INFORMATION TO USERS

SOME ASPECTS OF ROBUST ESTIMATION IN TIME SERIES ANALYSIS

By

Sanjoy Kumar Sinha

Your file Votre référence

Our file Notre référence

0-612-57354-0

Canada

DALHOUSIE UNIVERSITY

FACULTY OF GRADUATE STUDIES

The undersigned hereby certify that they have read and recommend to the Faculty of

Graduate Studies for acceptance a thesis entitled "Some Aspects of Robust Estimation

in Time Series Analysis"

by   Sanjoy Kumar Sinha

in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated: _____ August 4, 2000 _____

External Examiner —

Research Supervisor —

Research Supervisor —

Examining Committee —

—

—

ii
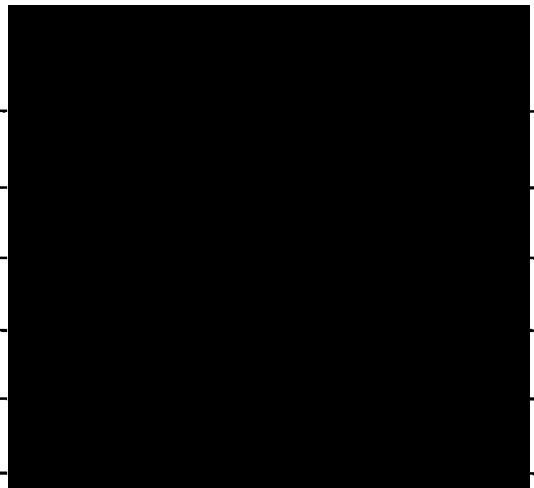
DALHOUSIE UNIVERSITY

Date: **August 2000**

Author: **Sanjoy Kumar Sinha**

Title: **Some Aspects of Robust Estimation in Time Series Analysis**

Department: **Mathematics and Statistics**

Degree: **Ph.D.**   Convocation: **October**   Year: **2000**

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.
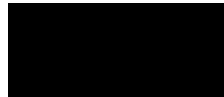
<hr>
Signature of Author

To the Memories of
My Father
and Mother

Contents

Acknowledgements

I would like to express my deepest gratitude and thanks to my supervisors Dr. Chris Field and Dr. Bruce Smith for their tireless efforts and patience in making this thesis become a reality. I would also like to thank Dr. David Brillinger, Dr. Keith Thompson and Dr. R. P. Gupta for reading this thesis and providing some insightful comments which improved the quality of the thesis. It is certainly a privilege to thank my brother-in-law Dr. Brajendra Sutradhar of Memorial University of Newfoundland for his enormous inspirations towards the progress of my thesis. I gratefully acknowledge the financial support provided by the Department of Mathematics and Statistics, Dalhousie University for my Ph. D. program.

**Some Aspects of Robust Estimation in
Time Series Analysis**

Sanjoy Kumar Sinha

Doctor of Philosophy
Department of Mathematics and Statistics
Dalhousie University
2000

Abstract

In this thesis, a number of robust methods have been developed for estimating the parameters in a time series setting. To estimate the power spectrum of an $ARMA$ process, an M estimation method has been introduced which maximizes the robust likelihood function of the discrete Fourier transforms of the process. This robust method is useful in estimating the parameters of the continuous spectrum $ARMA$ process by downweighting the influence of possible discrete spectrum harmonic components on the data. The proposed M estimation method has been applied to some actual time series data sets of sea level records, where a strong presence of tidal (harmonic) components is observed along with the continuous spectrum surge process. Here robust estimation of the power spectrum of the surge process has been considered assuming that the surge follows an $ARMA$ process.

A GM estimation technique has been introduced for the robust estimates of the parameters in a nonlinear regression setting with autoregressive errors. The asymptotic properties of the GM estimators have been studied in some detail. For choosing the appropriate order of an autoregressive process, a robust criterion has also been suggested, which uses the robust version of the Akaike Information Criterion. The proposed GM estimation and model selection criterion have been applied to a ozone data set which appears to have nonlinear relationship with some meteorological variables. As the data are collected sequentially over time, there appears to be a significant serial correlation in the errors. A nonlinear regression model with autoregressive errors has been fitted to this data set for the joint estimates of the regression parameters and the autoregressive parameters.

A new class of influence functions in the frequency domain has been introduced in this thesis. Like Hampel's influence function, this frequency influence function appears to have some close relationship with the asymptotic variance of a time series functional.

Chapter 1

Introduction

1.1 Robust Estimation in Time Series Analysis

So far, most of the robustness theories have concentrated on the development of robust statistical procedures for the case where the observations are independent. The theory on robustness in the time series setting has received less concentration, which is probably due to the increased technical problems imposed by a serial dependence in the data. Martin (1981) gives an overview of robust methods for time series. Some of the important developments of robust statistical procedures in the time series setting may be found in Kleiner, Martin and Thompson (1979), Martin (1983), Martin and Yohai (1986). Kleiner, Martin and Thompson (1979) and Martin (1983) discuss the robust estimation of the spectral density of a time series process. Martin and Yohai (1986) introduce a definition for influence functionals of parameter estimates in time series models. These influence functions are similar to Hampel's influence functions for the i.i.d. setting.

In this thesis, we develop some robust methods for analyzing real time series data, where the ordinary classical methods fail to give us satisfactory results in the presence of influential observations. In chapter 2, we develop a robust M estimation method for estimating the power spectrum of an $ARMA$ process. This robust technique has been constructed in order to downweight the effect of harmonic components mixed with the $ARMA$ process. Instead of following a pure $ARMA$ process, many oceanographic

data follow a mixed process which can be modeled as an $ARMA$ process mixed with some harmonic components. In section 2.1, we introduce an oceanographic data set which consists of sea level records at Halifax, Nova Scotia. The principal feature of the data set is the semi-diurnal tide, with period of approximately 12 hours 25 minutes. which is known in the oceanographic literature as the M2 (Moon - 2 cycles per day) tidal constituent. In addition, the spring-neap cycle is apparent, with relatively small amplitude tides near the beginning and end of the record, as compared to the middle of the record. We will presume that the observed sea level record consists of the sum of two parts, the tide and the surge, the latter being the residual obtained after removing the tide from the record. The tide consists of a number of harmonic components. In this thesis, our interest is in the structure of the surge. We assume that the surge follows an $ARMA$ process and our goal is to estimate the power spectrum of the $ARMA$ process. A natural approach would be to remove the harmonic components from the data using a suitable method and then study the residual for estimating the power spectrum of the $ARMA$ process. It will be discussed in chapter 2 that the residual series often contains remnants of harmonic components which are typically not distributed in a stationary fashion. As a consequence, the classical LS or ML methods often fail to provide efficient estimates of the power spectrum of the $ARMA$ process.

An alternative approach, and the one taken here, is to downweight the harmonic components mixed with the $ARMA$ process when estimating the power spectrum of the $ARMA$ process. In particular, we will assume that the $ARMA$ process has a continuous spectrum, and will allow for the possibility of contaminating harmonic constituents in the process. In the frequency domain, this presumes that the spectrum of the observed data consists of the sum of a continuous (surge) spectrum, and that this may be contaminated by a discrete spectrum (tidal) process. Our goal will be to downweight the influence of the discrete spectrum process on our estimate of the continuous spectral component.

An estimate of the power spectrum of an $ARMA$ process can be obtained by maximizing the approximate likelihood function of the discrete Fourier transforms

which are approximately independent complex normals with mean 0 and variance as a function of the power spectrum. Whittle (1953) introduced this Gaussian fitting procedure for the estimation of a finite dimensional parameter of a stationary process. The resulting approximate ML estimates of the $ARMA$ components can be used for estimating the corresponding power spectrum of the $ARMA$ process. The Whittle type estimates of the $ARMA$ components may be severely affected by any contaminating harmonic components mixed with the original $ARMA$ process, and so we will use a robustified likelihood method which downweights possible spikes in the periodograms due to the presence of harmonic components in the data. In section 2.2, we discuss the robust development of the approximate likelihood function of the discrete Fourier transforms. We show that the robust likelihood function can be related to the Huber's least favorable distribution. The resulting M estimate ensures the Fisher consistency of an estimator.

In section 2.3, we discuss the asymptotic properties of the proposed M estimates. We investigate the infinitesimal behavior of the M estimates based on Hampel's influence function. It is shown that the M estimates obtained by maximizing the proposed robust likelihood function have bounded influence function. Note that the estimation procedure divides naturally into two parts, in the first of which we discuss the robust estimation of the power spectrum of an $ARMA(p,q)$ process assuming the orders p and q are known a priori. In the second part, a robust model selection criterion has been proposed for choosing the appropriate orders of the process. In section 2.4, we discuss the robust development of the Akaike Information Criterion as a technique of selecting the orders of an $ARMA$ process. In section 2.5, the proposed robust estimation and model selection criterion have been applied to some actual data sets on sea level records at Halifax, Sydney and Yarmouth harbors of Nova Scotia, Canada. Section 2.6 discusses the results presented in chapter 2.

