a minimum at $\theta = -1$ and a maximum at $\theta = 1$. When $-0.25 \le w \le 0.25$, the concentrated likelihood has local minima at $\theta = \pm 1$. Given w, the mean likelihood estimate is the weighted mean value of θ, where the weighting is proportional to the likelihood. Figure 2 shows a graphic of the MLE and MeLE for every possible value of

the statistic W. It has to be noticed that the MLE is stretched to the boundaries ± 1 whereas the MeLE is pulled toward 0.

Remark 3.5: Figure 1 also shows the effect of the likelihood function on the initial prior, implicitly assumed uniform on $(-1,1)$. Before data collection, the mean likelihood estimate of θ is 0. After data collection, the statistic $W = -Z_1 Z_2 / (Z_1^2 + Z_2^2)$ is calculated and the posterior distribution is obtained. If w is negative the density is shifted in the direction of $\theta = -1$. On the other hand, if w is positive the density is shifted in the direction of $\theta = 1$. Asymptotically, the likelihood will concentrate around the true value of θ (ref.: chapter 1, proposition 3.16).

Remark 3.6: I will now derive the mean square error of the estimators as functions of θ. In Bayesian vocabulary, the mean square error corresponds to the risk with the square error loss function. The average mean square error corresponds to the Bayes risk. It is shown in chapter 1, proposition 3.13 that the MeLE minimizes the average mean square error.

I first derive the distribution of W because the MeLE and MLE are function of it.

Proposition 3.7: Let θ be the true value of the MA(1) parameter, then the probability density function of W is

$$f_W(w) = \frac{2\sqrt{1+\theta^2+\theta^4}}{\pi\sqrt{1-4w^2}\,(1+\theta^2-2\theta w)}, \quad |w| \le 0.5 \tag{3.4}$$

and

$$\int f_W(w)dw = \frac{2}{\pi} Arctan\left[\frac{(1+\theta^2)\ Tan\left[\frac{1}{2}\ Arcsin(2w)\right] - \theta}{\sqrt{1+\theta^2+\theta^4}}\right]. \tag{3.5}$$

Proof: The joint density of Z_1, Z_2 is proportional to the right hand side of equation (3.1). Consider the transformation:

$$W = \frac{-Z_1 Z_2}{Z_1^2 + Z_2^2},$$

$$V = Z_1^2 + Z_2^2.$$

Then we can set

$$Z_1 = \frac{1}{2}\sqrt{V}\ \{\sqrt{1+2W} + \sqrt{1-2W}\},$$

$$Z_2 = \frac{1}{2}\sqrt{V}\ \{\sqrt{1-2W} - \sqrt{1+2W}\},$$

with Jacobian:

$$|J| = \frac{1}{2}\sqrt{1-4W^2}.$$

Integration with respect to V from 0 to ∞ gives the desired result. Differenting both sides of (3.5) gives (3.4). Note that (Z_1, Z_2) and $(-Z_1, -Z_2)$ are mapped to the same values of (W, V), and similarly for $(-Z_1, Z_2)$ and $(Z_1, -Z_2)$.

Equation (3.5) allows us to calculate the distribution of $\hat{\theta}$.

Proposition 3.8: Let θ be the true value of the MA(1) parameter, then the probability density function (p.d.f.) of $\hat{\theta}$ is a mixture of a discrete distribution on the two points ± 1 and a continuous distribution on $(-1,1)$:

$$f_{\hat{\theta}}(x \mid \theta) = \frac{\sqrt{1+\theta^2+\theta^4}\,(1-x^2)}{\pi\sqrt{1+x^2+x^4}\,[\,(1+\theta^2)(1+x^2)-\theta x\,]}.$$

Proof: From equation (3.3), the probability to obtain an estimate equal to -1 is the integral (3.5) evaluated from -.25 to -.5. Similarly, the probability to obtain an estimate equal to +1 is the integral (3.5) evaluated from +.25 to +.5. These two probabilities are in columns 2 and 3 of Table 1. These probabilities are the same as those from Table 1 of Cryer and Ledolter (1981). In the interval (-.25,.25) the transformation (3.3) from W to $\hat{\theta}$ is one to one and the inverse transformation is:

$$W = \frac{\hat{\theta}}{2(\hat{\theta}^2+1)}. \tag{3.6}$$

Equation (3.6) can be used with (3.5) to calculate the cumulative distribution of $\hat{\theta}$. That is for $-1 < x < 1$:

$$\begin{aligned} P\{\hat{\theta} \le x\} &= P\{\hat{\theta} = -1\} + P\{-1 < \hat{\theta} \le x\} \\ &= P\{\hat{\theta} = -1\} + P\left\{-\frac{1}{4} < W \le \frac{x}{2(x^2+1)}\right\}. \end{aligned}$$

Remark 3.9: The distribution of $\hat{\theta}$ is now completely specified. Table 1 gives the mean, variance, bias and mean square error of $\hat{\theta}$ for various values of θ. The moments from the continuous distribution of $\hat{\theta}$ are calculated with Simpson's rule of numerical integration.

Remark 3.10: Table 2 gives the moments of the mean likelihood estimator $\bar{\theta}$. The moments are calculated from the distribution of W with $\bar{\theta}$ being a function of it. For example the mean is:

$$\mathcal{E}_{\theta}(\bar{\theta}) = \int_{-0.5}^{0.5} \frac{\int_{-1}^{1} \alpha L(\alpha|w)d\alpha}{\int_{-1}^{1} L(\alpha|w)d\alpha} f(w|\theta)dw.$$

It should be noticed that the covered probability area in Table 2 is around .998 and does not exactly equal 1. Equation (3.4) shows that the density of W goes to ∞ at the end points $\pm.5$, and therefore the outer integral is calculated from -.499998 to .499998.

Remark 3.11: A comparison of Table 1 and Table 2 shows that the MLE is uniformly less biased than the mean likelihood estimator but its MSE is uniformly larger. This is because the mean likelihood estimator has a smaller variance due to a smaller range of possible values (see Figure 2). On the other hand, there are always positive probabilities that the MLE take the values ± 1, which contribute greatly to its variance. The results from Table 1 and Table 2 demonstrate the following proposition.

Proposition 3.12: The MSE of the MeLE is uniformly smaller than that of the MLE for the parameter θ of a MA(1) process with sample size $n = 2$.

4. The moments of the MLE and MeLE of the parameter θ for a MA(1) process with sample size $n = 2, 5, 10, 25, 50$ and 100.

In this section I calculate the moments of the MeLE and MLE by a simulation study. The estimators are for $\theta = 0, \pm 0.2, \pm 0.4, \pm 0.6, \pm 0.8, \pm 0.95, \pm 1$, and for samples sizes $n = 2, 5, 10, 25, 50$ and 100. The number of replicates is $R = 1000$.

Simulation of a MA(1) process. A sample of size $n+1$ from an $U[0,1]$ distribution is generated with the Fortran function RSUPER (McLeod, 1982). This sample from an uniform distribution is transformed to a sample form a $N(0,1)$ distribution with the Box-Muller method, programmed in the subroutine RNNORM (McLeod, 1982). A sample $z=(z_1,\ldots,z_n)$ of size n from a MA(1) process is obtained with the transformation: $z_t = a_t - \theta a_{t-1}$, where a_t is $N(0,1)$ and θ is fixed and known.

Likelihood calculation. The concentrated likelihood for θ involves the calculations of the determinant $|\Omega(\theta)|$ and of the unconditional sum of squares $S(\theta) = z' \Omega(\theta)^{-1} z$. The determinant is given in proposition 2.4 with $|\Omega(1)| = |\Omega(-1)| = n + 1$. The value of $S(\theta)$ is calculated with a modified version of the subroutine SARMAS (McLeod and Sales, 1983) where the determinant and the value of modified sum of squares are not calculated. The concentrated likelihood and its product with θ are calculated for θ from -1 to 1 with increments of size 0.04. This gives 51 evaluations of the likelihood.

Computation of the MeLE and MLE. The MeLE is calculated by Simpson's rule of numerical integration. The value of θ where the concentrated likelihood is maximum, among the 51

evaluations, gives an initial value for the maximum likelihood estimate (MLE). This value for the MLE is then used as the initial value for the maximization algorithm DALL (Ishiguro and Akaike, 1989) to obtain the final MLE. This gives 2 values of the MLE obtained with 2 different methods. Denote the mean likelihood estimate by $\bar{\theta}$, and two values of the MLE by $\hat{\theta}_i$ for the MLE obtained with the grid search among the 51 evaluations of the concentrated likelihood (i.e. the initial estimate), and by $\hat{\theta}$ for the final MLE obtained with DALL.

Summary statistics. The mean and the mean square error are calculated for the three estimators over the ***1000*** replicates. For $\hat{\theta}_i$ the number of estimates in the intervals $[0.96, 1]$ and $[-1, -0.96]$ give the estimates of $Prob\{\hat{\theta}_i = +1\}$ and $Prob\{\theta_i = -1\}$. For $\hat{\theta}$, it is the number of estimates in the intervals $[0.985, 1]$ and $[-1, -0.985]$. The statistics are in Tables 3A, 3B, 3C and 3D. Table 3A gives the expected values of the three estimators $\bar{\theta}$, $\hat{\theta}_i$ and $\hat{\theta}$. Table 3B gives the mean square errors (MSE) of the estimators. Table 3C gives the probabilities that the MLE's are on the boundaries. For the MeLE, these probabilities are always zero (corollary 2.11). Table 3D gives the average mean square errors (AMSE) of the three estimators by sample size.

Analysis of the simulation. The results of the simulation can be compared with the theoretical results when $n = 2$. The same conclusions hold. The MLE's are less biased than the MeLE, but their MSE's are uniformly larger. The results of the simulation for the MLE compare very well with Table 1. All the 95% confidence intervals (not provided) of the expected values of the MLE's from the simulation contain the theoretical expectation of the MLE, and similarly for the estimates of the probabilities of the MLE's to be on the boundaries, except for $Prob\{\hat{\theta}_i = -1 \mid \theta = -0.2\}$, where the lower bound is .395. The confidence intervals of the

expected value of the MeLE contain the theoritical expectation of the MeLE from Table 2 for all θ.

BIAS: Figures 3a to 3f give the bias of the estimators by sample size. The following should be noticed. The bias decreases as the sample size increases (note that the range of the y-axis is smaller as the sample size increases). The MeLE is negatively biased because it is the Bayes estimator with a uniform prior. The expected value of the mean likelihood estimator is therefore pulled toward 0, the expected value of the prior, but this effect disappears as the sample size increases. The MeLE has greater bias than the MLE's except for $(n = 2, \theta = 0)$. $(n = 10, \theta = 0)$, $(n = 25, \theta = 0.2)$, $(n = 50, \theta = -0.8, -0.6, -0.4, -0.2, 0)$, and $(n = 100, \theta = -0.4, -0.2, 0)$ where the bias of the MeLE is smaller than that of the MLE's.

MEAN SQUARE ERROR: Figures 4a to 4f give the MSE of the estimators by sample size. The MeLE has a smaller MSE when θ is sufficiently inside the parameter space, actually in the range $[-0.8, 0.8]$ for all sample sizes under study. Also, the MSE's are smaller as the sample size increases since the estimators are all consistent. As the sample size increases, the MSE of the estimators follow each another more closely and at $n = 100$ there are barely any differences. The MLE's have smaller MSE than the MeLE at the end points $\theta = \pm 0.95$ and $\theta = \pm 1$ under investigation when $n > 2$. Table 3D shows that the MeLE has the smallest average mean square error, but for $n = 100$ where the difference is insignificant, and the difference in AMSE reduces as the sample size increases. It is clear form figures 4 that the MeLE should be used when the sample sizes are small or medium ($n \leq 50$).

5. Other ARMA models with a local maximum or minimum at the boundary in their Gaussian likelihood.

Proposition 5.1: Let z be an observed vector of size n from an ARMA(p,q) model with parameters $\beta = (\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q)$. Let $\Omega(\beta)$ be the covariance matrix of Z, and $\gamma_0(\beta)$ be the variance of the ARMA(p,q) process assuming $\sigma^2 = 1$. Then the concentrated likelihood function is independent of the variance of the process, i.e.:

$$L(\beta|z) \propto |P(\beta)|^{-\frac{1}{2}} \{z'P(\beta)^{-1}z\}^{-\frac{n}{2}},$$

where

$$\Omega(\beta) = \gamma_0(\beta)P(\beta).$$

Proof: First note that:

$$|\Omega(\beta)| = \{\gamma_0(\beta)\}^n |P(\beta)|$$
$$\Omega(\beta)^{-1} = \{\gamma_0(\beta)\}^{-1} P(\beta)^{-1}.$$

It follows that:

$$L(\beta|z) \propto |\Omega(\beta)|^{-\frac{1}{2}} \{z'\Omega(\beta)^{-1}z\}^{-\frac{n}{2}}$$

$$= \gamma_0(\beta)^{-\frac{n}{2}}|P(\beta)|^{-\frac{1}{2}} \{\gamma_0(\beta)^{-1} z'P(\beta)^{-1}z\}^{-\frac{n}{2}}$$

$$= |P(\beta)|^{-\frac{1}{2}} \{z'P(\beta)^{-1}z\}^{-\frac{n}{2}}.$$

Remark 5.2: Proposition 2.6 shows that the concentrated likelihood function of a MA(1) series has a local maximum or minimum at the boundaries $\theta = \pm 1$. A simpler proof of the proposition is to use the invariance of the lag 1 autocorrelation coefficient $\rho_1(\theta)$ under the reciprocal transformation, i.e. $\rho_1(\theta^{-1}) = \rho_1(\theta)$. The invariance of the autocorrelation coefficients under the reciprocal transformation of the parameters θ and Θ is also true for other ARMA models.

Example 5.3: The seasonal MA(1)$_s$ model is $z_t = (1 - \Theta B^s)a_t$. The non-zero autocorrelation is $\rho_s(\Theta) = -\dfrac{\Theta}{1+\Theta^2}$ and is unchanged when the parameter Θ is replaced by its reciprocal Θ^{-1}.

Example 5.4: The seasonal ARMA(0,1)(0,1)$_s$ model is $z_t = (1-\theta B)(1-\Theta B^s)a_t$. The non-zero autocorrelations are:

$$\rho_1(\theta,\Theta) = -\frac{\theta}{1+\theta^2},$$

$$\rho_{s-1}(\theta,\Theta) = \frac{\theta}{1+\theta^2}\,\frac{\Theta}{1+\Theta^2},$$

$$\rho_s(\theta,\Theta) = -\frac{\Theta}{1+\Theta^2},$$

$$\rho_{s+1} = \rho_{s-1}.$$

The autocorrelations remain unchanged when a parameter is changed to its reciprocal.

Example 5.5: The ARMA(1,1) model is $(1-\phi B)z_t = (1-\theta B)a_t$. The autocorrelations are:

$$\rho_1(\phi,\theta) = \frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\phi\theta},$$

$$\rho_k(\phi,\theta) = \phi\rho_{k-1}(\phi,\theta).$$

The autocorrelations are unchanged when θ is replaced by θ^{-1}.

6. Conclusions.

In this chapter I have studied MeLE of the parameter of a MA(1) process. I first reviewed why the MLE of the MA(1) parameter can be on the boundary of the parameter space, and I have given a proof of it by using the invariance of the lag one autocorrelation under the reciprocal transformation. The same proof is used to show that others ARMA(p,1)(P,1) models have local maxima or minima at the boundaries of the MA parameter space.

I have proposed to estimate the MA(1) parameter by MeLE to solve the boundary problem associated with the MLE. I have shown that the MSE of the MeLE is uniformly smaller than that of the MLE when the sample sizes are $n = 1$ and $n = 2$. I have conducted a simulation study with sample size $n = 2, 5, 10, 25, 50, 100$ to show that the MeLE has a smaller MSE than the MLE when the parameter is in the range $[-0.8, 0.8]$ for all sample sizes under study.

7. References.

Anderson,T.W. and Mentz,R.P. (1980), "On the Structure of the Likelihood Function of Autoregressive and Moving Average Models", **Journal of Time Series Analysis 1**, 83-94.

Anderson,T.W. and Takemura, A. (198?), "Why Do Noninvertible Estimated Moving Averages Occur?", **Journal of Time Series Analysis 7**, 235-254.

Ansley,C.F. and Newbold,P. (1980), "Finite Sample Properties for Autoregressive Moving Average Models", **Journal of Econometrics 13**, 159-183.

Barnard,G.O., Jenkins,G.M. and Winsten,G.B. (1962), "Likelihood Inference in Time Series", **JRSS-A 125**, 321-372.

Box,G.P. and Jenkins,G.M. (1976), **Time Series Analysis forecasting and control**, Holden-Day, Oakland.

Copas,J.B. (1966), "Monte Carlo Results for Estimation in Stable Markov Time Series", **JRSS-A 129**, 110-116.

Cryer,J.D. and Ledolter,J. (1981), "Small-sample properties of the maximum likelihood estimator in the first-order moving average model", **Biometrika 68**, 691-694.

Ishiguro,M and Akaike,H. (1989), "DALL: Davidon's Algorithm for Log Likelihood Maximization - A fortran subroutine for statistical model builder". **The Institute of Statistical Mathematics**, Computer Science Monograph #25, Tokyo.

Mcleod,I.A. (1982), "RSUPER and RNNORM", Fortran subroutines, Department of Statistical and Actuarial Sciences, **The University of Western Ontario**, London, Ontario, Canada.

McLeod,I.A. and Sales, P.R.H. (1983), "An Algorithm for Approximate Likelihood Calculation of ARMA and Seasonal ARMA Models", **Applied Statistics 31**, 211-223.

Quenneville,B. (1993), "CLIKMA1, J03MA1, MAONE, MELEMA1, SIMP, UNSSQ", Fortran subroutines, Time Series Research and Analysis Centre, **Statistics Canada**, Ottawa, Ontario, Canada.

TABLE 1: MA(1) process; Moments of the maximum likelihood estimator of θ for sample size 2.

θ	$P\{\hat{\theta}=+1\}$	$P\{\hat{\theta}=-1\}$	$\mathcal{E}(\hat{\theta})$	$\mathcal{E}(\hat{\theta}^2)-\mathcal{E}(\hat{\theta})^2$	$\mathcal{E}(\hat{\theta})-\theta$	$\mathcal{E}(\hat{\theta}-\theta)^2$
			mean	variance	bias	MSE
0.0	.33333	.33333	.00000	.71622	.00000	.71622
0.1	.36137	.30665	.05841	.71400	-.04159	.71573
0.2	.38943	.28241	.11411	.70775	-.08589	.71513
0.3	.41607	.26135	.16472	.69855	-.13528	.71685
0.4	.43996	.24387	.20841	.68788	-.19159	.72459
0.5	.46011	.23005	.24411	.67728	-.25589	.74276
0.6	.47595	.21971	.27151	.66797	-.32849	.77588
0.7	.48743	.21250	.29100	.66072	-.40900	.82800
0.8	.49487	.20793	.30347	.65581	-.49653	.90235
0.9	.49883	.20554	.31008	.65313	-.58992	1.00113
1.0	.50000	.20483	.31202	.65232	-.68798	1.12564

MSE: mean square error.

Corresponding quantities for $-\theta$ are as follow:

$$P\{\hat{\theta}=+1|-\theta\} = P\{\hat{\theta}=-1|\theta\},$$

$$P\{\hat{\theta}=-1|-\theta\} = P\{\hat{\theta}=1|\theta\},$$

$$\mathcal{E}\{\hat{\theta}|-\theta\} = -\mathcal{E}\{\hat{\theta}|\theta\},$$

$$Var\{\hat{\theta}|-\theta\} = Var\{\hat{\theta}|\theta\},$$

$$Bias\{\hat{\theta}|-\theta\} = -Bias\{\hat{\theta}|\theta\},$$

$$MSE\{\hat{\theta}|-\theta\} = MSE\{\hat{\theta}|\theta\}.$$

θ	n	MLE +1	MLE_I +1	MLE -1	MLE_I -1
-1.00	50	0.000	0.000	0.662	0.783
-0.95	50	0.000	0.000	0.551	0.681
-0.80	50	0.000	0.000	0.141	0.193
-0.60	50	0.000	0.000	0.011	0.019
-0.40	50	0.000	0.000	0.003	0.003
-0.20	50	0.000	0.000	0.000	0.000
0.00	50	0.000	0.000	0.000	0.000
0.20	50	0.000	0.000	0.000	0.000
0.40	50	0.000	0.000	0.000	0.000
0.60	50	0.023	0.027	0.000	0.000
0.80	50	0.133	0.171	0.000	0.000
0.95	50	0.522	0.674	0.000	0.000
1.00	50	0.663	0.806	0.000	0.000
-1.00	100	0.000	0.000	0.733	0.919
-0.95	100	0.000	0.000	0.401	0.687
-0.80	100	0.000	0.000	0.024	0.048
-0.60	100	0.000	0.000	0.001	0.001
-0.40	100	0.000	0.000	0.000	0.000
-0.20	100	0.000	0.000	0.000	0.000
0.00	100	0.000	0.000	0.000	0.000
0.20	100	0.000	0.000	0.000	0.000
0.40	100	0.000	0.000	0.000	0.000
0.60	100	0.000	0.000	0.000	0.000
0.80	100	0.024	0.046	0.000	0.000
0.95	100	0.383	0.685	0.000	0.000
1.00	100	0.741	0.933	0.000	0.000

TABLE 3A: MA(1) process; Mean[1] of the MLE, the inital value for the MLE (MLE_I), and the MeLE of θ.

θ	n	MLE	MLE_I	MeLE
-1.00	2	-0.3073	-0.2984	-0.0583
-0.95	2	-0.3055	-0.2968	-0.0581
-0.80	2	-0.2967	-0.2882	-0.0563
-0.60	2	-0.2793	-0.2707	-0.0530
-0.40	2	-0.2103	-0.2041	-0.0416
-0.20	2	-0.1392	-0.1343	-0.0248
0.00	2	-0.0226	-0.0218	-0.0025
0.20	2	0.1065	0.1040	0.0212
0.40	2	0.2094	0.2030	0.0383
0.60	2	0.2731	0.2647	0.0502
0.80	2	0.3021	0.2929	0.0563
0.95	2	0.3179	0.3086	0.0588
1.00	2	0.3218	0.3122	0.0593
-1.00	5	-0.6238	-0.6066	-0.3049
-0.95	5	-0.6195	-0.6015	-0.3041
-0.80	5	-0.5960	-0.5789	-0.2938
-0.60	5	-0.5262	-0.5124	-0.2575
-0.40	5	-0.3830	-0.3727	-0.1926
-0.20	5	-0.2156	-0.2104	-0.1059
0.00	5	-0.0073	-0.0072	-0.0080
0.20	5	0.1865	0.1822	0.0904
0.40	5	0.3575	0.3479	0.1787
0.60	5	0.5106	0.4964	0.2478
0.80	5	0.6174	0.5991	0.2904
0.95	5	0.6531	0.6324	0.3034
1.00	5	0.6583	0.6378	0.3045
-1.00	10	-0.8224	-0.8061	-0.5552
-0.95	10	-0.8188	-0.8030	-0.5545
-0.80	10	-0.7624	-0.7497	-0.5280
-0.60	10	-0.6193	-0.6098	-0.4426
-0.40	10	-0.4364	-0.4306	-0.3165
-0.20	10	-0.2273	-0.2247	-0.1668
0.00	10	-0.0088	-0.0081	-0.0076
0.20	10	0.2195	0.2172	0.1512
0.40	10	0.4216	0.4164	0.3012
0.60	10	0.6061	0.5979	0.4303
0.80	10	0.7567	0.7434	0.5199
0.95	10	0.8076	0.7927	0.5492
1.00	10	0.8160	0.8003	0.5514
-1.00	25	-0.9286	-0.9116	-0.7826
-0.95	25	-0.9207	-0.9044	-0.7770
-0.80	25	-0.8183	-0.8094	-0.7072
-0.60	25	-0.6254	-0.6220	-0.5498
-0.40	25	-0.4135	-0.4118	-0.3698
-0.20	25	-0.2096	-0.2090	-0.1846
0.00	25	0.0010	0.0015	0.0023
0.20	25	0.2112	0.2110	0.1900
0.40	25	0.4213	0.4200	0.3768
0.60	25	0.6366	0.6331	0.5578
0.80	25	0.8300	0.8204	0.7137
0.95	25	0.9270	0.9101	0.7810
1.00	25	0.9376	0.9201	0.7864

1: The variances for the estimates are given by ***(MSE / 1000)***.

θ	n	MLE	MLE_1	MeLE
-1.00	50	-0.9651	-0.9444	-0.8822
-0.95	50	-0.9509	-0.9335	-0.8732
-0.80	50	-0.8316	-0.8272	-0.7752
-0.60	50	-0.6291	-0.6279	-0.5918
-0.40	50	-0.4215	-0.4217	-0.3979
-0.20	50	-0.2146	-0.2147	-0.2029
0.00	50	-0.0084	-0.0087	-0.0080
0.20	50	0.1983	0.1980	0.1869
0.40	50	0.4048	0.4046	0.3823
0.60	50	0.6151	0.6134	0.5781
0.80	50	0.8243	0.8199	0.7674
0.95	50	0.9509	0.9349	0.8699
1.00	50	0.9651	0.9450	0.8792
-1.00	100	-0.9806	-0.9580	-0.9314
-0.95	100	-0.9576	-0.9442	-0.9168
-0.80	100	-0.8113	-0.8095	-0.7858
-0.60	100	-0.6086	-0.6078	-0.5904
-0.40	100	-0.4070	-0.4063	-0.3950
-0.20	100	-0.2050	-0.2052	-0.1990
0.00	100	-0.0024	-0.0027	-0.0021
0.20	100	0.2012	0.2009	0.1953
0.40	100	0.4048	0.4049	0.3930
0.60	100	0.6085	0.6074	0.5905
0.80	100	0.8128	0.8106	0.7871
0.95	100	0.9571	0.9443	0.9166
1.00	100	0.9818	0.9593	0.9319

TABLE 3B: MA(1) process; MSE[2] of the MLE, the inital value for the MLE (MLE_I), and the MeLE of θ.

