1998

The asymptotic quasi-likelihood method: practical criteria and applications in fractional ARIMA models

Riccardo Biondini
University of Wollongong

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author.

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.

Recommended Citation

The Asymptotic Quasi-likelihood Method:
Practical Criteria and Applications in Fractional ARIMA Models

Riccardo Biondini

A thesis submitted in accordance with the requirements for the award of Honours Masters of Science (in Statistics) at the University of Wollongong
Declaration

In accordance with the rules and regulations of the University of Wollongong, I hereby state that the work described herein is my own original work except where due references are made, and has not been submitted for a degree at any other university or institution.
Acknowledgements

I would like to express my sincere gratitude to my parents, Sergio and Maria-grazia. I cannot thank them enough for being so supportive, especially through the past few years.

I dedicate this thesis in memory of my grandparents, Rina and Cristiano Angonese who gave me love and inspiration throughout my youth. They will always hold a special place in my heart.

I would also like to express my appreciation and gratitude to my supervisor Dr Yan-Xia Lin for all her help with this thesis, it would not have been possible without her guidance and suggestions. I feel fortunate to have had the chance to learn so much from her profound knowledge of statistics. A special thank you to the statistics staff for their help, from which I have benefited greatly.

‘Dedicata ai nonni, dimenticati mai’

Riccardo Biondini
Notation

\( ' \) denotes transposition
\( \text{annd} \) denotes asymptotically nonnegative definite
\( \text{AQL} \) denotes asymptotic quasi-likelihood
\( \text{AQLM} \) denotes asymptotic quasi-likelihood method
\( \text{AQS} \) denotes asymptotic quasi-score
\( \text{E}(X) \) denotes expected value of \( X \)
\( \dot{f} \) denotes differentiation of \( f_t(\theta) \) with respect to the components of \( \theta \)
\( g_t \) denotes a predictable process
\( G_T(\theta) \) denotes the estimating function where \( \theta \) is an unknown parameter
\( \text{i.i.d} \) denotes independently and identically distributed
\( \text{LS} \) denotes least squares
\( \text{ML} \) denotes maximum likelihood
\( O_{\text{A}}-\text{optimal} \) denotes the optimal estimating function for asymptotic samples
\( O_{\text{F}}-\text{optimal} \) denotes the optimal estimating function for fixed samples
\( \text{OLS} \) denotes ordinary least squares
\( \text{p.d} \) denotes positive definite
\( \text{QL} \) denotes quasi-likelihood
\( \text{QLM} \) denotes quasi-likelihood method
\( \text{QS} \) denotes quasi-score
\( \text{R/S} \) denotes rescaled adjusted range
\( U(\theta) \) denotes the score function where \( \theta \) is an unknown parameter
\( \text{Var}(X) \) denotes variance of \( X \)
\( \text{WLS} \) denotes weighted least squares
Contents

1 Introduction 1

1.1 Overview ..................................................... 2

2 Quasi-likelihood 7

2.1 Introduction .................................................. 7

2.2 Quasi-score Estimating Functions ....................... 9

2.3 Optimality Criteria ......................................... 11

3 Asymptotic Quasi-likelihood 15

3.1 Introduction .................................................. 15

3.2 Asymptotic Quasi-score Estimating Functions .......... 16

3.3 Methodology ................................................. 19

4 A Possible Practical Procedure in Applying the Asymptotic Quasi-
6.2 Simulations ....................................................... 73
6.3 Conclusion .......................................................... 76

7 Conclusion ............................................................. 77

Appendices .............................................................. 81
A Martingales ............................................................ 81
B Consistency of the Asymptotic Quasi-likelihood Estimate 82
C S-plus Programs ..................................................... 84
## List of Tables

4.1 OLS and AQL estimates (for four possible $g_t$’s) for Example 1. 36

4.2 OLS and AQL estimates (for two possible $g_t$’s) for Example 2. 39

4.3 Data for Example 3. 41

4.4 WLS, OLS and AQL estimates for Example 3. 41

4.5 OLS and AQL estimates for one hundred simulations of Example 1. 42

4.6 OLS and AQL estimates for one hundred simulations of Example 2. 43

5.1 OLS estimates for Example 4. 60

5.2 OLS and AQL estimates for Example 5. 61

5.3 OLS and AQL estimates for Example 6. 63

5.4 OLS and AQL estimates for the fractional ARIMA($1,d,0$) model. 68

5.5 OLS and AQL estimates for the fractional ARIMA($0,d,1$) model. 68

5.6 Fractional ARIMA($1,d,1$) model (with positive coefficients). 69
5.7 Fractional ARIMA(1,d,1) model (with negative coefficients) . . . . . 69

6.1 OLS and AQLM estimates (for two possible $g_t$'s) for Example 7 . . . 74

6.2 OLS and AQLM estimates (for two possible $g_t$'s) for Example 8 . . . 75
List of Figures

4.1 \( y_1^2 \) (hard line) and two possible \( g_t \)'s (dotted lines) for Example 1. . . . 34

4.2 \( y_1^2 \) (hard line) and two possible \( g_t \)'s (dotted lines) for Example 1. . . . 35

4.3 \( y_1^2 \) (hard line) and two possible \( g_t \)'s (dotted lines) for Example 2. . . . 38

4.4 \( X_t^2 \) (hard line) and a possible \( g_t \) (dotted line) for Example 3. . . . . . 47

5.1 \( X_t^2 \) (hard line) and three possible \( g_t \)'s (dotted lines) for Example 5. . . 67
Chapter 1

Introduction

This thesis is concerned with a practical procedure of applying the Asymptotic Quasi-likelihood Method (AQLM) and involves the application of the AQLM in linear models as well as in estimating the fractional differencing parameter in fractional ARIMA($p,d,q$) models. The Quasi-likelihood Method (QLM) is an inference method which unifies the traditional methods of maximum likelihood (ML) and least squares (LS). The ML method, introduced by Fisher, is dependent upon knowledge of the entire form of the underlying distribution. The method of LS, developed by Gauss, focuses on minimising the sum of squares. The QLM follows the framework of the ML method but depends on weaker conditions as well as having a broader application than the ML method. The QLM is a common inference tool since it solves certain problems which may not be solved efficiently via the traditional methods of ML and LS. The AQLM, applied when the QLM is deemed inappropriate due to a lack of information on the process or due to
the appearance of nuisance parameters, is rapidly becoming a popular inference method especially in the field of stochastic processes.

There are two alternative directions in inference for stochastic processes. Firstly, there is the question of finite-sample ($O_F$) optimality properties as opposed to asymptotics ($O_A$). Secondly, there is the question of methods which make no assumptions as to the true underlying distribution, i.e. second order methods such as LS and QL and more general semi-parametric methods.

1.1 Overview

Godambe and Heyde (1987) have shown that the QLM gives estimates with good properties irrespective of the form of the underlying distribution. Earlier work on the QLM has been made available by Wedderburn (1974) and McCullagh (1980). The QLM, from the viewpoint of estimating functions, is described in Chapter 2 and the notion of quasi-score (QS) estimating functions is introduced along with the optimality criteria involving such QS functions.

Heyde and Gay (1989) pointed out that the problem of exact theory is that a QL estimator may contain nuisance parameters. They introduced the AQLM in which the criteria for optimality are not satisfied exactly but hold in an asymptotic sense. The asymptotic theory provided by these two indicates whether there is loss of information when an unknown parameter is replaced by a consistent estimator and gives circumstances when this is asymptotically irrelevant.
The topic of AQL has also been addressed by Chen and Heyde (1994), Desmond (1991), Heyde (1988, 1997), Heyde and Gay (1989), Hutton, Ogunyemi and Nelson (1991), Lin (1993, 1995) and Mvoi, Lin and Biondini (1997) and references therein. A review of the AQLM is contained in Chapter 3 along with an introduction to the concept of asymptotic quasi-score (AQS) estimating functions and the methodology behind the application of the AQLM to linear models.

Linear models are very popular in practical situations. Examples of such applications may be found in Weisberg (1985) and references therein. In this thesis the focus is primarily on the model:

\[ y_t = f_t(\theta) + M_t, \quad t = 1, 2, ..., T, \]  

(1.1)

where \( \theta \in \Theta \) is an unknown parameter, \( f_t(\theta) \) is a linear predictable process of \( \theta \), \( \{M_t\} \) is a sequence of martingale differences such that \( \text{E}(M_t|\mathcal{F}_{t-1}) = 0 \) and the nature of \( \text{E}(M_t^2|\mathcal{F}_{t-1}) \) is unknown but finite. To estimate the parameter \( \theta \) in such a model, the ordinary least squares (OLS) method can generally provide a very good estimate subject to the \( \{M_t\} \) being mutually uncorrelated and the variances of the \( \{M_t\} \) being equal. However, these conditions do not always hold in reality. To reduce the effect from unequal variances of random error, a weighted least squares (WLS) method is needed and proper weights need to be allocated. When \( \{y_t\} \) is a sequence of independent observations from a population, the weights can sometimes be estimated via groups of observations. In each group of observations, all of the observations are associated with approximately the same value of \( f_t(\theta) \) (see Weisberg, 1985). However when \( \{y_t\} \) is a single path of realisation of a
process, it seems that it is not possible to use the same method to obtain an estimate of the weights. To cope with this problem an inference procedure, called the AQLM, is derived by Lin (1995) and discussed by Mvoi, Lin and Biondini (1997) and references therein. The AQLM given by Lin (1995) is different from that of Heyde and Gay (1989) although they share the same name. Lin (1995) also proved the AQLM is asymptotically optimal. Mvoi, Lin and Biondini (1998) prove the consistency of the AQL estimate for linear models. This thesis will focus on some techniques in applying the practical asymptotic procedure. For the theoretical discussion on the properties of the AQLM see Lin (1995) and Mvoi, Lin and Biondini (1998).

A practical procedure based on AQL methodology is outlined in Chapter 4 to estimate the unknown parameters in linear stochastic models of the form of model (1.1). To obtain a good estimate of \( \theta \) via the AQL procedure if the finite dimensional distributions of \( \{M_t\} \) are unknown, an appropriate predictable process \( g_t \) should be determined. Criteria for determining \( g_t \) are introduced which, if satisfied, provide more accurate estimates of the parameters via the AQLM.

In practice, for given \( \{y_t\} \), several predictable processes can be determined. The choice of \( g_t \), however, will affect the accuracy of the estimate of \( \theta \). One topic of this thesis is to discuss the criteria of selecting a proper \( g_t \) for a given data set and investigate the way different choices of \( g_t \) affect inference results. To simplify the discussion, in this thesis all \( g_t \)'s are given in autoregressive form which include, if available, independent variables and are determined by the Box-Jenkins method.
(see Box, Jenkins and Reinsel, 1994). Several criteria are presented to help choose a proper $g_t$ and these criteria are applied to both data generated via several models and real-life data using a two-stage estimation procedure.

In Chapter 5 investigation is carried out on one particular application of the AQLM. Attention is focussed on rescaled adjusted range ($R/S$) analysis given a fractional ARIMA($p, d, q$) model with finite variance where the aim is to estimate the intensity of long-range dependence of the particular series. This is done through what is commonly referred to as the Hurst parameter, $H$, a measure of self-similarity of a given time-series. This thesis will examine the effectiveness of applying the method of AQL to $R/S$ analysis instead of the conventional method of OLS. A comparison is made between the methods of OLS and AQL.

The accuracy of both the methods of OLS and AQL in linear models is further investigated in Chapter 6 where the sample size is relatively small. A comparison of the methods both between themselves and with the previous estimates (obtained with a larger sample size) are made.

A brief summary of properties of martingales is outlined in Appendix A. Appendix B discusses the conditions under which the AQL estimate converges to the true parameter $\theta$. Appendix C contains the S-plus programs used in simulating data, determining possible $g_t$'s and obtaining estimates of the unknown parameters.

Most of the research in Chapter 4 appears in Biondini, Lin and Mvoi (1997). Some of the research in Chapter 5 appears in Biondini and Lin (1997). Appendix B is a very brief summary of extended research which may be found in Mvoi, Lin
Chapter 2

Quasi-likelihood

2.1 Introduction

When defining a likelihood function the form of distribution of the observations must be specified, but when defining a QL function it is specified entirely in terms of information on the first and second moments of the samples. The QLM was firstly developed by Wedderburn (1974), where the aim was to estimate the unknown parameters in regression models assuming some relationship between the mean and variance only (i.e. not a fully specified model). Later this method was further developed by Godambe and Heyde (1987), the focus being on how to achieve optimal estimating functions via the strong orthogonality properties of martingale estimating functions. It is this second approach which is of interest in this thesis. For an outline of the connections between these two different approaches in a semi-parametric framework see Desmond (1991) and Lin and
Heyde (1993) and references therein.

The method of QL is often relied upon when there is doubt upon the form of higher order moments which are difficult, if not impossible, to check. Thus the advantage of applying the QLM in estimating unknown parameters is that it is not necessary to know the distribution of the population except for some knowledge of the variance or conditional variance. The approach taken is where assumptions about the link and variance functions are made without trying to specify the entire distribution of the random variable.

In most respects a QL is very similar to an ordinary log-likelihood. For the estimation of the dispersion parameter $\sigma^2$, the behaviour of the QL estimate $\theta$ does not resemble that of a conventional likelihood. Properties of such QL estimators are typically obtained from an expansion of the QL in which the remainder term is asymptotically negligible. For an overview of the properties of the QL estimate see Godambe and Heyde (1987).

Within a one-parameter exponential family QL estimates coincide with ML estimates. It is of interest, however, to examine the relative efficiency of QL estimates under different sampling distributions (but obtained using the correct mean-variance relationship). This has been done by Firth (1987) who demonstrated that, in parametric regression models, QL estimates retain high efficiency under moderate departures from the appropriate exponential family distribution. Firth (1987) discussed three types of models; those with constant variances, those with constant coefficient of variation and those with over-dispersion with respect
to some exponential family.

Many of the ideas and procedures about fitting ordinary generalised linear models can be easily extended when the likelihoods are replaced by QL's. Nelder and Pregibon (1987) developed the extended QL which is derived from the theory of orthogonal estimating functions whereby the martingale structure and the filtering are replaced by a less general "conditioning".

### 2.2 Quasi-score Estimating Functions

Let \( \{y_t, 0 \leq t \leq T\} \) be a sequence of observations in discrete or continuous time drawn from a population with distribution \( F \) taking values in \( r \)-dimensional Euclidean space, the distribution of which is conditioned upon a parameter \( \theta \) taking values in an open subset \( \Theta \) of \( p \)-dimensional Euclidean space. Suppose that the probability distribution of \( \{y_t\} \) is given by \( \mathcal{P} = \{\mathcal{P}_\theta\} \), a union of distribution families, each family being indexed by the same parameter \( \theta \), and that \((\Omega, \mathcal{F}_t, \mathcal{P})\) is a complete probability space. Furthermore \( \{\mathcal{F}_t, t \geq 0\} \) is defined as a non-decreasing family of sigma fields which are generated by \( \{y_s\}, s \leq t \) and \( \mathcal{F}_0 = \bigcap_{t=1}^{T} \mathcal{F}_t \).

Defining \( \Psi_T \) as the class of all linearly unbiased (i.e. zero mean) square integrable estimating functions \( G_T(\theta) = G_T(\{y_t, 0 \leq t \leq T\}, \theta) \) for which \( \mathbb{E}G_T(\theta) = 0 \) for each \( \mathcal{P}_\theta \in \mathcal{P} \) with index \( \theta \) and letting \( \mathcal{G}_T \) be a subset of \( \Psi_T \). The element \( G_T(\theta) \), of \( \mathcal{G}_T \), is referred to as an estimating function and the equation \( G_T(\theta) = 0 \) is
referred to as an estimating equation. The estimating equations are vectors of dimension \( p \). In this estimating function space \( \mathcal{G}_T \), the element \( G_T(\theta) \) is differentiable and satisfies the conditions that \( E G_T(\theta) = 0 \) for each \( P_0 \), and to which the \( p \)-dimensional matrices \( E G_T(\theta) G_T'(\theta) \), \( E G_T(\theta) \) where \( G_T \) denotes differentiation of \( G_T \) with respect to the unknown parameter \( \theta \) and \( [G(\theta)]_T \) (the quadratic variation of \( G_T(\theta) \)) are non-singular for each \( T > 0 \), \( ^\prime \) denoting transposition. A point estimate for \( \theta \) based on \( G_T(\theta) \) is obtained by solving \( G_T(\theta) = 0 \) for the observed values of \( y \). \( G_T(\theta) \) is said to be unbiased if, for \( G_T(\theta) \in \mathcal{G}_T \), its expected value is equal to zero for all \( \theta \).

As Heyde (1997) noted it may be useful to adopt a martingale setting. Let \( U \) denote the underlying score function (i.e. the derivative of the exact likelihood). The estimating function space considered in these situations, \( \mathcal{G}_T \), is that containing all square integrable martingale estimating functions on \( (\Omega, \mathcal{F}_t, P) \). Since the score function \( U \) is a square-integrable martingale it is clear that, if the form of the score function is unknown, the best way to find an estimating function to approximate \( U \) is to find it from \( \mathcal{G}_T \). From \( \mathcal{G}_T \) the aim is to obtain an orthogonal martingale basis from which a standard QS estimating function can be written down directly. If a subspace of the square integrable martingale is large enough to contain the score function \( U \), the QS estimating function on that space will be exactly equal to \( U \). Lin and Heyde (1996) investigated the relationship between the score function, QS function, martingale space and the general linear space as well as showing why a selection of a QS estimating function from a martingale space is preferred to one from a non-martingale space.
2.3 Optimality Criteria

In general, discussions of optimality of the QL estimates and estimating functions have been concerned with whether fixed criteria hold. The estimating function approach concentrates on the function rather than the resulting estimator as there is often loss of information about the parameter in moving from estimating function to estimator. Optimality properties are for this particular estimating function. Heyde (1987), in combining data from different experiments, investigated optimal combinations of estimating functions rather than the resulting estimators. The properties of estimators derived from optimal estimating functions can be investigated in an asymptotic sense.

The most commonly used optimality criteria for the simultaneous estimation problem are those based on the following;

- the non-negative definiteness of the dispersion matrix (M-optimality);
- the trace of the dispersion matrix (T-optimality); and
- the determinant of the dispersion matrix (D-optimality).