In chapter 3, we discuss aspects of robust model selection in the time domain for choosing the correct order of an autoregressive process. Unlike the harmonic contamination process as assumed in chapter 2 in the analysis of sea level data, here

we deal with a different type of contamination process, where we assume that a certain proportion of the data may come from an arbitrary distribution rather than the original underlying distribution. Depending on the behavior of contamination, the resulting outliers may be termed as "innovation" or "additive" outliers in the time domain. Our goal is to find a robust method for choosing the correct order of an autoregressive process in the presence of outliers in the data. The Akaike Information Criterion (AIC) is widely used as a classical method for choosing the appropriate order of an autoregressive process. But it is well-known that the classical AIC is very sensitive to outliers in a time series process. Ronchetti (1997) suggests using a robust version of the classical AIC in order to avoid the affect of possible outliers in the process. In section 3.1, we introduce the robustness aspects of the model choice in an autoregressive process. In section 3.2, we point out some drawbacks of the robust Information Criterion as suggested by Ronchetti (1997), and, as a remedy, we suggest a modified version of this robust criterion. Section 3.3 discusses GM estimation of the autoregressive parameters briefly. In section 3.4, we carry out a simulation study to investigate the performance of the proposed robust Information Criterion. Section 3.5 gives the conclusions of the chapter.

In chapter 4, we develop a robust method for estimating the parameters in a nonlinear regression setting with autoregressive errors. Nonlinear regression models play an important role in many fields. The classical LS or ML estimates of nonlinear regression are often very sensitive to outliers in the data. Robust estimates of nonlinear regression have been studied by a number of authors. Most of the robust techniques are developed in the case of an i.i.d. setting of the residuals. But situations in which data are collected sequentially over time may result in substantial serial correlations in the errors. It usually occurs with economic data where the response y measures the state of a market at a particular time. Another typical example where serially correlated errors usually arise is in the modeling of growth curves to data on a single animal over a certain period. In many cases, we observe a nonlinear relationship between the response variable y and the vector of covariates $\mathbf{x}$. We develop a generalized M (GM) estimation method for the joint estimation of the regression parameters and the autoregressive parameters of a nonlinear regression model with autoregressive

errors. In section 4.2, we introduce the proposed GM estimation method. For an $AR(q)$ process of the errors, we minimize an objective function which is based on the robustified version of the conditional negative log-likelihood function of $y_{q+1}, \ldots, y_n$ for the given values $y_1, \ldots, y_q$.

We then study the infinitesimal behavior of the robust estimates based on the time series analogue of Hampel's influence function in section 4.3. It is shown that the resulting GM estimates have bounded influence functions under some regularity conditions. The influence functions are derived from the time series extension of Hampel's influence function as described in Martin and Yohai (1986). In section 4.4, we study the asymptotic properties of the robust estimates of nonlinear regression in some detail. We show that under some suitable assumptions, the proposed GM estimates of the regression parameters and the autoregressive parameters are asymptotically normally distributed with a certain mean vector and a covariance matrix. To choose the appropriate order of the autoregressive errors in a nonlinear regression setting, we extend the robust Akaike Information Criterion introduced in chapter 3 to the case of a nonlinear regression with autoregressive errors. This extended robust criterion is presented in section 4.5.

As an application of the proposed robust method, in section 4.6 we analyze a ground level ozone data set which appears to have nonlinear relationship with some meteorological variables as covariates. As the data are collected over equally spaced time, there appears to be a significant serial correlation in the errors. Moreover, the error process is found to be heteroscedastic with respect to the covariates. We develop a robust technique to model the variance of the heteroscedastic errors as a function of the covariates. Section 4.7 gives the conclusions of the chapter.

In chapter 5, we discuss a new class of influence functions introduced by Hastings (personal communication), which is referred to as the frequency influence function of a functional $g(F^*)$, where F^* is considered as the cumulative (non-normalized) power spectrum of a time series process $\{x_t\}$. The frequency influence function investigates the behavior of the functional $g(F^*)$ in terms of the point-mass perturbation of the

cumulative spectrum F^*. In section 5.2, we define the frequency influence function. Like Hampel's influence function, it is shown with some examples that the asymptotic variance of an estimator can be obtained as a function of its corresponding frequency influence function.

In chapter 6, we summarize our results and discuss some directions for further research.

1.2   Some Important Ideas in Robust Statistics

Robust statistics is concerned with the fact that many assumptions commonly made in statistics are not exactly true - they are mathematically convenient rationalizations of an often fuzzy knowledge or belief (Huber, 1981). The most common statistical procedures (in particular, those optimized for an underlying Gaussian distribution) are quite sensitive to slight deviations from the assumptions. Robustness, in a sense, signifies insensitivity to small deviations from the underlying assumptions. During the past few decades, the amount of statistical research devoted to robustness has increased considerably. Most of the research effort has focused on robust estimation in a parametric model.

The important pioneering work of Huber (1964) on the robust estimation of a location parameter is considered to be the basis for a theory of robust estimation. Huber introduced a class of M estimators which became a very useful tool in robust estimation. He also derived the consistency and asymptotic normality of the M estimators. These estimators are slight generalizations of maximum likelihood estimators. For a set of independent random variables x_i ($i = 1, \ldots, n$) from a one-parameter family of distributions, any estimate T obtained by solving $\sum \rho(x_i; T) = \min$ or $\sum \psi(x_i; T) = 0$, where ρ is an arbitrary function, $\psi(x; \theta) = (\partial/\partial\theta)\rho(x; \theta)$, is termed as an M estimate or maximum likelihood type estimate of the parameter θ. The choice $\rho(x; \theta) = -\log f(x; \theta)$, where f is the density of x, gives the classical ML estimate.

Huber then introduced the idea of gross error model. Instead of coming from a strict parametric model $F(x; \theta)$, he assumes that a known proportion ϵ of the data may come from an arbitrary unknown distribution $G(x; \theta)$, which result in gross errors. The underlying gross error model is thus considered as the ϵ-contaminated distribution: $F_\epsilon = (1 - \epsilon)F + \epsilon G$. Huber considers optimizing the worst that can happen over the neighborhood of the model, as measured by asymptotic variance of the estimator. The asymptotic variance of an M estimator defined by some function ψ at a distribution F_ϵ, which lies in some neighborhood $\mathcal{F}_\epsilon$ of the assumed model distribution F, is given by $V(\psi, F_\epsilon) = \int \psi^2 dF_\epsilon / \left(\int \psi' dF_\epsilon \right)^2$. For a location parameter, Huber's idea was to minimize the maximal asymptotic variance over $\mathcal{F}_\epsilon$, that is, to find the M estimator ψ_0 satisfying

$$\sup_{F_\epsilon \in \mathcal{F}_\epsilon} V(\psi_0, F_\epsilon) = \min_\psi \sup_{F_\epsilon \in \mathcal{F}_\epsilon} V(\psi, F_\epsilon). \tag{1.1}$$

This is achieved by finding the least favorable distribution F_0, that is, the distribution minimizing the Fisher information $J(F_\epsilon)$ over all $F_\epsilon \in \mathcal{F}_\epsilon$. Then $\psi_0 = -F_0''/F_0'$ gives the maximum likelihood estimator for this least favorable distribution. Note that for the case $F = \Phi$, where Φ is the standard normal distribution, the least favorable distribution F_0 has the density $f_0(x) = (1-\epsilon)(2\pi)^{-1/2} \exp(-\rho(x))$, where $\rho(x) = \int_0^x \psi_0(t) dt$ and $\psi_0(x)$ is the Huber's ψ function defined by $\psi_0(x) = \max(-k, \min(k, x))$. The values k and ϵ are connected through the equation $2\phi(k)/k - 2\Phi(-k) = \epsilon/(1 - \epsilon)$, where $\phi(x)$ represents the normal density.

Another important development in robust statistics is the introduction of the influence function (IF) in order to investigate the infinitesimal behavior of real-valued functionals such as $T(F)$. This influence function introduced by Hampel (1968, 1974) is considered as the most useful heuristic tool of robust statistics. It is a technique for studying the local stability of an estimator in terms of the point-mass contamination of the data or the underlying distribution. To define influence function, let us consider a random sample $(x_1, \ldots, x_n)$ from a distribution F. Assume that an estimator $T_n(x_1, \ldots, x_n)$ can be obtained from a functional $T = T(F)$ defined on a family of distributions by evaluating T at the empirical distribution function $F_n : T_n = T(F_n)$. Let $F_\epsilon = (1 - \epsilon)F + \epsilon \Delta_x$ be a contamination distribution, where Δ_x has all its mass

at x. Then Hampel's influence function is the directional derivative of the functional T at $\epsilon = 0$:

$$IF(x; T, F) = \lim_{\epsilon \to 0} \frac{T(F_\epsilon) - T(F)}{\epsilon} \qquad (1.2)$$

provided that the limit exists. This influence function is considered as an asymptotic as well as an infinitesimal tool. Note that if $F = F_{n-1}$ and $\epsilon = 1/n$, the influence function, before taking the limit, becomes the finite sample sensitivity curve of Tukey (1970); so that one can argue that an influence function measures approximately n times the change of T caused by an additional observation in x when T is applied to a large sample of size $n - 1$. In other words, it describes the effect of an infinitesimal contamination at the point x on the estimate, standardized by the mass of the contamination. Its maximum absolute value

$$\gamma^* = \sup_x |IF(x, T, F)| \qquad (1.3)$$

is termed as gross error sensitivity. It is useful in the construction of optimal estimates under the constraint of a bounded gross error sensitivity. Under some regularity conditions, $\sqrt{n}(T(F_n) - T(F))$ is asymptotically normal with mean 0 and variance $V(T, F) = \int IF(x, T, F)^2 dF(x)$. Thus the influence function allows an immediate and simple heuristic assessment of the asymptotic properties of an estimator as it leads to an explicit formula for the asymptotic variance.