θ	n	MLE	MLE_I	MeLE
-1.00	2	1.1029	1.0786	0.9083
-0.95	2	1.0370	1.0117	0.8170
-0.80	2	0.8876	0.8599	0.5749
-0.60	2	0.7498	0.7189	0.3217
-0.40	2	0.7208	0.6837	0.1519
-0.20	2	0.7200	0.6802	0.0552
0.00	2	0.7450	0.7018	0.0254
0.20	2	0.7397	0.6972	0.0573
0.40	2	0.7254	0.6881	0.1551
0.60	2	0.7897	0.7556	0.3260
0.80	2	0.9104	0.8804	0.5760
0.95	2	1.0463	1.0214	0.8166
1.00	2	1.1016	1.0775	0.9070
-1.00	5	0.4628	0.4542	0.5404
-0.95	5	0.4357	0.4256	0.4745
-0.80	5	0.3785	0.3630	0.3153
-0.60	5	0.3715	0.3502	0.1827
-0.40	5	0.4228	0.3981	0.1178
-0.20	5	0.4535	0.4272	0.0931
0.00	5	0.4670	0.4414	0.0893
0.20	5	0.4670	0.4403	0.0989
0.40	5	0.4274	0.4032	0.1270
0.60	5	0.3822	0.3616	0.1908
0.80	5	0.3689	0.3534	0.3192
0.95	5	0.4107	0.4022	0.4751
1.00	5	0.4367	0.4299	0.5405
-1.00	10	0.1403	0.1403	0.2377
-0.95	10	0.1269	0.1251	0.1974
-0.80	10	0.1317	0.1256	0.1201
-0.60	10	0.1686	0.1598	0.0873
-0.40	10	0.2046	0.1957	0.0896
-0.20	10	0.2310	0.2231	0.0988
0.00	10	0.2531	0.2453	0.1030
0.20	10	0.2296	0.2221	0.0982
0.40	10	0.1943	0.1868	0.0890
0.60	10	0.1576	0.1502	0.0886
0.80	10	0.1251	0.1194	0.1230
0.95	10	0.1284	0.1265	0.2004
1.00	10	0.1394	0.1392	0.2409
-1.00	25	0.0199	0.0208	0.0565
-0.95	25	0.0164	0.0156	0.0396
-0.80	25	0.0296	0.0268	0.0252
-0.60	25	0.0488	0.0467	0.0332
-0.40	25	0.0596	0.0588	0.0424
-0.20	25	0.0647	0.0642	0.0473
0.00	25	0.0640	0.0642	0.0480
0.20	25	0.0635	0.0633	0.0463
0.40	25	0.0617	0.0609	0.0412
0.60	25	0.0484	0.0465	0.0313
0.80	25	0.0267	0.0235	0.0220
0.95	25	0.0127	0.0118	0.0361
1.00	25	0.0143	0.0152	0.0526

2: The variances for the estimates of the MSE are approximately given by $(2\ MSE^2 / 1000)$.

θ	n	MLE	MLE_I	MeLE
-1.00	50	0.0038	0.0049	0.0156
-0.95	50	0.0039	0.0031	0.0082
-0.80	50	0.0125	0.0110	0.0086
-0.60	50	0.0187	0.0184	0.0150
-0.40	50	0.0222	0.0220	0.0188
-0.20	50	0.0234	0.0238	0.0206
0.00	50	0.0241	0.0241	0.0214
0.20	50	0.0239	0.0241	0.0212
0.40	50	0.0222	0.0223	0.0196
0.60	50	0.0199	0.0194	0.0160
0.80	50	0.0124	0.0112	0.0093
0.95	50	0.0036	0.0029	0.0088
1.00	50	0.0041	0.0051	0.0165
-1.00	100	0.0010	0.0021	0.0051
-0.95	100	0.0016	0.0010	0.0020
-0.80	100	0.0052	0.0050	0.0045
-0.60	100	0.0077	0.0077	0.0071
-0.40	100	0.0098	0.0099	0.0091
-0.20	100	0.0111	0.0113	0.0104
0.00	100	0.0114	0.0115	0.0107
0.20	100	0.0107	0.0108	0.0101
0.40	100	0.0094	0.0095	0.0089
0.60	100	0.0076	0.0076	0.0071
0.80	100	0.0052	0.0050	0.0045
0.95	100	0.0015	0.0010	0.0020
1.00	100	0.0009	0.0020	0.0050

TABLE 3C: MA(1) process; Probabilities[3] of the MLE, and the inital value for the MLE (MLE_I) of θ to be on the boundaries.

θ	n	MLE +1	MLE_I +1	MLE -1	MLE_I -1
-1.00	2	0.196	0.200	0.480	0.487
-0.95	2	0.198	0.201	0.479	0.483
-0.80	2	0.199	0.204	0.475	0.484
-0.60	2	0.203	0.207	0.475	0.483
-0.40	2	0.242	0.251	0.437	0.450
-0.20	2	0.279	0.280	0.415	0.426
0.00	2	0.341	0.347	0.359	0.363
0.20	2	0.396	0.405	0.302	0.308
0.40	2	0.433	0.446	0.250	0.256
0.60	2	0.490	0.495	0.229	0.232
0.80	2	0.498	0.505	0.221	0.223
0.95	2	0.504	0.508	0.203	0.208
1.00	2	0.502	0.510	0.203	0.205
-1.00	5	0.058	0.058	0.554	0.560
-0.95	5	0.060	0.060	0.553	0.558
-0.80	5	0.064	0.064	0.526	0.529
-0.60	5	0.071	0.072	0.458	0.459
-0.40	5	0.096	0.097	0.361	0.364
-0.20	5	0.128	0.130	0.265	0.267
0.00	5	0.173	0.174	0.181	0.184
0.20	5	0.254	0.258	0.132	0.133
0.40	5	0.350	0.354	0.089	0.091
0.60	5	0.449	0.453	0.072	0.072
0.80	5	0.560	0.564	0.060	0.061
0.95	5	0.610	0.612	0.053	0.054
1.00	5	0.610	0.617	0.052	0.053
-1.00	10	0.006	0.009	0.451	0.623
-0.95	10	0.006	0.009	0.449	0.617
-0.80	10	0.009	0.011	0.393	0.515
-0.60	10	0.013	0.015	0.269	0.341
-0.40	10	0.017	0.021	0.175	0.214
-0.20	10	0.026	0.032	0.103	0.116
0.00	10	0.054	0.067	0.062	0.072
0.20	10	0.103	0.121	0.028	0.032
0.40	10	0.155	0.195	0.016	0.018
0.60	10	0.237	0.321	0.011	0.012
0.80	10	0.385	0.511	0.007	0.009
0.95	10	0.418	0.608	0.009	0.009
1.00	10	0.431	0.630	0.009	0.010
-1.00	25	0.000	0.000	0.592	0.691
-0.95	25	0.000	0.000	0.565	0.647
-0.80	25	0.000	0.000	0.282	0.333
-0.60	25	0.000	0.000	0.092	0.115
-0.40	25	0.000	0.000	0.027	0.031
-0.20	25	0.000	0.000	0.012	0.013
0.00	25	0.001	0.002	0.001	0.004
0.20	25	0.005	0.008	0.001	0.001
0.40	25	0.030	0.039	0.000	0.000
0.60	25	0.106	0.121	0.000	0.000
0.80	25	0.316	0.365	0.000	0.000
0.95	25	0.590	0.664	0.000	0.000
1.00	25	0.616	0.717	0.000	0.000

3: The variances of the estimates are given by ***(p(1-p) / 1000)***, where ***p*** is the estimate of the probability.

θ	n	MLE +1	MLE_I +1	MLE -1	MLE_I -1
-1.00	50	0.000	0.000	0.662	0.783
-0.95	50	0.000	0.000	0.551	0.681
-0.80	50	0.000	0.000	0.141	0.193
-0.60	50	0.000	0.000	0.011	0.019
-0.40	50	0.000	0.000	0.003	0.003
-0.20	50	0.000	0.000	0.000	0.000
0.00	50	0.000	0.000	0.000	0.000
0.20	50	0.000	0.000	0.000	0.000
0.40	50	0.000	0.000	0.000	0.000
0.60	50	0.023	0.027	0.000	0.000
0.80	50	0.133	0.171	0.000	0.000
0.95	50	0.522	0.674	0.000	0.000
1.00	50	0.663	0.806	0.000	0.000
-1.00	100	0.000	0.000	0.733	0.919
-0.95	100	0.000	0.000	0.401	0.687
-0.80	100	0.000	0.000	0.024	0.048
-0.60	100	0.000	0.000	0.001	0.001
-0.40	100	0.000	0.000	0.000	0.000
-0.20	100	0.000	0.000	0.000	0.000
0.00	100	0.000	0.000	0.000	0.000
0.20	100	0.000	0.000	0.000	0.000
0.40	100	0.000	0.000	0.000	0.000
0.60	100	0.000	0.000	0.000	0.000
0.80	100	0.024	0.046	0.000	0.000
0.95	100	0.383	0.685	0.000	0.000
1.00	100	0.741	0.933	0.000	0.000

TABLE 3D: MA(1) process; Average mean square errors of the estimators by sample size, excluding the values $\hat{\theta} = \pm 0.95$.

SAMPLE SIZE	MLE[1]	MLE_I[2]	MeLE[3]
2	0.83572	0.80199	0.36898
5	0.42166	0.40205	0.23773
10	0.17957	0.17341	0.12511
25	0.04556	0.04463	0.04055
50	0.01702	0.01694	0.01660
100	0.00727	0.00749	0.00750

1: The MLE from DALL.

2: The MLE from the grid search (the initial value for the MLE).

3: The MeLE.

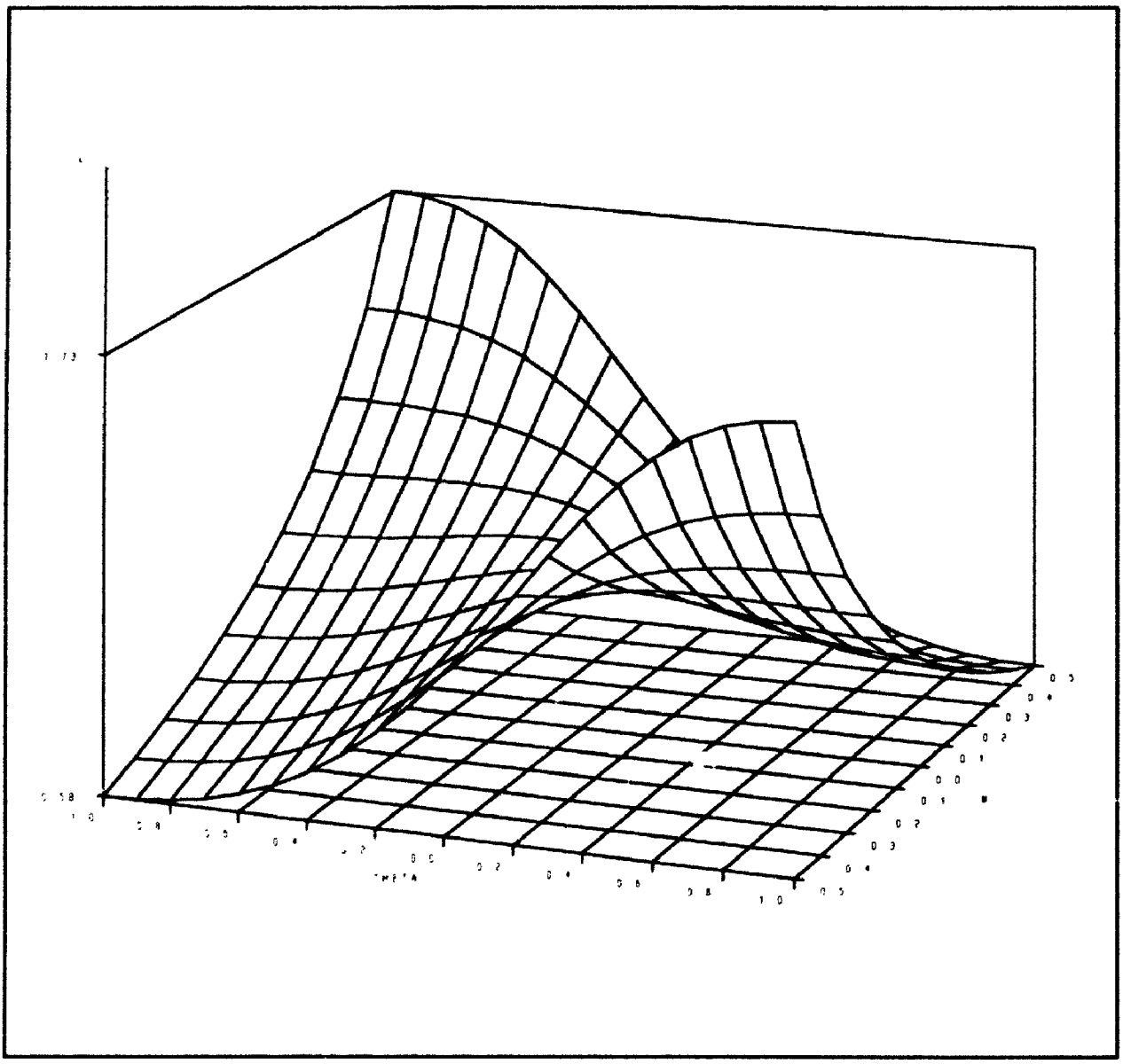

Figure 1: MA(1) process; Ensemble of concentrated likelihood functions, $L(\theta \mid w)$, for sample size $n = 2$.

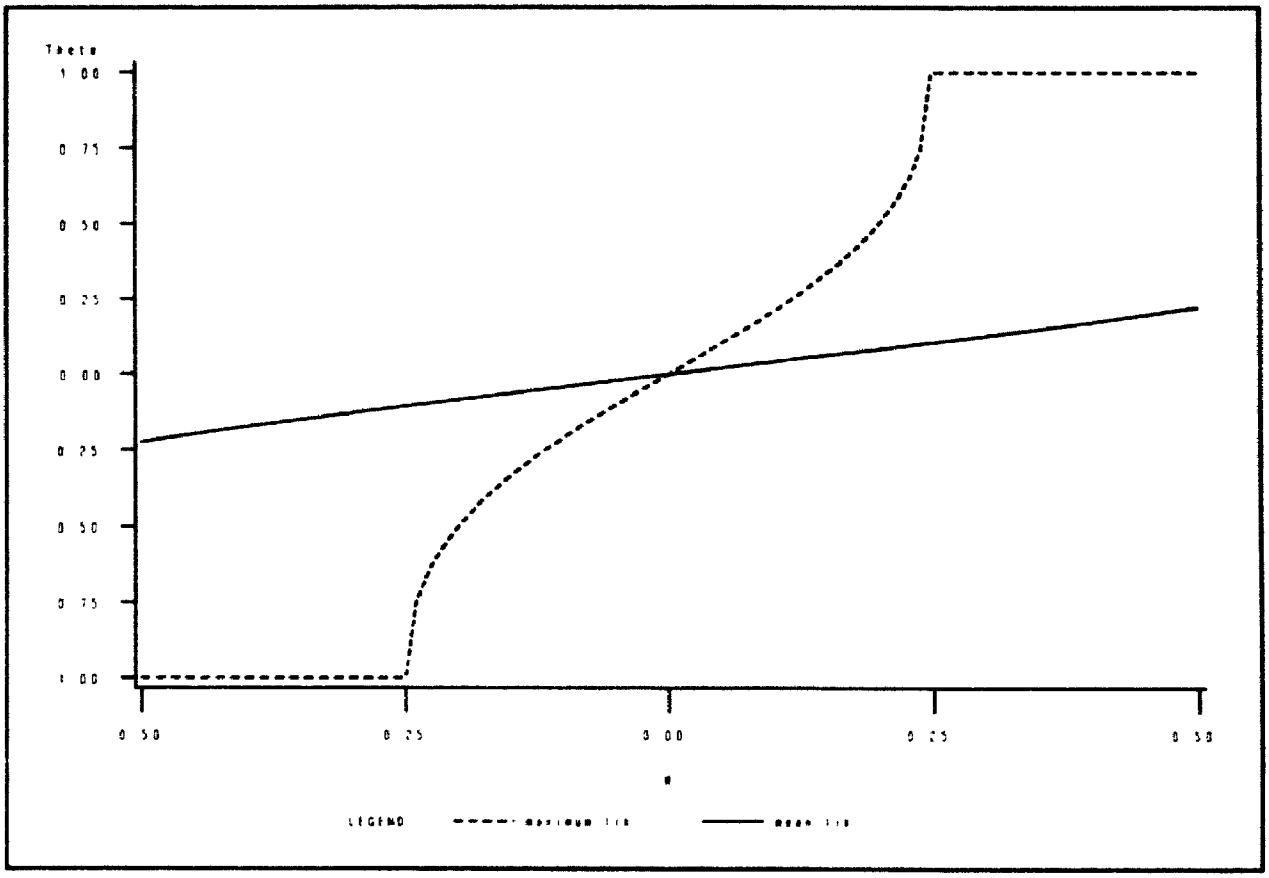

Figure 2: MA(1) process; MeLE and MLE of θ as a function of W for sample size 2.

Figures 3: MA(1) process; Bias of the MLE, the initial value for the MLE, and the MeLE of θ.

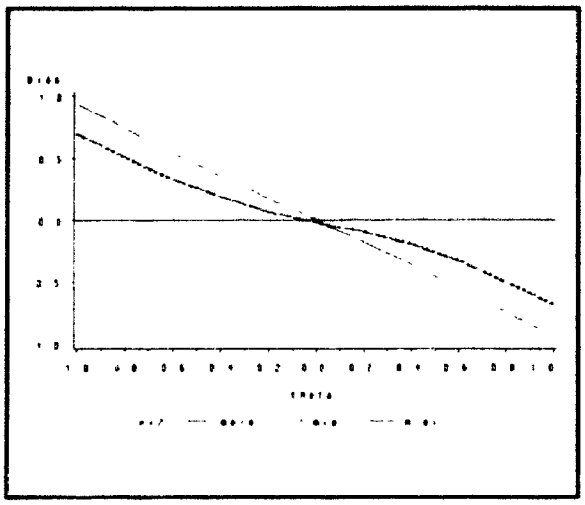

Figure 3a: $n=2$.

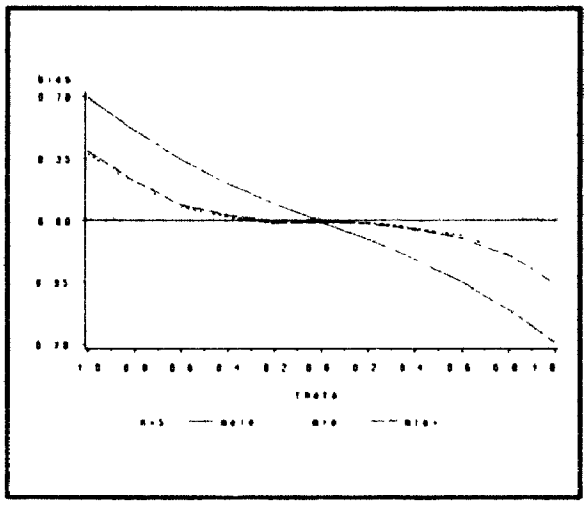

Figure 3b: $n=5$.

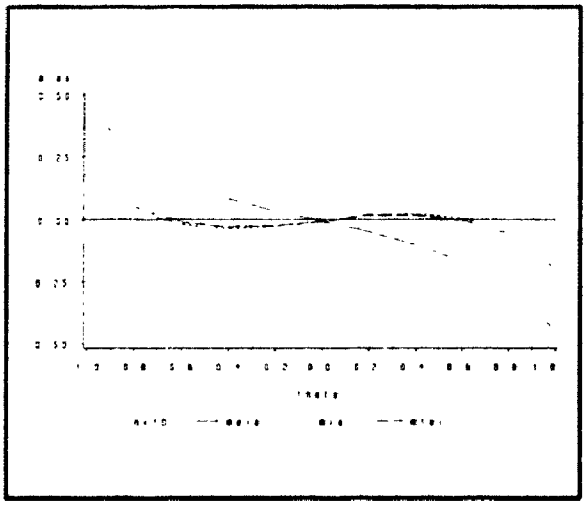

Figure 3c: $n=10$.

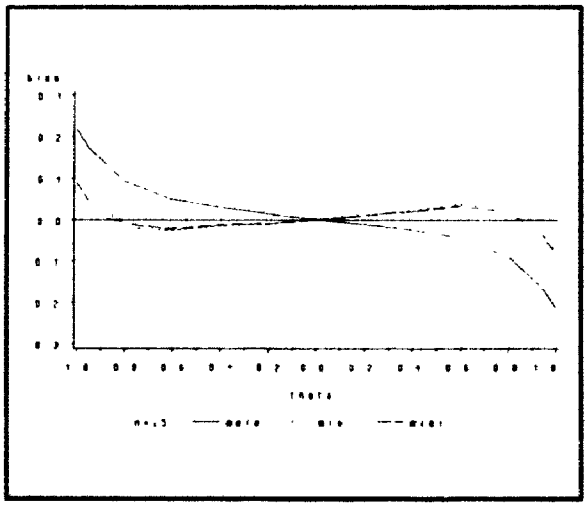

Figure 3d: $n=25$.

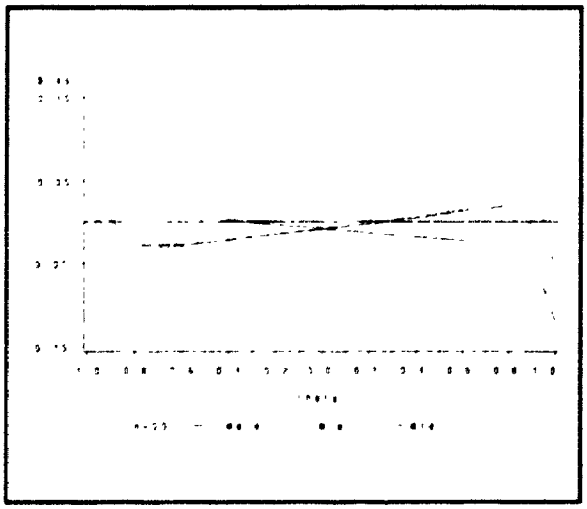

Figure 3e: $n=50$.

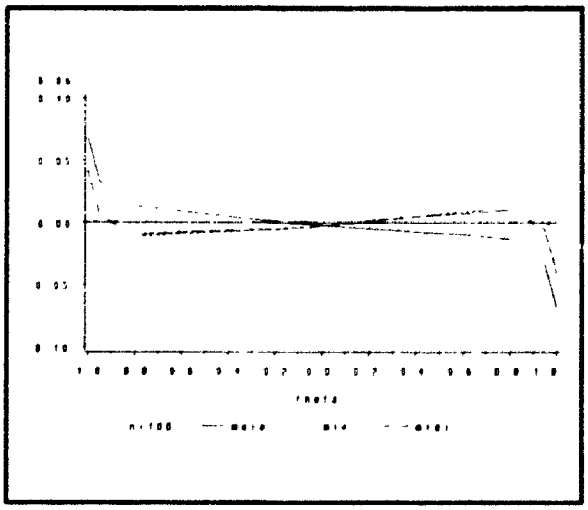

Figure 3f: $n=100$.

Figures 4: MA(1) process; MSE of the MLE, the initial value for the MLE, and the MeLE of θ.

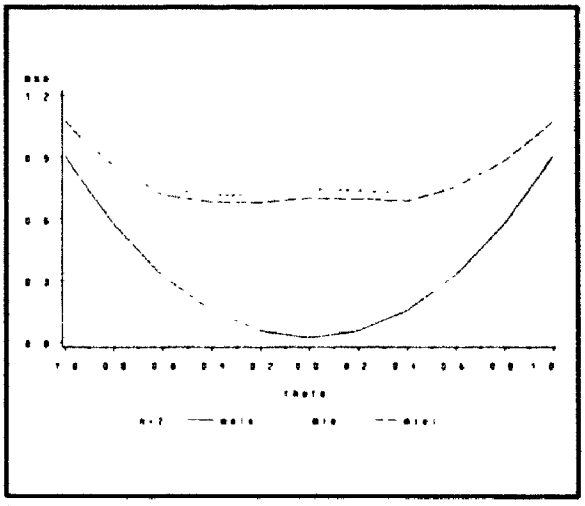

Figure 4a: $n=2$.

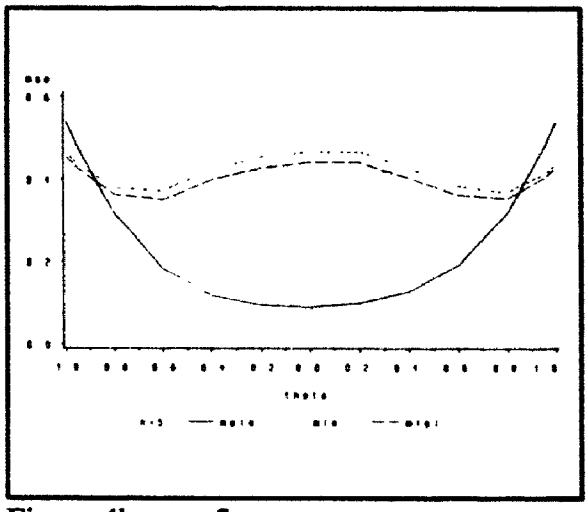

Figure 4b: $n=5$.

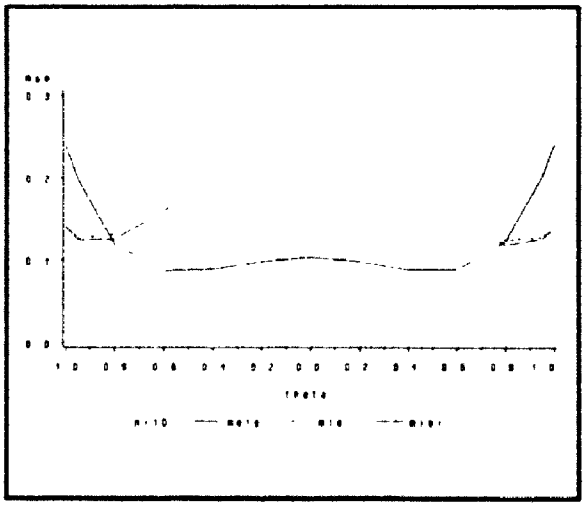

Figure 4c: $n=10$.

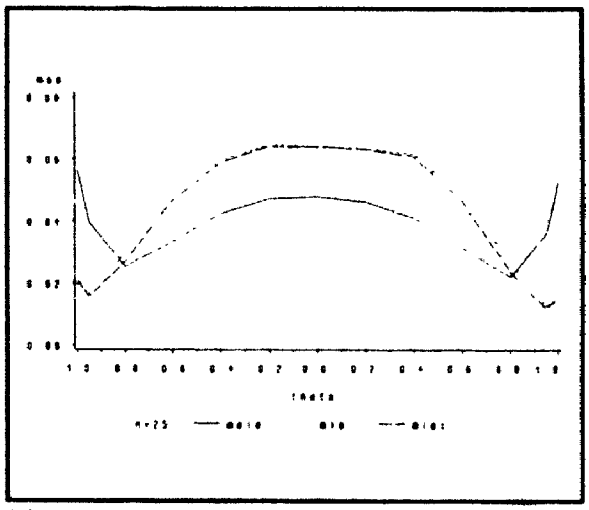

Figure 4d: $n=25$.