It is known that M-optimality implies the other two optimality criteria. In fact, Chandrasekar and Kale (1984) have shown that if $G_T^*(\theta)$ is optimal with respect to any one of the three above-mentioned criteria, it is also optimal with respect to the remaining two criteria.

Godambe and Heyde (1987) developed three equivalent criteria regarding optimal
estimating equations. Note that Definition 1 does not require the existence of $U$ and is employed when determining whether a particular function is in fact the QS function.

**Definition 1** $G^*_T(\theta)$ is the QS estimating function in $\mathcal{G}_T \subseteq \Psi_T$ iff

$$(E\dot{G}_T(\theta))^{-1}(EG_T(\theta)G'_T(\theta))(E\dot{G}'_T(\theta))^{-1} - (E\dot{G}^*_T(\theta))^{-1}(EG^*_T(\theta)G^*_T(\theta))(E\dot{G}^*_T(\theta))^{-1}$$

is non-negative definite for all $G_T(\theta) \in \mathcal{G}_T$, $\theta \in \Theta$ and $P_\theta \in \mathcal{P}$.

**Theorem 1** Suppose $G^*_T(\theta)$ is the QS estimating function in $\mathcal{G}_T \subseteq \Psi_T$. Then for some fixed matrix function $\alpha$ depending on $\theta$ and $T$,

$$E((\alpha UT - G^*_T(\theta))G'_T(\theta)) = E(G_T(\theta)(\alpha UT - G^*_T(\theta))^{'}) = 0$$

for all $G_T(\theta) \in \mathcal{G}_T$, $\theta \in \Theta$ and $P_\theta \in \mathcal{P}$.

**Theorem 2** Suppose $G^*_T(\theta)$ is the QS estimating function in $\mathcal{G}_T \subseteq \Psi_T$. Then for some fixed matrix function $\alpha$ depending on $\theta$ and $T$,

$$E((\alpha UT - G_T(\theta))(\alpha UT - G_T(\theta))^{'} - E((\alpha UT - G^*_T(\theta))(\alpha UT - G^*_T(\theta))^{'} - E(\alpha UT - G^*_T(\theta))(\alpha UT - G^*_T(\theta))^{'}$$

is non-negative definite for all $G_T(\theta) \in \mathcal{G}_T$, $\theta \in \Theta$ and $P_\theta \in \mathcal{P}$.

There are several equivalent definitions for a QS estimating function. Definition 1 may simply be re-written as;

**Definition 2** $G^*_T(\theta)$ is the QS estimating function in $\mathcal{G}_T \subseteq \Psi_T$ iff

$$(E\dot{G}^*_T(\theta))^{'}(EG^*_T(\theta)G^*_T(\theta))^{-1}(E\dot{G}^*_T(\theta)) - (E\dot{G}_T(\theta))^{'}(EG_T(\theta)G_T(\theta))^{-1}(E\dot{G}_T(\theta))$$
is non-negative definite for all \( G_T(\theta) \in \mathcal{G}_T, \theta \in \Theta \) and \( P_\theta \in \mathcal{P} \).

\( G^*_T(\theta) \) is defined only up to a constant multiple and is usually a non-linear function of \( \theta \) (thus producing more than one solution of the QS normal equation). The reasons for \( G^*_T(\theta) \) being optimal are now discussed. Under suitable regularity conditions the unbiased estimator obtained from the QL estimating function \( G^*_T(\theta) \) usually has minimal variance amongst the estimators obtained from all estimating functions in \( \mathcal{G}_T(\theta) \). Similarly in the same estimating function space, \( G^*_T(\theta) \) has minimum dispersion distance from the generally unknown score function. Thus the QS estimating function \( G^*_T(\theta) \) has the further important asymptotic property that confidence intervals for the unknown parameter \( \theta \) associated with \( G^*_T(\theta) \) are asymptotically shortest within \( \mathcal{G}_T \). Furthermore \( G^*_T(\theta) \) provides more Fisher information than all other estimating functions in \( \mathcal{G}_T \). In other words, the information matrix is maximised. As a result the QS estimating function is the optimal estimating function in \( \mathcal{G}_T \). For an extended discussion on both finite sample and asymptotic optimality and a comparison between both types of optimality in both parametric and non-parametric settings see Heyde (1988) and references therein. Wedderburn (1974) showed that the estimates of the parameters can be obtained by solving \( G_T(\theta) = 0 \) for any choices of link and variance functions even if they fail to correspond to a particular member of the exponential family.

In many cases the optimal estimating function \( G^*_T(\theta) \) will essentially be, under appropriate distributional assumptions, the true score estimating function and
hence will lead to likelihood estimation. Godambe (1960) showed that the score function is the optimal estimating function in parametric situations (i.e. situations where the form of the underlying distribution is known up to an unknown parameter). Kulkarni and Heyde (1987) propose a general method of constructing robust QL estimating functions for discrete time-series processes.

QS estimating functions enjoy the same basic optimality properties of both LS and ML procedures but only within the local estimating space $\mathcal{G}_T \in \Psi_T$. This means that when a QS estimating function is obtained, it is just necessary to emphasise the $\mathcal{G}_T$ from which the QS estimating function is chosen.
Chapter 3

Asymptotic Quasi-likelihood

3.1 Introduction

An unfortunate problem which may arise from the implementation of QS functions is that they may often lead to QL estimates being unattainable since $G_T^*(\theta)$ might involve nuisance parameters or the exact mathematical expression of $G_T^*(\theta)$ may be unknown. To combat this problem, Heyde and Gay (1989) and Lin (1995) introduced the notion of asymptotic quasi-likelihood (AQL) functions. The proposed methods may be applied when the sample size is either large or increasing (thus reducing the effect of nuisance parameters) and some information on $E(M_t^2|\mathcal{F}_{t-1})$ can be obtained from $\{y_t^2\}$. The method applied by Heyde, Gay and Lin is known as the asymptotic quasi-likelihood method (AQLM).
3.2 Asymptotic Quasi-score Estimating Functions

The first step in formulating the principle of asymptotic quasi-score (AQS) estimating functions involves introducing the notion of asymptotically non-negative definite matrices (denoted by \( \text{annd} \)). Given a matrix \( A = (a_{ij}) \), \( \|A\| \) is defined as the Frobenius norm of \( A \), and

\[
\|A\| = \left( \sum_{i} \sum_{j} a_{ij}^2 \right)^{1/2} = (\text{trace}(AA'))^{1/2}.
\]

Let \( \{A_n\} \) and \( \{B_n\} \) be two sequences of symmetric positive definite (p.d.) matrices. If there is a sequence of matrices \( \{D_n\} \) such that \( \{A_n - B_n + D_n\} \) is non-negative definite (i.e. \( \{A_n - B_n + D_n\} \geq 0 \) for every \( n \)) and \( \|D_n\| \to 0 \) as \( n \to \infty \), then \( A_n - B_n \) is \( \text{annd} \).

AQS estimating functions obtained under the approach given by Heyde and Gay (1989) are mainly suitable in cases where the exact expression of \( E(M_i^2|\mathcal{F}_{t-1}) \) is known. This thesis will focus on the cases where the nature of \( E(M_i^2|\mathcal{F}_{t-1}) \) is unknown.

Lin (1995) gives a new definition of the AQS sequence of estimating functions and discusses the relationship between the sequence of QS estimating functions and the sequence of AQS estimating functions where the estimating function space is fixed. The motivation for this restriction is to solve the problem of how to estimate the unknown parameters when the exact form of the QS estimating function is impossible to be correctly written down. The definition is as follows;
Definition 3 Suppose $G_{T,n}^*(\theta) \in \mathcal{G}_T \subseteq \Psi_T$, for all $n > 0$ such that

$$
(EG_T(\theta))^{-1}(EG_T(\theta)G_T^*(\theta))(EG_T(\theta))^{-1} - 
$$

$$
(EG_{T,n}^*(\theta))^{-1}(EG_{T,n}^*(\theta)G_{T,n}^*(\theta))(EG_{T,n}^*(\theta))^{-1}
$$

is and for all $G_{T,n}(\theta) \in \mathcal{G}_T$, as $n \to \infty$. \{G_{T,n}^*(\theta)\} is said to be an AQS sequence of estimating functions within $\mathcal{G}_T$ and the solution $\theta_{T,n}^*$ which satisfies the equation $G_{T,n}^*(\theta) = 0$ is called the AQL estimate within $\mathcal{G}_T$.

From Definition 3 it may be seen that if $G_T(\theta)$ is $O_F$-optimal (i.e. optimal in a fixed sample sense) within $\mathcal{G}_T$ for each $T$ then \{G_{T,n}^*(\theta)\} is an AQS sequence of estimating functions within $\mathcal{G}_T$. Lin (1995) obtains the following Theorem;

Theorem 3 Suppose \{G_T^*(\theta)\} is the QS estimating function in $\mathcal{G}_T \subseteq \Psi_T$. Then \{G_{T,n}^*(\theta)\} is an AQS sequence of estimating functions in $\mathcal{G}_T$ iff

$$
||(EG_{T,n}^*(\theta))^{-1}(EG_{T,n}^*(\theta)G_{T,n}^*(\theta))(EG_{T,n}^*(\theta))^{-1} - 
$$

$$
(EG_{T,n}^*(\theta))^{-1}(EG_{T,n}^*(\theta)G_{T,n}^*(\theta))(EG_{T,n}^*(\theta))^{-1}|| \to 0,
$$

as $n \to \infty$, for all $G_T(\theta) \in \mathcal{G}_T$, $\theta \in \Theta$ and $P_\theta \in P$.

Under the usual regularity conditions, Heyde and Lin (1991) have shown that the confidence region for true parameters may usually be determined by

$$
(EG_{T,n}^*(\theta))^{-1}(EG_{T,n}^*(\theta)G_{T,n}^*(\theta))(EG_{T,n}^*(\theta))^{-1},
$$

where $G_{T}^*(\theta)$ is a quasi-score (QS) estimating function. From Theorem 3 a confidence region for the unknown parameter $\theta$ may be obtained via the use of
asymptotic sequences of estimating functions when \( n \) is large or increasing and the estimate of the unknown parameters yielded via the AQLM is reasonably accurate.

Note that from Definition 3 it may sometimes be difficult to determine whether a sequence of estimating functions is an AQS sequence of estimating functions. However Lin (1995) provides the following sufficient condition for checking whether a sequence of estimating functions \( \{G_{T,n}^*(\theta)\} \) is an AQS sequence of estimating functions.

**Theorem 4** Suppose \( G_{T,n}^*(\theta) \in G_T \subseteq \Psi_T, \) for all \( n > 0. \) If

\[
\lim_{n \to \infty} \|E G_{T,n}^*(\theta) G_{T,n}^{*-1}(\theta)\| \geq \alpha > 0,
\]

and there exists a QS function \( G_T^*(\theta) \in G_T, \) then \( \{G_{T,n}^*(\theta)\} \) is an AQS sequence of estimating functions in \( G_T \) if, for all \( G_T(\theta) \in G_T \) and as \( n \to \infty, \)

\[
(E G_T(\theta)^{-1} E G_T(\theta) G_{T,n}^* G_{T,n}^{*-1}(\theta) \to K = \lim_{n \to \infty} (E G_{T,n}^*(\theta))^{-1} E G_{T,n}^*(\theta) G_{T,n}^{*-1}(\theta),
\]

where \( K \) is a non-singular matrix.

For a detailed proof of Theorem 4 see Lin (1995).
3.3 Methodology

Assuming that $y_t$ is square integrable, a regression-type model of the following form may be constructed;

$$y_t = E(y_t|\mathcal{F}_{t-1}) + y_t - E(y_t|\mathcal{F}_{t-1}) = f_t(\theta) + M_t,$$

where $f_t(\theta) = E(y_t|\mathcal{F}_{t-1})$ and $M_t = y_t - E(y_t|\mathcal{F}_{t-1})$ which may be simplified to model (1.1) where $\theta \in \Theta$ is an unknown parameter and $\Theta$ is an open parameter space, $f_t(\theta)$ is a linear predictable process of $\theta$, $\{M_t\}$ is a sequence of martingale differences such that $E(M_t|\mathcal{F}_{t-1}) = 0$ and the nature of $E(M_t^2|\mathcal{F}_{t-1})$ is unknown but finite.

Under the usual regularity conditions a QS estimating function in $G_T$ may be written down easily based on model (1.1) as follows;

$$G_T^*(\theta) = \sum_{t=1}^{T} \frac{\hat{f}_t(\theta)}{E(M_t^2|\mathcal{F}_{t-1})} M_t. \quad (3.1)$$

Looking at equation (3.1) a problem obviously arises when the nature of $\{M_t\}$ is unknown. Therefore the form of $E(M_t^2|\mathcal{F}_{t-1})$ is also unknown. The problem involved in applying this QL procedure in a practical situation is not so much in the existence of nuisance parameters but rather that the nature of the model may make it difficult to find the exact expression of the QS function $G_T^*(\theta)$. Interest is focussed on the estimate of $E(M_t^2|\mathcal{F}_{t-1})$. Lin (1995) showed that under some regularity conditions, if there exists a sequence of predictable processes $g_{n,t}$ such
that \( E(y_t^2 - g_{n,t}|\mathcal{F}_{t-1}) = E(\epsilon_{n,t}|\mathcal{F}_{t-1}) \) \( \rightarrow 0 \) as \( n \rightarrow \infty \) and for all \( 0 \leq t \leq T \), then

\[
G_T^*(\theta) = \sum_{t=1}^{T} \frac{\hat{f}_t(\theta)}{g_t - \hat{f}_t^2(\theta)} M_t = 0
\]

will be an AQS sequence of estimating functions in \( G_T \). This result led to Lin (1995) suggesting a practical procedure of obtaining the AQS function for model (1.1). The idea is now described. Assuming the usual regularity conditions, the objective is firstly to find a predictable process \( g_t \in \mathcal{F}_{t-1} \) such that \( y_t^2 = g_t + \epsilon_t \) (i.e. \( E(y_t^2|\mathcal{F}_{t-1}) = g_t + E(\epsilon_t|\mathcal{F}_{t-1}) \)) and that \( E(\epsilon_t|\mathcal{F}_{t-1}) \) (for \( n > 0 \)) is negligible and approximately stationary for all \( t \) then

\[
y_t^2 = f_t^2 + 2f_t M_t + M_t^2
\]

(3.2)

\[= g_t + \epsilon_t, \quad \text{say.}\]

Note that from (3.2)

\[f_t^2 + 2f_t M_t + M_t^2 = g_t + \epsilon_t\]

and taking conditional expectations of both sides gives

\[f_t^2 + E(M_t^2|\mathcal{F}_{t-1}) = g_t + E(\epsilon_t|\mathcal{F}_{t-1}).\]

Rearranging this equality gives

\[E(M_t^2|\mathcal{F}_{t-1}) = g_t - f_t^2 + E(\epsilon_t|\mathcal{F}_{t-1})\]

\[\approx g_t - f_t^2.\]

Noting that much of the information on the second conditional moment of the martingales, \( E(M_t^2|\mathcal{F}_{t-1}) \), may be obtained from \( y_t^2 \), the quantity \( g_t - f_t^2 \) may be
used to approach $E(M_t^2|\mathcal{F}_{t-1})$ and an AQS estimating function may be obtained. The search is essentially for a predictable process which may be used to estimate the quadratic process. It may be possible to check whether or not $g_t$ accurately approaches $y_t^2$ by plotting both $g_t$ and $y_t^2$. The search is for a predictable process $g_t$ such that the error $\epsilon_t$ is negligible and thus the quantity $E(M_t^2|\mathcal{F}_{t-1})$ may be approximated by $g_t - f_t^2$.

Note that $E(M_t^2|\mathcal{F}_{t-1})$ has a strong link with $\epsilon_t$, or $E(\epsilon_t|\mathcal{F}_{t-1})$. The information on $E(\epsilon_t|\mathcal{F}_{t-1})$ may partially come from the quantity $y_t^2 - g_t$. The unknown parameters may then be obtained by employing the AQS estimating function. Note that in this procedure $\epsilon_t$ must be negligible and stationary. Since only one observation is available at each time period it is required that the $\epsilon_t$ be asymptotic stationary. This means that, given $y_t^2$, if a $g_t$ cannot be found such that $y_t^2 - g_t$ is stationary the following method will fail. Note that if $\epsilon_t$ is stationary then

$$E(\epsilon_t) \approx \frac{1}{n} \sum_{t=1}^{T} \epsilon_t.$$  

However this does not mean that

$$E(\epsilon_t|\mathcal{F}_{t-1}) \approx \frac{1}{n} \sum_{t=1}^{T} \epsilon_t.$$  

It is known that

$$E(E(\epsilon_t|\mathcal{F}_{t-1})) = E(\epsilon_t) \approx \frac{1}{n} \sum_{t=1}^{T} \epsilon_t.$$  

Therefore for model (1.1), whether $E(M_t^2|\mathcal{F}_{t-1})$ may be well approximated by $g_t - f_t^2$ will depend on whether $E(y_t^2 - g_t|\mathcal{F}_{t-1})$ is small enough or not. For the
difference between \( E(M_t^2|\mathcal{F}_{t-1}) \) and \( g_t - f_t^2 \) to be small,

\[
\text{Var}(\varepsilon_t) \approx \frac{1}{n-1}(\varepsilon_t - \bar{\varepsilon}_t)^2
\]

must be as small as possible. As \( \text{Var}(\varepsilon_t) \) gets smaller the average of the difference of \( E(M_t^2|\mathcal{F}_{t-1}) \) and \( g_t - f_t^2 \) becomes smaller. Of course how accurate the AQL estimate is depends strongly on the quality of the approximation of \( y_t^2 \) by \( g_t \).