1.3 Robust Estimation in Linear Models

In the last two decades, there has been considerable research on robust estimation in linear models. In fact, the linear regression model is one of the most widely used tools in statistical analysis. The classical least squares (or maximum likelihood) method is the commonly used technique for the estimation of a linear model. But in spite of its mathematical beauty and computational simplicity, this estimator is found to be very sensitive to outliers. Most robust developments on the estimation of linear models are based on the generalizations of least squares or maximum likelihood methods. Some of the robust techniques are discussed in Huber (1981) and Hampel et al

(1986). Huber (1973, 1977) extended his results on robust estimation of a location parameter to the case of linear regression. The resulting M estimator of the regression parameter can bound the influence of the residual, but not the influence of position in the design space. To bound the joint influence of residual and position in the design space, generalized M (GM) estimators are proposed by a number of authors (see Mallows, 1975; Hampel, 1978; Krasker, 1980; Krasker and Welsch, 1982). These bounded influence estimators have breakdown points of at most $1/(p+1)$, where p is the number of predictor variables, (see Maronna, Bustos and Yohai, 1979), suggesting that they can be overwhelmed by a cluster of outliers. Several high breakdown point (HBP) estimators have been proposed by a number of authors that achieve breakdown points almost $1/2$ for each p. The high breakdown point estimators include the least median squares (LMS) of Rousseeuw (1984), the S estimator of Rousseeuw and Yohai (1984). These also include the MM estimator of Yohai (1987) and the τ estimator of Yohai and Zamar (1988), which combine good asymptotic efficiency under normal linear model with high breakdown point. The MM and τ estimators do not have bounded influence function.

To construct regression estimators that have bounded influence function and high breakdown points, Simpson, Ruppert and Carroll (1992) suggest using a one-step GM estimator. To find the one-step GM estimator, we can start with a high breakdown point estimator and perform one iteration of a Newton-Raphson algorithm towards solution of the GM estimating equation.

Recently, Ferry, Kelmansky, Yohai and Zamar (1999) introduce a new class of robust estimates in a linear regression setting. These estimates, termed as generalized τ (G_τ) estimates, are defined by minimizing the τ scale of the weighted residuals, where the weights penalize the high-leverage observations. These estimates are the generalization of the τ estimates introduced by Yohai and Zamar (1988). The G_τ estimates are obtained such that they have bounded influence functions. Like Yohai's MM and τ estimates, redescending ψ functions are used to obtain the generalized τ estimates. The authors argued that the G_τ estimates inherit the properties of high breakdown point and high efficiency. However, from their simulation study, it

is observed that in some cases the efficiency of the G_τ estimates is less than that of the LMS estimates although G_τ and LMS estimates have the same high breakdown point. It is well-known that LMS estimates have high breakdown point, but they are highly inefficient. In fact, there is always a trade-off between high efficiency and high breakdown point. The redescending ψ function used in the G_τ estimation results in a high breakdown point robust estimator. However, the use of redescending ψ functions in robust estimation is not free of controversy. The redescending ψ function may not give a unique solution to an estimating equation. So it is important to choose the initial estimates wisely in order to obtain the desired robust estimates (Hampel et al, 1986). Huber (1981) commented that redescending ψ functions are certainly beneficial if there are extreme outliers, but the improvement is relatively minor with respect to the asymptotic variance and is counterbalanced by an increase of the minimax risk.

In a recent paper, Rousseeuw and Hubert (1999) introduce a notion of depth in the regression setting. The regression depth of a fit θ to a given data set Z_n of size n, denoted by rdepth(θ, Z_n), is defined as the smallest number of observations of Z_n that would need to be removed to make θ a nonfit. Rousseeuw and Hubert define the "deepest regression" estimator $T_r^*(Z_n) = \arg\max_\theta \text{rdepth}(\theta, Z_n)$. T_r^* has breakdown point $1/3$, whereas LMS and LTS estimators always have breakdown point almost $1/2$. But unlike the deepest regression estimator T_r^*, the LMS and LTS estimators are not consistent for a model with skewness and heteroscedasticity. Rousseeuw and Hubert suggest using LMS and LTS estimators in cases where robustness and outlier detection are the most important. On the other hand, if there are not so many outliers and the emphasis is on the possibility of skewness and heteroscedasticity, the deepest regression estimator T_r^* may be the natural choice.

We can extend the robust estimation from linear to nonlinear regression. However, the asymptotic properties of robust estimators in a nonlinear regression setting needs to be carefully investigated. The GM estimator of a nonlinear regression parameter may not have a bounded influence function in both the response variable and the vector of covariates. In chapter 4, we discuss the asymptotic properties of the GM estimators in a nonlinear regression setting with autoregressive errors.

Chapter 2

Robust Spectrum Estimation in an ARMA(p, q) Model With Application to Sea Level Data

2.1 Introduction

Many physical oceanographic measurements show the influence of harmonic tidal components which, due to their relatively large size, can obscure other signals in the data. In this work, our focus is on estimating the characteristics of harbor seiches, which are resonant responses of harbors to external forcing.

The data which are analyzed consist of sea level records at Sydney, Halifax and Yarmouth, Nova Scotia, each consisting of 2048 observations collected at a rate of four per hour, beginning midnight January 1, 1997. The Halifax record is illustrated in Figure 2.1, in which the ordinate represents mm above chart datum. The principal feature of the data is the semi-diurnal tide, with period of approximately 12 hours 25 minutes, which is known in the oceanographic literature as the M2 (Moon - 2 cycles per day) tidal constituent. In addition, the spring-neap cycle is apparent, with relatively small amplitude tides near the beginning and end of the record, as compared to the middle of the record. The structure of the data is more clearly discernible in the periodogram, which shows a large spike at frequency near .12 (the M2 tidal
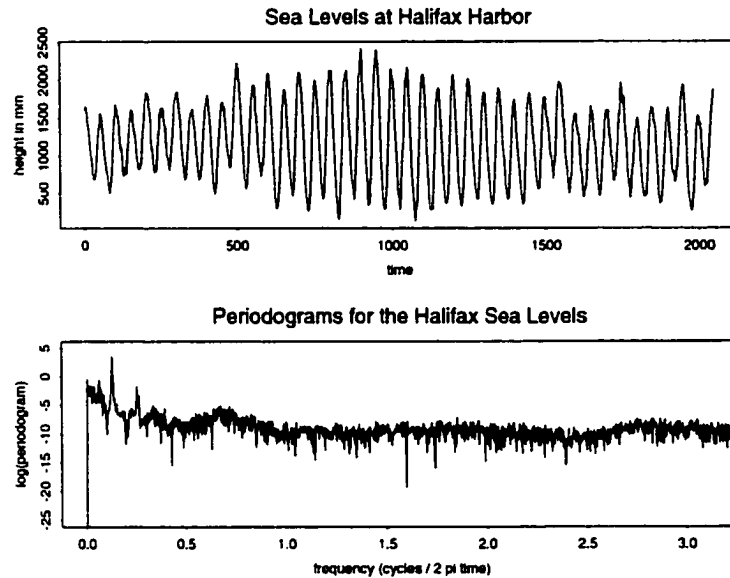
Figure 2.1: Sea level records at Halifax harbor and the periodograms

constituent), and another spike near frequency .25 (the M4 tidal constituent). The latter is a harmonic of M2. In addition, there appears to be a relatively wide band structure near frequency .75, corresponding to a period of just over two hours, and there is some weak evidence of a wide band structure near frequency 2.8.

Smith and Miyaoka (1999) studied the Halifax and Sydney records and concluded that the wide band spectral structures are compatible with tidally forced harbor seiches. A simplified physical model for a seiche can be found, for example, in Proudman (1953). We will presume that the observed sea level record consists of the sum of two parts, the tide and the surge, the latter being the residual obtained after removing the tide from the record. The tide consists of a number of harmonic components. The Canadian Hydrographic Service models the tide using the program of Foreman (1977). and typically includes on the order of 50-100 harmonic constituents, with the phase and amplitude of the constituents being determined by the motions of the earth, moon, and sun, among other factors. In addition, the model is non-stationary, and requires occasional "nodal correction".

In this work, our interest is in the structure of the surge, in which we will be

looking for characteristics of seiche motions. A natural approach would be to remove the tidal component from the data using Foreman's tidal model, and then study the residual. We have used this approach in the past and found that the residual series often contains remnants of tidal components which are typically not distributed in a stationary fashion. This would be expected if the tidal motions contain relatively short term non-stationaries, as Foreman's model assumes short to moderate term stationarity and includes the nodal corrections only at longer time scales.

Schuster (1898) suggested using the periodograms for estimating the continuous spectrum surge process. The surge spectrum at a certain frequency ω may be estimated by smoothing the periodograms near frequency ω, avoiding the periodograms for substantial peaks in the neighborhoods of the ω. An alternative approach, and the one taken here, is to downweight tidal components when estimating the surge. In particular, we will assume that the surge process has a continuous spectrum, and will allow for the possibility of contaminating harmonic constituents in the tidal process. In the frequency domain, this presumes that the spectrum of the observed data consists of the sum of a continuous surge spectrum, and that this may be contaminated by a discrete spectrum tidal process. Our goal will be to downweight the influence of the discrete spectrum process on our estimate of the continuous spectral component.