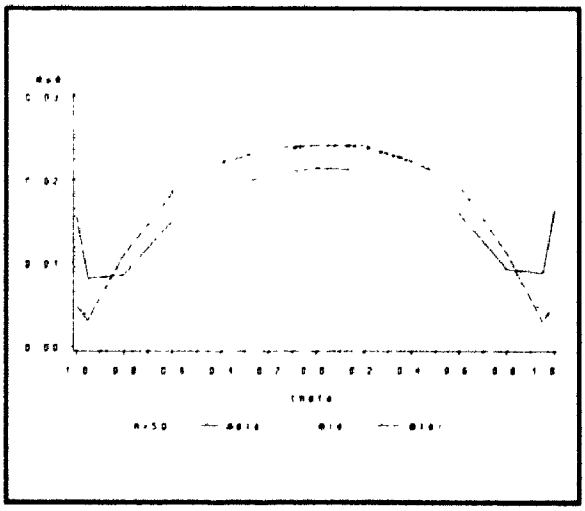

Figure 4e: $n=50$.

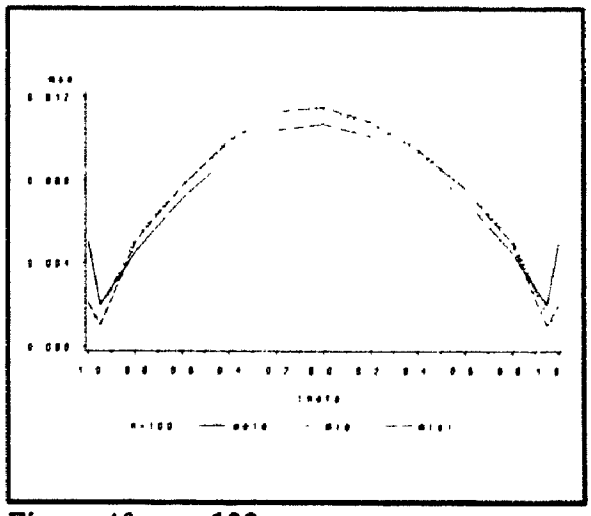

Figure 4f: $n=100$.

CHAPTER 3

MEAN LIKELIHOOD ESTIMATION OF FRACTIONALLY DIFFERENCED AUTOREGRESSIVE MOVING AVERAGE TIME SERIES MODELS

The fractionally differenced autoregressive moving average (FARMA) class of models generalizes the ARMA class where an extra fractional difference parameter f takes its value in the open interval $(-0.5, 0.5)$. The FARMA class reduces to the ARMA class when $f = 0$. This extra fractional difference parameter allows one to model a series that has long memory ($f > 0$) as opposed to short memory ($f = 0$ or $f < 0$).

This chapter deals with the practical aspects of FARMA models. I illustrate the FARMA process. I present the theory and methods needed to simulate FARMA series and to calculate the exact value of the likelihood function in the time domain. I discuss how to estimate the parameters of a FARMA model by the methods of mean likelihood estimation (MeLE) and maximum likelihood estimation (MLE). Finally, I run various simulations to calculate the mean square errors of the MeLE and MLE, and to compare the available estimates of the covariance matrix of the MeLE.

Throughout this chapter the theoretical mean of the FARMA process is assumed to be known and equal to zero. The estimation of the unknown mean of a stationary FARMA process is discussed in chapter 4.

Keywords: Fractionally differenced autoregressive moving average model; Autocovariance function; Durbin-Levinson recursion; Maximum and mean likelihood estimation; Monte Carlo integration; Powell's minimization algorithm; DALL maximization algorithm.

Foreword: Tables and figures are at the end of the current chapter. A reference to an equation is in the current chapter unless otherwise specified.

autocorrelations of the AR(1) process are all positive and decay exponentially. Consequently, the observations of the AR(1) process are positively correlated with small local trends.

Definition 2.8: The difference operator ∇^f, where f is any real number, is defined by the binomial series

$$\begin{aligned}\nabla^f &= (1-B)^f \\ &= \sum_{k=0}^{\infty} \binom{f}{k} (-1)^k B^k \\ &= 1 - fB + \frac{f(f-1)}{2}B^2 + \ldots + (-1)^n \frac{f(f-1)\ldots(f-n+1)}{n!}B^n + \ldots.\end{aligned}$$

Definition 2.9: The process $\{ x_t \}$ is a fractionally differenced process if it satisfies $\nabla^f x_t = a_t$, where f is in the open interval $(-0.5, 0.5)$, and $\{ a_t \}$ is a sequence of independent $N(0, \sigma_a^2)$ variables.

Remark 2.10: The fractionally differenced process will be denoted by FARMA(0,f,0) for reasons that will be obvious in definition 2.14.

Proposition 2.11: The fractionally differenced process has the following properties:

a) The process is stationary and invertible if $-0.5 < f < 0.5$.

This chapter is organized as follow. In section 2, I give the definition of an ARMA process of order (p,q), of a fractionally differenced process, and of a FARMA(p,f,q) process. I also give the corresponding autocovariance functions (ACF), and examples of such processes: an ARMA(1,0) series with parameter $\phi = 0.8$, a fractionally differenced series with parameter $f = 0.4$, and a FARMA(1,f,0) series with parameters $\phi = 0.8$ and $f = 0.4$. In section 3, I use the Durbin-Levinson recursion to derive a convenient expression for the joint distribution of a vector of size n from a FARMA(p,f,q) process. From this expression, I derive the algorithms to simulate a FARMA series and to calculate the exact likelihood function of the parameters of a FARMA series. In section 4, I use methods from numerical analysis to calculate the MLE and the MeLE. I give a transformation of the ARMA parameter space of dimension $p+q$ into the hypercube $(-1,1)^{p+q}$ where I carry out Monte Carlo integration. I give the algorithm for the transformation and the necessary steps to do the numerical integration over the parameter space of an ARMA model. In section 5, I discuss related algorithms and methods to simulate series, to calculate the likelihood function, to find the MLE, and to integrate over the stationary and invertible region of an ARMA(p,q) process.

I run various simulations in section 6. In section 6.1, I run a simulation to calculate the bias and MSE of the MeLE and the MLE of the fractional difference parameter in a FARMA(1,f,0) model, where the values of the autoregressive parameter and the noise variance are known. I show that the MSE of the MeLE is smaller than the MSE of the MLE for the values of $f = -0.4, -0.2, 0, 0.2, 0.4$ under investigation. In section 6.2, I run a simulation to compare various methods to calculate the MeLE and the MLE of the parameters of a FARMA(1,f,0) model and a FARMA(0,f,1) model. I compare the approximate likelihood calculations of Li and McLeod (1986) with the exact likelihood calculations. I also compare three algorithms to find the MLE.

The three algorithms are: a random grid search method, Powell's (1964) minimization algorithm, and DALL (Ishiguro and Akaike, 1989) maximization algorithm. I compute the MeLE by Monte Carlo integration. Unfortunately, Powell's or DALL algorithms are numerically unstable with the approximate likelihood calculations of Li and McLeod (1986). My conclusions are the following: 1) For smaller MSE's, the exact likelihood calculations instead of the approximate likelihood calculations should be used to compute the MeLE or to compute the MLE with a random grid search. On the other hand, the approximate likelihood calculations are much faster. 2) The MeLE instead of the MLE should be used when the approximate likelihood calculations are used. 3) The MeLE has a smaller MSE than the MLE, computed with Powell's algorithm, when the parameters are sufficiently inside the parameter space. Consequently the MeLE should be used in those cases. 4) I found that DALL is not reliable in 4 out of 25 cases. In the others cases, the MLE, computed with DALL, and the MeLE compete very well. In light of the previous conclusions, I suggest using the MeLE when the parameters are sufficiently inside the parameter space. I conclude section 6.2 with comments on the practical advantages of the MeLE over the MLE. In section 6.3, I run a simulation to compare the two estimates of the covariance matrix of the MeLE that are available in a practical application. The first estimate is the likelihood covariance matrix of the MeLE which is computed simultaneously with the MeLE. The second estimate is the inverse of the Fisher information matrix evaluated at the MeLE. I show that the diagonal elements of the likelihood covariance matrix of the MeLE are the closest to the sampling variances of the MeLE. I give the conclusions of this chapter in section 7.

2. The fractionally differenced autoregressive moving average process.

Definition 2.1: The process $\{ u_t, t = 0, \pm 1, \pm 2, \ldots \}$ is an autoregressive moving average process of order (p,q), ARMA(p,q), if it satisfies

$$u_t - \phi_1 u_{t-1} - \ldots - \phi_p u_{t-p} = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q},$$

where $\{ a_t \}$ is a sequence of independent Normal variables with mean zero and variance σ_a^2, $N(0, \sigma_a^2)$.

Remark 2.2: The roots of $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$ must lie outside the unit circle to ensure that the process is stationary and invertible. These conditions define the parameter space for an ARMA process. Denote by E_p the set of all points $(\alpha_1, \ldots, \alpha_p) \in \mathbb{R}^p$ such that the roots of the polynomial

$$\alpha(B) = 1 - \alpha_1 B - \ldots - \alpha_p B^p$$

are outside the unit circle.

Proposition 2.3: The invertibile region, E_q, of the MA(q) model coincides with that of the stationary region of the AR(q) model.

Example 2.4: For the ARMA(1,0) and ARMA(0,1) models the region is the open interval $E_1 = (-1,1)$. For the ARMA(2,0) and ARMA(0,2) models the region E_2 is the triangle $-1 < y < 1$, $|x| < 1 - y$ in the x-y plane. For the ARMA(1,1) model the region is the cartesian product of $E_1 \times E_1$, which is the square $(-1,1) \times (-1,1)$, though the line $\phi = \theta$ gives an over parametrized model.

Proposition 2.5: The autocovariance generating function $\gamma^u(B) = \sum_{j=-\infty}^{\infty} \gamma_j^u B^j$ of the ARMA(p,q) process is $\gamma^u(B) = \sigma_a^2 \psi^u(B) \psi^u(B^{-1})$ where $\psi^u(B) = \frac{\theta(B)}{\phi(B)}$.

Proof: See Box-Jenkins, 1976, equation A3.1.3.

Remark 2.6: The coefficients of the polynomial $\psi^u(B)$ are called the ψ-weights of the corresponding ARMA model. McLeod (1975) gives an algorithm for the calculation of the theoretical autocovariance function of an ARMA model that uses only a finite number of the ψ-weight.

Example 2.7: *The AR(1) process.* Figure 1 (located at the end of the chapter) gives a sequence $\{a_t\}$ of 100 observations. Figure 2 shows an AR(1) series, $u_t = \phi u_{t-1} + a_t$ with $\phi = 0.8$. The sequence $\{a_t\}$ is the same in both figures 1 and 2. Figure 3 gives the theoretical autocorrelation function (ACF) of an AR(1) process with $\phi = 0.8$ The autocorrelation at lag k of the AR(1) process is $\rho_k^u = \phi^k$. The variance of the AR(1) process is $\gamma_0^u = (1-\phi^2)^{-1}$. The

autocorrelations of the AR(1) process are all positive and decay exponentially. Consequently, the observations of the AR(1) process are positively correlated with small local trends.

Definition 2.8: The difference operator ∇^f, where f is any real number, is defined by the binomial series

$$\begin{aligned}\nabla^f &= (1-B)^f \\ &= \sum_{k=0}^{\infty} \binom{f}{k} (-1)^k B^k \\ &= 1 - fB + \frac{f(f-1)}{2}B^2 + \ldots + (-1)^n \frac{f(f-1)\ldots(f-n+1)}{n!}B^n + \ldots.\end{aligned}$$

Definition 2.9: The process $\{ x_t \}$ is a fractionally differenced process if it satisfies $\nabla^f x_t = a_t$, where f is in the open interval $(-0.5, 0.5)$, and $\{ a_t \}$ is a sequence of independent $N(0, \sigma_a^2)$ variables.

Remark 2.10: The fractionally differenced process will be denoted by FARMA(0,f,0) for reasons that will be obvious in definition 2.14.

Proposition 2.11: The fractionally differenced process has the following properties:

a) The process is stationary and invertible if $-0.5 < f < 0.5$.

b) When the process is stationary and invertible, and $\sigma_a^2 = 1$, the autocovariance function of the process is:

$$\begin{aligned} \gamma_0^x &= \frac{\Gamma(1-2f)}{\{\Gamma(1-f)\}^2}, \\ \gamma_k^x &= \gamma_{k-1}^x \frac{k-1+f}{k-f}, \quad k \geq 1 \\ \gamma_{-k}^x &= \gamma_k^x. \end{aligned} \tag{2.1}$$

c) The coefficients of $\psi^x(B) = (1-B)^{-f}$ give the ψ-weights of the fractionally differenced process.

Hosking (1981) and Granger and Joyeux (1980) have derived these properties of the fractionally differenced process.

Remark 2.12: Equations (2.1) provide a suitable algorithm to calculate the autocovariances of the fractionally differenced process. The variance requires the evaluation of the gamma function. For $1 \leq x \leq 2$, let $z = x - 1$, then the gamma function is approximated by

$$\Gamma(x) \approx 1 + \sum_{i=1}^{8} \beta_i z^i,$$

where the β_i's are provided in Abramowitz and Stegun (1968), (equation 6.1.36). For $0 < x < 1$,

$$\Gamma(x) = \frac{\Gamma(1+x)}{x}$$

and for $x > 2$, let $z = x - n$, where n is the integer part of x. Then:

$$\Gamma(x) = (n-1+z)(n-2+z)...(1+z)\Gamma(1+z).$$

Example 2.13: *The FARMA(0,f,0) process.* Figure 4 shows a FARMA(0,f,0) series with $f = 0.4$. The sequence $\{ a_t \}$ is the same as in the previous examples. Figure 5 displays the ACF of the process. A comparison of figures 1 and 4 shows that the FARMA(0,f,0) series has local trends whereas the trend of the series $\{ a_t \}$ is an horizontal line at zero. The autocorrelations of the FARMA(0,f,0) process with $f > 0$ are all positive and decrease hyperbolically as compared to those of an ARMA process, which decrease exponentially.

Definition 2.14: The process $\{ y_t \}$ is a fractionally differenced autoregressive moving average, FARMA(p,f,q), if it satisfies $\phi(B)\nabla^f y_t = \theta(B)a_t$. Here $\phi(B)$ and $\theta(B)$ are the autoregressive and moving average operators respectively, and ∇^f is the fractional difference operator. $\{ a_t \}$ is a sequence of independent $N(0, \sigma_a^2)$ variables.

Proposition 2.15: The FARMA(p,f,q) process is stationary and invertible if $-0.5 < f < 0.5$ and the roots of $\phi(B)$ and $\theta(B)$ lie outside the unit circle.

Proposition 3.3: The joint density of $Y = (Y_0, \ldots, Y_{n-1})$ from a FARMA(p,f,q) process with mean zero and, parameters $\beta = (\phi_1,\ldots,\phi_p,f,\theta_1,\ldots,\theta_q)$ and σ_a^2 is:

$$f_Y(y \mid \beta, \sigma_a^2) = \left(2\pi\sigma_a^2\right)^{-\frac{n}{2}} \left(\prod_{t=0}^{n-1} v_t^* \right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma_a^2} \sum_{t=0}^{n-1} \frac{(y_t - m_t)^2}{v_t^*} \right].$$

where:

$$m_0 = 0$$

$$m_t = \mathcal{E}(y_t \mid y_{t-1},\ldots,y_0) = \sum_{j=1}^{t} \phi_{t,j} y_{t-j}$$

$$v_0 = \sigma_a^2 \gamma_0^y$$

$$v_t = Var(y_t \mid y_{t-1},\ldots,y_0) = \sigma_a^2 \gamma_0^y \prod_{j=1}^{t} (1 - \phi_{j,j}^2) = \sigma_a^2 v_t^*.$$

Proof: The joint distribution of the random vector Y is:

$$f_Y(y \mid \beta, \sigma_a^2) = f_Y(y_0,\ldots,y_{n-1} \mid \beta, \sigma_a^2)$$
$$= f_Y(y_0 \mid \beta, \sigma_a^2)\, f_Y(y_1 \mid y_0; \beta, \sigma_a^2) \ldots f_Y(y_{n-1} \mid y_0,\ldots,y_{n-2}; \beta, \sigma_a^2).$$

The marginal distribution of Y_0 is $Y_0 \sim N(0, \sigma_a^2 \gamma_0^y)$ The conditional distribution of $Y_t \mid Y_{t-1}, \ldots, Y_0$ is normal with conditional mean m_t and conditional variance v_t where the coefficients $\phi_{t,j}$ $j=1,\ldots,t$, $t=1,\ldots,n-1$ are obtained from the Durbin-Levinson recursion.

I now give the algorithm to simulate a FARMA(p,f,q) series. I use that algorithm to simulate the series of examples 2.7, 2.13 and 2.17.

Algorithm 2.18: The algorithm to calculate the autocovariances of a FARMA(p,f,q) process is described by equation (2.2). If $p = q = 0$, the autocovariances are those of a FARMA(0,f,0). If $f = 0$, the autocovariances of the FARMA(p,0,q) process are those of an ARMA(p,q) process. If $p = 0$, $f \neq 0$ and $q > 0$ the summation in (2.2) is from $-q$ to q since the autocorrelations of an ARMA(0,q) process are zero after lag q. When $p > 0$ the autocorrelations of the ARMA(p,q) process decay exponentially to zero and the summation in (2.2) is taken from $-M$ to M, where M is large enough. One way to pick M is to select an error level (for example $\epsilon = 10^{-5}$) and to select M such that:

$$|\gamma_0^y - \gamma_0^y(M)| < \epsilon, \tag{2.3}$$

where

$$\gamma_0^y(M) = \sigma_a^2 \gamma_0^u \gamma_0^x \sum_{j=-M}^{M} \rho_j^u \rho_j^x. \tag{2.4}$$

Proposition 2.19: A lower bound for M is:

$$M^* = \frac{\log \epsilon_2 (1 - |G^*|)}{\log |G^*|} + 1, \tag{2.5}$$

where

$$\epsilon_2 = \frac{\epsilon}{2p\ \sigma_a^2\ \gamma_0^u\ \gamma_0^x \max(|A_1|,\ldots,|A_p|)},$$

$$|G^*| = \max(G_1,\ldots,G_p).$$

The constants $A_1,\ldots,A_p,G_1,\ldots,G_p$ are such that:

$$\rho_j^u = \sum_{i=1}^{p} A_i G_i^j, \tag{2.6}$$

where $G_1^{-1},G_2^{-1},\ldots,G_p^{-1}$ are the roots of the characteristic equation

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p = 0,$$

and $|G_i| < 1$ by stationarity.

Proof: Since

$$\gamma_0^y - \gamma_0^y(M) = 2\sigma_a^2 \gamma_0^u \gamma_0^x \sum_{j=M+1}^{\infty} \rho_j^u \rho_j^x,$$

it follows that (2.3) holds if:

$$\left| \sum_{j=M+1}^{\infty} \rho_j^a \rho_j^x \right| < \sum_{j=M+1}^{\infty} |\rho_j^a| < \frac{\epsilon}{2\sigma_a^2 \gamma_0^a \gamma_0^x}. \tag{2.7}$$

Substitution of (2.6) into (2.7) for the autocorrelations of the ARMA(p,q) process gives:

$$\begin{aligned} \sum_{j=M+1}^{\infty} \left| \sum_{l=1}^{p} A_l G_l^j \right| &\le \sum_{j=M+1}^{\infty} \sum_{l=1}^{p} |A_l| \, |G_l^j| \\ &\le p \max(|A_1|,\ldots,|A_p|) \sum_{j=M+1}^{\infty} |G^*|^j \\ &\le p \max(|A_1|,\ldots,|A_p|) \frac{|G^*|^{M+1}}{1 - |G^*|}. \end{aligned} \tag{2.8}$$

Equation (2.5) now follows from (2.8) and (2.7).

3. Simulation and likelihood calculation of FARMA(p,f,q) series.

Remark 3.1: In this section, I use the Durbin-Levinson recursion to derive a convenient expression for the joint distribution of a vector $Y = (Y_0, \ldots, Y_{n-1})$ of size n from a FARMA(p,f,q) process. The distribution is determined by the mean (assumed to be equal to zero), by the variance and by the autocorrelation from lag 1 to $n-1$. This is because the underlying distribution of the noise sequence $\{ a_t \}$ is assumed to be Normal. Next I derive the algorithms to simulate a FARMA series and to calculate the likelihood function of the parameters.

Algorithm 3.2: The Durbin-Levinson recursion (Box-Jenkins, 1976, equations A3.2.7 and A3.2.8) is used to calculate the partial autocorrelation function (PACF) $\pi = \{\pi_1, \pi_2, \ldots\} = \{\phi_{1,1}, \phi_{2,2}, \ldots\}$ from the autocorrelation function (ACF) $\rho = \{\rho_1, \rho_2, \ldots\}$.

Initialization: Let $N_1 = \phi_{1,1} = \pi_1 = \rho_1$, $D_1 = 1$.

Step 1: Calculate the partial autocorrelation coefficient.

$$N_t = \rho_t - \sum_{j=1}^{t-1} \phi_{t-1,j} \rho_{t-j},$$

$$D_t = D_{t-1} - \frac{N_{t-1}^2}{D_{t-1}},$$

$$\pi_t = \phi_{t,t} = \frac{N_t}{D_t}.$$

Step 2: Calculate the partial linear regression coefficients.

$$\phi_{t,j} = \phi_{t-1,j} - \phi_{t,t}\phi_{t-1,t-j}, \quad j = 1, \ldots, t-1.$$

Proposition 3.3: The joint density of $Y = (Y_0, \ldots, Y_{n-1})$ from a FARMA(p,f,q) process with mean zero and, parameters $\beta = (\phi_1,\ldots,\phi_p,f,\theta_1,\ldots,\theta_q)$ and σ_a^2 is:

$$f_Y(y\,|\,\beta,\sigma_a^2) = \left(2\pi\sigma_a^2\right)^{-\frac{n}{2}} \left(\prod_{t=0}^{n-1} v_t^*\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=0}^{n-1}\frac{(y_t - m_t)^2}{v_t^*}\right].$$

where:

$$m_0 = 0$$

$$m_t = \mathcal{E}(y_t|y_{t-1},\ldots,y_0) = \sum_{j=1}^{t}\phi_{t,j}y_{t-j}$$

$$v_0 = \sigma_a^2\gamma_0^y$$

$$v_t = Var\,(y_t|y_{t-1},\ldots,y_0) = \sigma_a^2\gamma_0^y\prod_{j=1}^{t}(1-\phi_{j,j}^2) = \sigma_a^2 v_t^*.$$

Proof: The joint distribution of the random vector Y is:

$$f_Y(y\,|\,\beta,\sigma_a^2) = f_Y(y_0,\ldots,y_{n-1}\,|\,\beta,\sigma_a^2)$$

$$= f_Y(y_0\,|\,\beta,\sigma_a^2)\, f_Y(y_1|y_0;\beta,\sigma_a^2)\ldots f_Y(y_{n-1}|y_0,\ldots,y_{n-2};\beta,\sigma_a^2).$$

The marginal distribution of Y_0 is $Y_0 \sim N(0,\sigma_a^2\gamma_0^y)$ The conditional distribution of $Y_t \mid Y_{t-1}, \ldots, Y_0$ is normal with conditional mean m_t and conditional variance v_t where the coefficients $\phi_{t,j}\ j=1,\ldots,t,\ t=1,\ldots,n-1$ are obtained from the Durbin-Levinson recursion.

I now give the algorithm to simulate a FARMA(p,f,q) series. I use that algorithm to simulate the series of examples 2.7, 2.13 and 2.17.

Algorithm 3.4: The algorithm to simulate a sample $y = (y_0,\ldots,y_{n-1})$ of size n from a stationary FARMA process was suggested by Hosking (1984). The algorithm is:

Step 1: Given the parameters $\beta = (\phi_1,\ldots,\phi_p,f,\theta_1,\ldots,\theta_q)$ calculate the theoretical variance γ_0^y and theoretical autocorrelations $\{\rho_1^y,\ldots,\rho_{n-1}^y\}$ of the FARMA(p,f,q) process assuming $\sigma_a^2 = 1$.

Step 2: Generate y_0 from a $N(0, \sigma_a^2 \gamma_0^y)$ distribution.

Step 3: Calculate the conditional mean m_t and conditional variance v_t.

Step 4: Generate y_t from a $N(m_t, v_t)$ distribution.

Algorithm 3.5: The algorithm to calculate the exact likelihood function for the parameters of a FARMA(p,f,q) series, given an observed vector $y = (y_0,\ldots,y_{n-1})$, is:

Step 1: Given the parameters $\beta = (\phi_1,\ldots,\phi_p,f,\theta_1,\ldots,\theta_q)$ calculate the theoretical variance γ_0^y and theoretical autocorrelations $\{\rho_1^y,\ldots,\rho_{n-1}^y\}$ of the FARMA(p,f,q) process assuming $\sigma_a^2 = 1$.

Step 2: Calculate the one step ahead forecast errors as follows:

Initialization: Let $m_0 = 0$, $v_0^* = \gamma_0^y$ and $e_0 = y_0 / (\gamma_0^y)^{1/2}$.

Recursion: Calculate the conditional means m_t and conditional variances v_t^* using the Durbin-Levinson recursion. Calculate the one step ahead forecast errors $e_t = \dfrac{y_t - m_t}{\sqrt{v_t^*}}$.

Step 3: The likelihood is:

$$L(\beta,\sigma_a^2|y) \propto \left(\sigma_a^2\right)^{-\frac{n}{2}} \left(\prod_{t=0}^{n-1} v_t^*\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=0}^{n-1} e_t^2\right].$$

Remark 3.6: Note that the algorithms to simulate a FARMA process and to calculate the exact likelihood function are similar. First the theoretical autocorrelations $\rho' = (\rho'_1, \ldots \rho'_{n-1})$ and variance γ'_0 of the FARMA(p,f,q) process are calculated given the vector of parameters $\beta = (\phi_1, \ldots, \phi_p, f, \theta_1, \ldots, \theta_q)$ and $\sigma_a^2 = 1$. Next, the Durbin-Levinson recursion is used to calculate the conditional means and variances. To simulate a process, a set of n independent observations from a $N(0,1)$ distribution are linearly transformed by multiplying by the square root of the conditional variances, and by adding the conditional means. To calculate the likelihood function, it is the observations that are linearly transformed by subtracting the conditional means and dividing by the square root of the conditional variances. In both cases, the data transformation and the calculation of the conditional means and variances are done simultaneously. Though Hosking (1984) has proposed the algorithm to simulate the FARMA(p,f,q) process, he did not recognize that the same algorithm can be used to calculate the exact likelihood function by going backward.