Thus the determination of \( g_t \) via plot and data analysis is very important. The value \( g_t - f_t^2 \) estimates a positive quantity \( E(M_t^2|\mathcal{F}_{t-1}) \) and should therefore be positive. This is not always the case in practical situations. The determination of \( g_t \) may result in \( g_t - f_t^2 \) being negative for some values of \( t \). In such situations, the absolute value of this quantity is considered. In practice, the QS normal equation that has been used is of the following form

\[
\hat{G}_T^*(\theta) = \sum_{t=1}^{T} \frac{\hat{f}_t(\theta)M_t}{|g_t - f_t^2(\theta)|} = 0.
\]

Since \( g_t - f_t^2(\theta) \) is used to estimate \( E(M_t^2|\mathcal{F}_{t-1}) \) and \( E(M_t^2|\mathcal{F}_{t-1}) \) is always positive, it is reasonable to use \( |g_t - f_t^2(\theta)| \) rather than \( g_t - f_t^2(\theta) \) in the QS normal equation. For the new form of the QS normal equation, Mvoi, Lin and Biondini (1998) have proved that when \( f_t(\theta) \) is a linear function of \( \theta \), under certain conditions, the AQL estimate is a good estimate of the true parameter. Also, this estimate is consistent as sample size is increasing (see Appendix B). According to the proof given by Mvoi, Lin and Biondini (1998), based on the AQL procedure, a good estimate of \( \theta \) will be obtained if

\[
\frac{E(M_t^2|\mathcal{F}_{t-1})}{|g_t - f_t^2|}
\]
is bounded. Therefore, there is a $k_1 > 0$ and $k_2 > 0$ such that

$$k_2 \leq \frac{E(M_t^2|\mathcal{F}_{t-1})}{|g_t - f_t^2|} \leq k_1.$$ 

The smaller the difference between $k_1$ and $k_2$, the quicker the iteration will converge. A problem arises when $|g_t - f_t^2|$ is too close to zero. To ensure the iteration converges in such circumstances, $|g_t - f_t^2|$ is replaced by $|g_t - f_t^2| + c$, where $c$ is a suitable positive constant. Therefore, if a good estimate is expected to be obtained, the ratio

$$\frac{E(M_t^2|\mathcal{F}_{t-1})}{|g_t - f_t^2| + c} \quad (3.3)$$

should be bounded between two finite values for each $t$, $c$ denoting a suitable positive constant. This constant must not be too large, further investigation is needed to determine the appropriate value of $c$. The above ratio may be re-written in the following form;

$$\frac{E(M_t^2|\mathcal{F}_{t-1})}{|g_t - f_t^2| + c} = \frac{E(y_t^2|\mathcal{F}_{t-1}) - f_t^2}{|g_t - f_t^2| + c} = \frac{g_t - f_t^2 + c}{|g_t - f_t^2| + c} + \frac{E(\epsilon_t|\mathcal{F}_{t-1}) - c}{|g_t - f_t^2| + c}. \quad (3.4)$$

It is easily seen that if $\frac{E(\epsilon_t|\mathcal{F}_{t-1}) - c}{|g_t - f_t^2| + c}$ is bounded then $\frac{E(M_t^2|\mathcal{F}_{t-1})}{|g_t - f_t^2| + c}$ is also bounded because the quantity $\frac{g_t - f_t^2 + c}{|g_t - f_t^2| + c}$ is bounded between -1 and 1, i.e. $-1 \leq \frac{g_t - f_t^2 + c}{|g_t - f_t^2| + c} \leq 1$. If $|g_t - f_t^2|$ is not negligible and $E(\epsilon_t|\mathcal{F}_{t-1})$ is not much larger than $|g_t - f_t^2|$, $c$ is not needed for the ratio (3.3) to become bounded. If $|g_t - f_t^2|$ is negligible and $E(\epsilon_t|\mathcal{F}_{t-1}) >> |g_t - f_t^2|$, then from (3.4), a proper $c$ can make

$$\frac{E(\epsilon_t|\mathcal{F}_{t-1}) - c}{|g_t - f_t^2| + c}$$
bounded, so the ratio (3.3) becomes bounded. The closer the ratio is to 1, the quicker the convergence of the AQL estimates and the more accurate these estimates will be (for theoretical details see Mvoi, Lin and Biondini, 1998). The resulting AQL estimates are found to be also very accurate provided the QL estimates are accurate. Even though in practical situations this ratio is unavailable, by the inclusion of a suitable positive constant $c$ in the denominator in (3.3) we can assure that the ratio (3.3) for each $t$ is much smaller than it would be if this constant was not taken into consideration. Since the ratio is no longer large this ensures convergence. It must be noted that this constant must not be too large as $|g_t - f_t^2| + c$ will be dominated by the value of $c$ and the value of the AQL estimates will be closer to the ordinary least squares (OLS) estimates.

In this thesis the predictable process $\{g_t\}$ is usually obtained by fitting an autoregressive model on $\{y_t^2\}$ using, if available, independent variables. Different choices of $g_t$ result in different estimates of the parameters. The next chapter will discuss the criteria used in the selection of $g_t$. 
Chapter 4

A Possible Practical Procedure in Applying the Asymptotic Quasi-likelihood Method

4.1 Introduction

Assume that the observed process is given by (1.1). A QS estimating function can be determined based on (1.1) and has the expression (3.1). The QL estimate of \( \theta \) is obtained by solving the QS normal equation \( G_{\theta}^\prime(\theta) = 0 \). When \( f_t(\theta) \) is a linear function of \( \theta \), the QLM always provides a good estimate of \( \theta \) subject to \( E(M_t^2|\mathcal{F}_{t-1}) \) being known. If \( E(M_t^2|\mathcal{F}_{t-1}) \) is not known it must be estimated.

It was seen in the previous chapter that if, assuming certain regularity conditions,
a predictable process \( g_t \in \mathcal{F}_{t-1} \) can be found such that \( y_t^2 = g_t + \epsilon_t \) and that \( E(\epsilon_t|\mathcal{F}_{t-1}) \) (for \( n > 0 \)) is negligible and approximately stationary for all \( t \) then

\[
E(M_t^2|\mathcal{F}_{t-1}) - f_t^2(\theta) \approx g_t - f_t^2(\theta).
\]

4.2 Objectives

The AQLM can usually yield a good estimate of the unknown parameter(s) and simultaneously the prediction of \( E(M_t^2|F_{t-1}) \) can be obtained subject to knowledge of \( g_t \). To obtain the \( g_t \) from given information of \( y_t^2 \), it has been found that, in practice, the autoregression technique discussed by Box and Jenkins (see Box, Jenkins and Reinsel, 1994) is a simple method to use and attention is therefore limited to \( g_t \) obtained via autoregressive methods and including, if available, independent variables.

The objectives of this chapter are summarised as follows;

1. Which criteria are applied to determine the appropriate \( g_t \) to approach \( y_t^2 \)?

2. What is the relationship between \( g_t \) and the estimate \( \hat{\theta} \)?

3. What is the mean and variance for the estimates of \( \hat{\theta} \) for different \( g_t \)?
4.3 Criteria in Selection of $g_t$

In this section criteria are listed in determining a predictable process $g_t$ to approach $y_t^2$ and then logic to each point is provided. The criteria discussed here will be applied to both simulated and real-life data to demonstrate the practicality of the AQLM. The criteria in selecting predictable processes which yield accurate AQL estimates can be summarised as follows;

- Examine the time-series plot of $g_t$ and $y_t^2$, $g_t$ should be chosen such that it accurately approaches $y_t^2$.

- Examine the stationarity of $\epsilon_t$ (where $\epsilon_t = y_t^2 - g_t$), $g_t$ should be chosen such that $\epsilon_t$ for that particular $g_t$ is approximately stationary and negligible. The correlation between $g_t$ and $\epsilon_t$ should not be very large.

As mentioned previously, it would be expected that the better that $g_t$ approaches $y_t^2$, the more accurate the AQL estimates of the parameters are likely to be. This is because of the relationship between these two quantities, i.e.

$$ E(y_t^2|\mathcal{F}_{t-1}) - f_t^2(\theta) \approx g_t - f_t^2(\theta). $$

If, however, a $g_t$ cannot be found such that $g_t - f_t^2(\theta)$ is close to $E(M_t^2|\mathcal{F}_{t-1})$, a $g_t$ may be found such that this $g_t$ "mimics" $y_t^2$. By "mimics" it is meant that the change in successive observations of $y_t^2$ should be subsequently accounted for by $g_t$. Therefore, even if the graph of $g_t$ does not accurately approach $y_t^2$ it should model the pattern of the $y_t^2$. It will be seen from Figure 4.3 that the graph of $g_1$
is less accurate than the graph of \( g_2 \) due to the larger vertical distance between \( g_t \) and \( y_t^2 \). In these cases a constant may be added (or subtracted) to improve both the graphical approximation of \( y_t^2 \) by \( g_t \) and the accuracy of the corresponding AQL estimates.

In examining the adequacy of the predictable process \( g_t \), an analysis of the residuals \( \epsilon_t \) should be carried out. Analysis of the autocorrelation and partial autocorrelation functions of the residuals is of utmost importance. It is preferred that minimal correlation between \( g_t \) and \( \epsilon_t \) exists. Once the model is fitted, the residuals \{\( \epsilon_t \)\} should be a sequence of uncorrelated random variables with constant mean (0 in this case) and constant variance and the autocorrelation and partial autocorrelation functions of \{\( \epsilon_t \)\} should ideally be negligible.

It appears as though the sample mean of \( \epsilon_t \) would not help determine whether or not \( g_t \) adequately approaches \( y_t^2 \), the reason being that the sample mean for such a statistic will be very close to zero if the predictable process overestimates and then underestimates the true series \( y_t^2 \), the positive values of \( \epsilon_t \) might “balance” out the negative values of \( \epsilon_t \) thus producing a low mean value of \( \epsilon_t \).

As noted in Appendix B the ratio

\[
\frac{\text{E}(M_t^2 \mid \mathcal{F}_{t-1})}{|g_t - f_t^2|}
\]

should be bounded between two finite values for each \( t \). The above ratio may be re-written in the following form;

\[
\frac{\text{E}(M_t^2 \mid \mathcal{F}_{t-1})}{|g_t - f_t^2|} = \frac{\text{E}(y_t^2 \mid \mathcal{F}_{t-1}) - f_t^2}{|g_t - f_t^2|}
\]
\[
= \frac{|g_t - f_t^2| + E(\epsilon_t|F_{t-1})}{|g_t - f_t^2|} = 1 + \frac{E(\epsilon_t|F_{t-1})}{|g_t - f_t^2|}.
\]

If \(|g_t - f_t^2|\) is close to 0 and/or \(E(\epsilon_t|F_{t-1}) \gg |g_t - f_t^2|\) then the ratio (4.1) becomes larger leading to estimates which are not as accurate as they would be if this ratio was closer to 1. The closer the ratio is to 1, the quicker the convergence of the AQL estimates and the more accurate these estimates will be (for theoretical details see Mvoi, Lin and Biondini (1997). The resulting AQL estimates are also found to be very accurate providing the QL estimates are accurate. Even though in practical situations this ratio is unavailable, by adding a suitable constant \(c\) to the denominator in (4.1), the resultant ratio for each \(t\) is assured to be much smaller than it would be if this constant was not taken into consideration. Since the ratio is no longer large, this ensures convergence. It must be noted that this constant must not be too large as \(|g_t - f_t^2| + c\) will be dominated by the value of \(c\) and the value of the AQL estimates will be closer to the OLS estimates.

### 4.4 The Two-stage Estimation Procedure

Once \(g_t\) is determined the following algorithm was applied to find the parameter estimates via the two-stage estimation procedure using the AQLM;

1. Start with preliminary estimates of \(\hat{\theta}\). These initial values are determined via the method of OLS.
2. Substitute the $\hat{\theta}$ into the $g_t - f_t^2(\theta)$ term in the AQS normal equation and solve this equation for $\theta$; i.e.

\[ \text{i.e. } \tilde{G}_T^*(\theta) = \sum_{t=1}^{T} \frac{\dot{f}_t(\theta)}{|g_t - f_t^2(\theta)|} M_t = 0. \]

3. Let the solutions obtained in step (2) be the updated estimates of $\hat{\theta}$.

4. Repeat steps (2) and (3) until the sequence of estimates is convergent.

However, sometimes due to $|g_t - f_t^2(\theta)|$ being too close to zero the iteration may not converge quickly, a positive constant $c$ is added to $|g_t - f_t^2(\theta)|$.

As was seen in Chapter 3, a suitable constant $c$ may need to be added to $|g_t - f_t^2(\theta)|$. In the simulations conducted herein the initial value of $c$ is 0 and this value is increased to 0.01 if the AQL estimates do not appear to converge. The value of $c$ continues to increase by increments of 0.01 until the AQL estimates converge. It is important to note that $c$ must be increased only to the point of immediate convergence, a choice of $c$ which is significantly greater than that necessary for convergence might lead to the AQL estimates becoming closer to those obtained via the method of OLS and thus compromise the effectiveness of the procedure outlined here.

### 4.5 Application of Criteria

If the quantity $E(M_t^2|\mathcal{F}_{t-1})$ is known, the QLM can be used to estimate $\theta$ via (3.1). In practice this quantity is unknown and thus must be estimated. If
E(M_t^2 | \mathcal{F}_{t-1}) is non-constant, OLS is not a suitable method for estimating the unknown parameters. To demonstrate the criteria listed in the previous section and show how these criteria are to be applied, two simulated data sets are analysed in this section as well as a real-life data set which shows the practicality of the AQLM.

For a comparison of the accuracy of both the AQLM, the QLM and the method of OLS a quantity $S$ is introduced. $S$ is defined as

$$S = \sqrt{(\theta_0 - \hat{\theta}_0)^2 + (\theta_1 - \hat{\theta}_1)^2 + \ldots + (\theta_p - \hat{\theta}_p)^2},$$

where $\theta = (\theta_0, \ldots, \theta_p)^T$ is an unknown parameter and $\hat{\theta} = (\hat{\theta}_0, \ldots, \hat{\theta}_p)^T$ is an estimate of $\theta$. Since this value is determined by the true value of the parameters, it is a quantity which cannot be calculated using real-life data.

**Example 1** 240 data values are generated from the following model:

$$y_t = 0.3 + 0.5y_{t-1} + M_t, \quad t \geq 2,$$

where $M_t = N_t - 0.5(y_{t-1}^2 + y_{t-2}^2)$ and given $y_s, s \leq t$, $N_t$ has the Poisson distribution with rate $0.5(y_{t-1}^2 + y_{t-2}^2)$.

Accepting that the true model is

$$y_t = \theta_0 + \theta_1y_{t-1},$$

then following the practical procedure mentioned previously and noting that $\theta_0$ and $\theta_1$ are the unknown parameters, the AQS estimating function $G_T^*(\theta)$ gives:

$$G_T^*(\theta) = \sum_{t=1}^{240} \left( \frac{1}{y_{t-1}} \right) \frac{y_t - (\theta_0 + \theta_1y_{t-1})}{g_t - (\theta_0 + \theta_1y_{t-1})^2},$$

31
with

\[ E(M_t^2|\mathcal{F}_{t-1}) \approx g_t - (\theta_0 + \theta_1 y_{t-1})^2. \]

By letting \( G_T^*(\theta_0, \theta_1) = 0 \) the approximate roots of the AQS estimating function may be obtained.

For this random sample of 240 data values four different predictable processes \( g_t \) are analysed, the first \( g_t \) is that based on the ARIMA(1,0,0) model (denoted by \( g_{1t} \)), the second based on the ARIMA(1,1,0) model (denoted by \( g_{2t} \)), the third \( g_t \) analysed is that based on the ARIMA(2,0,0) model (denoted by \( g_{3t} \)) and the last is based on the ARIMA(2,1,0) model (denoted by \( g_{4t} \)). The predictable processes are listed below:

\[
\begin{align*}
g_{1t} &= 6.743 + 0.193 y_{t-1}^2, \\
g_{2t} &= 0.517 y_{t-1}^2 + 0.483 y_{t-2}^2, \\
g_{3t} &= 5.864 + 0.168 y_{t-1}^2 + 0.131 y_{t-2}^2, \\
g_{4t} &= 0.340 y_{t-1}^2 + 0.291 y_{t-2}^2 + 0.368 y_{t-3}^2.
\end{align*}
\]

The plots of \( y_t^2 \) and four possible \( g_t \)’s are shown in Figures 4.1 and 4.2. \( g_{2t} \) and \( g_{4t} \) are examined because it appears as though \( y_t^2 \) is non-stationary from the time-series plots. This is known from the generation of the data. The plots of the autocorrelation and partial autocorrelation functions of \( y_t^2 \) however do not reveal any significant non-stationarity. The second and fourth predictable processes, \( g_{2t} \) and \( g_{4t} \) (see Figure 4.2), are found to graphically approach the quantity \( y_t^2 \) better than if the two other predictable processes are chosen. The first and third
predictable processes also appear to be good if the constant is not taken into consideration but rather only the square of the autoregressive terms. If this is done however, the predictable process \( g_t \) will be poor at approaching \( y_t^2 \) when the observed values of \( y_t^2 \) are relatively large. Also, the correlation between \( g_t \) and \( \epsilon_t \) is much smaller when \( g_{2t} \) or \( g_{4t} \) are chosen.

From Figure 4.1 the ARIMA(1,0,0) model is not good at approaching \( y_t^2 \) if the constant term is taken into account. Similarly it can also be seen that the ARIMA(2,0,0) model is also not good at approaching \( y_t^2 \). The \( \{\epsilon_t\} \) are approximately stationary for each of the four \( g_t \)'s but only negligible for the predictable processes where differencing is taken into consideration. From Table 4.1 it is seen that the most accurate AQL estimates occur when the second predictable process is used. The \( S \)-values for the AQL estimates when each predictable process is considered are 0.111, 0.009, 0.080 and 0.092 respectively. In the next section it will be seen that the ARIMA(1,1,0) model (i.e. \( g_{2t} \)) is always better at graphically approaching \( y_t^2 \) and in the vast majority of simulations provides a much better estimate of the unknown parameters than a predictable process which does not consider possible non-stationarity. The correlation between \( g_t \) and \( \epsilon_t \) is minimised when either \( g_{2t} \) or \( g_{4t} \) are selected to approach \( y_t^2 \). In this simulation, the \( S \)-value associated with the estimates obtained when the ARIMA(1,1,0) process is used is much smaller than when each of the other predictable processes are applied and are surprisingly much smaller than when the QLM is invoked.
Figure 4.1: $y_t^2$ (hard line) and two possible $g_t$'s (dotted lines) for Example 1.
Figure 4.2: $y_t^2$ (hard line) and two possible $g_t$'s (dotted lines) for Example 1.
Table 4.1: OLS and AQL estimates (for four possible $g_t$'s) for Example 1.

Example 2 240 data values are generated from the following model:

$$y_t = 0.2 + 0.6y_{t-1} + 0.8x_t + M_t, \ t \geq 2,$$

where given $y_s, s \leq t, M_t$ is generated from the normal distribution with mean 0 and variance $0.3(y^2_{t-1} + y^2_{t-2})$, i.e. $E(M_t|\mathcal{F}_{t-1}) = 0$ and $E(M_t^2|\mathcal{F}_{t-1}) = 0.3(y^2_{t-1} + y^2_{t-2})$. $x_t$ is a standard normal random variable and is independent of $y_s, s \leq t$.