To start we assume that the surge process $\{x(t)\}$ follows a differential equation

$$x^{(p)}(t) + \alpha_1 x^{(p-1)}(t) + \ldots + \alpha_p x(t) = \xi(t) \tag{2.1}$$

where $x^{(p)}(t) = \left(\frac{\delta}{\delta t}\right)^p x(t)$, $\alpha_1, \ldots, \alpha_p$ are constants, and $\xi(t)$ represents a forcing term. As we only have data at regularly spaced time points, we consider a discretized version of (2.1) in which derivatives are replaced by divided difference approximations of the appropriate order. Where the differences are taken "backwards in time" (for example, $\frac{\delta}{\delta t} x(t) \approx [x(t) - x(t - \delta)]/\delta$), the resulting approximation to $x(t)$ is linear in past and present observations, and is therefore equivalent to an autoregressive process. We accommodate the possibility that the forcing $\xi(t)$ is not white noise by allowing the discrete time approximation of $\xi(t)$ to itself follow an autoregressive model.

Thus we consider the surge data x_t $(t = 0, \ldots, n - 1)$ to follow an $ARMA(p, q)$ model

$$x_t + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} = u_t + \zeta_1 u_{t-1} + \ldots + \zeta_q u_{t-q} \tag{2.2}$$

where u_t $(t = 0, \ldots, n - 1)$ is a purely random process and ϕ_{j_1} $(j_1 = 1, \ldots, p)$, ζ_{j_2} $(j_2 = 1, \ldots, q)$ are constants.

Our model for the sea level process Y is

$$Y = X + Z \tag{2.3}$$

where $Z_t = \sum_{s=1}^{K} A_s \cos(\lambda_s t + \varphi_s)$ is the tidal component Z of the record at time t, consisting of one or more harmonic constituents. For a stationary tide we would assume that the phase angles φ_s $(s = 1, \ldots, K)$ are independently uniformly distributed over $(-\pi, \pi)$. We will assume that the surge and tide processes are independent of one another.

The linearity of (2.3) leads us to an analogous relation for power spectra as

$$f_{YY}(\omega) = f_{XX}(\omega) + f_{ZZ}(\omega) \tag{2.4}$$

where

$$f_{XX}(\omega) = \frac{\sigma^2 \left| 1 + \sum_{j_2=1}^{q} \zeta_{j_2} e^{-ij_2\omega} \right|^2}{2\pi \left| 1 + \sum_{j_1=1}^{p} \phi_{j_1} e^{-ij_1\omega} \right|^2}, \tag{2.5}$$

σ^2 is the variance of the white noise process $\{u_t\}$, and $f_{ZZ}(\omega)$ has infinite spikes at frequencies $\lambda_s \pm 2\pi J$ $(s = 1, \ldots, K)$ for all integers J, and is zero at all other frequencies.

As will be discussed in section 2.2, the periodogram $I(\omega_j)$ of the data $\{y_t, t = 0, \ldots, n - 1\}$ at the Fourier frequencies $\omega_j = 2\pi j/n$ $(j = 1, \ldots, [(n - 1)/2])$ are approximately independent $f_{YY}(\omega_j)\chi^2_{(2)}/2$ random variables, which are multiples of $\chi^2_{(2)}$

random variables, and when normalized by the surge spectrum, are approximately i.i.d., except at Fourier frequencies in the vicinity of the harmonic tidal components where the periodograms will be of a higher order of magnitude.

From the frequency domain standpoint, this is the classic scenario calling for robust estimation. We have a number of data points (the periodograms at Fourier frequencies) among which there are a certain number of contaminated values (the periodograms at frequencies in the vicinity of the harmonic tidal components), and we would like to downweight the contaminating values when estimating the smooth (surge) spectral component.

An estimate of the power spectrum of an $ARMA$ process can be obtained by maximizing the approximate likelihood function of the discrete Fourier transforms which are approximately independent complex normals with mean 0 and variance as a function of the power spectrum. Whittle (1953) introduced this Gaussian fitting procedure for the estimation of a finite dimensional parameter of a stationary process. The resulting approximate ML estimates of the $ARMA$ components can be used for estimating the corresponding power spectrum of the $ARMA$ process. The Whittle type estimates may be severely affected by any contaminating harmonic components as we will show, and so we will use a robustified method which downweights possible spikes in the periodograms due to the harmonic components mixed with the original $ARMA$ process. The resulting estimates can be considered as M estimates, and when applied to data from an uncontaminated $ARMA$ process, are very close to Whittle's approximate MLE.

To demonstrate this, we consider an $ARMA(1,1)$ process contaminated with a harmonic process as follows:

$$y_t + .8y_{t-1} = u_t + .4u_{t-1} + .5\cos(.5t + \varphi_1) + .5\cos(1.5t + \varphi_2) \qquad (2.6)$$

where u_t are assumed to be independent $N(0,1)$ and φ_l ($l = 1, 2$) are independent uniform$(-\pi, \pi)$, independent of u_t. We draw a random sample of size $n = 200$ using model (2.6). In Figure 2.2, we plot the periodograms of the random sample at all

Periodograms for a contaminated ARMA(1,1) process
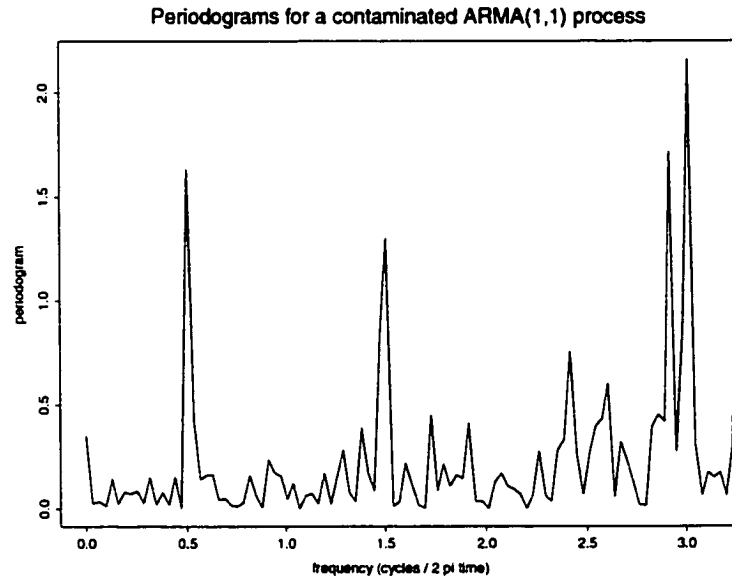
Figure 2.2: Periodograms of an $ARMA(1,1)$ process contaminated with harmonic components

frequencies ω_j $(j = 0, \ldots, (n-1)/2)$. From the plot, we observe two unusual spikes at frequencies .5 and 1.5 which are due to the harmonic components in model (2.6). The spikes in the right end are due to the $ARMA$ components in the model. This data set can be modeled as in equation (2.3) assuming that the amplitudes in the harmonic components remain constant over time. But in real life data, the amplitudes may change over time which makes the estimation of the harmonic components very difficult. In the next section, we develop a robust procedure which estimates the power spectrum of an $ARMA$ process by downweighting such spikes in the periodograms.

Robust spectrum estimation has received attention from a number of authors. Kleiner, Martin and Thompson (1979) proposed two robust-resistant methodologies for the estimation of a spectral density (see also Martin, 1983). These methods are based on robust prewhitening and on weighting the resulting residuals. Prewhitening is a technique for reducing the dynamic range of the spectrum by filtering the data so as to transform the original time series into a white noise series of nearly flat spectrum. Their first method computes the prewhitened values iteratively. The second method fits a robust autoregressive model by iterated weighted least squares.

Martin (1983) discussed the robust version of the smoothed periodogram estimation for estimating the spectral density of a time series process. These robust methods provide protection against "innovation" or "additive" outliers in the time domain (cf. Martin, 1980) in the estimation of the power spectrum of a time series process. For innovation outliers, the process u_t in the $ARMA$ model (2.2) comes from a mixture distribution: $(1 - \epsilon)F + \epsilon G$, where G is any other distribution, instead of coming from an uncontaminated distribution F. For innovation outliers, we may have big jumps in a time series process y_t and the process will propagate through the jumps. For additive outliers, with probability $(1 - \epsilon)$ an $ARMA$ process y_t itself is observed, and with probability ϵ, y_t is an $ARMA$ process plus an error with distribution G. So for additive outliers, we may have big jumps in a time series process y_t, but the process will not propagate through the jumps. However, in our case, we have a different kind of contamination process, which is an $ARMA$ process mixed with some harmonic process. As we can see from the sea level record at Halifax, there appear to be no outliers in this process in the context of innovation or additive outliers. However, the harmonic components mixed with the $ARMA$ process result in spikes in the periodograms, which we can consider outliers in the frequency domain. And our goal is to estimate the power spectrum by downweighting such outliers in the frequency domain. Here it should be noted that the outliers in the time domain do not correspond to the outliers in the frequency domain. Our robust procedure deals only with the outliers in the frequency domain, which are the outcomes of the harmonic components mixed with the $ARMA$ process. If innovation or additive outliers in the time domain are present in an $ARMA$ process, rather than having spikes in some particular points in the periodograms, all of the points in the periodograms may be affected by such outliers. In such cases, we suggest using our proposed M estimation method after cleaning the data such that no innovation or additive outliers are present in the process. Martin (1983) discusses robust filters and smoother which can be used to obtain a cleaned data set.

In the next section, we discuss the robust estimation of the spectral density in some detail. In this work, the problem of robust estimation is divided into two parts. In the first part, we discuss the robust estimation of the power spectrum of an $ARMA(p, q)$

process provided that the orders p and q of the process are known a priori. In the second part, we propose a robust model selection procedure for choosing the appropriate orders of the process.

Choosing a suitable model is an important issue in time series analysis. The Akaike Information Criterion (AIC) is widely used for the model selection in an $ARMA(p,q)$ process (see Priestley, 1981). As this criterion is sensitive to influential observations, a robust version of this classical AIC has been considered by a number of authors (see Martin, 1980; Ronchetti, 1997). We use those concepts of robust model selection with some modifications.