4. Mean and maximum likelihood estimation of the parameters of a FARMA process.

Proposition 4.1: i) The maximum likelihood estimate (MLE) of σ_a^2 is:

$$\hat{\sigma}_a^2 = \frac{1}{n} \sum_{t=0}^{n-1} e_t^2.$$

ii) The concentrated likelihood for the FARMA parameters is:

$$l(\beta \mid y) \propto \left[\prod_{t=0}^{n-1} v_t^* \right]^{-\frac{1}{2}} \left[\sum_{t=0}^{n-1} e_t^2 \right]^{-\frac{n}{2}}.$$

Notice that $e_t, v_t^*, \hat{\sigma}_a^2$ are functions of the parameters β and the observations $y = (y_0, \ldots, y_{n-1})$.

Remark 4.2: The concentrated likelihood is obtained when the MLE of σ^2 is substituted into the likelihood function. I have shown in chapter 2, remark 2.5 that the integration of the likelihood with respect to $(1/\sigma_a)\, d\sigma_a$ is equivalent to the concentrated likelihood.

Remark 4.3: The MLE of the parameters $\beta = (\phi_1,\ldots,\phi_p,f,\theta_1,\ldots,\theta_q)$ is a point in the parameter space where the concentrated likelihood has its maximum value. Equivalently, it is a point where the function

$$L_c^*(\beta \mid y) = n\log\hat{\sigma}_a^2 + \sum_{t=0}^{n-1}\log v_t^*.$$

has its minimum value.

Definition 4.4: The mean likelihood estimate (MeLE) $\bar{\beta} = (\bar{\phi}_1,\ldots,\bar{\phi}_p,\bar{f},\bar{\theta}_1,\ldots,\bar{\theta}_q)$ of a vector of parameter β has ith element:

$$\bar{\beta}_i = \bar{\beta}_i(y) = \frac{\int \beta_i l(\beta \mid y)d\beta}{\int l(\beta \mid y)d\beta}.$$

The (i,j) element of the likelihood covariance matrix of the MeLE is:

$$Cov_L(\bar{\beta}_i,\bar{\beta}_j) = \frac{\int \beta_i\beta_j l(\beta \mid y)d\beta}{\int l(\beta \mid y)d\beta} - \bar{\beta}_i\bar{\beta}_j.$$

Proposition 4.5: The MeLE of the FARMA parameters have the following properties:

i) The MeLE are biased.

ii) The MeLE have the smallest average mean square error, where the average is over all the parameter space.

iii) The MeLE and the MLE have the same asymptotic distribution.

Remark 4.6: Monte Carlo integration is used to obtain the MeLE of the FARMA parameters because there is no analytical form for it. The integration of the concentrated likelihood function over the parameter space of the fractional difference parameter is easily done. It suffices to generate uniform random variables in the interval $(-0.5,0.5)$. However, it is not trivial to integrate the concentrated likelihood over the ARMA parameter space. I use the *change of variables* technique to map the ARMA parameter space of dimension $p+q$ into the hypercube $(-1, 1)^{p+q}$, where I carry out Monte Carlo integration. First I give the algorithm for the transformation. Next, I give the algorithm for Monte Carlo integration over the ARMA parameter space.

Remark 4.7: *The transformation of the ARMA parameter space and its Jacobian.* I have mentioned in Proposition 2.3 that the invertible region E_q of the MA(q) process coincides with the stationary region of the corresponding AR(q) process and consequently, consider only the stationary region of an autoregressive process for a moment.

Wise (1956), Barndorff-Nielsen and Schou (1973), Ramsey (1974), Anderson (1975) and Piccolo (1982) have studied the region E_p. Ramsey (1974) gives a proof of the Barndorff-Nielsen and Schou (1973) theorem which says that there is a one to one correspondence between the

coefficients $\phi = (\phi_1,\ldots,\phi_p)$ of a stationary AR(p) process and its partial autocorrelation function (PACF) $\pi = (\pi_1,\ldots,\pi_p)$ defined over the hypercube $F_p = (-1, 1)^p$. Consequently, the proposed transformation is to take a point in the partial autocorrelation space F_p and to find its image in the parameter space E_p along with the jacobian.

Algorithm 4.8: *The Barndorff-Nielsen and Schou transformation.* The algorithm allows one to calculate ϕ from π.

i) set $\phi_{1,1}=\pi_1$.

ii) for k=2,...,p, calculate

$$\begin{aligned}\phi_{k,j} &= \phi_{k-1,j} - \pi_k\phi_{k-1,k-j}, \quad j = 1,\ldots,k-1\\ \phi_{k,k} &= \pi_k\end{aligned}$$

iii) set $\phi = (\phi_1,\ldots,\phi_p) = (\phi_{p,1},\ldots,\phi_{p,p})$.

The Jacobian of the transformation, previously derived by Daniels (1956) and given in Piccolo (1982), is:

$$J = \left|\frac{\partial\phi}{\partial\pi}\right| = \prod_{k=2}^{p}(1-\pi_k)^{\left[\frac{k}{2}\right]}(1+\pi_k)^{\left[\frac{k-1}{2}\right]}, \tag{4.1}$$

where $[a]$ is the integer part of a.

Remark 4.9: Notice that $\phi = (\phi_{p,1},\phi_{p,2},\ldots,\phi_{p,p})$ is also the vector of the partial linear regression coefficients calculated from the Durbin-Levinson recursion after p iterations. The recursive calculation of the partial linear regression coefficients from the PACF has been derived

many times, see Monahan (1984) for references dating from 1922. It does not seem, however, that it has been recognized as being related to the Durbin-Levinson recursion.

Remark 4.10: *Integration over* $E_p \times E_q$. For the computation of the MeLE it is required to calculate integral of the form.

$$I = \int_{E_p \times E_q} f(\phi,\theta) d\phi d\theta,$$

where $f(\phi, \theta)$ is any function of the parameters (ϕ, θ) of the ARMA(p,q) process. This integral can be evaluated by a *change of variables* technique:

$$I = \int_{F_p \times F_q} f(\phi(\pi),\theta(\eta)) \left|\frac{\partial\phi}{\partial\pi}\right| \left|\frac{\partial\theta}{\partial\eta}\right| d\pi d\eta .$$

Algorithm 4.11: The steps for one evaluation of the integrand by Monte Carlo integration are:

If $p \neq 0$ then:

i) generate $\pi = (\pi_1,\ldots,\pi_p)$ where $\pi_i \sim U(-1,1)$,

ii) map $\pi = (\pi_1,\ldots,\pi_p)$ into $\phi = (\phi_1,\ldots,\phi_p)$ using the Barndorff-Nielsen and Schou transformation (algorithm 4.8),

iii) calculate the Jacobian.

If $q \neq 0$ then:

iv) generate $\eta = (\eta_1,\ldots,\eta_q)$ where $\eta_j \sim U(-1,1)$,

v) map $\eta = (\eta_1,\ldots,\eta_q)$ into $\phi = (\phi_1,\ldots,\phi_q)$, treating η as π in the Barndorff-Nielsen and Schou transformation (algorithm 4.8),

vi) calculate the Jacobian,

vii) set $\theta = (\theta_1,\ldots,\theta_q) = (\phi_1,\ldots,\phi_q)$.

Finally evaluate f at (ϕ, θ).

5. Discussion of related algorithms.

5.1 Simulation of series.

Algorithm 3.4 for simulating a FARMA(p,f,q) can simulate ARMA(p,q) and FARMA(0,f,0) series as special cases. McLeod and Hipel (1978 b) present a faster algorithm for exact simulation of ARMA(p,q) series. McLeod and Hipel (1978 a) present an algorithm for exact simulation of FARMA(0,f,0) series but the algorithm requires $O(n^2)$ storage locations whereas algorithm 3.4 requires only $O(n)$ storage locations.

5.2 Likelihood calculations.

The basic idea behind the calculation of the likelihood is to compute the values taken by the one-step-ahead forecast errors, e_t, and their variances, v_t, both of them obtained by calculating the conditional means and conditional variances. Carlin et al.(1985) report that they use exact likelihood calculations but do not present the numerical algorithms. Haslett and Raftery (1989) use the same strategy as here but approximate the partial linear regression coefficients in the Durbin-Levinson recursion instead of computing them exactly. Hosking (1984) and Li and McLeod (1986) present approximations to the exact likelihood.

The proposed algorithm 3.5 can be used to compute the exact likelihood of ARMA process by setting $f = 0$. Others exact procedures for ARMA models, which are usually faster, are given in Ansley (1979), Gardner et al. (1980) and Mélard (1984). Ansley (1979) first transforms the original observations to obtain a new process with a band covariance matrix whose Cholesky

decomposition can be readily computed. A second transformation is applied to obtain a sequence of independent normal variables. The Cholesky decomposition and the second transformation are done at the same time. Gardner et al. (1980) write the ARMA model in a state space form and use the Kalman filter to perform the calculations. Mélard (1984) improves on the algorithm of Gardner et al. (1980). Algorithm 3.5 is attractive to statisticians because it uses the concepts of conditional means and conditional variances of a multivariate normal distribution. The conditional mean is obtained by a regression on the previous values of the series and the regression coefficients are computed with the Durbin-Levinson recursion. This approach is essentially the same as Kalman filtering but without the need for a state space representation.

McLeod and Hipel (1978 a) present an exact procedure to calculate the likelihood function of a FARMA(0,f,0) process but, as in the simulation of time series, their procedure requires $O(n^2)$ storage locations whereas the algorithm 3.5 requires only $O(n)$ storage locations. Li (1981) suggests the same procedure as the one discussed here but applies it only to FARMA(0,f,0) series because he does not have an efficient algorithm to calculate the theoretical autocovariances of the FARMA(p,f,q) process with $p + q > 0$.

Algorithms 3.5 is $O(n^2)$ flops but it could be truncated to say at 500 lags to get an approximate $O(n)$ algorithm for a very large sample size (e.g. $n > 700$).

5.3 Maximum likelihood estimation.

Three numerical algorithms are used to find the maximum likelihood estimates. The first is the function minimization algorithm of Powell (1964). It is convenient to standardize the series to have zero mean and unit variance when this unconstrained minimization algorithm is used. When the vector of parameters β is found to be outside the range of admissible values, the value of the function is set to an appropriate value (for example, the value of the function at the initial value of the parameters). DALL: Davidon's algorithm for log-likelihood maximization (Ishiguro and Akaike, 1989) is the second algorithm used to find the maximum likelihood estimates of the FARMA parameters. Finally, the third algorithm to find the MLE is a by-product of the MeLE calculations. It is a random grid search, where the MLE is a point at which the concentrated likelihood takes its maximum value among all points used to calculate the MeLE.

5.4 Integration over the stationary and invertible region of an ARMA(p,q).

Jones (1987) uses the "Barndorff-Nielsen and Schou" transformation to randomly select a point in the ARMA parameter space according to a uniform distribution. His method requires the generation of partial autocorrelations independently distributed as appropriate beta random variables on the interval $(-1, 1)$, whose joint density is proportional to the Jacobian J given in (4.1). This is achieved by generating a total of $n_1 = \frac{1}{2}\{p(p+1) + q(q+1)\}$ uniform random variables and computing order statistics. He notes that for large values of p and q it may be worthwhile to replace his method by more sophisticated routines which cut down the computation time. An immediate advantage of the algorithm 4.8 is that it requires only $n_2 = p+q$ uniform random variables which are transformed to give a point in the parameter space of an ARMA

model. It is not necessary to generate the PACF according to beta distributions for the purpose of integration over the parameter space since the Jacobian of the transformation makes the necessary adjustments to give the desired uniform distribution.

Marriott and Smith (1992) is a recent reference that reviews parameter transformations and numerical integration for Bayesian procedures in ARMA time series models.

6. Simulation studies.

6.1 Mean and maximum likelihood estimation of the fractional difference parameter f in a FARMA(1,f,0) model given ϕ and $\sigma_a^2=1$.

Remark 6.1: In this section, I run a simulation study to calculate the bias and the mean square error (MSE) of the MeLE and the MLE of the fractional difference parameter f of a FARMA(1,f,0) model. I assume that the value of the autoregressive parameter ϕ is known, and that the noise variance is $\sigma_a^2=1$. The estimators are studied for the values of f = *-0.4, -0.2, 0, 0.2, 0.4*. The autoregressive parameter takes the values ϕ = *-0.9, -0.45, 0, 0.45, 0.9*. The sample size is n = *100*. The number of replicates is R = *400*, and was selected as a compromise between the simulation accuracy and the required computing time. There are *50* evaluations of the likelihood to compute the MeLE by Monte Carlo integration. I selected *50* evaluations after testing with other sizes such as *10*, *100* and *200*. With *50* evaluations, the error of the Monte Carlo integration is negligible, and its increase on the variance estimate is not significant.

Remark 6.2: *Simulation of the FARMA(1,f,0) series.* One objective of the simulation study is to investigate the effect of the parameter ϕ on the estimation of the fractional difference parameter f. In order to focus on this aspect, I generate R = *400* replicates of a sample of size n = *100* from a *N(0,1)*. Each replicate is transformed to a FARMA(1,f,0) series of length n = *100* by using algorithm 3.4, given the values of f and ϕ.

Remark 6.3: *Computation of the MeLE and the MLE.* The MeLE is calculated by Monte Carlo integration. I use ***400*** replicates of a sample of size ***50*** from a ***U(-0.5, 0.5)***. The first FARMA replicate uses the first replicate of ***U(-0.5, 0.5)*** to calculate its MeLE, the second FARMA replicate uses the second replicate, and so on. The random grid search method to calculate the MLE of f is the value of f where the likelihood is maximum among the 50 evaluations done to calculate the MeLE of f (this is a by-product of the MeLE computations). Another MLE is calculated with the subroutine DALL. This gives 2 values for the MLE.

Remark 6.4: *Summary statistics.* Table 1 A gives the bias of the estimators of the fractional difference parameter f and table 1 B gives the MSE of the estimators of f. Figures 8a to 8e give the absolute value of the bias of the estimators of f for each values of the parameter f by the values of the parameter ϕ. Figures 9a to 9e give the MSE of the estimators of f. The labels for the estimators are the following one: The MeLE is labelled as MeLE, the MLE from the random grid search method is labelled as MLE-R, and the MLE obtained with DALL is labelled as MLE-D.

Remark 6.5: *Analysis of the simulation.* The first striking feature is that the estimators are not affected by the different values of the parameter ϕ. For f fixed, the values of the bias and MSE are essentially the same whatever is the value of ϕ. The bias of the MeLE is smaller than that of the MLE-R when $f = -0.2, 0, .2$, that is when f is not too close to the boundary $f = \pm 0.5$. The MSE of the MeLE and the MLE-R are close to each another except that the MSE of the MeLE is uniformly smaller. In conclusion, both estimators of f are very close to each other in terms of MSE, and the MeLE is more biased when f is far from the value 0. A similar situation occurred in chapter 2 in the estimation of the parameter θ of an MA(1) process. When DALL

is used to compute the MLE, it can be seen that it performs very well at $f = 0$. This is because the value of $f = 0$ is used as the initial estimate of the parameter in the algorithm. The first evaluation of the likelihood is then done at the true value of the parameter, where the MLE should be. Consequently DALL algorithm does not need to move very far to find the maximum. This also shows how the choice of the starting values has an effect on the performance of the algorithm DALL.

6.2 Comparison of methods to calculate the MeLE and MLE of the parameters of a FARMA model.

In this section, I run simulations to compare various methods to calculate the mean and maximum likelihood estimators of the parameters of a FARMA model. The following 6 estimators are compared:

RME: MeLE; exact likelihood calculations; Monte Carlo integration with 50 evaluations of the likelihood function.

RML: The MLE out of the 50 evaluations of the exact likelihood function in calculating RME.

PML: MLE; exact likelihood calculations; Powell's minimization algorithm.

DML: MLE; exact likelihood calculations; DALL maximization algorithm.

RIE: MeLE; Li-Mcleod (1986) approximate likelihood calculations; Monte Carlo integration with 50 evaluations of the likelihood function.

RIL: The MLE out of the 50 evaluations of the approximate likelihood function in calculating RIE.

The estimators are put in a convenient table format as follow:

MLE computation method	Grid search (of size 50)	Grid search (of size 50)	Powell's	DALL's
Likelihood Function (LF)	Exact LF	Approximate LF	Exact LF	Exact LF
MLE	RML	RIL	PML	DML
MeLE	RME	RIE		

Remark 6.6: The simulations use $R = 400$ replicates of a FARMA(p,f,q) series with sample size $n = 100$. The 6 estimators are calculated on every replicate. I use two FARMA(p,f,q) models. The first model is a FARMA(1,f,0) model where the parameters are $f = -0.4, -0.2, 0, 0.2, 0.4$ and $\phi = -0.9, -0.45, 0, 0.45, 0.9$. The second model is a FARMA(0,f,1) model where the parameters are $f = -0.4, -0.2, 0, 0.2, 0.4$ and $\theta = -0.9, -0.45, 0, 0.45, 0.9$. This gives a total of 25 sets of parameter values, of which 16 have at least one parameter close to the boundary. I selected these parameters values because I suspected from the previous results, that I have obtained, that it would put the MeLE in the least favourable position.

Remark 6.7: *Likelihood calculations.* The simulation assumes that the noise variance is unknown, in which case the concentrated likelihood function (proposition 4.1,ii) has to be calculated. The concentrated likelihood is calculated with the exact algorithm 3.5. The method of Li and McLeod (1986) for approximate calculations of the likelihood is also used. Professor McLeod gave me the computer programs to compute the approximate likelihood.

Remark 6.8: *The summary statistics.* There are two sets of tables. Tables 2 contains the summary statistics for the simulation on the FARMA(1,f,0) model and Tables 3, those on the FARMA(0,f,1) model. Tables 2 A and 3 A give the number of replicates, the number of replicates where Powell's minimization algorithm was successful, and the number of replicates where DALL maximization algorithm was successful. Powell's and DALL algorithms are not successful when the algorithms fail to converge for various reasons. Without going into too much detail, the most common reason is that one of the algorithm evaluates the likelihood function in a region close to the boundary of the parameter space. This may cause numerical problems if the likelihood function is evaluated at a parameter point outside the admissible region. Tables 2 B and 3 B give the bias of the estimators. Tables 2 C and 3 C give the mean square error (MSE) of the estimators for each parameters. The averages are taken over the successful replicates. The last line of tables 2 B and 3 B gives the average of the norm of the bias vector. The norm is defined as the Euclidian distance of the bias vector from the zero vector. The last two lines of table 2 C and 3 C give the average mean square error (AMSE) for the two parameters f and, ϕ or θ. The estimator is closer (in a Bayesian sense) to the true value in terms of MSE when the AMSE is smaller.

Main results from the simulation.

Comparison 6.9: *RME VS RIE, exact and approximate likelihood calculations for the computation of the MeLE.* The MSE of the two estimators are very close to each another. The AMSE of the estimators are smaller when the exact likelihood calculations are used. However, the approximate likelihood calculations are faster because they do not require calculation of the ACF of the FARMA model, which is particularly time consuming when an AR component is present. The recommendation is to use exact likelihood calculations for smaller MSE and to use approximate likelihood calculations to save on computing time.

Comparison 6.10: *RML VS RIL, exact and approximate likelihood calculations for the computation of the MLE with a random grid search.* The MSE of the estimators are usually smaller with the exact likelihood calculations. The AMSE of the estimator that uses the exact likelihood function are also smaller. The recommendation is to use exact likelihood calculations for smaller MSE and to use approximate likelihood calculations to save on computing time.

Remark 6.11: I found that Powell's minimization algorithm and DALL maximization algorithm are numerically unstable when they are used to compute the MLE with the approximate likelihood calculations of Li and McLeod (1986).

Comparison 6.12: *RML VS PML VS DML, a comparison of optimization algorithms to calculate the MLE.* DALL maximization algorithm gives the smallest AMSE in the estimation of the parameters f and ϕ in a FARMA(1,f,0) model. However, table 2 A shows that DALL has a poor success rates in some cases. For example, there are only 52 successful replicates at $f = -0.4$,

ϕ = *-0.9*. DALL does not work well when the true parameters are too close to the boundary in the FARMA(1,f,0) simulation. This is not the case for Powell's algorithm. In the FARMA(0,f,1) model, it is the random grid search method (RML) that gives the smallest AMSE for both parameters.

Comparison 6.13: *RIE VS RIL, MeLE against MLE with approximate likelihood calculations.* There are only 3 cases where the MLE has a smaller MSE than the MeLE. The 3 cases are for the estimation of f at $(f=-0.4,\ \phi=0.45)$ in the FARMA(1,f,0) model, and at $(f=-0.4,\ \theta=-0.45)$ and $(f=-0.4,\ \theta=0)$ in the FARMA(0,f,1) model. The recommendation is to use the MeLE if the approximate likelihood calculations are used.

Comparison 6.14: *RME VS RML, exact likelihood calculations, MeLE against the MLE obtained with a random grid search.* There are only 2 cases where the MLE has a smaller MSE than the MeLE. The two cases are for the estimation of the parameter f at $(f=-0.4,\ \phi=0.45)$ and at $(f=-0.4,\ \theta=-0.45)$. The recommendation is to use the MeLE over the MLE obtained with a random grid search method.

Comparison 6.15: *RME VS PML, exact likelihood calculations, MeLE against the MLE obtained with Powell's minimization algorithm.* There are 25 combinations of parameter values, 2 parameters in each simulation, and 2 kind of models. This give a total of 100 cases where the MSE of an estimator is calculated. There are 51 cases where the MSE of the MeLE is smaller than that of the MLE. Most of the cases are when at least one of the parameter is not too close to the boundary of the parameter space. For example the MSE of the MeLE of f in a FARMA(1,f,0) model is smaller than the MSE of the MLE when ϕ = *-0.45, 0, .45* and

$f = -0.2, 0, 0.2, 0.4$. Whenever $f = -0.4$ or $\phi = -0.9$ or $\phi = 0.9$ the MSE of the MLE of f is smaller. The AMSE of the MeLE are smaller than the AMSE of the MLE for all the parameters when Powell's algorithm is used. The recommendation is to use the MeLE over the MLE obtained with Powell's algorithm when the true values of the parameters are sufficiently inside the parameter space. Otherwise, the MLE should be used.

Comparison 6.16: *RME VS DML, exact likelihood calculations, MeLE against the MLE obtained with DALL maximization algorithm.* The MSE of the MeLE is smaller than the MSE of the MLE in 38 cases out of 100. It has to be remembered, however, that DALL did not perform well in the FARMA(1,f,0) model in the following 4 combinations of the parameters: $(f=-0.4, \phi=-0.9)$, $(f=-0.2, \phi=-0.9)$, $(f=0, \phi=-0.9)$ and $(f=0.4, \phi=0.9)$. Nevertheless, the two estimators compete well. The AMSE of the MeLE are smaller than those of the MLE. These two estimators are usually the best two estimators in terms of MSE. The MLE performs better than the MeLE in more cases because the majority of the parameter combinations are close to the boundary. In light of the previous comparisons, the recommendation is to use the MeLE if the parameters are sufficiently inside the parameter space.

Remark 6.17: *Practical advantages of the MeLE over the MLE.* I found that the MeLE is simpler to compute than the MLE. It uses Monte Carlo integration instead of a numerical optimization method. Monte Carlo integration is easy to use and to implement. For this method it is necessary to calculate the likelihood at many points in the parameter space and then to average them. A numerical optimization method is more complicated to put into computer code. Moreover, a numerical optimization method, such as Powell's or DALL, need starting values. These values may strongly influence the results. In theory, the starting values should be close

to the true values but in practice, the starting values are set to a predetermined point, say the origin for example, because the true values are unknown. Also, in the case of the estimation of the FARMA parameters, the optimization has to done within the parameter space and this requires a constrained optimization. A second advantage of the MeLE computation is that it gives an estimate of the variance of the estimator as a simple by-product. The computation for the variance is done at the same time as the estimates are computed. A final advantage is that, on the average, the MSE of the MeLE is smaller than that of the MLE. This is particularly advantageous when the parameter values can be anywhere in the parameter space. One is obviously not restricted to use either MLE or MeLE. For example, the MeLE can be used as the starting values for the optimization methods. Similarly, the MeLE can be improved with a second Monte-Carlo integration where the importance function (Chapter 1, Definition 2.22) is the normal approximation to the likelihood provided in Chapter 1, Proposition 3.16.i.

6.3 The estimate of the covariance matrix of the MeLE in the FARMA(1,f,0) model.

In this section, I run a simulation to compare two estimators of the variance of the MeLE of the parameters f and ϕ in the FARMA(1,f,0) model.

Remark 6.18: *The simulation.* The simulation uses $R = 400$ replicates of a FARMA(1,f,0) series with sample size $n = 100$. The parameters are $f = -0.4, -0.2, 0, 0.2, 0.4$ and $\phi = -0.9, -0.45, 0, 0.45, 0.9$. The MeLE is calculated with exact likelihood calculations and Monte Carlo integration (RME of the previous simulations).

Remark 6.19: *The estimates of the variance of the MeLE.* I propose two estimators for the covariance matrix of the MeLE, The first estimator is the likelihood covariance matrix of the MeLE, $Cov_L(\tilde{\beta})$ from definition 4.4, which is computed simultaneously with the MeLE. The second estimator is the inverse of the Fisher information matrix evaluated at the MeLE, $I^{-1}(\tilde{\beta})$. Li and McLeod (1980) have derived the Fisher information matrix for the general FARMA model. The covariance between the estimated ARMA parameters and the fractional difference parameter is in general a complicated expression but for the simple case of a FARMA(1,f,0) model the Fisher information matrix per observation is:

$$I(f,\phi) = \begin{bmatrix} \frac{\pi^2}{6} & \frac{-\log(1-\phi)}{\phi} \\ \frac{-\log(1-\phi)}{\phi} & \frac{1}{(1-\phi^2)} \end{bmatrix}.$$

The Fisher information matrix has to be calculated and inverted after the computation of the parameters.

Remark 6.20: *Reference values for the estimates of the covariance matrix of the MeLE.* The simulation provides two extra estimates for the variance of the MeLE that are not available in a practical application. The first reference is the variance of the MeLE over the replicates, denoted by $Var(\tilde{\beta})$. $Var(\tilde{\beta})$ is to be considered to be the true sampling variance of the MeLE. The second reference is the inverse of the Fisher information matrix evaluated at the true value of the parameter, denoted by $\Gamma^{1}(\beta)$. $\Gamma^{1}(\beta)$ is the asymptotic covariance matrix of the MLE.

Remark 6.21: *Analysis of the simulation, part 1: the MeLE of the parameters.* Table 4 A gives the mean values and the MSE of the MeLE over the replicates. The bias and MSE's are of the same magnitude as in the FARMA(1,f,0) simulation and do not require further comments.