For this random sample of 240 data values two different predictable processes are analysed. The first $g_t$ is that based on the ARIMA(1,0,0) model (denoted by $g_{1t}$) and the second based on the ARIMA(1,1,0) model (denoted by $g_{2t}$). The predictable processes are listed below;

$$g_{1t} = 2.059 + 0.539y^2_{t-1},$$

$$g_{2t} = 0.603y^2_{t-1} + 0.397y^2_{t-2}.$$

The coefficients of each $g_t$ for all predictable processes analysed are significant. $g_{2t}$ is examined because it appears as though $y^2_t$ is non-stationary from the time-series plots. The second predictable process $g_{2t}$ is found to graphically approach
the quantity $y_t^2$ better than if the other predictable process ($g_{1t}$) is chosen. It is the only predictable process where the estimates of both parameters are within two standard errors of the true values of the parameters (in fact both estimates are within one standard error of the true value).

In the sample, the first $g_t$ is based on the ARIMA(1,0,0) model and the second $g_t$ is based on the ARIMA(1,1,0) model. Therefore, a comparison will be made between both models with the first order autoregressive component, that with the differencing term ($g_{1t}$) and that without the differencing term ($g_{2t}$). From Figure 4.3 the latter predictable process is found to approach the quantity $y_t^2$ better than if the former predictable process is chosen. The $\{\epsilon_t\}$ are approximately stationary when either $g_{1t}$ and $g_{2t}$ are selected but are only negligible for the predictable process where differencing is taken into consideration (i.e. $g_{2t}$). The AQL estimates when each of the two $g_t$'s are chosen are shown in Table 4.2. However, the estimates when $g_t$ is based on the ARIMA(1,0,0) model result in an $S$-value of 0.234 whereas the $S$-value, when the second predictable process is selected, is equal to 0.054. The estimates of each parameter can be seen to be much more accurate when the second predictable process is selected. The first predictable process in this particular simulation even produces an $S$-value which is greater than that obtained via the method of OLS. The OLS method yields estimates of 0.274, 0.552 and 0.677 respectively which results in an $S$-value of 0.152. On the other hand the estimates via the QLM are 0.219, 0.657 and 0.796 respectively and the resultant $S$-value is 0.060. The AQL estimates when $g_{2t}$ is chosen as the predictable process are very similar to the QL estimates.
Figure 4.3: $y_t^2$ (hard line) and two possible $g_t$'s (dotted lines) for Example 2.
Table 4.2: OLS and AQL estimates (for two possible \( g_i \)’s) for Example 2.

<table>
<thead>
<tr>
<th>Method</th>
<th>( \hat{\theta}_0 )</th>
<th>( \hat{\theta}_1 )</th>
<th>( \hat{\theta}_2 )</th>
<th>S-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.274</td>
<td>0.552</td>
<td>0.677</td>
<td>0.152</td>
</tr>
<tr>
<td>AQLM((g_1))</td>
<td>0.348</td>
<td>0.530</td>
<td>0.629</td>
<td>0.234</td>
</tr>
<tr>
<td>AQLM((g_2))</td>
<td>0.200</td>
<td>0.590</td>
<td>0.747</td>
<td>0.054</td>
</tr>
<tr>
<td>QLM</td>
<td>0.219</td>
<td>0.657</td>
<td>0.796</td>
<td>0.060</td>
</tr>
</tbody>
</table>

Example 3 The experiment involves aiming a beam \( a \), having various values of incident momentum \( p_{a}^{lab} \) which are measured in the laboratory frame of reference, at a target containing protons and results in the emission of other particles. The quantity measured, \( y \), is the scattering cross-section of a particular particle. A quantity of more basic significance than \( p_{a}^{lab} \) is \( s \), the square of the total energy in the centre-of-mass frame of reference system. The quantity \( s \), under the conditions of the experiment is given by

\[
s = 2m_{p}p_{a}^{lab},
\]

where \( s \) is measured in \((GeV)^2\), where \( 1GeV = 1 \times 10^9 \) electron volts is the energy that an elementary particle reaches on being accelerated by an electric potential.
of one billion volts. The momentum $p_a^{ab}$ and the mass $m_p$ are measured in $GeV$, and $m_p = 0.938GeV$ for a proton.

Theoretical physicists believe that, under certain conditions (satisfied by this experiment), the cross-section $y$ is given by the model

$$y = \theta_0 + \theta_1 s^{-1/2} + \text{relatively small terms.}$$

Table 4.3 summarises the results of the experiment. At each $p_a^{ab}$, a very large number of particles $N_a$ were used so that the variance of the observed $y$ values could be accurately obtained from theoretical considerations. The square roots of these deviations are given in the fourth column of Table 4.3. The best approach to estimate $\theta_0$ and $\theta_1$ is by using the weighted least square (WLS) method subject to the estimated standard variances being known. Here the AQLM is applied to only one single path of realisation.

The estimates obtained via the AQLM are compared with those via the WLS and OLS methods and are given in Table 4.4. A possible predictable process is shown in Figure 4.4. The AQLM is applied to only one single path of realisation based on $g = 29705 + 856104(s^{-1/2})^2$. Although Weisberg does not state whether or not the data is ordered by time a look at the data will show that a strong linear relationship appears to exist between $s^{-1/2}$ and the estimated standard deviation and thus a time-ordered sequence may also appear plausible. The AQL estimates are closer to the WLS estimates than the OLS estimates for each of the two parameters. This indicates that in the availability of only one realisation of data and lacking any knowledge regarding the nature of the error, the AQLM
Table 4.3: Data for Example 3.

<table>
<thead>
<tr>
<th>$p_{lab}^{1/2}$ GeV/c</th>
<th>$s^{-1/2}$ GeV/c$^{-1}$</th>
<th>$y$ (µb)</th>
<th>Estimated Stand. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.345</td>
<td>367</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>0.287</td>
<td>311</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>0.251</td>
<td>295</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>0.225</td>
<td>268</td>
<td>7</td>
</tr>
<tr>
<td>12</td>
<td>0.207</td>
<td>253</td>
<td>7</td>
</tr>
<tr>
<td>15</td>
<td>0.186</td>
<td>239</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>0.161</td>
<td>220</td>
<td>6</td>
</tr>
<tr>
<td>30</td>
<td>0.132</td>
<td>213</td>
<td>6</td>
</tr>
<tr>
<td>75</td>
<td>0.084</td>
<td>193</td>
<td>5</td>
</tr>
<tr>
<td>150</td>
<td>0.060</td>
<td>192</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 4.4: WLS, OLS and AQL estimates for Example 3.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{\theta}_0$</th>
<th>$\hat{\theta}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WLS</td>
<td>148.47</td>
<td>530.84</td>
</tr>
<tr>
<td>OLS</td>
<td>135.00</td>
<td>619.71</td>
</tr>
<tr>
<td>AQLM</td>
<td>148.93</td>
<td>537.46</td>
</tr>
</tbody>
</table>

provides the best estimates.

4.6 Simulations

In this section one hundred simulations are performed via each of the two models mentioned previously and the results discussed.

The OLS and AQL estimates (i.e. average of the one hundred simulations) based
Table 4.5: OLS and AQL estimates for one hundred simulations of Example 1.

<table>
<thead>
<tr>
<th>Method</th>
<th>mean(\hat{\theta}_0) (st. err (\hat{\theta}_0))</th>
<th>mean(\hat{\theta}_1) (st. err (\hat{\theta}_1))</th>
<th>(\bar{S})</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.329(0.012)</td>
<td>0.441(0.013)</td>
<td>0.019</td>
</tr>
<tr>
<td>AQLM(g_1)</td>
<td>0.318(0.006)</td>
<td>0.462(0.011)</td>
<td>0.013</td>
</tr>
<tr>
<td>AQLM(g_2)</td>
<td>0.296(0.004)</td>
<td>0.493(0.009)</td>
<td>0.009</td>
</tr>
<tr>
<td>AQLM(g_3)</td>
<td>0.317(0.006)</td>
<td>0.466(0.011)</td>
<td>0.013</td>
</tr>
<tr>
<td>AQLM(g_4)</td>
<td>0.306(0.005)</td>
<td>0.486(0.011)</td>
<td>0.012</td>
</tr>
</tbody>
</table>

on the model in the first example are included in Table 4.5 along with the standard errors associated with the estimates. From the table it can be seen that the OLS estimates are very inaccurate for both parameters. The mean value of the estimates of \(\hat{\theta}_0\) and \(\hat{\theta}_1\) over one hundred simulations via the AQLM (when the second predictable process is used to approach \(y_i^2\)) are very accurate (0.296 and 0.493 respectively with standard errors of 0.004 and 0.009 respectively). The AQL also yields accurate estimates when \(g_4t\) is used to approach \(y_i^2\). Using the method of OLS, however, the mean of the estimates of the two unknown parameters are 0.329 and 0.441 respectively. The corresponding standard errors of these estimates are 0.012 and 0.013 respectively.

It is obvious from these results that the method of OLS is very inaccurate and since the QLM is unable to be applied the AQLM is very effective providing an appropriate \(g_t\) is chosen.

\(\bar{S}\) is defined as the average of the \(S\)-values from the one hundred simulations, the \(S\)-value being previously defined. The \(\bar{S}\)-value for the estimates via the AQLM (when \(g_{2t}\) is used) is 0.009 whereas the \(\bar{S}\)-value equals 0.019 for the OLS method.
The corresponding $\bar{S}$-value for each of the three remaining predictable processes are significantly lower than the resulting $\bar{S}$-value when the method of OLS is applied. $g_{2t}$ always approaches $y_t^2$ better than any other possible $g_t$. In the vast majority of simulations it provides more accurate estimates than is the case when any other predictable process is chosen.

The OLS and AQL estimates for the simulations based on the model in the second example are included in Table 4.6. From the table it can be seen that the OLS estimates are not as accurate as the AQL estimates when taking the mean of the one hundred sets of estimates. The estimates of $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\theta}_2$ when the method of OLS is invoked are 0.211, 0.541 and 0.844 respectively. The mean values of the estimates of $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\theta}_2$ over one hundred simulations via the AQLM when the predictable process $g_{2t}$ is chosen are 0.191, 0.591 and 0.808 respectively and the first predictable process $g_{1t}$ yields a mean value of the estimates of 0.205, 0.581 and 0.781 respectively. The use of either of the two predictable processes leads to more accurate estimates than the method of OLS. The mean of the estimates of the parameters when any of the predictable processes is chosen is within two standard errors of the true values of the unknown parameters but when OLS is invoked only the mean of the $\hat{\theta}_0$ lies within two standard errors of the true values.

<table>
<thead>
<tr>
<th>Method</th>
<th>mean$\hat{\theta}_0$(st. err $\hat{\theta}_0$)</th>
<th>mean$\hat{\theta}_1$(st. err $\hat{\theta}_1$)</th>
<th>mean$\hat{\theta}_2$(st. err $\hat{\theta}_2$)</th>
<th>$\bar{S}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.211(0.010)</td>
<td>0.541(0.009)</td>
<td>0.844(0.010)</td>
<td>0.018</td>
</tr>
<tr>
<td>AQLM($g_1$)</td>
<td>0.205(0.010)</td>
<td>0.581(0.011)</td>
<td>0.781(0.012)</td>
<td>0.019</td>
</tr>
<tr>
<td>AQLM($g_2$)</td>
<td>0.191(0.007)</td>
<td>0.591(0.009)</td>
<td>0.808(0.010)</td>
<td>0.015</td>
</tr>
</tbody>
</table>

Table 4.6: OLS and AQL estimates for one hundred simulations of Example 2.
The $\bar{S}$-value when the predictable process $g_{2t}$ is chosen is equal to 0.015, slightly smaller than when $g_{1t}$ is chosen (0.019) and that via the method of OLS (0.018).

From the graphs of $y_t^2$ and $g_t$ it is seen in each and every simulation of both models that information is given as to which predictable process will produce the better estimates of the unknown parameters. If $g_t$ approaches $y_t^2$ very well then the AQL estimates will be much more accurate than if this approximation was not as good. How accurate they will be will depend on the accuracy of the corresponding QL estimates. Obviously, if these QL estimates are not accurate the corresponding AQL estimates will, more than likely, be not as precise as they would be if the QL estimates were accurate.

For the higher order models, the approximation of $y_t^2$ by $g_t$ was not as good as it was for the lower order models. This could well be due to the fact that the $y_t$ is generated to be dependent upon $y_{t-1}$ and not dependent upon the observation of $y$ at higher order lags. However, lower order processes will be more responsive to outliers than higher order processes and therefore place more weight upon the previous observation of $y_t^2$. Higher order models tend to spread the weight upon previous values (e.g. the ARIMA(1,1,0) model will consider only the $y_{t-1}^2$ and $y_{t-2}^2$ terms whereas an ARIMA(2,1,0) process will consider the $y_{t-1}^2$, $y_{t-2}^2$ and $y_{t-3}^2$ terms). Higher order autoregressive processes will therefore produce a "smoothing" effect on the predictable process, thus it is preferred that a lower order $g_t$ be chosen. Taking into account that $y_t^2$ may possibly be non-stationary is important when selecting $g_t$. The graphs of $g_t$ and $y_t^2$ seem to be very similar when
the $g_t$ is based on the ARIMA(1,1,0) model. However, even if the $g_t$ is selected by not taking into consideration any possible non-stationarity the estimates via QL will usually be better than the OLS estimates.

4.7 Conclusion

The method of OLS does not yield accurate estimates of the unknown parameters because it assumes the residuals are uncorrelated and have equal variance. The QL estimates are accurate when the form of $E(M_t^2|\mathcal{F}_{t-1})$ is known, which in practical situations is not true. In the first model, the AQLM is much better than the method of OLS. It appears that if the residuals do not have a common variance the method of OLS will be very inaccurate indeed. In the second model, the OLS estimates are once again inaccurate (though not as inaccurate as they were in the first model where the errors are observations from the Poisson distribution). The estimates obtained via the AQLM will be much more accurate than the OLS estimates for this particular model providing a suitable predictable process $g_t$ is chosen. From the practical procedure outlined here, the AQLM takes account of the effect from errors and, in general, this method will always produce a better estimate of the parameters than will the method of OLS.

The quality of improvement in applying the AQLM can also be maximised by careful selection of a predictable process $g_t$. If this $g_t$ approaches the quantity $y_t^2$ very well then the AQL estimates will be more accurate. That is why this selection
is very important. The OLS estimates can be improved upon in most cases but to maximise this improvement the best possible $g_t$ must be chosen. Furthermore, when a constant $c$ is added the AQL estimates will converge immediately as the ratio $\frac{E(M_t^2|\mathcal{F}_{t-1})}{c+|g_t-\bar{f}_t^2|}$ in the AQS function (which is unable to be calculated in practice) will be closer to 1. There is no significantly large value of the ratio for any $t$. 
Figure 4.4: $X_t^2$ (hard line) and a possible $g_t$ (dotted line) for Example 3.
Chapter 5

Application of the Asymptotic Quasi-likelihood Method in Fractional ARIMA\((p, d, q)\) Models

5.1 Introduction

In this chapter, the application of the AQLM to fractional ARIMA\((p, d, q)\) processes is examined. Several examples will be given to show how the AQLM applies in Hurst’s \(R/S\) analysis and the comparison between the estimates of the Hurst parameter \(H\) obtained by the AQLM and the traditional method of OLS will be
Fractional ARIMA($p, d, q$) processes are defined by

$$\Phi(B)(1 - B)^d X_t = \Theta(B)\epsilon_t,$$

(5.1)

where the $\epsilon_t$'s are independently and identically distributed (i.i.d) normal random variables with mean 0, $B$ denotes the backward shift operator, $\Phi(B)$ denotes the autoregressive component(s) of the process and $\Theta(B)$ denotes the moving-average component(s) of the process. $(1 - B)^d$ denotes the fractional differencing operator and is defined as $\sum_{k=0}^{\infty} d C_k (-B)^k$. If $d$ is larger than or equal to 1, then the original series $X_t$ is not stationary and must be differenced $d$ times in order for it to become stationary. For example if $X_t$ is defined as in (5.1) with $d = 1.2$ then the differenced process $X_t - X_{t-1}$ is the stationary solution of (5.1) with $d = 0.2$.

The Hurst parameter ($H$) is a measure of the intensity of long-range dependence of a particular time-series. Fractional autoregressive integrated moving-average or fractional ARIMA($p, d, q$) processes (with $0 < d < \frac{1}{2}$, since the process is not stationary if $d \geq \frac{1}{2}$) are examples of asymptotically second-order self-similar processes with Hurst parameter $H = d + \frac{1}{2}$ (providing the process under consideration has finite variance) or $H = d + \frac{1}{\alpha}$ (if the process being analysed possesses infinite variance). Taqqu and Teverovsky (1996) examined fractional ARIMA($p, d, q$) models with both finite and infinite variance structures and found that the resulting estimates are not unduly influenced when either of the variance structures are considered. The attention, herein, is limited to finite variance structures. The alternative to modelling long-range dependence via self-similar processes is via time-series methods which would involve more parameters as the sample size.
increases thus making the analysis and interpretation of the results even more complicated.