In section 2.2, we discuss the proposed robust estimation of the power spectrum based on the approximate likelihood function of the discrete Fourier transforms. Section 2.3 discusses the influence function as well as the asymptotic variance of the estimators. Section 2.4 discusses the robust model selection criterion for choosing a suitable model. In section 2.5, the proposed robust estimation and model selection criterion have been applied to some actual data sets on sea level fluctuations at Halifax, Sydney and Yarmouth harbors of Nova Scotia, Canada. Section 2.6 gives the conclusions of the chapter.

2.2   Robust Estimation of the Power Spectrum

We know that for a stationary process y_t ($t = 0, \ldots, n-1$), the discrete Fourier transforms

$$d_y(\omega_j) = \sum_{t=0}^{n-1} y_t e^{-i\omega_j t} \tag{2.7}$$

($j = 1, \ldots, [(n-1)/2]$) are approximately independent complex normals with mean 0 and variance $2\pi n f_{yy}(\omega_j)$, where $\omega_j = 2\pi j/n$ (see Whittle, 1953; Brillinger, 1981; Brillinger, 1985). Here $f_{yy}(\omega_j)$ is the power spectrum of the stationary process y_t which takes the form (2.5) for an $ARMA(p,q)$ process as defined in equation (2.2). Treating the discrete Fourier transforms of the $ARMA(p,q)$ process y_t as exactly

independent complex normals with mean 0 and variance $2\pi n f_{yy}(\omega_j)$, the negative of the log-likelihood can be expressed, up to a constant, in the form

$$-\log L(\boldsymbol{\theta}) \;=\; \frac{1}{2}\sum_j \log f_{yy}(\omega_j|\boldsymbol{\theta}) + \frac{1}{2}\sum_j \frac{I(\omega_j)}{f_{yy}(\omega_j|\boldsymbol{\theta})} \tag{2.8}$$

where

$$I(\omega_j) \;=\; \frac{|d_y(\omega_j)|^2}{2\pi n} \tag{2.9}$$

is the periodogram of the process y_t at frequency ω_j and $\boldsymbol{\theta}^T = (\phi_1,\ldots,\phi_p,\zeta_1,\ldots,\zeta_q,\sigma)$ is the vector of parameters to be estimated. The approximate maximum likelihood (ML) estimate of $\boldsymbol{\theta}$ can be obtained by minimizing equation (2.8). This Gaussian fitting procedure was introduced in Whittle (1953) for the estimation of a finite dimensional parameter of a stationary process. Differentiating this negative log-likelihood with respect to $\boldsymbol{\theta}$ gives the estimating equation

$$\sum_j [I(\omega_j) - f_{yy}(\omega_j|\boldsymbol{\theta})]\frac{1}{\{f_{yy}(\omega_j|\boldsymbol{\theta})\}^2}\frac{\partial}{\partial \boldsymbol{\theta}}f_{yy}(\omega_j|\boldsymbol{\theta}) = 0 \tag{2.10}$$

which can be solved iteratively for the ML estimate of $\boldsymbol{\theta}$. This ML estimator will be statistically efficient if the time series process y_t follows the $ARMA(p,q)$ model without any contamination of the process. But as we mentioned earlier in section 2.1, instead of following a pure $ARMA$ process, many oceanographic data are expected to follow a contaminated process (i.e., an $ARMA$ process mixed with some harmonic components), which results in some unusual spikes in the plot of periodograms. We need to develop a robust procedure for estimating $\boldsymbol{\theta}$ by downweighting such spikes due to the harmonic components. Two times the negative log-likelihood of the discrete Fourier transforms of the $ARMA(p,q)$ process y_t can be expressed, up to a constant, as

$$-2\log L(\boldsymbol{\theta}) \;=\; \sum_j \log f_{yy}(\omega_j|\boldsymbol{\theta}) + \sum_j \frac{\{d_y^{\mathbb{R}}(\omega_j)\}^2 + \{d_y^{\mathbb{I}}(\omega_j)\}^2}{2\pi n\; f_{yy}(\omega_j|\boldsymbol{\theta})} \tag{2.11}$$

where $d_y^{\mathbb{R}}(\omega_j)$ and $d_y^{\mathbb{I}}(\omega_j)$ are the real and imaginary parts of the Fourier transform $d_y(\omega_j)$, respectively. Note that $d_y^{\mathbb{R}}(\omega_j)$ and $d_y^{\mathbb{I}}(\omega_j)$ are approximately independently normally distributed with mean 0 and variance $\pi n\, f_{yy}(\omega_j|\boldsymbol{\theta})$. Equation (2.11) can be rewritten as

$$-2\log L(\boldsymbol{\theta}) = \sum_j \log f_{yy}(\omega_j|\boldsymbol{\theta}) + \sum_j \frac{1}{2}r_{1j}^2 + \sum_j \frac{1}{2}r_{2j}^2 \qquad (2.12)$$

where $r_{1j} = \dfrac{d_y^{\mathbb{R}}(\omega_j)}{\sqrt{\pi n\, f_{yy}(\omega_j|\theta)}}$ and $r_{2j} = \dfrac{d_y^{\mathbb{I}}(\omega_j)}{\sqrt{\pi n\, f_{yy}(\omega_j|\theta)}}$ are approximately i.i.d. normals with mean 0 and variance 1. Note that the ML estimates obtained by minimizing the objective function $-2\log L(\boldsymbol{\theta})$ have unbounded influence functions (see section 2.3). To have an estimator with bounded influence function, following Huber's "Proposal 2" (see Huber, 1981) we consider a robust version of the likelihood function of equation (2.12) in the form

$$H_1(\boldsymbol{\theta}) = c\sum_j \log f_{yy}(\omega_j|\boldsymbol{\theta}) + \sum_j \rho_1(r_{1j}) + \sum_j \rho_1(r_{2j}) \qquad (2.13)$$

where c is a tuning constant and the function ρ_1 is defined by

$$\rho_1(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq k \\ \frac{1}{2}k^2 - k^2 \log\left(\frac{k}{|x|}\right) & \text{otherwise.} \end{cases} \qquad (2.14)$$

This ρ_1 function can be related to the Huber's least favorable distribution for scale. Huber (1981, p.120) replaces the standard normal distribution $\phi(x)$ by the least favorable distribution $f_0(x)$ of the form

$$f_0(x) = \begin{cases} (1-\epsilon)\,\phi(k_0)\,\left(\frac{k_0}{|x|}\right)^{k_0^2} & \text{for } |x| < k_0 \\ (1-\epsilon)\,\phi(x) & \text{for } k_0 \leq |x| \leq k_1 \\ (1-\epsilon)\,\phi(k_1)\,\left(\frac{k_1}{|x|}\right)^{k_1^2} & \text{for } |x| > k_1 \end{cases} \qquad (2.15)$$

For the choice $k_1 = -k_0 = k$, $-\log f_0(x)$ reduces to the ρ_1 function defined above except for a constant term. The constant k in the function ρ_1 can be chosen as $k = 1.345$ as suggested by Huber (see Huber, 1981). Note that the choice $k = \infty$

gives the classical ML estimate of θ. To find the robust estimate of θ, one can minimize the objective function $H_1(\theta)$ in equation (2.13) directly by using numerical methods. Equivalently, one can also solve iteratively the M estimating equations (which are the derivatives of $H_1(\theta)$ with respect to θ equating to zero)

$$\sum_j \left\{ \frac{1}{2}\psi^2(r_{1j}) + \frac{1}{2}\psi^2(r_{2j}) - c \right\} \frac{1}{f_{yy}(\omega_j|\theta)} \frac{\partial}{\partial\theta} f_{yy}(\omega_j|\theta) = 0 \qquad (2.16)$$

for the parameter θ starting from a set of initial values. To derive the above equation, we use the fact that $x \times \rho_1'(x) = \psi^2(x)$, where ρ_1' is the derivative of ρ_1 and the function ψ is the Huber's ψ function defined by $\psi(x) = \max(-k, \min(x, k))$. The tuning constant c is chosen such that the estimator of θ gives Fisher consistency. To define Fisher consistency, let us treat r_{1j} and r_{2j} as random samples from a distribution F. Consider the functional $T(F) = \theta$ defined on the space of distributions for r_1 and r_2. The functional $T(F)$ can be defined from equation (2.16) as the solution of

$$\int \left\{ \left(\frac{1}{2}\psi^2(r_1; T(F)) + \frac{1}{2}\psi^2(r_2; T(F)) - c(F) \right) \times \right.$$
$$\left. \times \left. \frac{1}{f_{yy}(\omega|T(F))} \frac{\partial}{\partial\theta} f_{yy}(\omega|\theta) \right|_{\theta=T(F)} \right\} dF(r_1, r_2) = 0. \qquad (2.17)$$

The estimator T of θ is said to be Fisher consistent if equation (2.17) is satisfied. It follows that the choice $c = E_F \psi^2$ ensures the Fisher consistency of the estimator of θ. In the next section, we show that the M estimator obtained by minimizing the objective function $H_1(\theta)$ in equation (2.13) or, equivalently, by solving the M estimating equation (2.16) has bounded influence function. We can relate equation (2.16) with the estimating equation suggested by Huber's "Proposal 2" for a scale parameter. If x is distributed as a normal with the scale parameter σ, then Huber's "Proposal 2" estimates σ by solving the equation

$$\sum_j \left[\psi^2 \left(\frac{x_j}{\sigma} \right) - c \right] = 0 \qquad (2.18)$$

where $c = E\psi^2 \left(\frac{x}{\sigma} \right)$. Note that if we choose $f_{yy}(\omega_j|\theta) = \theta = \sigma$, then equation (2.16) reduces to equation (2.18) except for the fact that the ψ function in equation (2.16)

is split into independent real and imaginary parts.