Remark 6.22: *Analysis of the simulation, part 2: the estimates of the covariance matrix of the MeLE of the parameters.* Tables 4 B and 4 C give the the variances of the MeLE of ϕ and f respectively. $I^{-1}(\tilde{\beta})$ and $Var_L(\tilde{\beta})$ are their respective averages over the replicates. $Var_L(\tilde{\beta})$ correspond to the diagonal elements of $Cov_L(\tilde{\beta})$. The main results are: 1) $I^{-1}(\tilde{\beta})$ is close to $I^{-1}(\beta)$ whenever the bias in the estimate is small. 2) $I^{-1}(\tilde{\beta})$ and $Var_L(\tilde{\beta})$ give similar estimates. 3) $Var_L(\tilde{\beta})$ is closer to $Var(\tilde{\beta})$ than $I^{-1}(\tilde{\beta})$ is, in the sense that the average of the squared differences is smaller for $Var_L(\tilde{\beta})$. 4) $I^{-1}(\tilde{\beta})$ and $I^{-1}(\beta)$ display similar patterns, and $Var(\tilde{\beta})$ and $Var_L(\tilde{\beta})$ display similar patterns. The recommendation is to use $Var_L(\tilde{\beta})$ over $I^{-1}(\tilde{\beta})$ because it is closer to $Var(\tilde{\beta})$. $Var_L(\tilde{\beta})$ is also easier to calculate than $I^{-1}(\tilde{\beta})$ for every FARMA model.

7. Conclusions.

In this chapter, I have illustrated the FARMA process. I have shown how to calculate the exact likelihood function of the parameters of a FARMA process and how to simulate such processes. I have investigated two estimators based on the likelihood function: the maximum and mean likelihood estimators. I have given an algorithm to integrate over the stationary and invertible region of the ARMA process to calculate the MeLE of the FARMA parameters. Moreover, I have conducted simulations to study the performance of the estimators in combination with various algorithms to calculate them. I found that the empirical results generally confirm what is to be expected from the theory: The MeLE has a smaller average mean square error than the MLE. Moreover, the simple method of Monte Carlo integration has proven to be very successful for computing the MeLE and its covariance matrix.

8. References.

Abramowitz, M. and Stegun, A. (1968), **Handbook of Mathematical Functions**, United States Bureau of the Census, Washington.

Anderson,O.D. (1975), "The Recursive Nature of the Stationarity and Invertibility Restraints on the Parameters of a Mixed Autoregressive-Moving Average Processes", **Biometrika** 62, p. 704-707.

Ansley, C.F. (1979), "An algorithm for the exact likelihood of a mixed autoregressive-moving average process", **Biometrika** 66, p. 59-65.

Barndorff-Nielsen, O. and Schou, G. (1973), "On the Parametrization of Autoregressive Models by Partial Autocorrelations", **J. Multivariate Anal.** 3, p. 408-419.

Box, G.P. and Jenkins, G.M. (1976), **Time Series Analysis forecasting and control**, Holden-Day, Oakland.

Carlin, J.B., Dempster, A.P. and Jonas, A.B. (1985), "On models and Methods for Bayesian Time Series Analysis", **Journal of Econometrics** 30, p. 67-90.

Daniels, H.E. (1956), "The Approximate Distribution of Serial Correlation Coefficients", **Biometrika** 43, p. 169-185.

Gardner, J.B., Harvey, A.C. and Phillips, G.D.A. (1980), "Algorithm AS154, An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering". **Appl. Statist.** 29, p. 311-322.

Geweke, J. and Porter-Hudak, S. (1983), "The Estimation and Application of Long Memory Time Series Models". **J. Time Series Analysis** 4, p. 221-238.

Granger, C.W.J. and Joyeux, R. (1980), "An Introduction to Long-Memory Time Series Models and Fractional Differencing", **Journal of Time Series Analysis** 1, p. 15-29.

Haslett, J. and Raftery, A.E. (1989), "Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource", **Appl. Statist.** 38, p. 1-50.

Hosking, J.R.M. (1981), "Fractional Differencing", **Biometrika** 68, p. 165-76.

Hosking, J.M.R. (1984), "Modelling Persistence in Hydrological Time Series Using Fractional Differencing", **Water Resources Research** 20, p. 1898-1908.

Hurst, H.E. (1951), "Long Term Storage Capacity of Reservoirs", **Trans. Amer. Soc. Civil Engrs.** 116, p. 770-799.

Hurst, H.E. (1956), "Methods of Using Long-Term Storage in Reservoirs". **Proc. Inst. Civil Engrs.** 1, p. 519-543.

Ishiguro, M. and Akaike, H. (1989), "DALL: Davidon's Algorithm for Log Likelihood Maximization - A FORTRAN subroutine for statistical model builder", **The Institute of Statistical Mathematics**, Computer Science Monographs #25, Tokyo.

Janacek, G.J. (1982), "Determining the Degree of Differencing for Time Series Via The Log Spectrum", **J. Time Series Analysis** 3, p. 177-183.

Jones, M.C.(1987). Randomly Choosing Parameters for the Stationarity and Invertibility Region of Autoregressive-Moving Average Models, **Appl. Statist.** 36, 134-138.

Li, W.K. (1981), "Topics in Time Series Modelling", Unpublished Ph.d. thesis, **University of Western Ontario**.

Li, W.K. and McLeod, A.I. (1986), "Fractional time series modelling", **Biometrika** 73, p. 217-21.

Mandelbrot, B.B. and Wallis, J.R. (1969a), "Computers Experiments with Fractional Gaussian Noises", **Water Resour. Res.** 5, p.228-267.

Mandelbrot, B.B. and Wallis, J.R. (1969b), "Some Long-Run Properties of Geophysical Records", **Water Resour. Res.** 5, p.321-340.

Marriott,J.M. and Smith, A.F.M. (1992). Reparametrization Aspects of Numerical Bayesian Methodology for Autoregressive Moving-Average Models, **Journal of Time Series Analysis** 13, 327-343.

Mélard, G. (1984), "Algorithm AS197. A fast Algorithm for the Exact Likelihood of Autoregressive-moving Average Models", **Appl. Statist.** 33, p. 104-114.

McLeod, A.I. (1975), "Derivation of the Theoretical Autocovariance Function of Autoregressive-Moving Average Time Series", **Appl. Statist.** 27, p. 255-256. Correction: Vol. 26 (1977), p. 194.

McLeod, A.I. and Hipel, K.W. (1978a), "Preservation of the Rescaled Adjusted Range", **Water Resources Research** 14, p. 491-518.

McLeod, A.I. and Hipel, K.W. (1978b), "Simulation Procedures for Box-Jenkins Models", **Water Resources Research** 14, p. 969-975.

Monahan,J.F. (1984), A Note of Enforcing Stationarity in Autoregressive-Moving Average Models, **Biometrika** 71, 402-404.

Piccolo, D. (1982), "The Size of the Stationarity and Invertibility Region of an Autoregressive-Moving Average Process", **Journal of Time Series Analysis** 3, p. 245-247.

Powell, M.J.D. (1964), "An efficient method for finding the minimum of a function of several variables without calculating derivatives", **The Computer Journal** 7, p. 155-162

Ramsey, F.L. (1974), "Characterization of the Partial Autocorrelation Function", **Ann. Statist.** 2, p. 1296-1301.

Wallis, J.R. and Matalas, N.C. (1970), "Small Sample Properties of H and K, Estimators of the Hurst Coefficient h", **Water Resour. Res.** 6, p. 1583-1594.

Wise, J. (1956), "Stationarity Conditions for Stochastic Processes of the Autoregressive and Moving-Average Type", **Biometrika** 43, p. 215-219.

TABLE 1 A: FARMA(1,f,0); Bias of the estimators of f given ϕ and $\sigma_a^2=1$.

f	ϕ	$\bar{f}-f$	$\hat{f}^1-f$	$\hat{f}^2-f$
-.40	-.90	.0282	-.00180	.0145
-.40	-.45	.0282	-.00215	.0120
-.40	.00	.0280	-.00227	.0132
-.40	.45	.0275	-.00276	.0150
-.40	.90	.0251	-.00348	.0136
-.20	-.90	-.00517	-.0162	.0405
-.20	-.45	-.00520	-.0168	.0411
-.20	.00	-.00528	-.0171	.0384
-.20	.45	-.00537	-.0172	.0388
-.20	.90	-.00546	-.0177	.0386
.00	-.90	-.00594	-.0163	-.00374
.00	-.45	-.00585	-.0164	-.00578
.00	.00	-.00592	-.0162	-.00629
.00	.45	-.00615	-.0164	-.00569
.00	.90	-.00684	-.0170	-.00561
.20	-.90	-.0110	-.0178	-.0529
.20	-.45	-.0108	-.0173	-.0566
.20	.00	-.0109	-.0176	-.0548
.20	.45	-.0110	-.0185	-.0538
.20	.90	-.0108	-.0186	-.0546
.40	-.90	-.0299	-.0248	-.0424
.40	-.45	-.0299	-.0247	-.0402
.40	.00	-.0306	-.0262	-.0442
.40	.45	-.0313	-.0270	-.0443
.40	.90	-.0295	-.0261	-.0413

$\bar{f}$: The MeLE.
$\hat{f}^1$: The MLE obtained with a reandom grid search (MLE-R).
$\hat{f}^2$: The MLE obtained with DALL (MLE-D)
The variances for the estimates are given by ***(MSE / 400)***.

TABLE 1 B: FARMA(1,f,0); MSE of the estimators of f given ϕ and $\sigma_a^2=1$.

f	ϕ	$\bar{f}$	$\hat{f}^1$	$\hat{f}^2$
-.40	-.90	.00439	.00536	.0109
-.40	-.45	.00438	.00540	.0106
-.40	.00	.00435	.00547	.0102
-.40	.45	.00429	.00531	.0106
-.40	.90	.00396	.00491	.00974
-.20	-.90	.00764	.00854	.0117
-.20	-.45	.00760	.00855	.0117
-.20	.00	.00755	.00821	.0117
-.20	.45	.00744	.00827	.0115
-.20	.90	.00698	.00768	.0108
.00	-.90	.00788	.00803	.00341
.00	-.45	.00785	.00813	.00380
.00	.00	.00781	.00803	.00348
.00	.45	.00772	.00792	.00385
.00	.90	.00721	.00752	.00302
.20	-.90	.00687	.00726	.0112
.20	-.45	.00686	.00728	.0107
.20	.00	.00687	.00732	.0104
.20	.45	.00685	.00728	.0105
.20	.90	.00643	.00680	.0102
.40	-.90	.00433	.00480	.00897
.40	-.45	.00439	.00468	.00819
.40	.00	.00448	.00489	.00893
.40	.45	.00464	.00512	.00918
.40	.90	.00424	.00479	.00818

$\bar{f}$: The MeLE .
$\hat{f}^1$: The MLE obtained with a reandom grid search (MLE-R).
$\hat{f}^2$: The MLE obtained with DALL (MLE-D)
The variances for the estimates of the MSE are approximately given by $(2\ MSE^2 / 400)$.

TABLE 2 A: FARMA(1,f,0); Actual number of replicates.

f	ϕ	n	R	Powell	DALL
-.40	-.90	100	400	399	52
-.40	-.45	100	400	400	400
-.40	.00	100	400	400	398
-.40	.45	100	400	400	399
-.40	.90	100	400	399	400
-.20	-.90	100	400	400	79
-.20	-.45	100	400	400	400
-.20	.00	100	400	400	400
-.20	.45	100	400	400	400
-.20	.90	100	400	400	400
.00	-.90	100	400	400	272
.00	-.45	100	400	400	400
.00	.00	100	400	400	400
.00	.45	100	400	400	400
.00	.90	100	400	400	396
.20	-.90	100	400	400	378
.20	-.45	100	400	400	400
.20	.00	100	400	400	400
.20	.45	100	400	399	398
.20	.90	100	400	400	308
.40	-.90	100	400	400	387
.40	-.45	100	400	400	395
.40	.00	100	400	400	393
.40	.45	100	400	400	396
.40	.90	100	400	400	225

n: sample size,
R: number of replicates,
Powell: number of successful replicates with Powell's minimization algorithm,
DALL: number of successful replicates with DALL maximization algorithm.

TABLE 2 B: FARMA(1,f,0); Bias of the estimators.

	RME	RML	PML	DML	RIE	RIL
-.40	.0855	.0798	.0220	-.0225	.0720	.0641
-.90	.0358	.0438	.0145	.0139	.0315	.0310
-.40	.0536	.0367	-.0135	-.0053	.0498	.0291
-.45	-.0153	-.0024	.0133	.0083	-.0131	-.0038
-.40	.0813	.0575	-.0202	-.0004	.0698	.0391
.00	-.0733	-.0522	.0008	-.0108	-.0649	-.0387
-.40	.1387	.0888	-.0230	.0381	.1334	.0901
.45	-.1489	-.1064	-.0165	-.0658	-.1476	-.1142
-.40	.3901	.3250	.1997	.4219	.4509	.4164
.90	-.3748	-.3271	-.2274	-.4182	-.4397	-.4152
-.20	-.0016	-.0074	-.0321	.0003	-.0059	-.0092
-.90	.0396	.0480	.0220	.0743	.0365	.0346
-.20	-.0110	-.0212	-.0626	-.0251	-.0215	-.0299
-.45	.0230	.0305	.0407	.0214	.0291	.0261
-.20	-.0118	-.0289	-.1024	-.0376	-.0177	-.0446
.00	.0134	.0257	.0736	.0241	.0188	.0397
-.20	.0369	-.0035	-.1470	-.0324	.0387	-.0063
.45	-.0549	-.0253	.0913	-.0014	-.0573	-.0202
-.20	.2590	.2090	.0358	.3355	.3286	.3131
.90	-.2513	-.2151	-.0728	-.3274	-.3212	-.3121
.00	-.0102	-.0088	-.0536	-.0033	-.0107	-.0110
-.90	.0390	.0489	.0244	.0378	.0365	.0344
.00	-.0264	-.0298	-.0803	-.0268	-.0284	-.0377
-.45	.0380	.0431	.0520	.0229	.0343	.0367
.00	-.0511	-.0547	-.1613	-.0477	-.0505	-.0598
.00	.0554	.0480	.1326	.0344	.0531	.0503
.00	-.0438	-.0790	-.2250	-.0526	-.0213	-.0608
.45	.0184	.0419	.1555	.0160	-.0060	.0266
.00	.1664	.1334	.0013	.1809	.1992	.1893
.90	-.1615	-.1454	-.0437	-.1830	-.1944	-.1865

	RME	RML	PML	DML	RIE	RIL
.20	-.0372	-.0398	-.0618	-.0201	-.0160	-.0142
-.90	.0471	.0565	.0252	.0234	.0384	.0367
.20	-.0304	-.0309	-.0906	-.0286	-.0217	-.0205
-.45	.0410	.0452	.0599	.0244	.0318	.0254
.20	-.0897	-.0766	-.2145	-.0548	-.0627	-.0514
.00	.0987	.0738	.1848	.0423	.0683	.0431
.20	-.1127	-.1233	-.3086	-.0807	-.0666	-.0664
.45	.0810	.0812	.2313	.0494	.0392	.0354
.20	.0637	.0514	-.0237	.0603	.0645	.0665
.90	-.0845	-.0881	-.0273	-.0941	-.0781	-.0760
.40	-.0769	-.0796	-.0751	-.0300	-.0671	-.0631
-.90	.0777	.0859	.0271	.0230	.0560	.0534
.40	-.0628	-.0604	-.1099	-.0367	-.0472	-.0357
-.45	.0552	.0492	.0752	.0295	.0621	.0503
.40	-.1472	-.1365	-.3789	-.0708	-.1121	-.0932
.00	.1659	.1388	.3475	.0650	.1239	.0954
.40	-.1737	-.1588	-.3833	-.0855	-.1551	-.1285
.45	.1571	.1320	.3003	.0580	.1384	.1116
.40	-.0603	-.0639	-.0708	.0047	-.0971	-.0933
.90	-.0363	-.0460	-.0037	-.0399	-.0283	-.0286
average norm:	.1265	.1149	.1494	.0993	.1252	.1139

The variances for the estimates are given by ***(MSE / 400)***.

TABLE 2 C: FARMA(1,f,0); Mean square error of the estimators.

	RME	RML	PML	DML	RIE	RIL
-.40	.0232	.0252	.0080	.0042	.0196	.0212
-.90	.0065	.0157	.0032	.0050	.0062	.0074
-.40	.0092	.0118	.0072	.0068	.0085	.0091
-.45	.0143	.0210	.0115	.0109	.0142	.0202
-.40	.0133	.0160	.0100	.0103	.0120	.0129
.00	.0220	.0290	.0178	.0190	.0215	.0293
-.40	.0301	.0278	.0164	.0198	.0290	.0276
.45	.0407	.0441	.0245	.0327	.0403	.0452
-.40	.1759	.1760	.1062	.2326	.2320	.2637
.90	.1754	.1976	.1291	.2475	.2325	.2788
-.20	.0181	.0222	.0096	.0094	.0166	.0217
-.90	.0064	.0171	.0035	.0104	.0058	.0072
-.20	.0111	.0167	.0174	.0123	.0120	.0174
-.45	.0161	.0249	.0166	.0138	.0170	.0224
-.20	.0131	.0225	.0338	.0219	.0148	.0255
.00	.0228	.0346	.0368	.0284	.0232	.0394
-.20	.0151	.0317	.0519	.0343	.0202	.0354
.45	.0221	.0443	.0422	.0396	.0249	.0423
-.20	.0884	.0960	.0253	.1586	.1318	.1670
.90	.0875	.1029	.0233	.1563	.1313	.1715
.00	.0175	.0205	.0122	.0062	.0172	.0231
-.90	.0063	.0174	.0037	.0045	.0061	.0074
.00	.0155	.0215	.0226	.0124	.0185	.0238
-.45	.0187	.0265	.0194	.0144	.0217	.0283
.00	.0225	.0348	.0688	.0271	.0260	.0452
.00	.0331	.0486	.0719	.0344	.0351	.0572
.00	.0193	.0508	.0920	.0390	.0242	.0585
.45	.0238	.0597	.0621	.0401	.0249	.0583
.00	.0436	.0489	.0171	.0620	.0552	.0755
.90	.0409	.0553	.0089	.0693	.0569	.0731

	RME	RML	PML	DML	RIE	RIL
.20	.0165	.0207	.0134	.0067	.0146	.0183
-.90	.0073	.0181	.0039	.0035	.0062	.0072
.20	.0129	.0172	.0246	.0108	.0150	.0205
-.45	.0200	.0274	.0212	.0141	.0199	.0259
.20	.0320	.0476	.1169	.0295	.0341	.0583
.00	.0486	.0669	.1173	.0384	.0489	.0770
.20	.0326	.0633	.1314	.0344	.0257	.0625
.45	.0292	.0599	.0834	.0355	.0241	.0580
.20	.0163	.0242	.0130	.0219	.0188	.0286
.90	.0161	.0315	.0048	.0228	.0159	.0207
.40	.0160	.0191	.0133	.0058	.0131	.0138
-.90	.0138	.0242	.0041	.0034	.0114	.0118
.40	.0121	.0147	.0259	.0077	.0096	.0110
-.45	.0266	.0314	.0234	.0136	.0251	.0301
.40	.0521	.0778	.2622	.0229	.0399	.0650
.00	.0809	.1065	.2669	.0391	.0662	.0972
.40	.0487	.0638	.1764	.0272	.0431	.0596
.45	.0473	.0639	.1085	.0301	.0446	.0639
.40	.0156	.0187	.0125	.0069	.0285	.0298
.90	.0091	.0204	.0023	.0087	.0126	.0140
AMSE(f):	.0308	.0396	.0515	.0332	.0352	.0478
AMSE(ϕ):	.0334	.0476	.0444	.0374	.0375	.0518

The variances for the estimates of the MSE are approximately given by $(2\ MSE^2 / 400)$.

TABLE 3 A: FARMA(0,f,1); Actual number of replicates.

f	θ	n	R	Powell	DALL
-.40	-.90	100	400	400	376
-.40	-.45	100	400	400	399
-.40	.00	100	400	400	399
-.40	.45	100	400	400	399
-.40	.90	100	400	400	365
-.20	-.90	100	400	400	397
-.20	-.45	100	400	400	400
-.20	.00	100	400	400	400
-.20	.45	100	400	400	400
-.20	.90	100	400	400	380
.00	-.90	100	400	400	398
.00	-.45	100	400	400	400
.00	.00	100	400	400	400
.00	.45	100	400	400	400
.00	.90	100	400	400	398
.20	-.90	100	400	400	397
.20	-.45	100	400	400	400
.20	.00	100	400	400	400
.20	.45	100	400	400	400
.20	.90	100	400	400	399
.40	-.90	100	400	400	388
.40	-.45	100	400	400	395
.40	.00	100	400	400	388
.40	.45	100	400	400	396
.40	.90	100	400	400	397

n: sample size.
R: number of replicates.
Powell: number of successful replicates with Powell's minimization algorithm.
DALL: number of successful replicates with DALL maximization algorithm.

TABLE 3 B: FARMA(0,f,1); Bias of the estimators.

	RME	RML	PML	DML	RIE	RIL
-.40	.0781	.0621	-.0180	.0045	.0701	.0597
-.90	.0628	.0501	-.0125	-.0042	.0674	.0577
-.40	.0901	.0651	-.0185	.0021	.0744	.0459
-.45	.0800	.0542	-.0088	.0021	.0739	.0504
-.40	.2079	.1402	.0458	.0128	.1805	.0985
.00	.2247	.1548	.0623	.0209	.1914	.1093
-.40	.2008	.1778	.2254	-.0076	.1674	.1251
.45	.1733	.1660	.2220	-.0040	.1466	.1195
-.40	.0663	.0601	.0514	.0056	.0574	.0490
.90	-.0419	-.0307	-.0697	.0096	-.0555	-.0494
-.20	.0258	.0164	-.0436	-.0096	.0264	.0181
-.90	.0459	.0326	-.0159	-.0103	.0527	.0402
-.20	.0269	.0029	-.0566	-.0150	.0235	-.0039
-.45	.0336	.0102	-.0299	-.0089	.0341	.0045
-.20	.0769	.0330	-.0723	-.0148	.0758	.0278
.00	.0874	.0430	-.0494	-.0045	.0906	.0408
-.20	.0830	.0577	-.0071	-.0682	.0619	.0288
.45	.0664	.0574	.0110	-.0661	.0459	.0267
-.20	-.0689	-.0636	-.0193	-.1400	-.0926	-.1060
.90	-.0939	-.0719	-.0531	-.1104	-.1076	-.1003
.00	.0170	.0096	-.0505	-.0106	.0271	.0201
-.90	.0474	.0348	-.0173	-.0103	.0486	.0350
.00	.0123	-.0060	-.0670	-.0167	.0190	-.0060
-.45	.0296	.0029	-.0360	-.0105	.0345	.0043
.00	.0312	-.0019	-.0968	-.0203	.0400	.0148
.00	.0488	.0216	-.0707	-.0098	.0564	.0311
.00	.0070	-.0045	-.1707	-.0686	.0061	.0041
.45	-.0035	-.0027	-.1489	-.0652	-.0051	.0087
.00	-.1840	-.1573	-.0617	-.4035	-.2251	-.2432
.90	-.1749	-.1325	-.0425	-.3691	-.2085	-.2070

	RME	RML	PML	DML	RIE	RIL
.20	.0120	.0071	-.0557	-.0111	.0253	.0215
-.90	.0537	.0401	-.0182	-.0097	.0559	.0442
.20	-.0032	-.0152	-.0744	-.0183	.0280	.0199
-.45	.0131	-.0043	-.0412	-.0121	.0381	.0266
.20	-.0038	-.0073	-.1102	-.0264	.0201	.0214
.00	.0036	.0045	-.0839	-.0164	.0266	.0333
.20	-.0526	-.0519	-.2174	-.0638	-.0381	-.0311
.45	-.0648	-.0502	-.1962	-.0560	-.0450	-.0267
.20	-.2932	-.2641	-.2033	-.5127	-.3459	-.3667
.90	-.2848	-.2362	-.1530	-.4964	-.3319	-.3411
.40	-.0378	-.0357	-.0709	-.0218	-.0317	-.0275
-.90	.0777	.0712	-.0218	-.0102	.0749	.0689
.40	-.0453	-.0412	-.0926	-.0300	-.0271	-.0224
-.45	-.0128	-.0157	-.0545	-.0218	.0039	.0028
.40	-.0584	-.0526	-.1328	-.0432	-.0301	-.0189
.00	-.0507	-.0407	-.1078	-.0356	-.0268	-.0171
.40	-.1021	-.0916	-.2441	-.0694	-.0749	-.0592
.45	-.1227	-.1023	-.2276	-.0609	-.0988	-.0681
.40	-.3417	-.3231	-.5834	-.4272	-.3715	-.3622
.90	-.3434	-.3098	-.5292	-.4240	-.3788	-.3489
average norm:	.1251	.1002	.1458	.1100	.1267	.1046

The variances for the estimates are given by ***(MSE / 400)***.

TABLE 3 C: FARMA(0,f,1); Mean square error of the estimators.