A self-similar process $X_t$ has the property that when aggregated (leading to a shorter time-series in which each point is the sum of multiple original points) the new series has the same autocorrelation function as the original series. Willinger et al's (1995) definition of self-similar processes will be adopted here. This definition is now outlined. Let $X = (X_t : t = 0, 1, 2, \ldots)$ be a covariance stationary (sometimes called wide-sense stationary) time-series with mean $\mu$, variance $\sigma^2$ and autocorrelation function $r_k, k = 0, 1, 2, \ldots$. A series is distributionally self-similar if the distribution of the aggregated time-series is the same (except for changes in scale) as that of the original. Hence, self-similar processes exhibit long-range dependence. Such processes have autocorrelation functions of the form

$$r_k \approx k^\beta L_1(k), \quad \text{as } k \to \infty,$$

where $\beta$ is a constant such that $0 < \beta < 1$ and $L_1$ is slowly decaying at infinity, thus indicating long-range dependence. In other words,

$$\lim_{k \to \infty} \frac{L_1(kx)}{L_1(k)} = 1, \quad \text{for all } x > 0.$$

For each $m = 1, 2, 3, \ldots$ $X^{(m)}$ is given by

$$X_k^{(m)} = \frac{1}{m} \sum_{i=km-m+1}^{km} X_i, \quad k = 1, 2, 3, \ldots$$

Note that for each $m$ the aggregated time-series $X^{(m)}$ defines a covariance stationary process; let $r_k^{(m)}$ denote the corresponding autocorrelation function. The process $X$ is (exactly) self-similar with self-similarity parameter $H = 1 - \beta/2$ if,
for all \( m = 1, 2, 3, \ldots \), the left-hand side of (5.3) has the same finite-dimensional
distribution as \( X \). It is (exactly second-order) self-similar with self-similarity pa-
rameter \( H = 1 - \beta/2 \) if, for all \( m = 1, 2, 3, \ldots \), the left-hand side of (5.3) has the
same variance and autocorrelation as \( X \). In terms of the aggregated processes
\( X^{(m)} \), this means that, for all \( m = 1, 2, 3, \ldots \),

\[
\text{Var}(X^{(m)}) = \sigma^2 m^{-\beta}, \quad \text{and}
\]

\[
r_k^{(m)} \to \frac{1}{2d^2(2-\beta)}, \quad \text{as } m \to \infty, \quad k = 0, 1, 2, \ldots,
\]

where \( d^2(f) \) denotes the second central difference operator applied to a function
\( f \) such that \( d^2(f(k)) = f(k+1) - 2f(k) + f(k-1) \). Therefore, an asymptotically
self-similar process has the property that, for large \( m \), the corresponding aggre-
gated time-series \( X^{(m)} \) has a fixed correlation structure, solely determined by
\( \beta \). Moreover, due to the asymptotic differencing and differentiating, \( r_k^{(m)} \) agrees
asymptotically with the correlation structure of \( X \) given by (5.2).

Intuitively, the most striking feature of (exactly or asymptotically) self-similar
processes is that their aggregated processes \( X^{(m)}_k \) possess a non-degenerate cor-
relation structure as \( m \to \infty \). In other words,

\[
\sum_{k=1}^{m} r_k^{(m)} \to \infty, \quad \text{as } m \to \infty, \quad k = 1, 2, 3, \ldots.
\]

This behaviour occurs when \( \frac{1}{2} < H < 1 \). It is in stark contrast to the more con-
ventional models, all of which have the property that their aggregated processes
\( X^{(m)} \) tend to second-order pure noise (as \( m \to \infty \)). In other words,

\[
\sum_{k=1}^{m} r_k^{(m)} \to 0, \quad \text{as } m \to \infty, \quad k = 1, 2, 3, \ldots
\]
Note that the above definition of self-similarity was chosen over the mathematically more convenient definition of a self-similar continuous-time stochastic process \( X = (X_t : t \geq 0) \) with mean zero and stationary increments, namely, for all \( a > 0 \),

\[
X_{ak} = a^H X_t,
\]

where the exponent \( H \) is the self-similarity parameter. Therefore \( X_{ak} \) and \( a^H X_t \) have the same finite-dimensional distributions.

Several heuristic methods are available in detecting and estimating the level of long-range dependence in a given set of observations. Some approaches are listed below;

1. Analysis of the variances of the aggregated processes;
2. Smooth periodogram analysis in the spectral domain;
3. Analysis of the \( R/S \) statistic for different block sizes.

Mandelbrot (1972) investigates the pitfalls of using correlation methods to estimate the Hurst parameter and discusses the \( R/S \) method in analysing time-series. The correlation methods are useful as purely diagnostic tools.

In the following section the rescaled adjusted range (\( R/S \)) procedure is outlined. \( R/S \) analysis is a robust procedure to analyse data that is suspected to be non-Gaussian. Mandelbrot and Wallis (1969) have shown that \( R/S \) analysis can detect long-range dependence in highly non-Gaussian time-series with large skewness.
and kurtosis. The statistic is robust when the data appear not to be independent of each other but rather seem to be dependent upon previous patterns, i.e. exhibit a certain memory.

Hurst's empirical power law relation is discussed in the next section. The range and the standard deviation in Hurst's rescaled adjusted range statistic is also defined and the effectiveness of $R/S$ analysis for the fractional ARIMA($p,d,q$) model is discussed. Also examined is the effect of a well-controlled short-range dependence structure on $R/S$ analysis (i.e. the order of $p$ and/or $q$ is not equal to 0). Simulations are performed to show what effect such a short-range dependence structure has on the accuracy of the resulting estimates of $H$. The estimates of $H$ via $R/S$ analysis using the AQLM are compared to those obtained where the method of OLS is employed and to those obtained by Taqqu and Teverovsky (1996). All simulations involve the Durbin-Levinson algorithm in S-plus.

### 5.2 The R/S Method

Harold Edwin Hurst was an English geophysicist who spent a lifetime studying the River Nile and the problems related to water storage. As a result of such an in-depth analysis he earnt the title "Abu Nil" (Father of the Nile). The objective was to maximise the use of the Nile waters (notably the Aswan Dam). One problem in particular involved the designing of a reservoir on the Nile, the aim being to determine how high the reservoir should be to contain all incoming
river water whilst simultaneously providing a constant outflow. Hurst, Black and Simaika (1965) took data on the annual flow of the Nile and other rivers, and computed the range \( R(n) \), defined as

\[
R(n) = \max_{1 \leq k \leq n} \sum_{j=1}^{k} (X_j - \bar{X}_n) - \min_{1 \leq k \leq n} \sum_{j=1}^{k} (X_j - \bar{X}_n),
\]

where \( n \) is the time-span considered and \( \bar{X}_n \) is the average of the \( n \) \( X \) values.

The standard deviation, denoted by \( S(n) \), is defined as

\[
S(n) = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (X_j - \bar{X}_n)^2}.
\]

Mathematically speaking, \( X_k \) is defined as the position at time \( k \) of a particle that walks at random with steps of unit length on a line. The sample size is \( N \) and \( n \) is the time-span considered.

Hurst, Black and Simaika (1965) found that many naturally occurring time-series appear to be well represented by the relation

\[
\text{E}[R(n)/S(n)] = bn^H, \quad \text{as } n \to \infty,
\]

with Hurst parameter \( 0 < H < 1 \), and \( b \) is a finite positive constant that does not depend on \( n \).

To obtain more information on the nature of the quantity \( R(n)/S(n) \) a common approach is to partition the \( N \) sample values into \( m \) equal sub-samples \((X_1, \ldots, X_n), (X_{n+1}, \ldots, X_{2n}), \ldots, (X_{(m-1)n+1}, \ldots, X_{mn})\). So for a given \( n \), there are \( N/n \) different \( R/S \) values, the mean of which is the statistic analysed.

Mandelbrot and Wallis (1969) have shown that the division of \( R \) by \( S \) leads to robustness against extreme deviations from normality, including the infinite
variance syndrome. It is particularly robust with respect to heavy-tailed distributions. The biggest drawback, however, is the loss of efficiency under Gaussian models than is the case with ML estimators, and thus this method does not necessarily minimise the bias.

The modified $R/S$ statistic, introduced by Lo (1991), corrects Hurst’s classical $R/S$ statistic, allowing for the effects of possible short-term dependence. The resulting statistic is found to be invariant over the general class of short-memory processes but deviates for long-memory processes.

A straight line is then plotted in the log-log scale:

$$\log[R(n)/S(n)] = c + H\log(n) + \epsilon(n)$$

and the estimate of $H$ is obtained via the OLS method.

The quantity $\log[R(n)/S(n)]$ plotted against $\log(n)$ is known as the rescaled adjusted range plot (also called the pox diagram of $R/S$) where $n = n_1, n_2, \ldots$ and $n_1 < n_2 < \ldots$, $n$ denoting the window size. A typical rescaled adjusted range plot commences with a transient region representing the short-range dependence structure in the sample (in this transient region the $R/S$ statistic increases more rapidly than $n^\frac{1}{2}$ for small $n$ than it does for relatively larger $n$), thus the need to examine the accuracy of the estimates when different initial window sizes are considered. With smaller initial window sizes the power of $R/S$ analysis may be severely compromised. This behaviour, however, eventually settles down and fluctuates in a straight “street” with a certain asymptotic slope. For a fractional Gaussian process this graph should have slope $H$ as $n$ increases. If the model
under consideration has a well controlled short-range dependence structure (e.g. Fractional Gaussian ARIMA(0,d,0)) the R/S method will always provide a very precise estimate of the parameter $H$. Herein the emphasis is on models with additional short-range dependence components.

Note the similarity between (5.4) and (1.1), the standard method for estimating $H$ is based on the assumption that the $\{\varepsilon(n)\}$ are independently and identically distributed. However, this assumption is not necessarily true based on the R/S estimation procedure mentioned above. Therefore it would be expected that the OLS method may not provide a good estimate of $H$. Rather than applying the method of LS in the R/S analysis the AQLM is implemented to estimate $H$.

Bodruzzaman et al (1991) suggest an alternative procedure of applying the R/S method. It is a simpler version of the common method of applying the R/S method and will be adapted here. Defining a window as the segment of the particular time-series, the beginning of the window is not allowed to move but the size of the window is doubled every time the R/S ratio is calculated. This is in contrast to the usual application of the R/S method where, for each window size $n$, there are $N/n$ different R/S values ($N$ being the sample size), the mean of which is the statistic analysed. In Bodruzzaman et al’s (1991) method there is only one observed value of the R/S statistic for each $n$, simplifying the resulting analysis. The estimated slope is the estimate of the Hurst parameter. By applying this method (when $N=8192$) twelve values of $R/S (R(n))/S(n)$, where $n_t = 2^{t+1}$ and $t = 1, 2, ...12$) are obtained from the original 8192 observations. 4 observa-
tions (i.e. \( t = 1, 2, \ldots, 12 \)) are used to commence with but the results via \( R/S \)
analysis when the initial window size is 8 \( (t = 2, 3, \ldots, 12) \) and 16 \( (t = 3, 4, \ldots 12) \)
respectively are also examined. This is done because, as mentioned before, the
usual rescaled adjusted range plot commences with a transient region in which
the quantity \( R/S \) grows faster than \( n^{\frac{1}{2}} \) for small \( n \). Therefore small values of
\( n \) should be discarded when calculating the slope so as not to unduly influence
the resulting estimate of \( H \). An objective is to reduce any bias which may ex­
ist from the initial transient region of the \( R/S \) plot (as mentioned previously).
The emphasis is not on the short-range dependence structure but on the nature
of long-range dependence in the sample. Mandelbrot and Taqqu (1979) suggest
\( n \approx 10 \).

Attention is focussed on the power of \( R/S \) analysis on the fractional ARIMA\( (p, d, q) \)
model. Under the scheme of OLS the \( R/S \) method can be affected by a variety
of factors, namely;

1. the range of \( d \),

2. the order of the autoregressive and/or moving-average components and their
   respective coefficients, and

3. the fact that the \( \{\epsilon(n)\} \) in equation (5.4) may possess non-constant variance
   and are thus not independent.

In the absence of short-range dependence, the estimates of the Hurst exponent \( H \)
via \( R/S \) analysis are found to be biased towards 0.72. More specifically, in using
the empirical Hurst law the estimate of $H$ when the true value of $H$ is less than 0.72 tends to be overestimated and the estimate of $H$ when the true value of $H$ is greater than 0.72 tends to be underestimated. Where short-range dependence does exist the estimate of the Hurst parameter is biased towards 0.8.

The second point has been addressed by Taqqu and Teverovsky (1996). They applied the $R/S$ method to data simulated from a fractional ARIMA($p, d, q$) model and found that this estimator does not work as well when the order of either $p$ or $q$ is not zero. Using $R/S$ analysis in situations where short-range dependence is also present leads to biases in the final estimate of the parameter $d$ (and $H$). Another interesting result that Taqqu and Teverovsky found in analysing a process with a short-term dependence structure is that, if the coefficient chosen for $p$ and/or $q$ is negative, there will be significantly less induced bias in the estimate than would be the case if the coefficients are positive. $R/S$ analysis is biased in this case even though the estimator is still efficient.

The third point is an important reason for introducing the AQLM to the $R/S$ procedure. The usual method of applying OLS linear regression to the data transformed via $R/S$ analysis will not provide a precise estimate of $H$ when the residuals possess non-constant variance. The method of OLS places more restrictions regarding the errors and ignores correlated $\epsilon_t$. The AQLM, in such circumstances, would appear to be effective as it makes less assumptions regarding the structure of the errors.

Note that fractional ARIMA($p, d, q$) processes (with $0 < d < \frac{1}{2}$) exhibit long-range
dependence where the parameter $d$ determines the level of long-range dependence whilst short-range dependence is modelled through the parameters $p$ and $q$. The effectiveness of several estimators used by Taqqu and Teverovsky (1996) to estimate $d$ decreases when there is an additional short-term dependence structure (i.e. when either the order of $p$ and/or $q$ is not equal to zero). The results for the $R/S$ analysis are dependent on the number of sub-intervals and the minimum and maximum lags chosen.

The effect of different initial window lengths on the estimates will now be examined.

Example 4 Data is simulated from the fractional ARIMA($1,0.22,0$) process

$$(1 - 0.8B)(1 - B)^{0.22}X_t = \epsilon_t,$$

where $\epsilon_t$ is white noise. Note that the true value of $H$ equals 0.72. One hundred data sets are simulated from the above model. Taking the initial window size to be 4, 8 and 16 respectively the average of the one hundred estimates of $H$ are given in Table 5.1. It is seen that the estimate becomes more precise when the initial observations are discarded before applying the method of OLS. The standard error of each initial window size is comparable but the value of $\bar{S}$ is larger for the initial window size of 4. When the initial window size is 16 the value of $\bar{S}$ is even smaller. Generally the initial window length (or lag), $n$, is taken to be about 10. As can be seen, increasing the initial window length may well lead to improved estimates of $H$. This poses the interesting question of how to determine the initial window length, the answer to which is not very clear.
It is now shown how to apply the AQLM to the data and obtain precise estimates of $H$ via the simulating of data from the fractional ARIMA(2,0.3,0) process.

**Example 5** The ARIMA model is specified as

$$(1 - 0.2B - 0.6B^2)(1 - B)^{0.3}X_t = \epsilon_t,$$

where $B$ denotes the backward shift operator and $\epsilon_t$ is white noise.

The process is stationary and the value of $H$ must be estimated, which from the selection of $d$ is known to be 0.8 (since $H = d + \frac{1}{2}$). The data was analysed using the $R/S$-statistic and the estimate of the Hurst parameter is obtained via both the method of OLS and the AQLM. 8192 data values were simulated from this model and by applying $R/S$ analysis to the data the 8192 data values are transformed to twelve data points. Considering model (5.4) where $t = 2, 3, ..., 12$ and $X_t = \log[R(n_t)/S(n_t)]$ (where $n_t = 2^{t+1}$ and $f_t(\theta) = C + H \log n_t$). Based on the procedure of AQL mentioned previously, three possible $g_t$'s are determined based on this sample of twelve data points $X_t$. The predictable processes are listed below;

$$g_1 = -0.743 - 0.312X_{t-1}^2 + 0.850(\log n_t)^2,$$
Table 5.2: OLS and AQL estimates for Example 5.

<table>
<thead>
<tr>
<th>Method</th>
<th>OLS(4)</th>
<th>OLS(8)</th>
<th>OLS(16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.128</td>
<td>0.138</td>
<td>0.103</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>AQL($g_1$)</th>
<th>AQL($g_2$)</th>
<th>AQL($g_3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.175</td>
<td>0.066</td>
<td>0.039</td>
</tr>
</tbody>
</table>

\[
g_2 = -1.426 - 0.477X_{t-1}^2 - 0.427X_{t-2}^2 + 1.721(\log n_t)^2,
\]

\[
g_3 = -1.714 - 0.536X_{t-1}^2 - 0.486X_{t-2}^2 - 0.094X_{t-3}^2 + 1.282(\log n_t)^2.
\]

These three possible $g_t$'s are displayed in Figure 5.1. Note that $g_{1t}$ commences at the first lag, $g_{2t}$ commences at the second lag and $g_{3t}$ commences at the third lag. According to the criterion discussed in Chapter 4 for selecting $g_t$, it can be seen from Figure 5.1 that $g_{1t}$ is not as good at approaching $X_t^2$ as the other two predictable processes. It can be seen that there is very little difference between $g_{2t}$ and $g_{3t}$. Turning the attention to the \( \{X_t^2 - g_t\} \), they can be accepted as approximately stationary and negligible for each of the three $g_t$'s. From Table 5.2 it is seen that the most precise AQL estimate occurs when the third predictable process is used (0.839) followed by the $g_{2t}$ (0.866). When the first predictable process is chosen the resulting estimate is 0.975. The OLS estimate is 0.928 (when the initial window size is 4).
This example shows the possibility of improving the estimate of $H$ via the AQLM is achieved by choosing a better $g_t$ to approach the quantity $X_t^2$ without serious reservations concerning the initial window length. By using the traditional method of OLS to the transformed data, an objective is to reduce any bias which may exist from the initial transient region of the $R/S$ plot (as mentioned previously). Via the method of AQL the estimates will not be unduly affected by any observed values in the transient region. It seems as though there is minimal improvement in the estimates in this particular example when increasing the initial window size from 4 to 8 but when the initial window size changes from 8 to 16 respectively, the OLS estimate of $H$ improves from 0.938 to 0.903.

5.3 Comparison of the Least Squares and Asymptotic Quasi-likelihood Methods Using Fractional ARIMA$(p, d, q)$ Data

In this section the accuracy of the estimate of $H$ via the OLS method and the method of AQL will once again be compared. The situations under which the AQLM will be much more precise than the method of OLS will also be discussed.

Example 6 One hundred simulations are performed from the fractional ARIMA$(2,0.3,0)$ process

$$(1 - 0.2B - 0.6B^2)(1 - B)^{0.3}X_t = \epsilon_t.$$
<table>
<thead>
<tr>
<th>Method</th>
<th>mean $\hat{H}$ (st. err $\hat{H}$)</th>
<th>$\hat{S}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS(4)</td>
<td>0.858(0.004)</td>
<td>0.143</td>
</tr>
<tr>
<td>OLS(8)</td>
<td>0.837(0.004)</td>
<td>0.124</td>
</tr>
<tr>
<td>OLS(16)</td>
<td>0.814(0.004)</td>
<td>0.103</td>
</tr>
<tr>
<td>AQL($g_3$)</td>
<td>0.803(0.005)</td>
<td>0.097</td>
</tr>
</tbody>
</table>

Table 5.3: OLS and AQL estimates for Example 6.