We can estimate $\boldsymbol{\theta}$ by directly optimizing the objective function $H_1(\boldsymbol{\theta})$ in equation (2.13) or by solving the M estimating equation (2.16) using Newton-Raphson method. The 1st step of the Newton-Raphson method solves the equation

$$\boldsymbol{\theta} = \boldsymbol{\theta}_0 + M_0^{-1} q_0 \qquad (2.19)$$

iteratively starting from an initial estimate $\boldsymbol{\theta}_0$ of $\boldsymbol{\theta}$, where

$$q_0 = \sum_j \left\{ \frac{1}{2}\psi^2(r_{1j}) + \frac{1}{2}\psi^2(r_{2j}) - c \right\} \frac{1}{f_{yy}(\omega_j|\boldsymbol{\theta}_0)} \frac{\partial}{\partial\boldsymbol{\theta}} f_{yy}(\omega_j|\boldsymbol{\theta}) \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

$$M_0 = \sum_j \left\{ \frac{1}{2}\psi(r_{1j})\psi'(r_{1j})r_{1j} + \frac{1}{2}\psi(r_{2j})\psi'(r_{2j})r_{2j} \right\} \times$$

$$\times \frac{\left(\frac{\partial}{\partial\boldsymbol{\theta}} f_{yy}(\omega_j|\boldsymbol{\theta})\right) \left(\frac{\partial}{\partial\boldsymbol{\theta}} f_{yy}(\omega_j|\boldsymbol{\theta})\right)^T}{\{f_{yy}(\omega_j|\boldsymbol{\theta}_0)\}^2} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

However, we prefer direct optimization of the objective function $H_1(\boldsymbol{\theta})$ to the Newton-Raphson method of solving the M estimating equation as in some cases the Newton-Raphson method failed to give us convergence in iteration. The *S-PLUS* function "nlminb" was used to minimize $H_1(\boldsymbol{\theta})$ in order to find the robust M estimate of $\boldsymbol{\theta}$. In the next section, we discuss the influence function as well as the asymptotic variance of the M estimators.

2.3 Influence Function of the M Estimator

Hampel (1974) introduced the influence function (IF) in order to investigate the infinitesimal behavior of real-valued functionals such as $T(F)$. Influence function can be considered as a technique of studying the local stability of an estimator in terms of the effect of point-mass perturbation of the data or the underlying distribution (Simpson et al, 1992). Recall the functional $T(F)$ of equation (2.17) defined on the space of distributions for r_1 and r_2. Let F_ϵ be a point-mass contamination of the target model F : $(1 - \epsilon)F + \epsilon\Delta_{r_1,r_2}$ for $0 \le \epsilon \le 1$. Here Δ_{r_1,r_2} is a delta function

which has point-mass 1 at (r_1, r_2) and 0 elsewhere. Then following Hampel et al (1986), we can define the influence function of T as

$$IF(r_1, r_2; T, F) = \lim_{\epsilon \to 0} \frac{T(F_\epsilon) - T(F)}{\epsilon} \qquad (2.20)$$

which is the directional derivative of T at $\epsilon = 0$. Using this definition, the influence function of the M estimator can be found by inserting $F_\epsilon = (1 - \epsilon)F + \epsilon \Delta_{r_1, r_2}$ for F into equation (2.17) and taking the derivative with respect to ϵ at $\epsilon = 0$. The influence function is obtained as

$$IF(r_1, r_2; T, F) = [-E_F(\dot{\Psi}_1)]^{-1} \Psi_1 \qquad (2.21)$$

where the functions Ψ_1 and $\dot{\Psi}_1$ are defined by

$$\Psi_1 = \left\{ \frac{1}{2}\psi^2(r_1) + \frac{1}{2}\psi^2(r_2) - c(F) \right\} \frac{1}{f_{yy}(\omega|T(F))} \frac{\partial}{\partial \theta} f_{yy}(\omega|\theta) \bigg|_{\theta = T(F)}$$

and

$$
\begin{aligned}
\dot{\Psi}_1 &= \frac{\partial}{\partial \theta} \Psi_1 \bigg|_{\theta = T(F)} \\
&= \frac{\partial}{\partial \theta} \left\{ \left(\frac{1}{2}\psi^2(r_1) + \frac{1}{2}\psi^2(r_2) - c(F) \right) \frac{\frac{\partial}{\partial \theta} f_{yy}(\omega|\theta)}{f_{yy}(\omega|\theta)} \right\} \bigg|_{\theta = T(F)} \\
&= - \left\{ \frac{1}{2}\psi(r_1)\psi'(r_1)r_1 + \frac{1}{2}\psi(r_2)\psi'(r_2)r_2 \right\} \times \\
&\qquad \times \frac{\left(\frac{\partial}{\partial \theta} f_{yy}(\omega|\theta) \right) \left(\frac{\partial}{\partial \theta} f_{yy}(\omega|\theta) \right)^T}{\{ f_{yy}(\omega|\theta) \}^2} \bigg|_{\theta = T(F)}
\end{aligned}
$$

We denote

$$IF(r_1, r_2; T, F) = M_1^{-1} \Psi_1 \qquad (2.22)$$

where $M_1 = -E_F(\dot{\Psi}_1)$. Note that as the function Ψ_1 is bounded as a function of (r_1, r_2), the influence function of the M estimator of θ is also bounded. However,

if we choose $k = \infty$ for the Huber's ψ function, which corresponds to the ML estimation, the estimate will give unbounded influence function as the function Ψ_1 is unbounded for this choice of k. Following Hampel et al (1986), it can be shown under mild regularity conditions that if $\hat{\theta}$ is an M estimator of θ, then $\sqrt{n}(\hat{\theta} - \theta)$ is asymptotically multivariate normal with mean vector 0 and a covariance matrix V. For a rigorous mathematical treatment, one can see Boos and Serfling (1980), Huber (1967, 1981), Clarke (1983, 1986), Heritier and Ronchetti (1994) as basic references. The asymptotic variance V of the M estimator can be obtained from the influence function as

$$V(T, F) \approx E(IF\, IF^T) = M_1^{-1} Q_1 M_1^{-1} \quad \text{(say)} \tag{2.23}$$

where

$$Q_1 = E_F(\Psi_1 \Psi_1^T)$$

$$= E_F \left\{ \frac{1}{2}\psi^2(r_1) + \frac{1}{2}\psi^2(r_2) - c(F) \right\}^2 \left. \frac{\left(\frac{\partial}{\partial \theta} f_{yy}(\omega|\theta)\right) \left(\frac{\partial}{\partial \theta} f_{yy}(\omega|\theta)\right)^T}{\{f_{yy}(\omega|\theta)\}^2} \right|_{\theta = T(F)}$$

The variance function $V(T, F)$ can be approximated by replacing the distribution function F with its empirical distribution function F_n as $\hat{V} = \hat{M}_1^{-1} \hat{Q}_1 \hat{M}_1^{-1}$, where

$$\hat{M}_1 = \frac{1}{n} \sum_j \left\{ \frac{1}{2}\psi(r_{1j})\psi'(r_{1j})r_{1j} + \frac{1}{2}\psi(r_{2j})\psi'(r_{2j})r_{2j} \right\} \times$$

$$\times \left. \frac{\left(\frac{\partial}{\partial \theta} f_{yy}(\omega_j|\theta)\right) \left(\frac{\partial}{\partial \theta} f_{yy}(\omega_j|\theta)\right)^T}{\{f_{yy}(\omega_j|\theta)\}^2} \right|_{\theta = \hat{\theta}}$$

$$\hat{Q}_1 = \frac{1}{n} \sum_j \left\{ \frac{1}{2}\psi^2(r_{1j}) + \frac{1}{2}\psi^2(r_{2j}) - c \right\}^2 \times$$

$$\times \left. \frac{\left(\frac{\partial}{\partial \theta} f_{yy}(\omega_j|\theta)\right) \left(\frac{\partial}{\partial \theta} f_{yy}(\omega_j|\theta)\right)^T}{\{f_{yy}(\omega_j|\theta)\}^2} \right|_{\theta = \hat{\theta}}$$

In the next section, we discuss the robustness aspects of model selection based on the classical Akaike Information Criterion (AIC).