	RME	RML	PML	DML	RIE	RIL
-.40	.0155	.0156	.0059	.0053	.0128	.0143
-.90	.0113	.0115	.0034	.0033	.0124	.0132
-.40	.0161	.0147	.0076	.0076	.0144	.0137
-.45	.0244	.0259	.0124	.0132	.0261	.0291
-.40	.0630	.0774	.0507	.0144	.0543	.0504
.00	.0952	.1249	.0843	.0279	.0808	.0845
-.40	.0533	.0774	.1108	.0143	.0418	.0507
.45	.0446	.0743	.1038	.0220	.0381	.0556
-.40	.0132	.0164	.0144	.0088	.0126	.0148
.90	.0096	.0111	.0105	.0049	.0104	.0127
-.20	.0137	.0179	.0112	.0077	.0150	.0193
-.90	.0079	.0084	.0034	.0035	.0084	.0089
-.20	.0134	.0168	.0164	.0116	.0153	.0199
-.45	.0200	.0231	.0143	.0145	.0197	.0233
-.20	.0266	.0368	.0261	.0205	.0311	.0426
.00	.0453	.0611	.0312	.0320	.0495	.0661
-.20	.0223	.0549	.0582	.0359	.0252	.0563
.45	.0226	.0621	.0718	.0395	.0261	.0637
-.20	.0153	.0273	.0204	.0377	.0201	.0333
.90	.0171	.0206	.0127	.0375	.0204	.0268
.00	.0124	.0156	.0123	.0076	.0147	.0194
-.90	.0085	.0088	.0034	.0035	.0077	.0081
.00	.0147	.0191	.0183	.0110	.0170	.0219
-.45	.0184	.0217	.0144	.0142	.0205	.0254
.00	.0179	.0273	.0311	.0192	.0271	.0388
.00	.0328	.0445	.0309	.0304	.0395	.0551
.00	.0222	.0501	.0670	.0450	.0294	.0644
.45	.0260	.0588	.0688	.0493	.0313	.0726
.00	.0492	.0704	.0387	.1830	.0653	.0980
.90	.0433	.0504	.0230	.1741	.0581	.0835

	RME	RML	PML	DML	RIE	RIL
.20	.0117	.0150	.0123	.0065	.0105	.0144
-.90	.0096	.0098	.0034	.0033	.0098	.0104
.20	.0100	.0154	.0184	.0093	.0127	.0184
-.45	.0149	.0202	.0142	.0135	.0180	.0239
.20	.0125	.0210	.0322	.0151	.0171	.0267
.00	.0233	.0357	.0307	.0260	.0294	.0455
.20	.0225	.0380	.0827	.0354	.0248	.0446
.45	.0308	.0501	.0781	.0463	.0346	.0583
.20	.1054	.1363	.1205	.3124	.1423	.1994
.90	.1032	.1225	.1093	.3073	.1380	.1967
.40	.0082	.0097	.0118	.0045	.0069	.0080
-.90	.0186	.0199	.0033	.0032	.0199	.0209
.40	.0084	.0104	.0177	.0063	.0070	.0090
-.45	.0179	.0238	.0138	.0119	.0188	.0257
.40	.0109	.0129	.0318	.0098	.0069	.0097
.00	.0249	.0317	.0301	.0203	.0228	.0305
.40	.0216	.0251	.0883	.0208	.0171	.0192
.45	.0382	.0456	.0834	.0341	.0365	.0407
.40	.1527	.1816	.4344	.2855	.1784	.2144
.90	.1557	.1802	.4178	.2870	.1842	.2112
AMSE(f):	.0293	.0401	.0536	.0454	.0328	.0449
AMSE(θ):	.0346	.0459	.0509	.0490	.0384	.0517

The variances for the estimates of the MSE are approximately given by $(2\ MSE^2 / 400)$.

TABLE 4 A: FARMA(1,f,0): Average and MSE of the MeLE of f and ϕ in the second simulation.

f	ϕ	$\bar{f}$	$\bar{\phi}$	MSE($\bar{f}$)	MSE($\bar{\phi}$)
-0.4	-0.90	-0.3210	-0.8650	.0205	.0057
-0.4	-0.45	-0.3380	-0.4720	.0106	.0165
-0.4	0.00	-0.3170	-0.0623	.0138	.0209
-0.4	0.45	-0.2570	0.3010	.0311	.0420
-0.4	0.90	-0.0186	0.5360	.1670	.1630
-0.2	-0.90	-0.2060	-0.8620	.0181	.0057
-0.2	-0.45	-0.2040	-0.4290	.0127	.0188
-0.2	0.00	-0.2160	0.0195	.0128	.0240
-0.2	0.45	-0.1660	0.4020	.0153	.0231
-0.2	0.90	0.0430	0.6650	.0821	.0779
0.0	-0.90	-0.0204	-0.8620	.0195	.0057
0.0	-0.45	-0.0307	-0.4090	.0163	.0239
0.0	0.00	-0.0580	0.0681	.0230	.0365
0.0	0.45	-0.0527	0.4800	.0202	.0235
0.0	0.90	0.1620	0.7390	.0432	.0400
0.2	-0.90	0.1660	-0.8550	.0171	.0066
0.2	-0.45	0.1570	-0.4020	.0171	.0258
0.2	0.00	0.0963	0.1130	.0367	.0562
0.2	0.45	0.0706	0.5470	.0373	.0311
0.2	0.90	0.2700	0.8080	.0180	.0169
0.4	-0.90	0.3170	-0.8280	.0175	.0131
0.4	-0.45	0.3310	-0.3760	.0124	.0283
0.4	0.00	0.2350	0.1850	.0583	.0894
0.4	0.45	0.2160	0.6130	.0537	.0481
0.4	0.90	0.3480	0.8440	.0152	.0171

The variances of the averages are given by $(MSE / 400)$.
The variances for the MSE are approximately given by $(2\,MSE^2 / 400)$.

TABLE 4 B: FARMA(1,f,0); Estimates of the variance of the MeLE of ϕ in the second simulation.

f	ϕ	$Var(\bar{\phi})$	$Var_L(\bar{\phi})$	$I^{-1}(\bar{\phi})$	$I^{-1}(\phi)$
-0.4	-0.90	.0044	.0014	.0028	.0020
-0.4	-0.45	.0160	.0095	.0115	.0119
-0.4	0.00	.0170	.0168	.0236	.0255
-0.4	0.45	.0199	.0251	.0434	.0553
-0.4	0.90	.0309	.0518	.0563	.0078
-0.2	-0.90	.0042	.0018	.0028	.0020
-0.2	-0.45	.0184	.0130	.0126	.0119
-0.2	0.00	.0236	.0245	.0271	.0255
-0.2	0.45	.0208	.0351	.0517	.0553
-0.2	0.90	.0228	.0320	.0501	.0078
0.0	-0.90	.0043	.0019	.0028	.0020
0.0	-0.45	.0222	.0153	.0132	.0119
0.0	0.00	.0318	.0315	.0298	.0255
0.0	0.45	.0226	.0390	.0571	.0553
0.0	0.90	.0140	.0169	.0427	.0078
0.2	-0.90	.0046	.0016	.0030	.0020
0.2	-0.45	.0235	.0158	.0133	.0119
0.2	0.00	.0434	.0412	.0328	.0255
0.2	0.45	.0218	.0367	.0585	.0553
0.2	0.90	.0085	.0061	.0305	.0078
0.4	-0.90	.0079	.0015	.0035	.0020
0.4	-0.45	.0229	.0141	.0140	.0119
0.4	0.00	.0552	.0453	.0362	.0255
0.4	0.45	.0214	.0234	.0563	.0553
0.4	0.90	.0140	.0022	.0208	.0078

Note:

$$\frac{1}{25}\sum_{f,\phi}(Var_L(\bar{\phi}) - Var(\bar{\phi}))^2 = 0.73255\times10^{-4}$$

$$\frac{1}{25}\sum_{f,\phi}(I^{-1}(\bar{\phi}) - Var(\bar{\phi}))^2 = 3.55464\times10^{-4}$$

$$\frac{1}{25}\sum_{f,\phi}(I^{-1}(\phi) - Var(\bar{\phi}))^2 = 3.36654\times10^{-4}$$

TABLE 4 C: FARMA(1,f,0); Estimates of the variance of the MeLE of f in the second simulation.

ϕ	f	$Var(\bar{f})$	$Var_L(\bar{f})$	$I^{-1}(\bar{f})$	$I^{-1}(f)$
-0.90	-0.4	.0142	.0041	.0066	.0065
-0.90	-0.2	.0180	.0074	.0066	.0065
-0.90	0.0	.0190	.0068	.0066	.0065
-0.90	0.2	.0159	.0057	.0067	.0065
-0.90	0.4	.0106	.0022	.0068	.0065
-0.45	-0.4	.0068	.0062	.0090	.0091
-0.45	-0.2	.0127	.0101	.0094	.0091
-0.45	0.0	.0154	.0115	.0097	.0091
-0.45	0.2	.0153	.0103	.0098	.0091
-0.45	0.4	.0076	.0053	.0100	.0091
0.00	-0.4	.0069	.0095	.0146	.0155
0.00	-0.2	.0125	.0156	.0169	.0155
0.00	0.0	.0196	.0209	.0191	.0155
0.00	0.2	.0259	.0258	.0218	.0155
0.00	0.4	.0310	.0266	.0260	.0155
0.45	-0.4	.0107	.0189	.0305	.0421
0.45	-0.2	.0141	.0270	.0402	.0421
0.45	0.0	.0174	.0351	.0492	.0421
0.45	0.2	.0206	.0333	.0552	.0421
0.45	0.4	.0199	.0226	.0585	.0421
0.90	-0.4	.0213	.0454	.0535	.0249
0.90	-0.2	.0230	.0342	.0569	.0249
0.90	0.0	.0170	.0213	.0558	.0249
0.90	0.2	.0131	.0100	.0470	.0249
0.90	0.4	.0125	.0032	.0363	.0249

Note:

$$\frac{1}{25}\sum_{f,\phi}(Var_L(\bar{f}) - Var(\bar{f}))^2 = 0.86569\times10^{-4}$$

$$\frac{1}{25}\sum_{f,\phi}(I^{-1}(\bar{f}) - Var(\bar{f}))^2 = 4.32970\times10^{-4}$$

$$\frac{1}{25}\sum_{f,\phi}(I^{-1}(f) - Var(\bar{f}))^2 = 1.88276\times10^{-4}$$

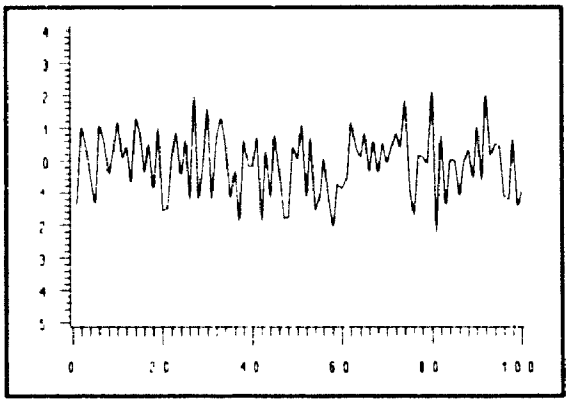

Figure 1: 100 observations from a $N(0,1)$.

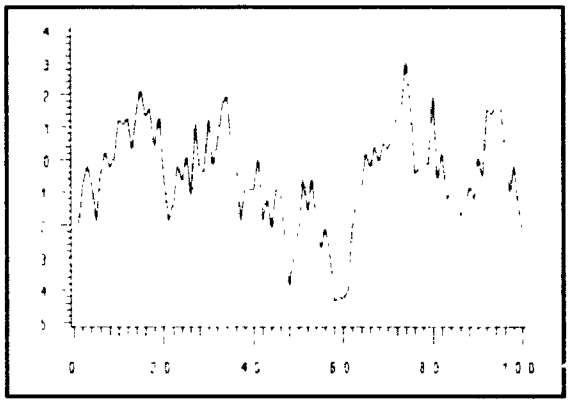

Figure 2: An AR(1) series with $\phi=0.8$.

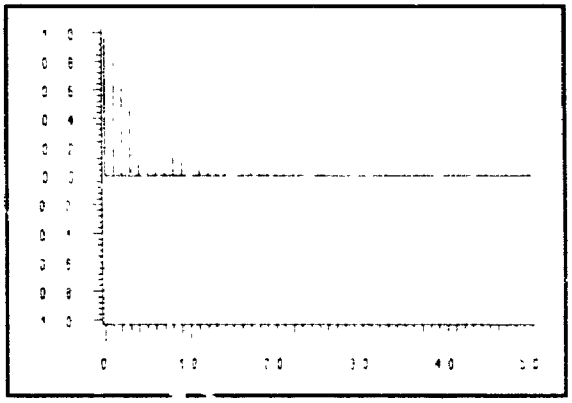

Figure 3: ACF of an AR(1) process with $\phi=0.8$ ($\gamma_0=2.778$).

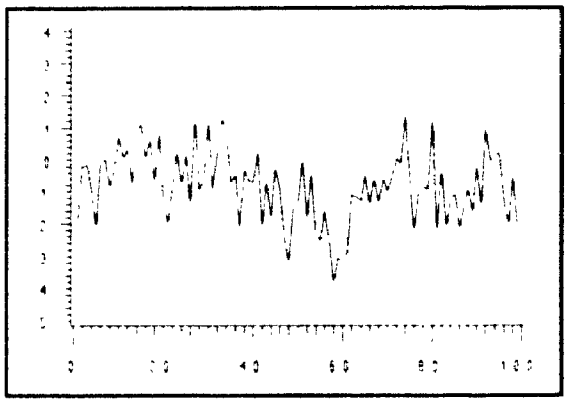

Figure 4: A FARMA(0,f,0) series with $f=0.4$.

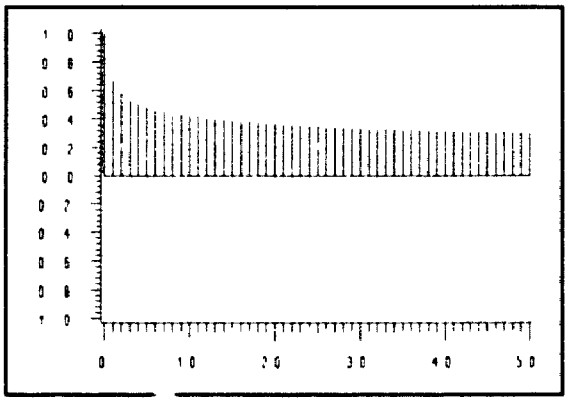

Figure 5: ACF of a FARMA(0,f,0) process with $f=0.4$ ($\gamma_0=2.070$).

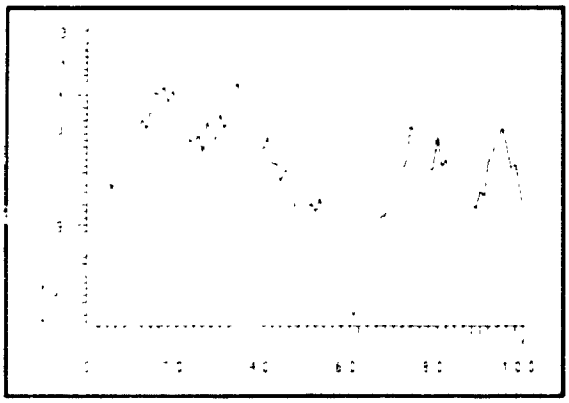

Figure 6: A FARMA(1,f,0) series with $\phi=0.8$ and $f=0.4$.

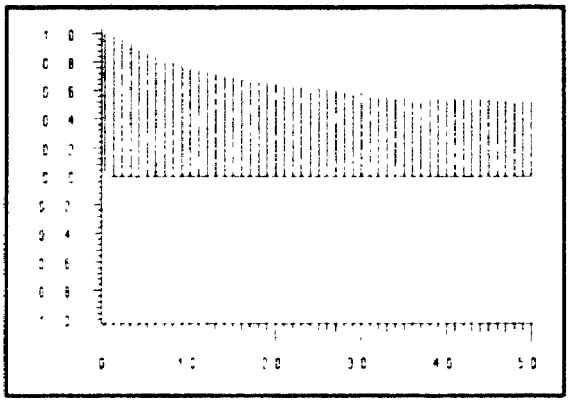

Figure 7: ACF of a FARMA(1,f,0) process with $\phi=0.8$ and $f=0.4$ ($\gamma_0=30.079$).

Figures 8: FARMA(1,f,0) process; Absolute bias of the estimators of f given the values of ϕ and $\sigma_a^2=1$.

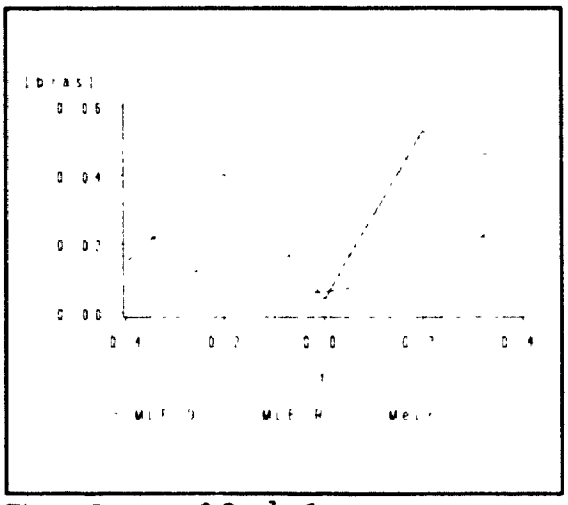

Figure 8a: $\phi=-0.9$, $\sigma_a^2=1$.

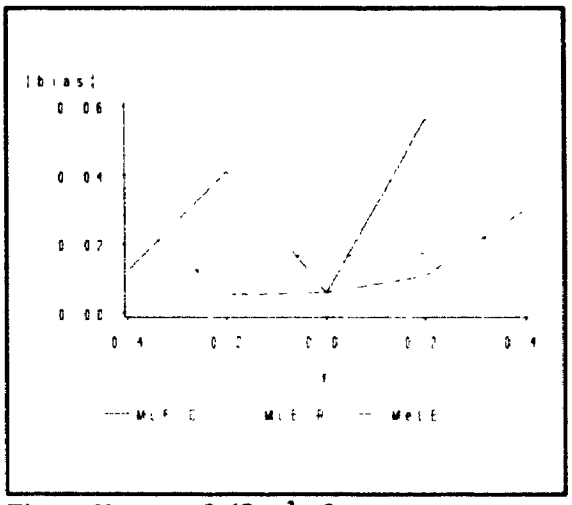

Figure 8b: $\phi=-0.45$, $\sigma_a^2=1$.

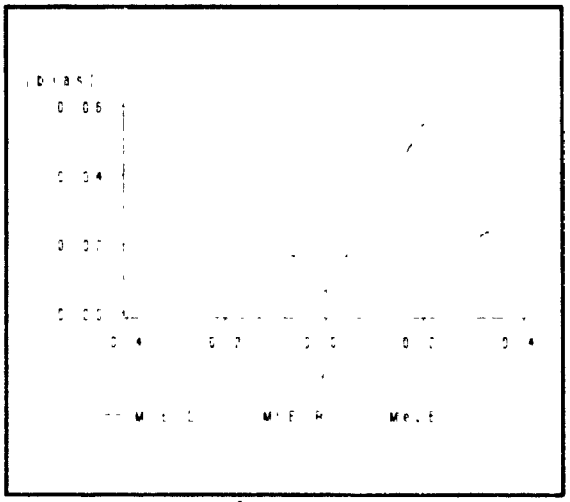

Figure 8c: $\phi=0$, $\sigma_a^2=1$.

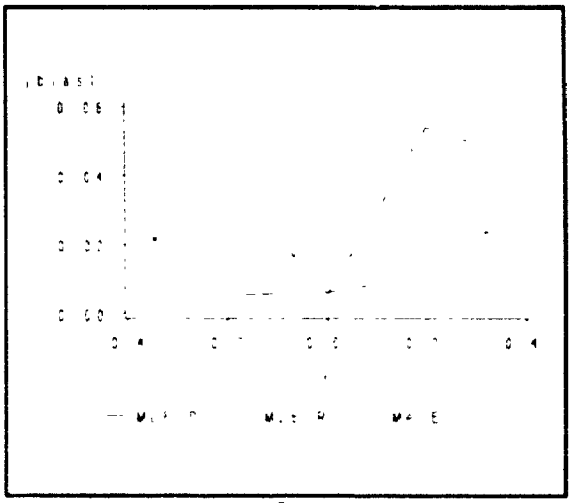

Figure 8d: $\phi=0.45$, $\sigma_a^2=1$.

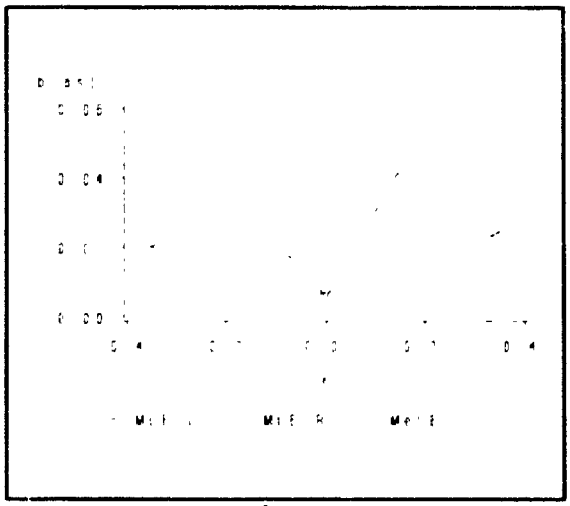

Figure 8e: $\phi=0.9$, $\sigma_a^2=1$.

Figures 9: FARMA(1,f,0) process; Mean Square Error of the estimators of f given the values of ϕ and $\sigma_a^2=1$.

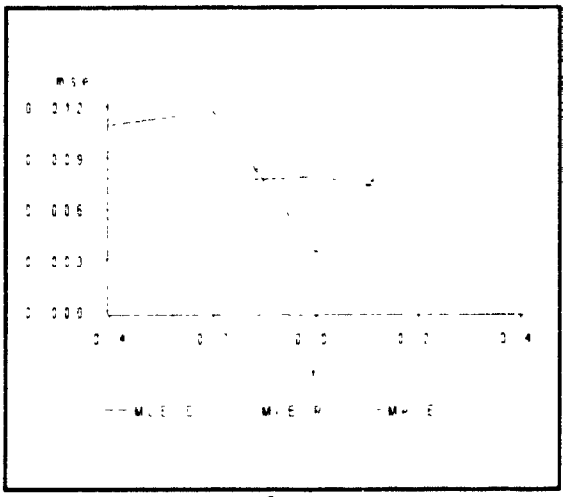

Figure 9a: $\phi=-0.9$, $\sigma_a^2=1$.

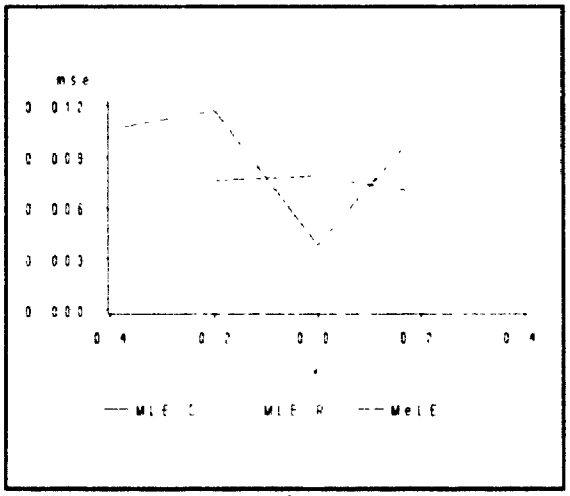

Figure 9b: $\phi=-0.45$, $\sigma_a^2=1$.

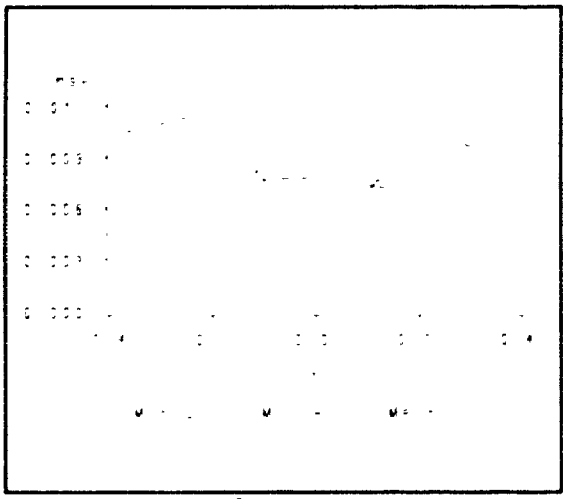

Figure 9c: $\phi=0$, $\sigma_a^2=1$.

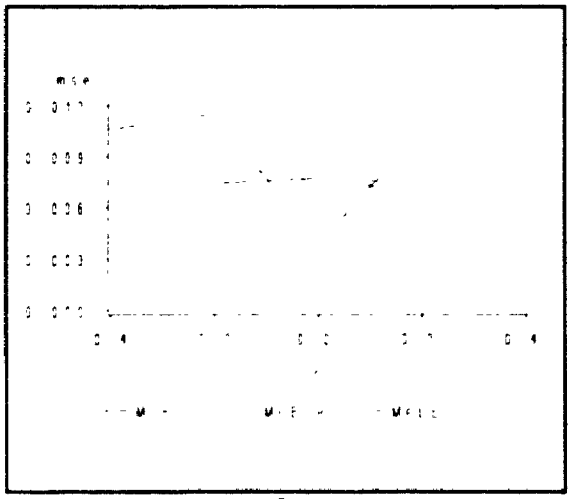

Figure 9d: $\phi=0.45$, $\sigma_a^2=1$.

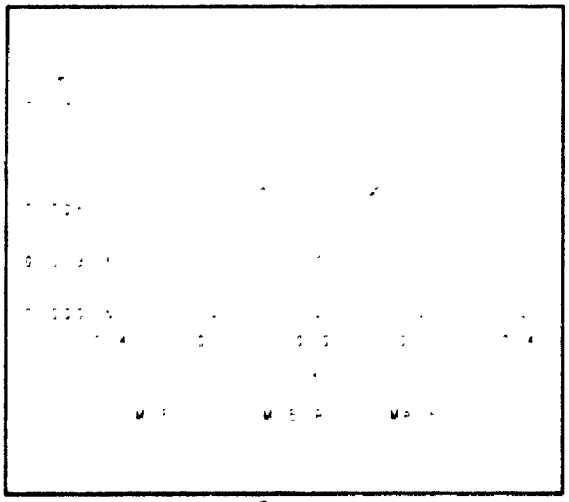

Figure 9e: $\phi=0.9$, $\sigma_a^2=1$.

CHAPTER 4

THE INTEGRATED FRACTIONALLY DIFFERENCED AUTOREGRESSIVE MOVING AVERAGE PROCESS

The Integrated fractionally differenced autoregressive moving average (IFARMA) process is an integrated FARMA process, which generalizes the ARIMA process where the order of differencing is restricted to be an integer. The IFARMA class of models offers the possibility to estimate the degree of difference jointly with the other parameters: the autoregressive parameters, the moving average parameters, the mean, the white noise variance and, the initial conditions.

In this chapter I examine the problem of selecting an appropriate IFARMA model for a given series. I use the minimum AIC and BIC strategy to select a model among a pre-specified group of competing models, which may include stationary and non-stationary models. I solve the problem of defining a likelihood function that unifies both the stationary and non-stationary cases. In the stationary case, the likelihood is well defined. In the non-stationary case, I make additional assumptions about the pre-observation period to obtain a well defined likelihood. Finally, I illustrate my model building strategy with four real series.

Keywords: AIC; BIC; Maximum Likelihood Estimation

1. Introduction.

The fractionally differenced autoregressive moving average (FARMA) class of model was introduced in chapter 3. The FARMA class is a generalisation of the ARMA class where the order of differencing is extended to $(-0.5, 0.5)$. Like the ARIMA processes are obtained by integrating ARMA processes, the integrated fractionally differenced autoregressive moving average (IFARMA) processes are obtained by integrating FARMA processes. An IFARMA process is thus the generalisation of an ARIMA process for which the order of differencing can take any value in $[-0.5, \infty)$.

I organize this chapter as follows. In section 2, I give the definition of a FARMA process with an unknown mean which generalizes the zero mean stationary FARMA process studied in chapter 3. I derive the maximum likelihood estimates of the mean and the white noise variance, and I obtain a well defined likelihood for the remaining parameters.