From Table 5.3 it may be seen that when taking the mean value of the one hundred simulations it is found that the method of OLS yields a value of 0.814 with a standard error of 0.004 (when an initial window size of 16 is taken) and the method of AQL with a predictable process of the form

$$ g = \theta_0 + \theta_1 (\log n_t)^2 + \theta_2 X_{t-1}^2 + \theta_3 X_{t-2}^2 + \theta_4 X_{t-3}^2 + \theta_5 (\log n_t), \quad (5.5) $$

where $\log n_t$ is the logarithm of the window size, yields a mean value of 0.803 with a standard error of 0.005. When an initial window size of 8 is taken, the mean of the OLS estimates is equal to 0.837 with a standard error of 0.004.

The AQLM is by far the most precise method in this example. It is seen that the mean value is within one standard error of the true value when the predictable process is of the form of (5.5) whereas this is not the case when the OLS method is applied, the estimated value is more than two standard errors from the true value. The $\hat{S}$ value is also smaller when the AQLM is used rather than the method of OLS.

In the following the ARIMA($p, d, q$) model is discussed for different values of $d$ and the resulting estimates of $H$ via both the method of OLS and the method of
AQL are compared. One hundred simulations of sample size 8192 are performed for each model varying the values of \(d\) from 0 to 0.4 in increments of 0.1. The order of the parameters \(p\) and \(q\) may either be 0 or 1. If the order of both \(p\) and \(q\) are equal to 1 then the coefficients of the model are either \(\phi_1 = 0.3\) and \(\theta_1 = 0.7\) or \(\phi_1 = -0.3\) and \(\theta_1 = -0.7\) respectively. Otherwise \(\phi_1 = 0.5\) or \(\theta_1 = 0.5\). A comparison is made between the mean of the OLS estimates and the mean of the AQL estimates using a predictable process of the form (5.5). The results are reported in Tables 5.4-5.7 at the end of this section. The initial window length for all these simulations was 8.

When \(p = 1\) and \(q = 0\) the AQL estimates are very precise. The OLS method only yields precise estimates when \(d = 0.3\), otherwise this method overestimates the value of \(d\) when the true value is less than 0.3 and underestimates the true value when \(d = 0.4\). At lower values of \(d\) the AQLM is much more effective than the method of OLS. The induced bias is much lower and the \(S\)-value is lower. In fact the bias when \(d\) changes from 0 to 0.2 in increments of 0.1 is 0.056, 0.028 and 0.014 respectively via the AQLM whereas this bias is 0.095, 0.071 and 0.042 respectively using the method of OLS. For \(d > 0.2\) there is slightly less induced bias via the method of OLS than the AQLM and the MSE is also lower via the method of OLS.

When \(p = 0\) and \(q = 1\) there is the reverse trend in the resulting estimates where the true value is always underestimated via both the method of OLS and the method of AQL. The AQLM yields much more accurate estimates than the
method of OLS at all values of $d$. The standard deviation, however, of these estimates is higher than that when the method of OLS is applied. The biases obtained by Taqqu and Teverovsky (1996) are very large compared to these results. They obtain mean biases (via OLS) of -0.113, -0.122 and -0.141 when the value of $d$ varies from 0.2 to 0.3 and finally to 0.4. The mean biases obtained via OLS here are -0.067, -0.073 and -0.079 respectively. The biases obtained via the method of AQL are usually always less than those obtained via OLS.

When $p = 1$ and $q = 1$ (with $\phi_1 = 0.3$ and $\theta_1 = 0.7$), the estimate of $d$ is always less than the true value. The biases are once again less than those obtained by Taqqu and Teverovsky (1996). For all $d$, the OLS estimate is further from the true value than the AQL estimate. When $d = 0.3$, the bias via the AQLM is -0.052 which compares favourably to the bias of -0.157 obtained by Taqqu and Teverovsky. The mean square errors are also much smaller in the simulations than those obtained by Taqqu and Teverovsky. The $\tilde{S}$-values are much smaller for the AQLM than the resulting $\tilde{S}$-values using the method of OLS.

When $p = 1$ and $q = 1$ (with $\phi_1 = -0.3$ and $\theta_1 = -0.7$), $d$ is in fact overestimated via each method when $d = 0$ and $d = 0.1$ and underestimated via both methods when $d = 0.3$ and $d = 0.4$. This is in contrast to the case when $\phi_1 = 0.3$ and $\theta_1 = 0.7$, where $H$ was always overestimated via each method at different values of $d$. Compared to the case when both coefficients were positive there is much less bias induced in this instance. To highlight the differences, when $d = 0.2$, the biases induced when the coefficients were positive were -0.060 and 0.095 for the
AQLM and the method of OLS respectively whereas when the coefficients were negative the biases were -0.005 and 0.012 respectively.

The algorithm used by Taqqu and Teverovsky (1996) leads to OLS estimates which are more biased than those obtained using the Durbin-Levinson algorithm at smaller values of $d$. Some of the bias may come from the fact that Taqqu and Teverovsky commence with an initial window length of 5 compared to 8 in this analysis. However, the $R/S$ procedure is strictly adhered to by these authors whereas a simplified method is used here to calculate the $R/S$ ratio, there is only one window and the window length varies.
Figure 5.1: $X^2_t$ (hard line) and three possible $g_t$'s (dotted lines) for Example 5.
<table>
<thead>
<tr>
<th></th>
<th>$d=0$</th>
<th>$d=0.1$</th>
<th>$d=0.2$</th>
<th>$d=0.3$</th>
<th>$d=0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AQL</td>
<td>OLS</td>
<td>AQL</td>
<td>OLS</td>
<td>AQL</td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>0.055</td>
<td>0.095</td>
<td>0.028</td>
<td>0.071</td>
<td>0.015</td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.065</td>
<td>0.035</td>
<td>0.074</td>
<td>0.037</td>
<td>0.074</td>
</tr>
<tr>
<td>$\sqrt{MSE}$</td>
<td>0.085</td>
<td>0.101</td>
<td>0.078</td>
<td>0.080</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>AQL</td>
<td>OLS</td>
<td>AQL</td>
<td>OLS</td>
<td>AQL</td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>-0.018</td>
<td>0.004</td>
<td>-0.052</td>
<td>-0.042</td>
<td>-0.027</td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.082</td>
<td>0.041</td>
<td>0.076</td>
<td>0.042</td>
<td></td>
</tr>
<tr>
<td>$\sqrt{MSE}$</td>
<td>0.083</td>
<td>0.041</td>
<td>0.092</td>
<td>0.059</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4: OLS and AQL estimates for the fractional ARIMA$(1,d,0)$ model.

<table>
<thead>
<tr>
<th></th>
<th>$d=0$</th>
<th>$d=0.1$</th>
<th>$d=0.2$</th>
<th>$d=0.3$</th>
<th>$d=0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AQL</td>
<td>OLS</td>
<td>AQL</td>
<td>OLS</td>
<td>AQL</td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>-0.022</td>
<td>-0.044</td>
<td>-0.054</td>
<td>-0.063</td>
<td>-0.027</td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.061</td>
<td>0.033</td>
<td>0.072</td>
<td>0.041</td>
<td>0.074</td>
</tr>
<tr>
<td>$\sqrt{MSE}$</td>
<td>0.064</td>
<td>0.055</td>
<td>0.090</td>
<td>0.076</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>AQL</td>
<td>OLS</td>
<td>AQL</td>
<td>OLS</td>
<td>AQL</td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>-0.024</td>
<td>-0.073</td>
<td>-0.025</td>
<td>-0.079</td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.089</td>
<td>0.048</td>
<td>0.084</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td>$\sqrt{MSE}$</td>
<td>0.092</td>
<td>0.087</td>
<td>0.086</td>
<td>0.089</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.5: OLS and AQL estimates for the fractional ARIMA$(0,d,1)$ model.
Table 5.6: Fractional ARIMA(1, \(d, 1\)) model (with positive coefficients).

<table>
<thead>
<tr>
<th></th>
<th>(d=0)</th>
<th>(d=0.1)</th>
<th>(d=0.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.054</td>
<td>-0.056</td>
<td>-0.060</td>
</tr>
<tr>
<td>Bias</td>
<td>-0.070</td>
<td>-0.081</td>
<td>-0.095</td>
</tr>
<tr>
<td>(\hat{\sigma})</td>
<td>0.071</td>
<td>0.068</td>
<td>0.064</td>
</tr>
<tr>
<td>(\sqrt{MSE})</td>
<td>0.089</td>
<td>0.081</td>
<td>0.088</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(d=0.3)</th>
<th>(d=0.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.052</td>
<td>-0.055</td>
</tr>
<tr>
<td>Bias</td>
<td>-0.103</td>
<td>-0.113</td>
</tr>
<tr>
<td>(\hat{\sigma})</td>
<td>0.072</td>
<td>0.089</td>
</tr>
<tr>
<td>(\sqrt{MSE})</td>
<td>0.088</td>
<td>0.104</td>
</tr>
</tbody>
</table>

Table 5.7: Fractional ARIMA(1, \(d, 1\)) model (with negative coefficients).

<table>
<thead>
<tr>
<th></th>
<th>(d=0)</th>
<th>(d=0.1)</th>
<th>(d=0.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>0.028</td>
<td>0.021</td>
<td>-0.005</td>
</tr>
<tr>
<td>Bias</td>
<td>0.057</td>
<td>0.036</td>
<td>0.012</td>
</tr>
<tr>
<td>(\hat{\sigma})</td>
<td>0.066</td>
<td>0.075</td>
<td>0.079</td>
</tr>
<tr>
<td>(\sqrt{MSE})</td>
<td>0.071</td>
<td>0.058</td>
<td>0.037</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(d=0.3)</th>
<th>(d=0.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.016</td>
<td>-0.033</td>
</tr>
<tr>
<td>Bias</td>
<td>-0.014</td>
<td>-0.045</td>
</tr>
<tr>
<td>(\hat{\sigma})</td>
<td>0.086</td>
<td>0.080</td>
</tr>
<tr>
<td>(\sqrt{MSE})</td>
<td>0.087</td>
<td>0.086</td>
</tr>
</tbody>
</table>
5.4 Conclusion

The method of AQL, applied here in $R/S$ analysis for estimating the differencing parameter $d$ in a fractional ARIMA($p,d,q$) model (i.e. $R/S$ analysis with the final estimate of $d$ coming from the application of the AQLM) will only affect the results if the variances are not equal (i.e. the bias with the application of $R/S$ analysis using the AQLM does not come from unequal variances). This method seems to be very effective when there exists a short-range dependence structure, and is much more effective than when the method of OLS is applied. The AQLM outperforms the OLS method at low values of $d$ when the ARIMA(1,$d$,0) model is considered. When considering the ARIMA(0,$d$,1) model the estimates of $H$ via both methods are comparable when $d = 0.1$ but the method of AQL is clearly much more precise when other values of $d$ are considered.

When the ARIMA(1,$d$,1) model (with positive coefficients) is considered the method of AQL is much more effective than OLS. However if the ARIMA(1,$d$,1) model (with negative coefficients) is considered the AQLM is much more precise than the method of OLS at low values of $d$. The estimates in general via both methods are much closer to the true value than is the case when the coefficients are positive.

In summary, the method of AQL is better than the method of OLS when $d$ is small in ARIMA(1,$d$,0) models, for most $d$ in ARIMA(1,$d$,1) models (with negative coefficients), for all $d$ in ARIMA(1,$d$,1) models (with positive coefficients) and for all $d$ in ARIMA(0,$d$,1) models.
Notwithstanding the model investigated, the AQLM leads to substantially less bias than the method of OLS at low values of $d$ (0, 0.1 and 0.2). As the process tends to Brownian noise ($d=0$) the AQLM is far superior to the method of OLS.

As seen, the AQLM will only affect the results if the variances are not equal. The bias, in this instance, is not a result of unequal variances. Another advantage is seen to lie in the fact that the resulting estimates are not unduly affected even with small initial window lengths. The method works well notwithstanding the observed values in the transient region of the sample.
Chapter 6

Small Sample Sizes and the Asymptotic Quasi-likelihood Method

6.1 Introduction

It was seen in Chapter 4 how effective the AQLM is when dealing with large sample sizes. The attention now turns to the loss of efficiency of the method for smaller sample sizes. The models mentioned in Examples 1 and 2 are considered. In Chapter 4, the sample size for each simulation was 240 but for each simulation in the following analyses the sample size is 60 and one hundred independent samples are simulated. It is well known that the accuracy of the estimates will be affected by the sample size. In this chapter interest is focussed on the effectiveness
of the AQL procedure compared to the OLS procedure whilst the sample size is relatively small.

6.2 Simulations

Example 7 60 data values are generated from the following model:

\[ y_t = 0.3 + 0.5y_{t-1} + M_t, \quad t \geq 2, \]

where \( M_t = N_t - 0.5(y_{t-1}^2 + y_{t-2}^2) \) and given \( y_s, \quad s \leq t \), \( N_t \) has the Poisson distribution with rate \( 0.5(y_{t-1}^2 + y_{t-2}^2) \).

The results are presented in Table 6.1. The means of the parameter estimates are, in general, not as accurate as when the sample size is larger. Even when the OLS method is applied the estimates are not accurate. They are, surprisingly, inferior to the AQLM estimates notwithstanding the predictable process chosen. The second predictable process \( g_{2t} \) yields the most accurate estimates, the reasons for which were explained in Chapter 4. The \( S \) is lowest when \( g_{2t} \) is chosen and the standard errors of the corresponding estimates is lowest.

Comparing the estimates of the parameters with the corresponding estimates when the sample size is 240, it can be seen from Tables 4.5 and 6.1 that the performance of the OLS method decreases considerably (the \( S \)-value increases from 0.019 to 0.024). The mean estimate of \( \hat{\theta}_1 \) in particular changes from 0.441 to 0.396 (the true value is 0.5). The estimates via the AQLM, however, are
very inaccurate when compared with the estimates obtained previously. The
standard errors and the $\bar{S}$-values are almost doubled when the sample size is
reduced from 240 to 60. When the predictable processes $g_3$ and $g_4$ are applied,
the estimates are very inaccurate and are approximately 0.05 further from the
ture value than when the sample size is 240. When $g_1$ and $g_2$ are the predictable
processes chosen, the estimates are approximately 0.03 further from the true value
than with the larger sample size. The AQLM, however, provides more accurate
estimates notwithstanding this smaller sample size.

Example 8 Data are generated from the following model:

$$y_t = 0.2 + 0.6y_{t-1} + 0.8x_t + M_t, \quad t \geq 2,$$

where given $y_s, s \leq t, M_t$ is generated from the normal distribution with mean 0
and variance $0.3(y_{t-1}^2 + y_{t-2}^2)$, i.e. $E(M_t|\mathcal{F}_{t-1}) = 0$ and $E(M_t^2|\mathcal{F}_{t-1}) = 0.3(y_{t-1}^2 + y_{t-2}^2)$. $x_t$ is a standard normal random variable, and is independent of $y_s, s \leq t$.

The results are presented in Table 6.2. The estimators, in general, are much less
accurate than was the case when the sample size was 240. The second predictable

<table>
<thead>
<tr>
<th>Method</th>
<th>mean $\hat{\theta}_0$ (st. err $\hat{\theta}_0$)</th>
<th>mean $\hat{\theta}_1$ (st. err $\hat{\theta}_1$)</th>
<th>$\bar{S}$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.352(0.012)</td>
<td>0.396(0.018)</td>
<td>0.024</td>
</tr>
<tr>
<td>AQLM($g_1$)</td>
<td>0.344(0.010)</td>
<td>0.431(0.021)</td>
<td>0.025</td>
</tr>
<tr>
<td>AQLM($g_2$)</td>
<td>0.325(0.008)</td>
<td>0.478(0.016)</td>
<td>0.018</td>
</tr>
<tr>
<td>AQLM($g_3$)</td>
<td>0.361(0.010)</td>
<td>0.402(0.019)</td>
<td>0.025</td>
</tr>
<tr>
<td>AQLM($g_4$)</td>
<td>0.348(0.011)</td>
<td>0.436(0.017)</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Table 6.1: OLS and AQLM estimates (for two possible $g_t$'s) for Example 7.
<table>
<thead>
<tr>
<th>Method</th>
<th>$\text{mean}\hat{\theta}_0(\text{st. err } \hat{\theta}_0)$</th>
<th>$\text{mean}\hat{\theta}_1(\text{st. err } \hat{\theta}_1)$</th>
<th>$\text{mean}\hat{\theta}_2(\text{st. err } \hat{\theta}_2)$</th>
<th>$\bar{S}$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.246(0.021)</td>
<td>0.500(0.014)</td>
<td>0.739(0.017)</td>
<td>0.032</td>
</tr>
<tr>
<td>AQLM($g_1$)</td>
<td>0.286(0.019)</td>
<td>0.513(0.016)</td>
<td>0.713(0.016)</td>
<td>0.033</td>
</tr>
<tr>
<td>AQLM($g_2$)</td>
<td>0.199(0.014)</td>
<td>0.545(0.014)</td>
<td>0.751(0.012)</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Table 6.2: OLS and AQLM estimates (for two possible $g_i$’s) for Example 8.

process still provides the best estimates. Whilst the mean estimate of $\hat{\theta}_0$ is accurate, the mean estimates of $\hat{\theta}_1$ and $\hat{\theta}_2$ are much less accurate (the difference between the true value and the estimate is approximately 0.05). The AQLM still performs well but is much less effective than was the case with larger sample sizes. It however provides more accurate estimates than the method of OLS.