2.4 Robust Model Selection

In sections 2.2 and 2.3, we discussed the robust estimation of the power spectrum in terms of the parameter estimates of an $ARMA(p,q)$ process provided that the orders p and q of the process are given a priori. In practice, we need to find p and q in order to obtain a suitable model for the process y_t. Choosing a suitable model is an important issue in time series analysis. As a classical method, one can consider the Akaiki Information Criterion (AIC) (see Akaiki, 1973) for choosing a correct model. The classical AIC amounts to choosing the model that minimizes $-2 \log L(\theta) + 2m$, where m is the dimension of the vector of parameters θ. For more details, see Priestley (1981). This procedure can be generalized by replacing $2m$ by κm for a given value of κ (Bhansali and Downham, 1977). As the classical AIC is not robust, we consider using a robust version of this Information Criterion. Ronchetti (1997) discusses different aspects of model selection based on the robust version of the log-likelihood function. Here we adopt a similar idea to define the robust AIC based on the robust version of the approximate likelihood function of the discrete Fourier transforms. Using the robustified negative log-likelihood function as defined in equation (2.13), we define a robust AIC in the form

$$RAIC_1(m) = c \sum_j \log f_{yy}(\omega_j|\hat{\theta}) + \sum_j \rho_1(r_{1j}|\hat{\theta}) + \sum_j \rho_1(r_{2j}|\hat{\theta}) + 2\alpha_m \qquad (2.24)$$

where m is the dimension of θ. The α_m can be chosen as $\alpha_m = m\frac{E[\psi^2(r)]}{E[\psi'(r)]}$ (see Ronchetti, 1997). Notice that if we choose $k = \infty$, then the $RAIC_1$ in equation (2.24) reduces to the classical AIC. In our complex variable case, the α_m may be estimated by

$$\hat{\alpha}_m = m\frac{\sum_j [\psi^2(r_{1j}) + \psi^2(r_{2j})]}{\sum_j [\psi'(r_{1j}) + \psi'(r_{2j})]}. \qquad (2.25)$$

We minimize this $RAIC_1(m)$ with respect to m for selecting the best models in some real data sets described in the next section.

2.5 Application: Analysis of Sea Level Data

We apply the proposed model selection and M estimation method discussed in the previous sections to three sets of data on sea level records which were collected at the harbors Halifax, Sydney and Yarmouth in Nova Scotia, Canada. Each data set consists of 2048 observations collected at a rate of 4 per hour beginning midnight January 1, 1997. The sea levels were measured as heights in mm above chart datum. The sea level record at Halifax harbor has been plotted in Figure 2.1. The sea levels at Sydney and Yarmouth harbors have been plotted in Figure 2.3. In section 2.1, we discuss the behavior of the sea level record at Halifax. Similar harmonic patterns are observed in the sea levels at Sydney and Yarmouth. But, as we pointed out earlier, these harmonic components are very hard to model properly due to the variation in the amplitudes and phases over time. Here we consider fitting an $ARMA(p, q)$ model to each data set using the proposed M estimation method which is not affected by the harmonic components mixed with the $ARMA$ process. For ease of computation, we consider the standardized process y_t defined by $y_t = (x_t - \bar{x})/1000$, where x_t is the original process of the sea level data and $\bar{x}$ is the mean of the process x_t. To find a suitable model for each of the three data sets, we consider fitting the $ARMA(p, q)$ model for each $p = 0, 1, \ldots, 10$ and $q = 0, 1, \ldots, 10$. For each combination of p and q, we estimate the corresponding parameters of the $ARMA(p, q)$ process by minimizing the objective function $H_1(\theta)$ in equation (2.13). The S-PLUS function "nlminb" has been used to minimize this objective function. We then compute the robust model selection criterion $RAIC_1(m)$ defined in equation (2.24). In total, for each data set 121 models have been fitted and their corresponding $RAIC_1$s have also been found in the search of the correct model. The model with the minimum $RAIC_1$ is considered to be the best model. The function $RAIC_1$ is found to be minimum at $ARMA(5, 7)$ for the Halifax data, $ARMA(8, 3)$ for the Sydney data and $ARMA(6, 5)$ for the Yarmouth data. The $RAIC_1$s of some selected models have been shown in Table 2.1 for all the three data sets.
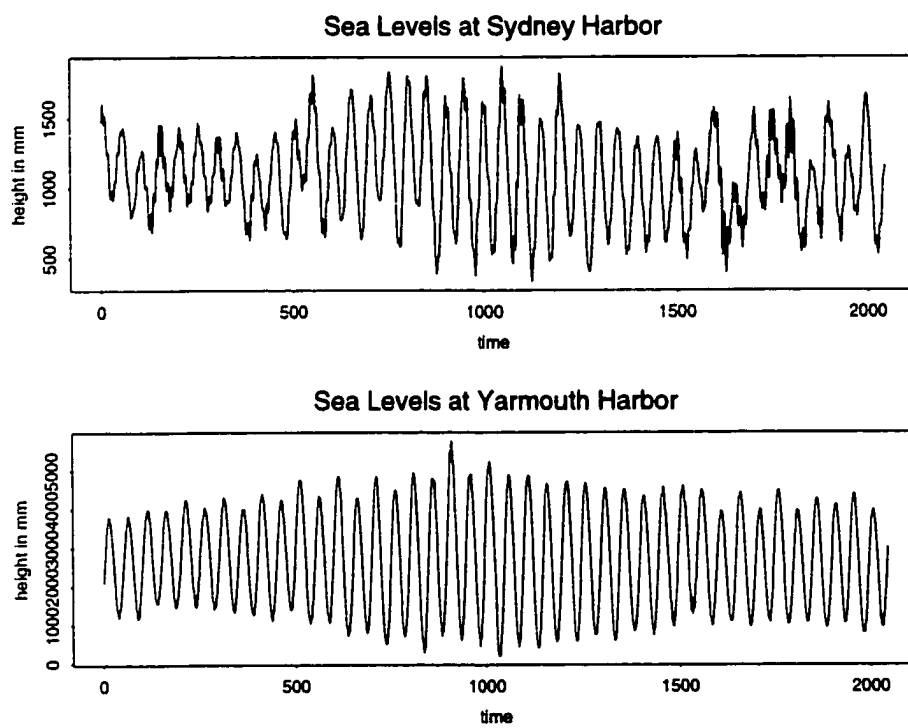
Figure 2.3: Plots of the sea level data

Table 2.1: Values of $RAIC_1$ for some selected models for all the three data sets

Model	Halifax	Sydney	Yarmouth
$ARMA(1,1)$	-10285.18	-11105.91	-9150.37
$ARMA(2,1)$	-10311.95	-11111.27	-9142.99
$ARMA(2,2)$	-10310.62	-11234.48	-9209.16
$ARMA(5,5)$	-10383.75	-11463.24	-9296.58
$ARMA(5,7)$	**-10406.41**	-11487.89	-9294.82
$ARMA(6,5)$	-10382.26	-11457.49	**-9300.97**
$ARMA(7,4)$	-10393.31	-11465.95	-9299.74
$ARMA(8,3)$	-10364.08	**-11495.32**	-9295.24
$ARMA(8,8)$	-10400.92	-11492.56	-9296.19
$ARMA(10,10)$	-10391.48	-11492.38	-9296.56

In Figure 2.4, we plot the periodograms with corresponding robust spectrum estimates assuming different $ARMA$ models for each data set. In the top panel of Figure 2.4, the log of the periodograms as well as the log of the corresponding spectrum estimates have been plotted for the Halifax sea levels considering the four models $ARMA(1,1)$, $ARMA(5,5)$, $ARMA(5,7)$ and $ARMA(8,8)$ which are shown respectively from top to bottom in this panel. Each spectrum has been displaced by 3 units to provide a clear view of the spectrum lines. For $ARMA(1,1)$ model, the estimated spectrum line does not seem to be fitting well as it fails to model the bumps in the periodograms of the process. When we increase the orders of the $ARMA$ process (see the spectrum line of $ARMA(5,5)$, for example), we observe the expected bumps in the estimated spectrum line corresponding to the bumps in the periodograms, which indicates a better fit. Recall that the $RAIC_1$ shows the best model for the Halifax data at $ARMA(5,7)$ and for this model, it is clear that the bumps in the periodograms have been "nicely" modeled by the estimated spectrum line. If we increase the orders of the $ARMA$ process further, the shape of the spectrum line does not seem to be changing significantly as we see from the estimated spectrum line of $ARMA(8,8)$. We follow the same procedure to plot the periodograms and the corresponding spectrum estimates for the Sydney and Yarmouth sea level data sets.