In section 3, I give the definition for an IFARMA process. I define the likelihood function by having additional assumptions for the pre-observations period. I consider two cases where the pre-observations period is a fixed vector of constants, and where the pre-observations period is a vector of random variables with a diffuse prior. In the first case, I end up with a well defined likelihood function, and in the second case I show that a suitably normed limit of the likelihood function reduces to the likelihood function obtained for the first case. My approach is to model a finite random vector of size n from a multivariate normal distribution with an unknown mean vector and covariance matrix. The mean vector is assumed to satisfy the d-th difference equation, where the first d values (the fixed pre-observations) are unknown. The covariance matrix is

assumed to depend on a set of unknown parameters, which characterizes the covariance matrix of a stationary zero mean FARMA process.

In section 4, I obtain the AIC and BIC of an IFARMA process. In section 5, I describe the model building strategy with IFARMA processes. I use the minimum AIC or BIC strategy to select a model among a pre-specified class of competing models, which may include stationary and non-stationary models. I illustrate the model building strategy with *Series A* and *B* of Box-Jenkins (1976), where I select a stationary FARMA model with an unknown mean for *Series A*, and an ARIMA(0, 1, 0) model for *Series B*. I also apply the model building strategy to the series of *Average Annual Temperature, Central England, 1723-1970*, where I select either a stationary FARMA(2, f, 0) with the first autoregressive parameter constraints to zero, or a stationary FARMA(0, f, 0) model according to the minimum AIC or BIC respectively. The final example is with the series of *Lake Michigan-Huron, highest monthly mean level for each calendar year, 1860-1955*, where the minimum AIC criterion selects a non-stationary IFARMA(2, d+f, 0) model, and the minimum BIC criterion selects a stationary ARMA(1, 1) model.

In section 6, I show the consistency of the likelihood maximization procedure for the estimation of the degree of difference d of an IFARMA(p, d+f, q) process. In section 7, I give the concluding remarks.

My contribution over previously published work is the extension of the FARMA class of model to the IFARMA class. I am also able to estimate the degree of differencing by the maximum of a suitably well defined likelihood function. Consequently, my model building strategy does not require any differencing of the data by the user prior to the estimation of the parameters. The

given definition of the likelihood function of an IFARMA model generalizes the likelihood of a stationary FARMA model, and generalizes the definition of the likelihood function of an ARIMA model by including the unknown means. I show that the likelihood that I derive is the full likelihood whereas that of Box-Jenkins (1976) is a marginal likelihood in the sense of Kalbfleish and Sprott (1970) (this was already pointed out by Kohn and Ansley, 1986). Finally, the definition of the AIC and BIC come out naturally from the definition of the full likelihood function. This approach is a major improvement over the arbitrary an ad hoc definition of the AIC of an ARIMA model provided by Ozaki (1977).

2. The fractionally differenced autoregressive moving average process with an unknown mean.

Definition 2.1: The process $\{ w_t, t = 0,\pm1,\pm2,... \}$ is a fractionally differenced autoregressive moving average process, FARMA(p,f,q), with an unknown mean μ if it satisfies:

$$w_t = \mu + \frac{\theta(B)}{\phi(B)\nabla^f} a_t. \tag{2.1}$$

B is the backshift operator $(B^i y_t = y_{t-i})$.

$$\begin{aligned} \phi(B) &= 1-\phi_1 B-...-\phi_p B^p, \\ \theta(B) &= 1-\theta_1 B-...-\theta_q B^q, \\ \nabla^f &= (1-B)^f = \sum_{k=0}^{\infty} \binom{f}{k} (-1)^k B^k \end{aligned} \tag{2.2}$$

are the autoregressive polynomial, the moving average polynomial, and the difference operator. $\{ a_t \}$ is an independently and identically distributed normal sequence with mean zero and variance σ_a^2, $N(0, \sigma_a^2)$.

Proposition 2.2: The process is stationary and invertible if f is in the open interval $(-0.5,0.5)$, and if $\phi(B)$ and $\theta(B)$ have their roots outside the unit circle with no common roots.

Remark 2.3: The FARMA(0,f,0) process with $f = -0.5$ is stationary but it is not invertible. Hosking (1981) has derived the properties of the FARMA(0,f,0) process with parameter $f = -0.5$. The process has variance $\gamma_0 = 4/\pi$, autocorrelation function $\rho_k = -1/(4k^2-1)$, and partial autocorrelation $\phi_{k,k} = -1/(2k+1)$.

Proposition 2.4: The joint density of $w = (w_1,\ldots,w_n)$ from a FARMA(p,f,q) process with parameters μ, σ_a^2, $\beta = (\phi_1,\ldots,\phi_p,f,\theta_1,\ldots,\theta_q)$ is:

$$g_W(w_1,\ldots,w_n;\mu,\sigma_a^2,\beta) = (2\pi)^{-\frac{n}{2}}\left(\sigma_a^2\right)^{-\frac{n}{2}}\left(\prod_{t=1}^{n} v_t^*\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma_a^2}\sum_{t=1}^{n}\frac{(w_t-m_t)^2}{v_t^*}\right\} \tag{2.3}$$

where:

$$\begin{aligned} m_1 &= \mu, \\ m_t &= \mu + \sum_{j=1}^{t-1}\phi_{t-1,j}\,(w_{t-j}-\mu), \\ v_1 &= \sigma_a^2\gamma_0, \\ v_t &= (1-\phi_{t-1,t-1}^2)v_{t-1}, \\ v_t^* &= v_t/\sigma_a^2. \end{aligned} \tag{2.4}$$

In (2.4) the regression coefficients $\phi_{t-1,j}$, $j = 1, \ldots, t-1$, $t = 1, \ldots, n$, are given by the Durbin-Levinson recursion (Chapter 3, Algorithm 3.2), and γ_t, $t = 0,\ldots,n-1$, are the autocovariances of a zero-mean FARMA(p, f, q) process with $\sigma_a^2 = 1$. Note also that the conditional variances v_t, $t \geq 1$, do not depend on the data but only on the parameters β and σ_a^2.

Proof: The joint density of the random vector w is:

$$g_w(w_1,\ldots,w_n) = g_w(w_1)\, g_w(w_2|w_1) \ldots g_w(w_n|w_{n-1},\ldots,w_1). \tag{2.5}$$

the marginal density of w_t and the conditional densities of $\{w_t|w_{t-1},\ldots,w_1,\ t=2,\ldots,n\}$ are:

$$\begin{aligned} w_1 &\sim N(\mu, \sigma_a^2\gamma_0) \\ (w_t|w_{t-1},\ldots,w_1) &\sim N(m_t, v_t). \end{aligned} \tag{2.6}$$

Proposition 2.5: i) Given data $w = (w_1,\ldots,w_n)$, the maximum likelihood estimate (MLE) of the mean is:

$$\begin{aligned} \hat{\mu} &= \frac{\hat{\mu}_n}{\hat{\mu}_d}, \\ \hat{\mu}_n &= \sum_{t=1}^{n} \frac{(w_t - \sum_{j=1}^{t-1} \phi_{t-1,j} w_{t-j})\ (1 - \sum_{j=1}^{t-1} \phi_{t-1,j})}{v_t^*}, \\ \hat{\mu}_d &= \sum_{t=1}^{n} \frac{(1 - \sum_{j=1}^{t-1} \phi_{t-1,j})^2}{v_t^*}\ . \end{aligned} \tag{2.7}$$

ii) The MLE of the white noise variance is:

$$\hat{\sigma}_a^2 = \frac{1}{n}\left[\sum_{t=1}^{n} \frac{(w_t - \sum_{j=1}^{t-1} \phi_{t-1,j} w_{t-j})^2}{v_t^*} - \hat{\mu}^2 \sum_{t=1}^{n} \frac{(1 - \sum_{j=1}^{t-1} \phi_{t-1,j})^2}{v_t^*}\right]. \tag{2.8}$$

iii) The concentrated likelihood function for the parameters $\beta = (\phi_1,\ldots,\phi_p,f,\theta_1,\ldots,\theta_q)$ is:

$$l(\beta \mid w) \propto \left(\hat{\sigma}_a^2\right)^{-\frac{n}{2}} \left(\prod_{t=1}^{n} v_t^*\right)^{-\frac{1}{2}}. \tag{2.9}$$

Proof: The proof is straightforward.

3. The integrated fractionally differenced autoregressive moving average process.

Definition 3.1: If d is a non-negative integer then $\{ z_t \}$ is an integrated fractionally differenced autoregressive moving average, IFARMA(p, d+f, q), process if:

$$w_t = \nabla^d z_t \tag{3.1}$$

is a zero mean FARMA(p, f, q) process.

Remark 3.2: The process is stationary if and only if $d = 0$, in which case it reduces to a FARMA(p, f, q) process.

Remark 3.3: Except when $d = 0$, the mean of $\{ z_t \}$ is not determined. When $d > 0$, an arbitrary polynomial trend of degree $d-1$ can be added to $\{ z_t \}$ without violating the second order properties of $\{ w_t \}$, but not those of $\{ z_t \}$. Consequently, it will be convenient to assume that $\mathcal{E}(w_t) = 0$, in which case additional assumptions will be needed to determine the joint distribution of a random vector $z = (z_1,\ldots,z_n)$ from an IFARMA(p, d+f, q) process.

Remark 3.4: In the non-stationary case with $d > 0$, the random variables are:

$$z_t = w_t + \sum_{i=1}^{d} \binom{d}{i} (-1)^{i+1} z_{t-i}. \tag{3.2}$$

The proposed approach is to assume that the joint density of $w = (w_1,...,w_n)$ is from a stationary FARMA(p,f,q) process with the mean assumed to be equal to zero. To obtain the joint distribution of $z = (z_1,...,z_n)$, some hypotheses have to be made about the pre-observations period $\alpha = (z_{1-d},...,z_0)$. Two important situations may occur: a) α is a fixed vector of constants, b) α is a random vector.

Theorem 3.5: Given data $z = (z_1,...,z_n)$, then under the assumption that $\alpha = (z_{1-d},...,z_0)$ is a fixed vector of constants:

i) The MLE of $\alpha = (z_{1-d},...,z_0)$ are the solutions of:

$$\nabla^d \hat{z}_t = \hat{w}_t, \tag{3.3}$$

where $(\hat{w}_1,...,\hat{w}_d)$ are the backcasts of the first d unobserved values of the vector $w = (w_1,...,w_n)$, and where $\hat{z}_t = z_t$, $1 \le t \le n$, $w_t = \nabla^d z_t$, $d+1 \le t \le n$.

ii) The MLE of σ_a^2 is:

$$\hat{\sigma}_a^2 = \frac{1}{n} \sum_{t=1}^{n-d} \frac{(w_{t+d} - \sum_{j=1}^{t-1} \phi_{t-1,j} w_{t+d-j})^2}{v_t^*}, \tag{3.4}$$

and does not depend on the MLE of α.

iii) The concentrated likelihood for the remaining parameters $\beta = (\phi_1,\ldots,\phi_p,f,\theta_1,\ldots,\theta_q)$ is:

$$l(\beta \mid z) \propto \left(\hat{\sigma}_a^2\right)^{-\frac{n}{2}} \left(\prod_{t=1}^{n} v_t^*\right)^{-\frac{1}{2}}, \tag{3.5}$$

and does not depend on the MLE of α.

Remark 3.6: The MLE of the white noise variance and the concentrated likelihood do not involve the estimates of the pre-observations period. The pre-observations $\alpha = (z_{1-d},\ldots,z_0)$ determine the first order moments of the IFARMA process, and the parameters σ_a^2, $\beta = (\phi_1,\ldots,\phi_p,f,\theta_1,\ldots,\theta_q)$ determine the second order moments of the IFARMA process, once the process is properly differenced.

Proof of proposition 3.5: To keep it simple, I first consider the case $d=1$. First write

$$w_t = z_t - z_{t-1},$$

where z_0 is unknown and fixed, and where w_t is the FARMA(p, f, q) process with mean $\mu = 0$. The Jacobian of the transformation is one and the joint density of $(w_1,\ldots,w_n)$ is that of a stationary random vector of size n from a FARMA(p, f, q) process. I now derive the MLE of z_0 and show how to get a likelihood for the remaining parameters that is independent of the MLE of z_0. The idea is to write:

$$g_W(w_1,w_2,\ldots,w_n;\beta,\sigma_a^2,\alpha) = g_W(w_1;\beta,\sigma_a^2,\alpha\,|w_2,\ldots,w_n)\; g_W(w_2,\ldots,w_n;\beta,\sigma_a^2)$$

$$= g_W(z_1-z_0;\beta,\sigma_a^2,\alpha\,|z_2-z_1,\ldots,z_n-z_{n-1})\; g_W(z_2-z_1,\ldots,z_n-z_{n-1};\beta,\sigma_a^2).$$

Here $g_W(z_2-z_1,\ldots,z_n-z_{n-1};\beta,\sigma_a^2,\alpha)$ is the density of a random vector of size $n-1$ from a zero-mean FARMA(p,f,q) process. Let Σ be the covariance matrix of a random vector of size n from a FARMA(p, f, q) process with mean zero and white noise variance $\sigma_a^2 = 1$, and Σ_{11} be the element in the first row and first column of it. The conditional density $g_W(z_1-z_0;\beta,\sigma_a^2,\alpha\,|z_2-z_1,\ldots,z_n-z_{n-1})$ is normal with mean and variance given by:

$$\mathcal{E}(w_1\,|\,w_2,\ldots,w_n) = \hat{w}_1 = \Sigma_{11}\Sigma_{22}^{-1}(w_2,\ldots,w_n)'$$

$$Var(w_1\,|\,w_2,\ldots,w_n) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

The joint density of $(z_1,\ldots,z_n)$ is therefore:

$$g_Z(z_1,\ldots,z_n\,|\,\beta,\sigma_a^2,z_0) = (2\pi)^{-\frac{n}{2}}\left(\sigma_a^2\right)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}\exp-\frac{1}{2\sigma_a^2}\left\{\frac{(z_1-z_0-\hat{w}_1)^2}{\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}+(w_2,\ldots,w_n)\Sigma_{22}^{-1}\begin{pmatrix}w_2\\ \ldots\\ w_n\end{pmatrix}\right\},$$

and defines the likelihood for the parameters β, σ_a^2 and fixed initial pre-observation $\alpha = z_0$ (remember: $w_t = z_t - z_{t-1}$, $t \geq 2$).

The MLE of the fixed pre-observation z_0 is:

$$\hat{z}_0 = z_1 - \hat{w}_1.$$

Consequently, the likelihood for the remaining parameters σ_a^2, $\beta = (\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q)$ is:

$$l(\beta,\sigma_a^2 \mid z_1,\ldots,z_n) = (2\pi)^{-\frac{n}{2}}\left(\sigma_a^2\right)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_a^2}(w_2,\ldots,w_n)\Sigma_{22}^{-1}\begin{pmatrix} w_2 \\ \ldots \\ w_n \end{pmatrix}\right\},$$

and does not depend on the estimate of z_0.

For the general case, $d>0$, the transformation

$$w_t = \nabla^d z_t \tag{3.6}$$

has unit Jacobian. In (3.6), $(w_1,\ldots,w_d)$ involves the unknown pre-observations $\alpha = (z_{1-d},\ldots,z_0)$. Let Σ_{11} be the first d rows and d columns of Σ. The density of $(w_1,\ldots,w_n)$ is:

$$g_W(w_1,\ldots,w_d,w_{d+1},\ldots,w_n;\beta,\sigma_a^2,\alpha) = g_W(w_1,\ldots w_d;\beta,\sigma_a^2,\alpha \mid w_{d+1},\ldots,w_n)\, g_W(w_{d+1},\ldots,w_n;\beta,\sigma_a^2). \tag{3.7}$$

Here $g_W(w_{d+1},...,w_n;\beta,\sigma_a^2)$ is the density of a random vector of size $n-d$ from a zero-mean FARMA(p,f,q) process, while the conditional density $g_W(w_1,...,w_d;\beta,\sigma_a^2,\alpha \mid w_{d+1},...,w_n)$ is that of normal random vector with mean and variance given by:

$$\begin{aligned} \mathcal{E}(w_1,...,w_d \mid w_{d+1},...,w_n) &= (\hat{w}_1,...,\hat{w}_d)' = \Sigma_{11}\Sigma_{22}^{-1}(w_{d+1},...,w_n)' \\ Var(w_1,...,w_d \mid w_{d+1},...,w_n) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned} \tag{3.8}$$

The MLE of the initial conditions are then provided by (3.3), and the likelihood function for the parameters σ_a^2, $\beta = (\phi_1,...,\phi_p,f,\theta_1,...,\theta_q)$ is:

$$l(\beta,\sigma_a^2 \mid z_1,...,z_n) = (2\pi)^{-\frac{n}{2}}(\sigma_a^2)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma_a^2}(w_{d+1},...,w_n)\Sigma_{22}^{-1}\begin{pmatrix} w_{d+1} \\ ... \\ w_n \end{pmatrix}\right\}, \tag{3.9}$$

which is independent of the estimate of the initial conditions.

Results ii) and iii) now follow from (3.9).

Remark 3.7: The MLE of $(w_1,...,w_d)$ in (3.3) are the backcast values of $(w_1,...,w_d)$, i.e. they are the forecasts of the past values of the series. The zero-mean stationary FARMA(p,f,q) process is time reversible because the distribution is Normal with a symmetric covariance matrix.

Remark 3.8: *Comparison with the Box-Jenkins approach.* In the ARIMA case, Box and Jenkins (1970) defines the likelihood of the parameters as the second term on the right hand side of equation (3.7). Box and Jenkins's likelihood is therefore a marginal likelihood in the sense of Kalbfleish and Sprott (1970) because, given the parameters $(\beta, \sigma_a^2, \alpha)$, the distribution of the sample $(z_1,\ldots,z_n;\beta,\sigma_a^2,\alpha)$ satisfies the following two conditions:

M1: The transformation (3.6) is non-singular and factorizes the distribution of the sample into two distribution functions (3.7).

M2: The first factor on the right side of (3.7) contains no available information concerning β and σ_a^2 in the absence of knowledge of α.

Theorem 3.5 shows that under the assumption that α is a fixed vector of constants there is no need to use a marginal likelihood because a full likelihood for the parameters β, σ_a^2 and α is available, out of which it is possible to estimate α. The likelihood in (3.5) is the concentrated likelihood for the vector of parameters β. It is necessary to have a full likelihood to obtain an information criterion that allows comparison between models with different numbers of pre-observation d.

I now consider the case where α is a random vector with a diffuse prior, and show that a properly normed limit of the likelihood reduces to the case where α is a fixed vector of constants.

Definition 3.9: The vector α is said to have a diffuse prior distribution if

$$\alpha = (z_{1-d},\ldots,z_0) \sim N(0,kI_d), \quad k\to\infty. \tag{3.10}$$

Theorem 3.10: Under the assumption that α is a random vector with a diffuse prior, the likelihood function for the parameters $\beta = (\phi_1,\ldots,\phi_p,f,\theta_1,\ldots,\theta_q)$ is the zero-function but:

$$\lim_{k\to\infty} k^{\frac{d}{2}}\, l(\beta\,|z) \propto \left(\hat{\sigma}_a^2\right)^{-\frac{n}{2}}\left(\prod_{t=1}^{n} v_t^*\right)^{-\frac{1}{2}}, \tag{3.11}$$

where the MLE of σ_a^2 is given by (3.4). (Recall that the v_t^* only depend on the parameter β).

Proof: Consider the transformation:

$$\begin{aligned} z_t &= z_t \qquad 1-d \le t \le 0, \\ w_t &= \nabla^d z_t \quad 1 \le t \le n \end{aligned} \tag{3.12}$$

with unit Jacobian. The joint distribution of $(z_{1-d},\ldots,z_0,w_1,\ldots,w_n)$ is:

$$g(z_{1-d},\ldots,z_0,w_1,\ldots,w_n;k,\beta,\sigma_a^2) = g(z_{1-d},\ldots,z_0;k)\,g(w_1,\ldots,w_n;\beta,\sigma_a^2\,|z_{1-d},\ldots,z_0). \tag{3.13}$$

Here $g(w_1,\ldots,w_n;\beta,\sigma_a^2\,|z_{1-d},\ldots,z_0)$ is the density of a random vector of size n from a zero-mean FARMA(p, f, q) process, and $g(z_{1-d},\ldots,z_0;k)$ is the Normal diffuse prior. As $k \to \infty$, the Normal diffuse prior is the zero-function and the log-likelihood does not exist. However,

$$\lim_{k\to\infty} k^{\frac{d}{2}}\, g(z_{1-d},\ldots,z_0;k) = (2\pi)^{-\frac{d}{2}},$$

and

$$\lim_{k\to\infty} k^{\frac{d}{2}}\, g(z_{1-d},\ldots,z_0,w_1,\ldots,w_n;k,\beta,\sigma_a^2) \propto g(w_1,\ldots,w_n;\beta,\sigma_a^2|z_{1-d},\ldots,z_0),$$

which reduces to the case where α is a vector of fixed constants (Theorem 3.5).

Remark 3.11: The parameters σ_a^2 and β determine the second order moments of the IFARMA(p, d+f, q) process and the likelihood for these parameters is the same under the two cases a) α is a fixed vector of constants, and b) α is a random vector. The difference between the two cases is that the first order moments of the IFARMA process are well determined in the fixed constants case. The fixed constants case is, in a way, a conditional approach in the sense that it models the non-stationary IFARMA(p, d+f, q) processes which go through α. The vector of fixed constants α plays, in the non-stationary case, the same role as the mean μ in the stationary case.

Remark 3.12: Kohn and Ansley have considered the diffuse prior case in the context of ARIMA modelling. They define the likelihood function as the density of $(w_{d+1},\ldots,w_n|z_1,\ldots,z_d)$ and have called it a marginal likelihood in the sense of Kalbfleish and Sprott (1970). In theorem 3.10 the likelihood function is defined as the density of $(w_1,\ldots,w_n|z_{1-d},\ldots,z_0)$. The prior distribution is then on the pre-observations period and the first d values of w_t are simply missing observations. The likelihood in theorem 3.10 is a marginal likelihood because (3.12) is a non-singular transformation such that the distribution of the sample factorizes into the distributions (3.13), where the first factor on the right side of (3.13) contains no available information concerning β and σ_a^2.

4. Information Criteria.

Definition 4.1: The Akaike (1974) Information Criterion (AIC) is defined as:

$$AIC = -2 \log (\text{maximized likelihood}) + 2 (\text{number of estimated parameters}).$$

Definition 4.2: The Bayesian Information Criterion (BIC), Schwarz (1978), is:

$$BIC = -2 \log(\text{maximized likelihood}) + \log n (\text{number of estimated parameters}).$$

Remark 4.3: Both AIC and BIC were designed to measure the fit of a model taking into account the number of parameters used. A model is judged better than another when its AIC or BIC is smaller.

Proposition 4.4: *The AIC and BIC of an IFARMA(p, d+f, q) process.* The estimated parameters are the p autoregressive parameters, the fractional difference parameter, the q moving average parameters, the white noise variance, the mean in the stationary case, and the d pre-observations values in the non-stationary case. The AIC and the BIC are:

$$\begin{aligned} AIC &= n \log \hat{\sigma}_a^2 + \sum_{t=1}^{n} \log v_t^* + 2(p + \delta_{0f} + q + 1 + \delta_{0\mu} + d) \\ BIC &= n \log \hat{\sigma}_a^2 + \sum_{t=1}^{n} \log v_t^* + \log n \ (p + \delta_{0f} + q + 1 + \delta_{0\mu} + d). \end{aligned} \tag{4.1}$$

where $\delta_{0f}=1$ if a fractional difference parameter or if d have to be estimated, 0 otherwise, and $\delta_{0\mu}=1$ if the mean has to be estimated and 0 otherwise.

Remark 4.5: In comparison to get the minimum for different choices of the number of parameters, the sample size n is always the same in the AIC and BIC given by (4.1). Also, the value of the determinant of the covariance matrix Σ of $(w_1,\ldots,w_n)$ is that of a random vector of size n from a stationary FARMA(p, f, q) process with $\sigma_a^2 = 1$, i.e.

$$\log|\Sigma| = \sum_{t=1}^{n} \log v_t^*.$$

Remark 4.6: The current practice in ARIMA modelling is to define the AIC as:

$$AIC_A = (n-d)\log\tilde{\sigma}_a^2 + \sum_{t=1}^{n-d} \log v_t^* + 2(p+q),$$

$$\tilde{\sigma}_a^2 = \frac{n}{n-d}\hat{\sigma}_a^2.$$

AIC_A is only suitable when all the competing models belong to the ARIMA(p, d, q) class with d fixed since the marginal likelihood is only defined for the ϕ and θ parameters. Ozaki (1977) has tried to generalize AIC_A to allow for different values of d. Ozaki's AIC is:

$$AIC_O = n\log\tilde{\sigma}_a^2 + \sum_{t=1}^{n-d} \log v_t^* + 2\frac{n}{n-d}(p+q),$$

and was derived by inflating the estimate of the white noise variance and the penalty factor to take into account the lose of the first d observations after differencing. On the other hand, the AIC in (4.1) comes out naturally from the definition of the likelihood function of an IFARMA(p,d+f,q) process and takes into account the unknown value of d.

5. Model building with IFARMA processes.

Remark 5.1: *Model building with ARIMA processes.* The original model building strategy with ARIMA processes was developed by Box and Jenkins (1976) and further improved over time by various researchers (e.g. Hipel, McLeod and Lennox, 1977). The basic strategy consists of three stages: identification, estimation, and diagnostic checks. The purpose of the identification stage is to determine if differencing is needed to produce a stationary ARMA(p, q) series, and to determine the order of p and q. At the estimation stage the parameters of the model are estimated by a suitable statistical method. Diagnostic checks are devised to test the model adequacy. The three stages are repeated until a model passes the diagnostic checks. Another way to select an appropriate model is to specify a variety of competing models and to choose the model with the smallest AIC or BIC. This is known as the minimum AIC or BIC strategy. The minimum AIC or BIC model also has to pass the diagnostic checks.

Remark 5.2: *Model building with IFARMA processes.* The model building strategy with IFARMA processes is similar to the ARIMA model building strategy. The strategy is to select a class of competing ARIMA(p, d, q) models and to add the corresponding IFARMA(p, d+f, q) models, where the degree of differencing will be estimated by a suitable statistical method. The minimum AIC or BIC strategy is used to select the appropriate model.