Comparing the estimates of the parameters with the corresponding estimates when the sample size equals 240, it can be seen from Tables 4.6 and 6.2 that the performance of the OLS method decreases considerably (the $\bar{S}$-value increases from 0.019 to 0.032). The mean estimate of $\hat{\theta}_0$ changes from 0.211 to 0.246, the mean estimate of $\hat{\theta}_1$ changes from 0.541 to 0.500 and the mean estimate of $\hat{\theta}_2$ becomes 0.739 (from 0.844 in the previous instance). The respective standard errors are almost twice those obtained with the larger sample size. The estimates via the AQLM, however, are inaccurate when compared with the estimates obtained previously (especially with respect to the estimation of $\hat{\theta}_1$ and $\hat{\theta}_2$) but they are still more accurate (when the second predictable process is chosen) than the estimates obtained via the method of OLS. The respective standard errors and the $\bar{S}$-values are almost doubled when the sample size is reduced from 240 to 60.
6.3 Conclusion

In both examples, the estimates were much less accurate via the AQLM when the sample size decreases from 240 to 60. Via the AQLM, the standard errors of each of the respective parameters as well as the $\bar{S}$-values almost double and the mean of the estimates is only mildly accurate. Though the power of the AQLM decreases with this smaller sample size it is still much more accurate (lower biases and lower standard errors for each of the parameters) than the method of OLS. Even with smaller sample sizes, the OLS method is not even comparable with the AQLM providing a $g_t$ is chosen in accordance with the criteria listed in Chapter 4. The AQLM is known to yield accurate estimates providing the sample size is large or increasing.
Chapter 7

Conclusion

In this thesis, an overview of both the QLM and the AQLM was given. A practical procedure was outlined in applying the AQLM via the selection of various predictable processes $g_t$, used to approach the quantity $y_t^2$. The $g_t$'s were obtained via autoregressive techniques including, if available, indicator variables. Criteria was applied in order to obtain reasonably accurate estimates of the parameters. These estimates were obtained via a two-stage estimation procedure and simulations were performed to investigate the effectiveness of such criteria. The AQLM was applied to both real-life and simulated data from both linear models and fractional ARIMA($p,d,q$) models and it was seen to be very accurate especially in cases were the sample size was either large or increasing.

The practical procedure in determining a suitable $g_t$ was applied to a fractional ARIMA($p,d,q$) data set and inferences made. It was seen that the biases and the MSE's incurred via the AQLM are generally lower than those obtained via the
method of OLS although the variances of such estimates are larger than those obtained when the method of OLS is applied. The AQLM was seen to provide better estimates than the method of OLS especially at lower values of $d$ (i.e. when the process tends towards Brownian noise, $d=0$).

Also discussed was the effect of small sample sizes on the resulting estimates via the AQLM. The estimates were less accurate as the sample size was rather small. In the cases where the errors were not i.i.d, the OLS method also suffered from the much smaller sample size. The AQLM, however, still provides more accurate estimates than the method of OLS.

It is evident that the AQLM provides a useful way of estimating parameters in linear models without making any assumptions as to the nature of the distribution of the error terms. The QLM cannot be used when the error term is unknown and the OLS method provides poor estimates when the error terms are not i.i.d. If it is known that the errors are not i.i.d then the AQLM should be used, this method will yield accurate estimates of the unknown parameters than would be the case via the application of the OLS method notwithstanding the fact that the sample size could be rather small. It is important to note that the accuracy increases as the sample size increases.

The QL estimates are accurate when the form of $E(M_i^2|F_{i-1})$ is known, which is never true. Using the AQLM, the effect from errors is accounted for thus leading to better estimates of the parameters than will the method of OLS. The latter method places more restrictions upon the errors, in practical situations these
restrictions are unlikely to hold.

It was also shown that the AQLM can also provide accurate estimates if a predictable process $g_t$ is chosen which satisfies certain criteria. If this $g_t$ approaches the quantity $y_t^2$ very well then the AQL estimates will be more accurate thus making the choice of predictable processes critical. In addition to this, if a constant $c$ is added to the quantity $g_t - f_t^2$ the AQL estimates will converge immediately as the ratio $\frac{E(M_t^2|\mathcal{F}_{t-1})}{c + |g_t - f_t^2|}$ will be bounded (even if this ratio cannot be calculated due to the form of $E(M_t^2|\mathcal{F}_{t-1})$ being unknown). If $g_t$ is close to $y_t^2$, then $g_t - f_t^2(\theta)$ will be close to $E(M_t^2|\mathcal{F}_{t-1})$ providing $E(\epsilon_t|\mathcal{F}_{t-1})$ (for $n > 0$) is negligible and approximately stationary for all $t$.

In smaller samples the AQLM is still more effective than the method of OLS but it is not as effective as with the larger sample size. In the $R/S$ analysis carried out there was generally less induced bias using the AQLM than was the case when the method of OLS is applied. However, there was less improvement than was the case when the linear models in Chapter 4 are considered. This is because $R/S$ analysis (using OLS) provides accurate estimates notwithstanding the distribution of the data. The OLS method provides accurate estimates when using the transformed data as much of the inherent bias in the original process is removed. The statistic is therefore robust when the data appear to not be independent of each other but rather seem to be dependent upon previous patterns, i.e. exhibit a certain memory.

Basic properties of martingales are included in Appendix A, Appendix B contains
some important properties of AQL estimates (including the consistency of the AQL estimate). Appendix C contains programs which will generate both data from simple linear models with either two or three unknown parameters and estimate the unknown parameters from such models via the automatic selection of several predictable processes. The third program in this Appendix generates data from the fractional ARIMA\((p,d,q)\) model, applies Hurst's \(R/S\) procedure and obtains an estimate of the parameter \(H\) using both the method of OLS and the AQLM (for each of several \(g_t\)'s).
Appendix A

Martingales

Let \((\Omega, \mathcal{F}, P)\) define a complete probability space. Furthermore \(\{\mathcal{F}_t, t \geq 0\}\) is defined as a non-decreasing family of sigma fields which are generated by \(\{y_s\}\), \(s \leq t\) and \(\mathcal{F}_0 = \bigcap_{t=1}^{T} \mathcal{F}_t\).

The cumulative sequence \(\{M_t\}\) is said to form a martingale with respect to \(\mathcal{F}_t\) if, for each \(t\),

- \(M_t\) is \(\mathcal{F}_t\)-measurable (i.e. the state of the process at \(t\) is observable over \([0, t])\);
- \(\mathbb{E}[|M_t|] < \infty\); and
- \(\mathbb{E}[M_t|\mathcal{F}_s] = M_s\) for all \(s < t\).

The last condition states that martingales are constant in the mean, i.e. \(\mathbb{E}[M_t] = \mathbb{E}[M_0]\) for each \(t\). This can be formally written as \(\mathbb{E}[M_t - M_s|\mathcal{F}_s] = 0\), \(s < t\).
Appendix B

Consistency of the Asymptotic Quasi-likelihood Estimate

Consider model (1.1) where the function \( f_t(\theta) \) is given by

\[
f_t(\theta) = a_{0t} + a_{1t}\theta_1 + a_{2t}\theta_2 + \cdots + a_{pt}\theta_p, \quad 0 < t \leq T.
\]

The coefficients of the parameters \( a_{0t}, a_{1t}, \ldots, a_{pt} \) are \( \mathcal{F}_{t-1} \)-measurable.

Lemma 1 For a given \( g_t \) and for all \( \theta \in \Theta \), \( \hat{G}_T = -D_T(\theta)O(1) \) where

\[
D_T(\theta) = \begin{pmatrix}
\sum_{t=1}^{T} \frac{a_{0t}^2}{|g_t - f_t(\theta)|} & 0 & 0 & \cdots & 0 \\
0 & \sum_{t=1}^{T} \frac{a_{1t}^2}{|g_t - f_t(\theta)|} & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & \cdots & 0 & \sum_{t=1}^{T} \frac{a_{pt}^2}{|g_t - f_t(\theta)|}
\end{pmatrix}
\]

and \( O(1) \) is a \( p \times p \) matrix such that \( \lim_{T \to \infty} \|O(1)\| < \infty \) and

\( \lim_{T \to \infty} \|O(1)^{-1}\| < \infty \), if the following conditions hold:
1. 
\[ U_{T,i} = \sum_{t=1}^{T} \frac{a_{it}^2}{|g_t - f_t^2(\theta)|} \to \infty \text{ as } T \to \infty, \ \theta \in \Theta; \ i = 1, 2, \ldots, p; \]

2. There are two positive numbers \( k_1 \) and \( k_2 \) such that
\[ k_1 \leq \frac{E(M_t \mid F_{t-1})}{|g_t - f_t^2(\theta)|} \leq k_2, \ \theta \in \Theta, \ 0 < t \leq T; \]

3. There exists a \( C > 0 \) such that
\[ \left( \frac{\sum_{t=1}^{T} \frac{a_{jt}^2}{|g_t - f_t^2(\theta)|}}{\sum_{j=1}^{t} \frac{a_{jt}^2}{|g_t - f_t^2(\theta)|}} \right)^{\frac{1}{2}} \leq C \text{ for all } t \geq 0 \text{ and } 0 < \delta < 1; \ i, k = 1, 2, \ldots, p; \]

4. For \( i = 1, 2, \ldots, p \) and \( \theta \in \Theta, \)
\[ \lim_{T \to \infty} \left( \frac{\max_{i} \sum_{t=1}^{T} \frac{a_{it}^2}{|g_t - f_t^2(\theta)|}}{\min_{i} \sum_{t=1}^{T} \frac{a_{it}^2}{|g_t - f_t^2(\theta)|}} \right) < \infty; \text{ and} \]

5. 
\[ \lambda_{\max} \left( \frac{\sum_{t=1}^{T} \frac{j_t f_t}{|g_t - f_t^2(\theta)|}}{\lambda_{\min} \left( \frac{\sum_{t=1}^{T} \frac{j_t f_t}{|g_t - f_t^2(\theta)|}}{\sum_{t=1}^{T} \frac{j_t f_t}{|g_t - f_t^2(\theta)|}} \right) \right) = O(1) \]
where \( \lambda_{\max} \left( \sum_{t=1}^{T} \frac{j_t f_t}{|g_t - f_t^2(\theta)|} \right) \) and \( \lambda_{\min} \left( \sum_{t=1}^{T} \frac{j_t f_t}{|g_t - f_t^2(\theta)|} \right) \) are the respective maximum and minimum eigenvalues of the matrix \( \sum_{t=1}^{T} \frac{j_t f_t}{|g_t - f_t^2(\theta)|} \).

Theorem 5 The AQL estimate \( \theta_T^* \) converges almost surely to the true parameter \( \theta_0 \) as \( T \to \infty \) if the conditions in Lemma 1 are satisfied and if there exists a \( k_3 > 0 \) such that \( k_3 \leq \frac{|g_t - f_t^2(\theta)|}{|g_t - f_t^2(\theta)|} \) for all \( t > 0 \) and for all \( \theta' \) satisfying \( 0 \leq \| \theta' - \theta_0 \| \leq \| \theta_T^* - \theta_0 \| \) when \( T \) is large enough.

Corollary 1 Assuming that \( G_t^*(\theta) \) is continuous on \( \theta \), a root exists for the AQS function if the conditions in Theorem 5 are satisfied.

For the detailed proofs of the above results see Mvoi, Lin and Biondini (1998).
Appendix C

S-plus Programs

The notation used in the programs are defined as follows;

<table>
<thead>
<tr>
<th>symbol</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y )</td>
<td>( y_t )</td>
</tr>
<tr>
<td>( ysq )</td>
<td>( y_t^2 )</td>
</tr>
<tr>
<td>( y_l )</td>
<td>( y_{t-1} )</td>
</tr>
<tr>
<td>( y1sq )</td>
<td>( y_{t-1}^2 )</td>
</tr>
<tr>
<td>( y2 )</td>
<td>( y_{t-2} )</td>
</tr>
<tr>
<td>( y2sq )</td>
<td>( y_{t-2}^2 )</td>
</tr>
<tr>
<td>( y3 )</td>
<td>( y_{t-3} )</td>
</tr>
<tr>
<td>( y3sq )</td>
<td>( y_{t-3}^2 )</td>
</tr>
<tr>
<td>( \theta_0 )</td>
<td>( \theta_0 )</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>( \theta_1 )</td>
</tr>
<tr>
<td>( \theta_2 )</td>
<td>( \theta_2 )</td>
</tr>
<tr>
<td>symbol</td>
<td>definition</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>g</td>
<td>predictable process $g_t$ which does not include differencing</td>
</tr>
<tr>
<td>gd</td>
<td>predictable process $g_t$ which includes differencing</td>
</tr>
<tr>
<td>c1</td>
<td>vector that estimates the second moment of the martingale</td>
</tr>
<tr>
<td></td>
<td>(i.e. $c_1 = g_t - f_t^2$, this quantity approximates $E(M_t^2</td>
</tr>
<tr>
<td>it</td>
<td>number of iterations before two-stage procedure converges</td>
</tr>
<tr>
<td>n</td>
<td>number of rows (i.e. number of observations in data set)</td>
</tr>
<tr>
<td>theta0ols</td>
<td>initial value of $\theta_0$ (obtained via OLS)</td>
</tr>
<tr>
<td>theta1ols</td>
<td>initial value of $\theta_1$ (obtained via OLS)</td>
</tr>
<tr>
<td>theta2ols</td>
<td>initial value of $\theta_2$ (obtained via OLS)</td>
</tr>
<tr>
<td>difftheta0</td>
<td>the difference between successive estimates of $\theta_0$</td>
</tr>
<tr>
<td>difftheta1</td>
<td>the difference between successive estimates of $\theta_1$</td>
</tr>
<tr>
<td>difftheta2</td>
<td>the difference between successive estimates of $\theta_2$</td>
</tr>
<tr>
<td>m</td>
<td>the starting element of $y$ for the iterations</td>
</tr>
</tbody>
</table>
The following S-plus program generates data from the following model:

\[ y_t = 0.3 + 0.5y_{t-1} + M_t, \quad t \geq 2, \]

where \( M_t = N_t - 0.5(y_{t-1}^2 + y_{t-2}^2) \) and given \( y_s, s \leq t, N_t \) has the Poisson distribution with rate \( 0.5(y_{t-1}^2 + y_{t-2}^2) \).

```splus
for (j in 1:100)
{
  y[1]_0
  M[2]_rpois(1,1)-1
  for (i in 2:241)
  {
    y[i]_0.3+0.5*y[i-1]+M[i]
    r_(0.5*((y[i])^2+(y[i-1])^2))
    M[i+1]_rpois(1,r)-r
  }
  n1_length(y)
  n_n1-1
  y1_y[-c(n)]
  y_y[-c(1)]
  y1ori_y1
  yori_y
  reg.lm_lm(y~y1)
  coeffols_coefficients(reg.lm)[1:2]
  theta0ols_coeffols[1]
  theta1ols_coeffols[2]
  for (m in 1:2)
  {
    y1_y1ori
    y_yori
    if (m==1)
    {
      y1_y[-c(n)]
      y_y[-c(1)]
      ysq_y~2
    }
  }

86
\begin{verbatim}
y1sq_y1^2
diffysq_diff(ysq)
diffy1sq_diff(ysq)
autoreg1.lm_lm(ysq~y1sq)
coeffreg_coefficients(autoreg1.lm)[1:2]
theta0reg_coeffreg[1]
theta1reg_coeffreg[2]
autoregdl.lm_lm(diffysq~diffy1sq)
coeffregd_coefficients(autoregdl.lm)[1:2]
theta0regd_coeffregd[1]
theta1regd_coeffregd[2]
}
else if (m==2)
{
y1_y[-c(1,n)]
y2_y[-c(n-1,n)]
y_y[-c(1,2)]
ysq_y^2
y1sq_y1^2
y2sq_y2^2
diffysq_diff(ysq)
diffy1sq_diff(ysq)
diffy2sq_diff(y2sq)
autoreg2.lm_lm(ysq~y1sq+y2sq)
coeffreg_coefficients(autoreg2.lm)[1:3]
theta0reg_coeffreg[1]
theta1reg_coeffreg[2]
theta2reg_coeffreg[3]
autoregdl2.lm_lm(diffysq~diffy1sq+diffy2sq)
coeffregd_coefficients(autoregdl2.lm)[1:3]
theta0regd_coeffregd[1]
theta1regd_coeffregd[2]
theta2regd_coeffregd[3]
}
else
{
y1_y[-c(1,2,n)]
y2_y[-c(1,n-1,n)]
y3_y[-c(n-2,n-1,n)]
y_y[-c(1,2,3)]
ysq_y^2
y1sq_y1^2
y2sq_y2^2
y3sq_y3^2
diffysq_diff(ysq)
diffy1sq_diff(ysq)
\end{verbatim}
\texttt{diffy2sq\_diff(y2sq)}
\texttt{diffy3sq\_diff(y3sq)}
\texttt{autoreg3.lm\_lm(ysq-yisq+y2sq+y3sq)}
\texttt{coeffreg\_coefficients(autoreg3.lm)[1:4]}  
\texttt{theta0reg\_coeffreg[1]}  
\texttt{theta1reg\_coeffreg[2]}  
\texttt{theta2reg\_coeffreg[3]}  
\texttt{theta3reg\_coeffreg[4]}  
\texttt{autoregd3.lm\_lm(diffysq-diffy1sq+diffy2sq+diffy3sq)}
\texttt{coeffregd\_coefficients(autoregd3.lm)[1:4]}  
\texttt{theta0regd\_coeffregd[1]}  
\texttt{theta1regd\_coeffregd[2]}  
\texttt{theta2regd\_coeffregd[3]}  
\texttt{theta3regd\_coeffregd[4]}  
\texttt{theta0\_theta0ols}
\texttt{theta1\_theta1ols}

\texttt{if (m==1)}
{  
\texttt{for (i in (m+2):(n-m))}  
{  
\texttt{g[i] = theta0reg + theta1reg * y[i-1]^2}  
\texttt{gd[i] = y[i-1]^2 + theta0regd + theta1regd * (y[i-1]^2 - y[i-2]^2)}  
  
\texttt{postscript(file="ar1.ps", horizontal=F)}  
\texttt{tsplot(y^2,g,gd,xlab="t",ylab="xsq,g",cex=1,las=1)}  
\}
\}
\texttt{else if (m==2)}
{  
\texttt{for (i in (m+2):(n-m))}  
{  
\texttt{g[i] = theta0reg + theta1reg * y[i-1]^2 + theta2reg * y[i-2]^2}  
\texttt{gd[i] = y[i-1]^2 + theta0regd + theta1regd * (y[i-1]^2 - y[i-2]^2)}\texttt{+theta2regd * (y[i-2]^2 - y[i-3]^2)}  
  
\texttt{postscript(file="ar2.ps", horizontal=F)}  
\texttt{tsplot(y^2,g,gd,xlab="t",ylab="xsq,g",cex=1,las=1)}  
\}
\}
\texttt{else}
{
for (i in (m+2):(n-m))
{
    g[i] = theta0reg + theta1reg*y[i-1] + theta2reg*y[i-2] + theta3reg*y[i-3] ^ 2
    gd[i] = y[i-1] + theta0regd + theta1regd*(y[i-1] - y[i-2]) + theta2regd*(y[i-2] - y[i-3]^2) + theta3regd*(y[i-3] - y[i-4]^2)
}

{ postscript(file="ar3.ps",horizontal=F)
  tsplot(y~2,g,gd,xlab="t",ylab="xsq,g",cex=1,las=1)
}

difftheta0_l
diffthetal_l
it_0
c_0

while ((difftheta0 > 0.0001) && (diffthetal > 0.0001))
{
    theta00 = theta0
    thetal1 = theta1

    a1_0
    a2_0
    a3_0
    a4_0
    a5_0

    for (i in (m+2):(n-m))
    {
        c1 = abs(g[i] - (theta0 + theta1*y[i-1])^2)
        a1 = a1 + y[i]/c1
        a2 = a2 + y[i-1]*y[i]/c1
        a3 = a3 + y[i-1]^2/c1
        a4 = a4 + y[i-1]/c1
        a5 = a5 + 1/c1
    }

    c2 = (a5*a3 - a4^2)
    theta0 = (a3*a1 - a4*a2)/c2
    thetal = (-a4*a1 + a5*a2)/c2

    while (abs(theta0)>1)
    {
        theta0 = theta0/2
    }
}
while (abs(theta1)>1) {
    theta1 = theta1/2
}
diffftheta0 = abs(theta00 - theta0)
diffftheta1 = abs(theta11 - theta1)

it = it+1
if (it==20) {
    c = c+0.1
    theta0 = theta0ols
    theta1 = theta1ols
    it = 0
}
if (c==1) {
    diffftheta0 = 0
    diffftheta1 = 0
}
thetafinal[4*m-3] = theta0
thetafinal[4*m-2] = theta1

diffftheta0 = 1
diffftheta1 = 1
it = 0
c = 0

while ((diffftheta0 > 0.001) && (diffftheta1 > 0.001)) {
    theta00 = theta0
    theta11 = theta1

    a1 = 0
    a2 = 0
    a3 = 0
    a4 = 0
    a5 = 0

    for (i in (m+2):(n-m-1)) {
        c1 = c+abs(gd[i]-(theta0+theta1+y[i-1])^2)
        a1 = a1+y[i]/c1
        a2 = a2+y[i-1]*y[i]/c1
    }
}
a3_a3+y[i-1]^2/c1
a4_a4+y[i-1]/c1
a5_a5+1/c1
}

c2_((a5*a3-a4)^2)
theta0_2*(a3*a1-a4*a2)/c2
thetal_(-a4*a1+a5*a2)/c2
while (abs(theta0)>1)
{
theta0_theta0/2
}
while (abs(thetal)>1)
{
theta1_thetal/2
}
diffttheta0_abs(theta00-theta0)
diffttheta1_abs(thetal1-thetal1)
it_it+1
if (it==20)
{
  c_c+0.1
  theta0_theta0ols
  thetal_thetalols
  it_0
}
if (c==1)
{
diffttheta0_0
diffttheta1_0
}
thetafinal[4*m-1]_theta0
thetafinal[4*m]_thetal
thetafinal[13]_theta0ols
thetafinal[14]_thetalols
cat(round(thetafinal,3),file="linear1.dat", fill=T,append=T,\n  "\n")
}
The following S-plus program generates data from the following model:

\[ y_t = 0.2 + 0.6y_{t-1} + 0.8x_t + M_t, \quad t \geq 2, \]

where given \( y_s, s \leq t \), \( M_t \) is generated from the normal distribution with mean 0 and variance \( 0.3(y_{t-1}^2 + y_{t-2}^2) \), i.e. \( E(M_t | F_{t-1}) = 0 \) and \( E(M_t^2 | F_{t-1}) = 0.3(y_{t-1}^2 + y_{t-2}^2) \). \( x_t \) is a standard normal random variable, and is independent of \( y_s, s \leq t \).

```splus
for (j in 1:100)
{
  y[1]_0
  x[1]_0
  M[2]_rnorm(1,0,1)
  for (i in 2:240)
  {
    x[i]_rnorm(1,0,1)
    y[i]_0.2+0.6*y[i-1]+0.8*x[i]+M[i]
    r_(0.3*((y[i])^2+(y[i-1])^2))
    M[i+1]_rnorm(1,0,r^0.5)
  }
  n1_length(y)
  n_n1-1
  y1_y[-c(n)]
  y_y[-c(1)]
  x_x[-c(1)]

  y1ori_y1
  yori_y
  xori_x

  reg.lm_lm(y~y1+x)
  coeffols_coefficients(reg.lm)[1:3]
  theta0ols_coeffols[1]
  theta1ols_coeffols[2]
  theta2ols_coeffols[3]

  for (m in 1:3)
  {
    y1_y1ori
  
```
if (m==1) {
  y1_y[-c(n)]
  y_y[-c(1)]
  x_x[-c(1)]
  ysq_y^2
  yisq_yi^2
  diffysq_diff(ysq)
  diffy1sq_diff(y1sq)
  autoreg1.1m_lm(ysq~y1sq)
  coeffreg_coefficients(autoreg1.1m)[1:2]
  thetaOreg_coeffreg[1]
  theta1reg_coeffreg[2]
  autoregd1.1m_lm(diffysq~diffy1sq)
  coeffregd_coefficients(autoreg1.1m)[1:2]
  theta0regd_coeffregd[1]
  theta1regd_coeffregd[2]
}
else if (m==2) {
  y1_y[-c(1,n)]
  y2_y[-c(n-1,n)]
  y_y[-c(1,2)]
  x_x[-c(1,2)]
  ysq_y^2
  yisq_yi^2
  y2sq_y2^2
  diffysq_diff(ysq)
  diffy1sq_diff(y1sq)
  diffy2sq_diff(y2sq)
  autoreg2.1m_lm(ysq~y1sq+y2sq)
  coeffreg_coefficients(autoreg2.1m)[1:3]
  theta0reg_coeffreg[1]
  theta1reg_coeffreg[2]
  theta2reg_coeffreg[3]
  autoregd2.1m_lm(diffysq~diffy1sq+diffy2sq)
  coeffregd_coefficients(autoreg2.1m)[1:3]
  theta0regd_coeffregd[1]
  theta1regd_coeffregd[2]
  theta2regd_coeffregd[3]
}
else {
}
if (m==1) {
  for (i in (m+2):(n-m)) {
    g[i] = theta0reg+y[i-1]^2
    gd[i] = y[i-1]^2+theta0reg+theta1reg*(y[i-1]^2-y[i-2]^2)
  }
}

else if (m==2) {

for (i in (m+2):(n-m))
{
  g[i] = theta0reg + theta1reg*y[i-1]^2 + theta2reg*y[i-2]^2
  gd[i] = y[i-1]^2 + theta0regd + theta1regd*(y[i-1]^2 - y[i-2]^2)
  + theta2regd*(y[i-2]^2 - y[i-3]^2)
}

postscript(file="ar2.ps",horizontal=F)
tsplot(y^2,g,gd,xlab="t",ylab="xsq,g",cex=1,las=1)
}
else
{
  for (i in (m+2):(n-m))
  {
    g[i] = theta0reg + theta1reg*y[i-1]^2 + theta2reg*y[i-2]^2
    + theta3reg*y[i-3]^2
    gd[i] = y[i-1]^2 + theta0regd + theta1regd*(y[i-1]^2 - y[i-2]^2)
    + theta2regd*(y[i-2]^2 - y[i-3]^2) + theta3regd*(y[i-3]^2 - y[i-4]^2)
  }
  postscript(file="ar3.ps",horizontal=F)
tsplot(y^2,g,gd,xlab="t",ylab="xsq,g",cex=1,las=1)
}

diffttheta0_1
diffttheta1_1
diffttheta2_1
it_0
c_0

while (((diffttheta0 > 0.001) && (diffttheta1 > 0.001)
  && (diffttheta2 > 0.001))
{
  theta00_theta0
  theta11_theta1
  theta22_theta2
  a1_0
  a2_0
  a3_0
  a4_0
  a5_0
  a6_0
  a7_0
for (i in (m+2):(n-m))
{
    c1 = c1 + abs(g[i]-(theta0+theta1*y[i-1]+theta2*x[i])^2)
    a1 = a1 + y[i] / c1
    a2 = a2 + y[i-1] * y[i] / c1
    a3 = a3 + y[i] * x[i] / c1
    a4 = a4 + x[i]^2 / c1
    a5 = a5 + y[i-1] * x[i] / c1
    a6 = a6 + y[i-1]^2 / c1
    a7 = a7 + x[i] / c1
    a8 = a8 + y[i-1] / c1
    a9 = a9 + l / c1
}

C2 = (a9*a6*a4 - a9*a5^2 - a8^2*a4 + 2*a8*a7*a5 - a7^2*a6)
theta0 = (a1*a6*a4 - a1*a5^2 - a2*a8*a4 + a2*a7*a5 + a3*a8*a5 - a3*a7*a6) / c2
theta1 = (-a1*a8*a4 + a1*a7*a5 + a2*a9*a4 - a2*a8*a7 - a3*a9*a5 + a3*a8*a7) / c2
theta2 = (a1*a8*a5 - a1*a7*a6 - a2*a9*a5 + a2*a8*a7 + a3*a9*a6 - a3*a8^2) / c2

while (abs(theta0) > 1)
{
    theta0 = theta0 / 2
}

while (abs(theta1) > 1)
{
    theta1 = theta1 / 2
}

while (abs(theta2) > 1)
{
    theta2 = theta2 / 2
}

difftheta0 = abs(theta00 - theta0)
difftheta1 = abs(theta11 - theta1)
difftheta2 = abs(theta22 - theta2)

it = it + 1
if (it == 20)
{
    c = c + 0.1
    theta0 = theta00

96
if (c==1)
{
difftheta0_0
diffthetal_0
difftheta2_0
}
}
thetafinal[6*m-5]_theta0
thetafinal[6*m-4]_theta1
thetafinal[6*m-3]_theta2

difftheta0_1
diffthetal_1
difftheta2_1
it_0
c_0

while ((difftheta0 > 0.001) && (diffthetal > 0.001) && (difftheta2 > 0.001))
{
theta00_theta0
theta11_theta1
theta22_theta2

for (i in (m+2):(n-m))
{
c1_c+abs(gd[i]-(theta0+theta1*y[i-1]+theta2*x[i])^2)
a1_a1+y[i]/c1
a2_a2+y[i-1]*y[i]/c1
a3_a3+y[i]*x[i]/c1
a4_a4+x[i]^2/c1
a5_a5+y[i-1]*x[i]/c1
a6\_a6+y\[i-1]\_2/c1
a7\_a7+x[i]/c1
a8\_a8+y[i-1]/c1
a9\_a9+1/c1
}

c2\_\left(\begin{array}{c}
(a9*a6*a4-a9*a5\_2-a8\_2*a4+2*a8*a7*a5-a7\_2*a6)
\end{array}\right)/c2
theta0\_\left(\begin{array}{c}
(a1*a6*a4-a1*a5\_2-a2*a8*a4+a2*a7*a5+a3*a8*a5-a3*a7*a6)
\end{array}\right)/c2
theta1\_\left(\begin{array}{c}
(-a1*a8*a4+a1*a7*a5+a2*a9*a4-a2*a7\_2-a3*a9*a5+a3*a8*a7)
\end{array}\right)/c2
theta2\_\left(\begin{array}{c}
(a1*a8*a5-a1*a7*a6-a2*a9*a5+a2*a8*a7+a3*a9*a6-a3*a8\_2)
\end{array}\right)/c2

while (abs(theta0)>1)
{
theta0\_theta0/2
}

while (abs(theta1)>1)
{
theta1\_theta1/2
}

while (abs(theta2)>1)
{
theta1\_theta1/2
}

difftheta0\_abs(theta00-theta0)
difftheta1\_abs(theta01-theta1)
difftheta2\_abs(theta22-theta2)

it\_it+1
if (it==20)
{
c\_c+0.1
theta0\_theta0;0ls
theta1\_theta1;0ls
theta2\_theta2;0ls
it\_0
}
if (c==1)
{
difftheta0\_0
difftheta1\_0
difftheta2\_0
}
}
thetafinal[6*m-2]_theta0
thetafinal[6*m-1]_theta1
thetafinal[6*m]_theta2
}
thetafinal[19]_theta0ols
thetafinal[20]_theta1ols
thetafinal[21]_theta2ols

cat(round(thetafinal,3),file="linear2.dat",fill=T,append=T,"\n")
}
The following S-plus program generates data from fractional ARIMA\((p, d, q)\) processes of the form

\[
\Phi(B)(1 - B)^d Y_j = \Theta(B) \epsilon_j
\]

where the \(\epsilon_j\)'s are independently and identically distributed (i.i.d) normal random variables with mean 0, \(B\) denotes the backward shift operator, \(\Phi(B)\) denotes the autoregressive component(s) of the process and \(\Theta(B)\) denotes the moving-average component(s) of the process. \((1 - B)^d\) denotes the fractional differencing operator.

Note that in this program \(p = 2, d = 0\) and \(q = 0\), with the coefficients of the autoregressive components being 0.2 and 0.6 respectively.

\[
\text{frac_arima.fracdiff.sim}(\text{model}=\text{list}(d=0, \text{ar}=c(0.2, 0.6), \text{mu}=0), n=8192)
\]

\[
\text{for (p in 1:11)} \{
\text{n}_2^{\text{(p+2)}}
\text{window_frac}[\text{c}(1:n)]
\text{var[p]}_\text{var}(\text{window})
\text{xori[p]}_\text{log10}(n)
\text{w}[1]_\text{window[1]}-\text{mean}(\text{window})
\text{for (k in 1:(n-1))} \{
\text{w[k+1]}_\text{w[k]}+\text{window[k+1]}-\text{mean}(\text{window})
\}
\text{range[p]}_\text{max}(0,w)-\text{min}(0,w)
\}
\text{yorixori.df_data.frame(yori,xori)}
\text{x_xori}
\text{y_yori}
\text{n_length(y)}
\text{reg.lm_lm(y~x)}
\text{reg.lm}
\]
coeffols_coefficients(reg.lm)[1:2]
theta0ols_coeffols[1]
theta1ols_coeffols[2]

for (m in 1:3)
{
  x_xori
  y_yori

  if (m==1)
  {
    y1_y[-c(n)]
    y_y[-c(1)]
    x_x[-c(1)]
    xsq_x^2
    ysq_y^2
    y1sq_y1^2
    y2sq_y2^2
    autoreg1.lm_lm(ysq~xsq+ylsq+x)
    coeffreg_coefficients(autoreg1.lm)[1:4]
    theta0reg_coeffreg[1]
    theta1reg_coeffreg[2]
    theta2reg_coeffreg[3]
    theta3reg_coeffreg[4]
  }
  else if (m==2)
  {
    y1_y[-c(1,n)]
    y2_y[-c(n-1,n)]
    y_y[-c(1,2)]
    x_x[-c(1,2)]
    xsq_x^2
    ysq_y^2
    y1sq_y1^2
    y2sq_y2^2
    autoreg2.lm_lm(ysq~xsq+y1sq+y2sq+x)
    autoreg2.lm
    coeffreg_coefficients(autoreg2.lm)[1:5]
    theta0reg_coeffreg[1]
    theta1reg_coeffreg[2]
    theta2reg_coeffreg[3]
    theta3reg_coeffreg[4]
    theta4reg_coeffreg[5]
  }
  else
  {

101
\[ y_1_y[-c(1,2,n)] \]
\[ y_2_y[-c(1,n-1,n)] \]
\[ y_3_y[-c(n-2,n-1,n)] \]
\[ y_y[-c(1,2,3)] \]
\[ x_x[-c(1,2,3)] \]
\[ xsq_x^2 \]
\[ ysq_y^2 \]
\[ y_1sq_y_1^2 \]
\[ y_2sq_y_2^2 \]
\[ y_3sq_y_3^2 \]
autoreg3.lm.lm(ysq^xsq+y_1sq+y_2sq+y_3sq+x)
autoreg3.lm
coeffreg._coefficients(autoreg3.lm)[1:6]
theta0reg._coeffreg[1]
theta1reg._coeffreg[2]
theta2reg._coeffreg[3]
theta3reg._coeffreg[4]
theta4reg._coeffreg[5]
theta5reg._coeffreg[6]
}
theta0._theta0.ols
theta1._theta1.ols
x_xori
y_yori
g_NULL
if (m==1)
{
for (i in 2:n)
{
g[i]_theta0reg+theta1reg*x[i]^2+theta2reg*y[i-1]^2
+theta3reg*x[i]
}
}
else if (m==2)
{
for (i in 3:n)
{
g[i]_theta0reg+theta1reg*x[i]^2+theta2reg*y[i-1]^2
+theta3reg*y[i-2]^2+theta4reg*x[i]
}
}
else
{
for (i in 4:n)
{
g[i] = theta0reg + theta1reg*x[i] + theta2reg*y[i-1] + theta3reg*y[i-2] + theta4reg*y[i-3] + theta5reg*x[i]
}

difftheta0 = 1
difftheta1 = 1
it = 0
C = 0
while ((difftheta0 > 0.0001) && (difftheta1 > 0.0001)) {
    theta0 = theta0
    theta1 = theta1
    a1 = 0
    a2 = 0
    a3 = 0
    a4 = 0
    a5 = 0
    for (i in (m+1):n) {
        c1 = (g[i] - (c + theta0 + theta1*x[i] + theta2reg*y[i-1] + theta3reg*y[i-2] + theta4reg*y[i-3] + theta5reg*x[i])^2)
        a1 = a1 + y[i] / c1
        a2 = a2 + x[i] * y[i] / c1
        a3 = a3 + x[i]^2 / c1
        a4 = a4 + x[i] / c1
        a5 = a5 + 1 / c1
    }
    c2 = a5*a3 - a4^2
    theta0 = (a3*a1 - a4*a2) / c2
    theta1 = (-a4*a1 + a5*a2) / c2
    difftheta0 = abs(theta0 - theta0)
    difftheta1 = abs(theta1 - theta1)
    it = it + 1
    if (it == 100) {
        c = c + 0.1
        theta0 = theta0ols
        theta1 = theta1ols
        it = 0
    }
}
thetafinal[4] = theta1ols
cat(thetafinal, file="loop.200", fill=T, append=T, "\n")
Bibliography