Remark 5.3: *Maximum likelihood estimation in IFARMA processes.* The algorithm to estimate the degree of differencing of an IFARMA(p, d+f, q) process takes regular differences of the original series z_t. The resulting series, say w_t, is either assumed to be from a stationary zero-mean FARMA(p, f, q) process if the number of regular differences is greater than zero, or to be

has the following asymptotic normal distribution:

$$\sqrt{n}\,(\gamma^*(0) - \gamma(0)) \sim N(0,v)$$

$$v = 2\sum_{k=-q}^{q} \gamma(k)^2$$

$$\gamma(0) = (1+\theta_1^2+\ldots+\theta_q^2)\sigma^2$$

$$\gamma(i) = (\theta_i+\theta_1\theta_{i+1}+\ldots+\theta_{q-i}\theta_q)\sigma^2,\ i = 1,\ldots,q$$

$$\gamma(i) = 0,\ i > q.$$

Proof: This follows from Theorem 7.3.2 of Brockwell and Davis (1987).

Proof of Theorem 6.1:

i) Apply Corollary 6.4 with

$$X_t = \nabla^b Z_t = \sum_{j=0}^{b} \binom{b}{j} (-1)^j Z_{t-j}$$

$$\theta_j = \binom{b}{j} (-1)^j$$

$$\gamma(0) = \sigma^2 \sum_{j=0}^{b} \binom{b}{j}^2 = \sigma^2 \binom{2b}{b}.$$

ii) Let

$$X_t = \nabla^{b+1} Z_t = \nabla^b (Z_t - Z_{t-1}) = \nabla^b a_t,$$

once differenced series. This means that taking one regular difference was an over-adjustment. On the other hand, the best model, according to the AIC or BIC, is the IFARMA(0, d+f , 0) with the degree of difference estimated at 0.39987. This means that the number of regular difference is zero and that the fractional difference parameter is estimated at 0.39987. This is a stationary model. Moreover, the diagnostic checks indicate that the model is appropriate. The Dickey-Fuller (1979) unit-root test:

$$H_0: \ \rho = 1$$

$$H_1: \ \rho < 1$$

for the model:

$$y_t = \mu + \rho y_{t-1} + e_t,$$

gives estimated values $\hat{\rho} = 0.5723$, $\hat{\tau} = -7.2457$, and consequently the test is significant at the 1% level (i.e. reject the null hypothesis). (The results for the Dickey-Fuller unit-root test are calculated from MHTS, McLeod, 1993). Figure 1b provides a graph of the values of

$$2 \log l(d+f|z_1,\ldots,z_{197})$$

for $d + f = -0.5$ *to* 2 *by* 0.1. The maximum, i.e. an approximation to the MLE, is at $d + f = 0.4$. Figure 1b also show that the likelihood is a continuous function of $d + f$, because there is no abrupt change at either $d + f = 0.5$ or $d + f = 1.5$.

Example 5.5: *Series B of Box-Jenkins.* Figure 4 displays Series B of Box-Jenkins (1976). This is clearly a non-stationary series. The SACF of the once differenced series is given in figure 5. This suggests that an ARIMA(0,1,0) model could be adequate. Box-Jenkins also suggest the ARIMA (0,1,1) model. The corresponding models in the IFARMA class are the IFARMA(0, d+f, 0) and the IFARMA(0, d+f, 1). Table 2 gives the estimated values of the parameters, and the values of the AIC and BIC. The IFARMA(1, d+f, 0) model is also included for comparison purpose. The estimated values for the different models yield the conclusion that the series is an integrated white noise sequence of order 1. The best model in term of AIC is the ARIMA(0,1,1) but a difference in AIC-values less than 2 may usually be ignored in the interest of a simpler model and consequently, the ARIMA(0, 1, 0) could be preferred over the ARIMA(0, 1, 1). The ARIMA(0,1,0) model is best according to the BIC. The IFARMA(0, d+f, 0) model has an estimated value of $\boldsymbol{d+f = 1.05994}$. This gives an estimated fractional difference value of $\boldsymbol{f = 0.05994}$ conditional on $\boldsymbol{d = 1}$. Consequently, this estimated value of $\boldsymbol{f = 0.05994}$ is not significantly different from zero, and so the BIC criterion selects an ARIMA(0,1,0) model. Diagnostic checks applied to these models suggest that the models are all acceptable. The Dickey-Fuller unit-root test has estimated values $\hat{\rho} = 0.9991$, $\hat{\tau} = -0.2805$, and the test is not significant at the 10% level (i.e. accept the null hypothesis).

Example 5.6: *Average Annual Temperature, Central England, 1723-1970.* Figure 6 displays the series, and figure 7 displays the SACF. The Dickey-Fuller unit-root test has estimated values $\hat{\rho} = 0.1487$, $\hat{\tau} = -13.4891$, and the test is significant at the 1% level (i.e. reject the null hypothesis). Jiménez (1988) has fitted a stationary AR(2) model to this series according to a minimum AIC criteria. Table 3 gives the estimated values of the parameters and the values of

the AIC and BIC for a competing class of models. The minimum AIC model is the stationary IFARMA(2, d+f, 0) with $\phi_1=0$, and the minimum BIC model is the stationary FARMA(0, f, 0) model.

Example 5.7: *Lake Michigan-Huron, highest monthly mean level for each calendar year, 1860-1955.* Figure 8 displays the series, and figure 9 displays the SACF. The Dickey-Fuller unit-root test has estimated values $\hat{\rho} = 0.8508$, $t = -2.8814$, the test is significant at the 10% level but not at the 5% level, and consequently there is insufficient data to either reject or accept the null hypothesis. Table 4 gives the estimated values of the parameters and the values of the AIC and BIC for a competing class of models. The minimum AIC model is the non-stationary IFARMA(2, d+f, 0) where $d + f = 0.6092$, and the minimum BIC model is the stationary ARIMA(1, 0, 1) model. As in the Dickey-Fuller test, there is insufficient data to discriminate between a stationary or a non-stationary model if both the minimum AIC and BIC criterion are used. The same conclusion was reached by Roberts (1990), where he found that the ARIMA(2, 1, 0) and the ARIMA(1, 0, 1) equally fit the data with a slight edge to the mean stationary ARIMA(1, 0, 1) model.

6. Consistency of the likelihood maximization procedure for the estimation of the degree of difference *d* of an IFARMA(p,d+f,q) process.

Theorem 6.1:

i) Let $\{...,z_{-1},z_0,z_1,...\}$ be i.i.d. from a $N(0,\sigma^2)$ distribution. Then

$$S(b) = \frac{1}{n}\sum_{t=1}^{n}\{\nabla^b z_t\}^2$$

is an increasing function of

$$b \in \{0,1,2,...,B\}\ ,\ B < \infty\ ,\ \frac{B}{n} \to 0,$$

with

$$S(b) \to \binom{2b}{b}\sigma^2$$

in probability as $n \to \infty$.

ii) Let $\{\ a_t,\ t = 1,2,...\}$ be i.i.d. from a $N(0,\sigma^2)$ distribution, and define

$$z_t = \sum_{i=0}^{t} a_i\ ,\ a_0 \textit{ fixed and unknown.}$$

Then

$$S(0) = O(n)$$

$$S(b) \to \binom{2(b-1)}{b-1} \sigma^2 \, , \; b \geq 1.$$

as $n \to \infty$.

iii) Let $\{z_t\}$ be from an IFARMA(0, d+f, 0) process with fixed pre-observations. Then

$$S(b) \text{ is at leat } O(n), \; b < d$$

$$S(d+k) \to \binom{2k}{k} \sigma^2 \, , \; k \geq 0.$$

as $n \to \infty$.

Remark 6.2: The consistency of the likelihood maximization procedure to the estimation of the degree of difference d of an IFARMA(p,d+f,q) process is a direct consequence of the next corollary.

Corollary 6.3: Let $\{z_t\}$ be from an IFARMA(p, d+f, q) process with fixed pre-observations. Then

$$S(b) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\nabla^b z_t - \sum_{j=1}^{t-1} \phi_{t-1,j} \nabla^b z_{t-j}}{v_t^*} \right)^2$$

satisfies iii) of Theorem 6.1.

Proof: This is because the one step ahead forecast errors are i.i.d. from a Normal distribution, i.e.:

$$e_t = \frac{\nabla^d z_t - \sum_{j=1}^{t-1} \phi_{t-1,j} \nabla^d z_{t-j}}{v_t^*} \sim N(0, \sigma^2).$$

Remark: The next corollary is necessary to prove Theorem 6.1.

Corollary 6.4: If $\{ X_t \}$ is the finite moving average

$$X_t = \sum_{j=0}^{q} \theta_j Z_{t-j}$$

$$\theta_0 = 1,$$

$$Z_t \sim \textit{i.i.d. } N(0, \sigma^2).$$

Then the sample variance estimator,

$$\gamma^*(0) = \frac{1}{n} \sum_{t=1}^{n} X_t^2$$

has the following asymptotic normal distribution:

$$\sqrt{n}\,(\gamma^*(0) - \gamma(0)) \sim N(0,v)$$

$$v = 2\sum_{k=-q}^{q} \gamma(k)^2$$

$$\gamma(0) = (1+\theta_1^2+\ldots+\theta_q^2)\sigma^2$$

$$\gamma(i) = (\theta_i+\theta_1\theta_{i+1}+\ldots+\theta_{q-i}\theta_q)\sigma^2,\ i = 1,\ldots,q$$

$$\gamma(i) = 0,\ i > q.$$

Proof: This follows from Theorem 7.3.2 of Brockwell and Davis (1987).

Proof of Theorem 6.1:

i) Apply Corollary 6.4 with

$$X_t = \nabla^b Z_t = \sum_{j=0}^{b} \binom{b}{j} (-1)^j Z_{t-j}$$

$$\theta_j = \binom{b}{j} (-1)^j$$

$$\gamma(0) = \sigma^2 \sum_{j=0}^{b} \binom{b}{j}^2 = \sigma^2 \binom{2b}{b}.$$

ii) Let

$$X_t = \nabla^{b+1} Z_t = \nabla^b (Z_t - Z_{t-1}) = \nabla^b a_t,$$

where

$$a_t \sim \textit{i.i.d. } N(0,\sigma^2).$$

Then, if follows from part i) that

$$S(b) \sim \binom{2(b-1)}{b-1} \sigma^2 \, , \; b \geq 1.$$

For $S(0)$ it suffices to note that

$$\mathcal{E}(Z_t^2) = a_0^2 + t\sigma^2,$$

and consequently:

$$\mathcal{E}(S(0)) = a_0^2 + \sigma^2 \frac{(n+1)}{2} = O(n).$$

iii) This follows from ii).

7. Conclusions.

In this chapter I provide a model building strategy to analyze stationary and non-stationary time series through the IFARMA class of model. I give the necessary theoretical details and from a practical point of view I apply the IFARMA methodology to *Series A* and *B* of Box-Jenkins (1976), to the series of *Average Annual Temperature, Central England* and to the series *of Lake Michigan-Huron, Highest monthly mean level for each calendar year*. For those 4 series, the values of the AIC and the BIC under the different models are close to one another. My objective is to show that the IFARMA class of model provides an alternative to the ARIMA class where the degree of difference can be estimated by maximum likelihood estimation. This is, by itself, of major importance in time series modelling.

Throughout this chapter I use MLE but it is straightforward to use Mean Likelihood Estimation (MeLE) by integrating the combined difference parameter $\delta = d+f$ over the interval $(-.5,10)$. The choice of the upper bound of $\delta = d+f = 10$ is based on the empirical knowledge that time series have a degree of difference smaller than that. The integer values of $d = 0,1,2$ are the most frequent values used in ARIMA modelling.

Another research topic is to extend the model building strategy for stationary and non-stationary processes to seasonal models, where the pre-observations period will determine the seasonal means.

8. References.

Akaike,H. (1974). "A New Look at the Statistical Model Identification", **IEEE Trans. Automat. Contr.**, AC-19(6), 716-723.

Box,G.E.P. and Jenkins, G.M. (1976), **Time Series Analysis: Forecasting and control**, Holden-Day, San Francisco.

Brockwell, P.J. and Davis, A.D. (1987), **Time Series: Theory and Methods**, Springer-Verlag, New-York.

Dickey,D.A. and Fuller, W.A. (1979), "Distribution of Estimators for Autoregressive Time Series with a Unit Root", **JASA 74**, 427-431.

Hipel,K.W., McLeod,A.I. and Lennox,W.C. (1977), "Advances in Box-Jenkins Modelling", **Water Resources Research** 13, 567-586.

Jiménez Mosquera,C.J. (1988), "Advances in Time Series Analysis with Hydrological Applications", Unpublished Ph.D. thesis, **The University of Western Ontario**, London, Ontario, Canada.

Kalbfleish,J.D. and Sprott,D.A. (1970), "Application of Likelihood Methods to Models Involving Large Numbers of Parameters", (with discussion), **JRSS-B** 32, 175-208.

Kohn,R. and Ansley,C.F. (1986), "Estimation, Prediction, and Interpolation for ARIMA Models With Missing Data", **JASA** 81, 751-761.

McLeod,A.I. (1993), "MHTS: McLeod Hipel Time Series package", **The University of Western Ontario**, London, Ontario, Canada.

Roberts,H.V. (1990), **Data Analysis for Managers with MINITAB**, The Scientific Press, San-Francisco.

Ozaki,T. (1977), "On the Order Determination of ARIMA Models", **Appl. Statist.** 26, 290-301.

Schwarz,G. (1978). "Estimating the Dimension of a Model", **The Annals of Statistics** 6, 461-464.

Table 1: Parameter estimates for *Series A* ARIMA and IFARMA models, and values of the AIC and BIC.

Model	Parameters	MLE	AIC	BIC
ARIMA(0,1,1)	θ	0.69961	-446.50	-436.65
IFARMA(0,d+f,0)	$d+f$	0.39987	-450.31	-440.46
	μ	17.07863		
IFARMA(1,d+f,0)	ϕ	-0.04299	-448.49	-435.35
	$d+f$	0.42124		
	μ	17.07222		
IFARMA(0,d+f,1)	$d+f$	0.41909	-448.45	-435.32
	θ	0.03687		
	μ	17.07305		

Table 2: Parameter estimates for *Series B* ARIMA and IFARMA models, and values of the AIC and BIC.

Model	Parameters	MLE	AIC	BIC
ARIMA(0,1,0)			1465.36	1473.18
ARIMA(0,1,1)	θ	-0.08630	1464.57	1476.30
IFARMA(0,d+f,0)	$d+f$	1.05994	1464.94	1476.68
IFARMA(1,d+f,0)	ϕ	0.05943	1466.37	1482.01
	$d+f$	1.02627		
IFARMA(0,d+f,1)	$d+f$	1.02640	1466.33	1481.97
	θ	-0.06073		

Table 3: Parameter estimates for the series of *Average Annual Temperature, Central England, 1723-1970*, ARIMA and IFARMA models, and values of the AIC and BIC.

Model	Parameters	MLE	AIC	BIC
ARIMA(2,0,0)	ϕ_1	0.1188	-261.17	-247.12
	ϕ_2	0.2000		
	μ	9.2185		
IFARMA(0,d+f,0)	$d+f$	0.1485	-258.29	-247.75
	μ	9.2317		
IFARMA(2,d+f,0)	ϕ_1	0.0471	-259.38	-241.81
	ϕ_2	0.1681		
	$d+f$	0.0689		
	μ	9.2246		
IFARMA(2,d+f,0)	$\phi_1=0$		-261.23	-247.18
	ϕ_2	0.1513		
	$d+f$	0.1054		
	u	9.2280		

Table 4: Parameter estimates for the series of *Lake Michigan-Huron, highest monthly mean level for each calendar year, 1860-1955*, ARIMA and IFARMA models, and values of the AIC and BIC.

Model	Parameters	MLE	AIC	BIC
ARIMA(2,1,0)	ϕ_1	0.1422	-72.291	-62.034
	ϕ_2	-0.2558		
ARIMA(1,0,1)	ϕ_1	0.7898	-73.690	-63.433
	θ_1	-0.2903		
	μ	81.4354		
IFARMA(2,d+f,0)	ϕ_1	0.4506	-75.698	-62.876
	ϕ_2	-0.2134		
	$d+f$	0.6092		
IFARMA(1,d+f,1)	ϕ_1	0.1735	-74.881	-61.993
	θ_1	-0.3499		
	$d+f$	0.5272		
FARMA(2,f,0)	ϕ_1	0.6484	-72.783	-59.961
	ϕ_2	-0.1944		
	f	0.3933		
	μ	82.0658		
FARMA(1,f,1)	ϕ_1	0.2834	-72.641	-59.819
	θ_1	-0.3523		
	f	0.4014		
	μ	82.1557		

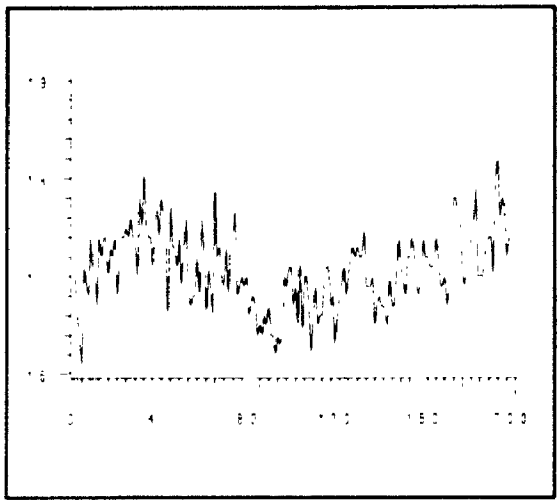

Figure 1a: *Series A* of Box-Jenkins (1976).

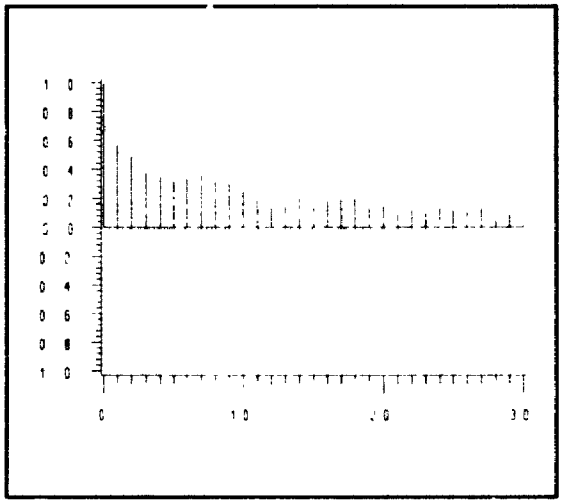

Figure 2: Sample autocorrelation function of *Series A*.

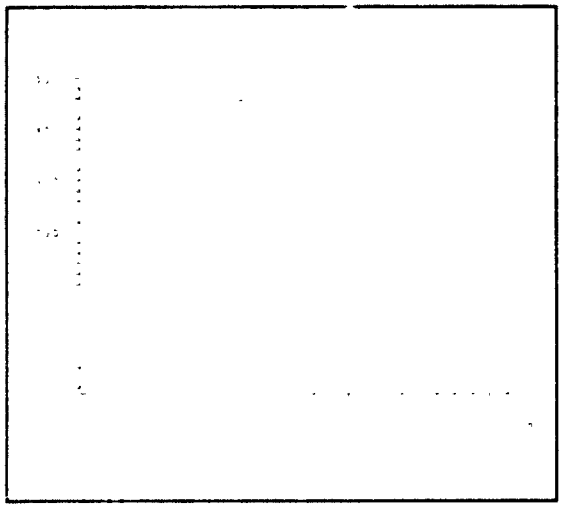

Figure 1b: ***2 log l(d+f)*** of an IFARMA(0,d+f,0) process evaluated at ***d + f = -0.5 to 2 by 0.1.***

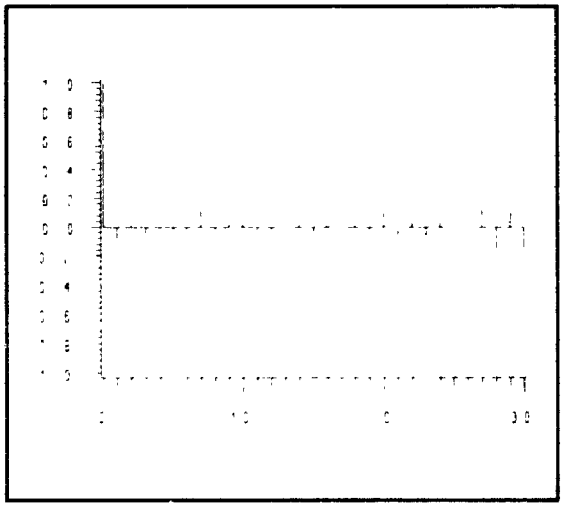

Figure 3: Sample autocorrelation function of *Series A* with one difference.

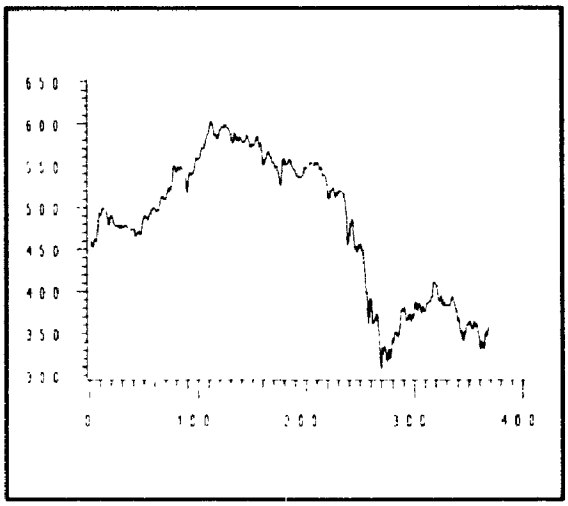

Figure 4: *Series B* of Box-Jenkins (1976).

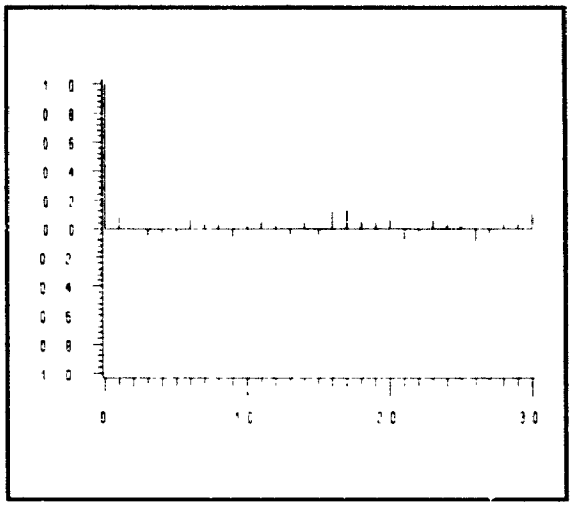

Figure 5: Sample autocorrelation function of *Series B* with one difference.

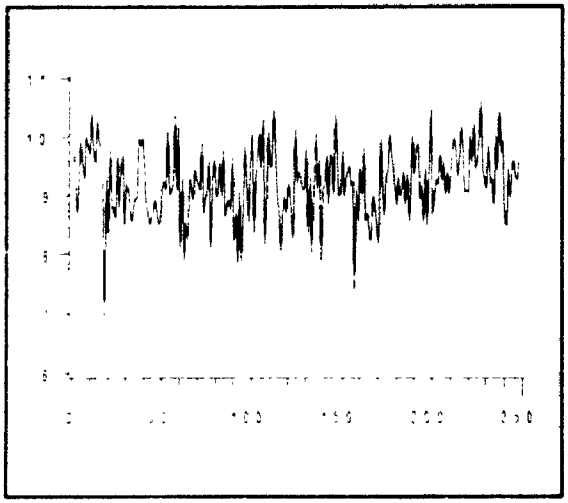

Figure 6: *Average Annual Temperature, Central England, 1723-1970.*

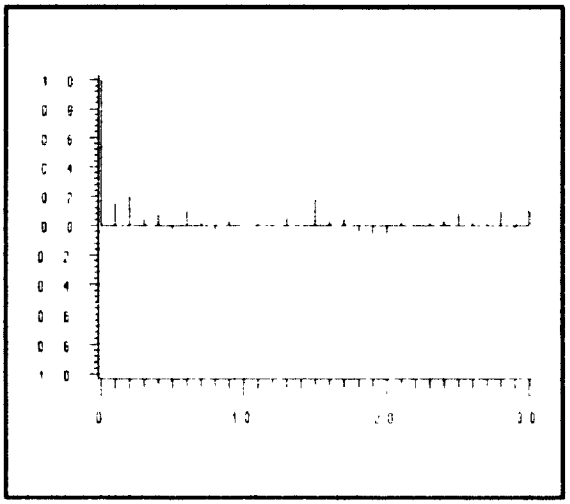

Figure 7: Sample autocorrelation function of the series of *Average Annual Temperature, Central England.*

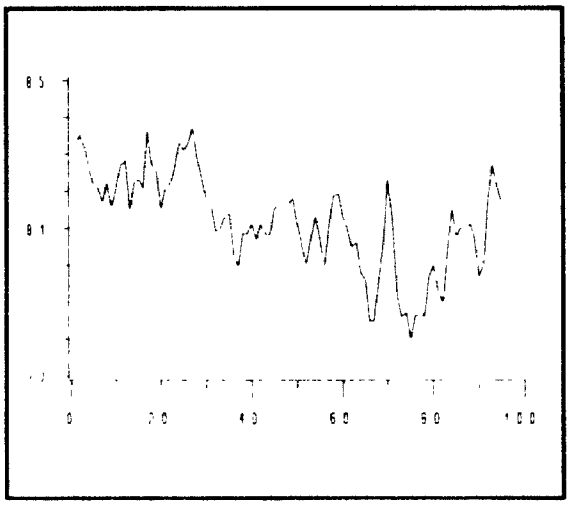

Figure 8: *Lake Michigan-Huron, highest monthly mean level for each calendar year, 1860-1955.*

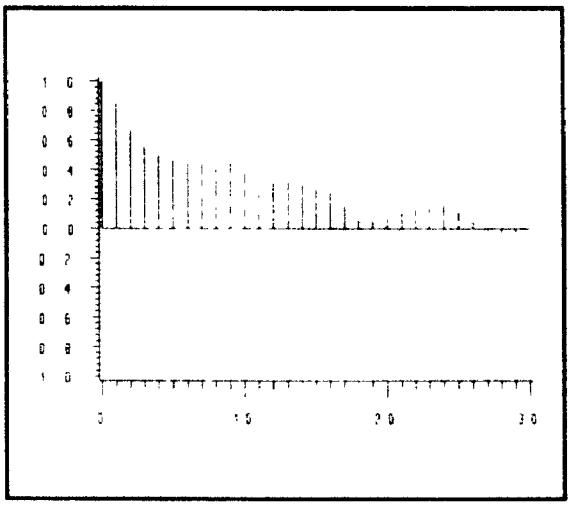

Figure 9: Sample autocorrelation function of the series *Lake Michigan-Huron, highest monthly mean level for each calendar year.*