In the middle panel of Figure 2.4, the periodograms and the corresponding estimated spectrum lines have been plotted for the Sydney sea levels considering the four models $ARMA(1,1)$, $ARMA(5,5)$, $ARMA(8,3)$ and $ARMA(8,8)$ which are shown respectively from top to bottom in the plot. Here also we observe that for the best fitted model $ARMA(8,3)$, the spectrum estimates are quite satisfactory as the bumps in the periodograms are nicely modeled by the estimated spectrum line at $ARMA(8,3)$. On the other hand, with $ARMA(1,1)$, the spectrum line does not seem to be fitting well as it cannot model the bumps in the periodograms, which is observed from the top line of the panel. At $ARMA(8,3)$ fitting, we observe a bump in the spectrum line at frequency near .1, which seems to be reasonable as we also observe a bump in the periodograms at this frequency. In the bottom panel of Figure 2.4, the periodograms and the corresponding estimated spectrum lines have been plotted for the Yarmouth sea levels considering the four models $ARMA(1,1)$, $ARMA(5,5)$, $ARMA(6,5)$ and $ARMA(8,8)$ which are shown respectively from top to bottom. Here we observe that at the best fitted model $ARMA(6,5)$, the estimated spectrum line gives a bump at a frequency near 2.7, which is not present in $ARMA(1,1)$ and $ARMA(5,5)$ fitting. The shapes of the estimated spectrum lines of $ARMA(6,5)$ and $ARMA(8,8)$ models appear to be very similar.
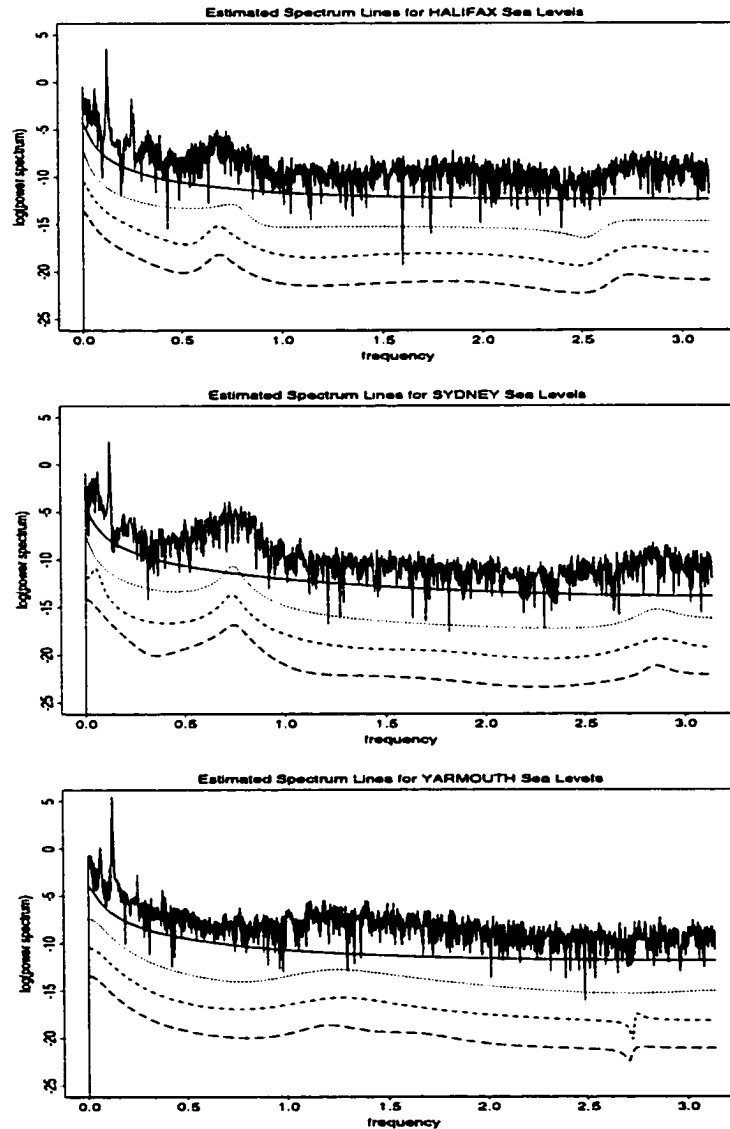
Figure 2.4: Plots of the log-periodograms and log of the corresponding spectrum estimates for the sea level records at Halifax, Sydney and Yarmouth. The four spectrum lines (from top to bottom) in each panel are estimated considering respectively the models - 1st line: $ARMA(1,1)$ for each data set; 2nd line: $ARMA(5,5)$ for each data set; 3rd line: $ARMA(5,7)$ for Halifax data, $ARMA(8,3)$ for Sydney data, $ARMA(6,5)$ for Yarmouth data; 4th line: $ARMA(8,8)$ for each data set. (Each line is displaced by 3 units.)

To have some idea about the potential influential discrete Fourier transforms, we consider finding the weights of the Fourier transforms at different frequencies. We define

$$v_j = \frac{1}{2}\frac{\psi(r_{1j})}{r_{1j}} + \frac{1}{2}\frac{\psi(r_{2j})}{r_{2j}} \qquad (2.26)$$

as the weight corresponding to the frequency at point $j = 1, \ldots, [(n-1)/2]$. Notice that if we put $\psi(r) = r$, then $v_j = 1$, and the M estimators become the classical ML estimators. In Figure 2.5, we plot the weights of the M estimates at different models for each of the three data sets. In the top three panels of Figure 2.5, the log-periodograms and the log of the spectrum estimates have been shown as described earlier. Their corresponding weight functions have been plotted in the three panels of the second row in Figure 2.5. From these plots it is observed that the points corresponding to the huge spikes (possibly due to the harmonic or tidal components) in the periodograms are severely downweighted as they should be. Another interesting pattern in the weight functions can be observed from the $ARMA(1,1)$ fitting. As the spectrum estimates of the $ARMA(1,1)$ fit cannot model the bumps in the periodogram plots, the points corresponding to these bumps are also severely downweighted. But when we plot the weight functions of the best fitted models for each of the three data sets, we observe a "significant" change in their values. From the bottom three panels of Figure 2.5, it is clear that as the spectrum estimates model the bumps in the periodograms, their corresponding weight functions are not severely downweighted any more. On the other hand, the weights corresponding to the huge spikes in the periodograms are close to zero as we have seen in the $ARMA(1,1)$ case. In the weight functions of the three data sets, we observe a common behavior at a certain frequency. At frequency near .126, the weight function is close to zero for each of the three data sets. The frequency .126 corresponds to a period of about 12.5 hours and represents the fundamental lunar component of the tide.

Now, we consider the fact that if there are no influential points in the discrete Fourier transforms, the M estimation method and the classical ML method should give similar results. To check this, we first show the M estimates and the ML estimates (with standard errors in parenthesis) of the parameters of an $ARMA(1,1)$
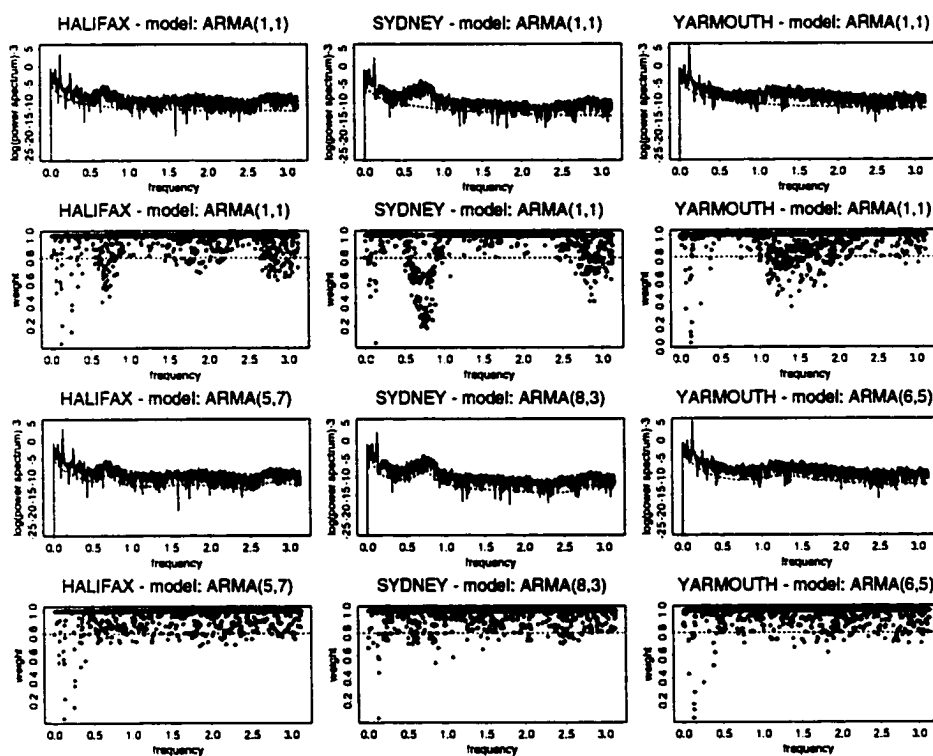
Figure 2.5: Weights of the Fourier transforms at different frequencies considering an $ARMA(1,1)$ model and a best fitted model for each of the three data sets of sea level records.

model in Table 2.2 for all the three data sets. From the table, it is seen that the ML estimates are "significantly different" from the M estimates for each of the three data sets. We also observe a significant difference between the standard errors of the parameter estimates in the M estimation and the ML estimation. We then consider fitting the model after deleting the influential observations in the discrete Fourier transforms. We take a point as influential if its corresponding weight is less than .80. After deleting the influential Fourier transforms, we estimate the parameters of the $ARMA(1,1)$ model using both M estimation and ML estimation methods. The estimates (with standard errors in parenthesis) are shown in Table 2.3. From the table, it is observed that the M estimates and the ML estimates are very close (as is expected) after deleting the influential Fourier transforms. Also note that the ML estimates have smaller standard errors as compared to the M estimates when the influential tidal frequencies are omitted. From this viewpoint, one may think of simply removing the periodograms with substantial peaks from the fitting procedure and then use the standard maximum likelihood estimation. However, the use of weight functions in the M estimation procedure gives us a precise idea about the tidal frequencies and we can incorporate the tidal information in our robust procedure without any effect on the estimates of the surge spectrum.

Table 2.2: Parameter estimates (with standard errors in parenthesis) of the $ARMA(1,1)$ model for all the three data sets

Method	Parameter	Halifax	Sydney	Yarmouth
M	ϕ_1	-.9794 (.00403)	-.9735 (.00741)	-.9700 (.00498)
estimation	ζ_1	-.2609 (.02447)	.1155 (.02869)	-.1231 (.02051)
	σ	.0363 (.00058)	.0244 (.00043)	.0533 (.00082)
ML	ϕ_1	-1.0155 (.04962)	-1.0209 (.03794)	-.9877 (.19651)
estimation	ζ_1	3.0391 (1.21016)	.4410 (.06361)	.6626 (.08936)
	σ	.0221(.01188)	.0457 (.00338)	-.1196 (.02551)

Table 2.3: Parameter estimates (with standard errors in parenthesis) of the $ARMA(1,1)$ model for all the three data sets after deleting the possible influential Fourier transforms

Method	Parameter	Halifax	Sydney	Yarmouth
M	ϕ_1	-.9854 (.00350)	-.9794 (.00579)	-.9768 (.00222)
estimation	ζ_1	-.3036 (.02314)	.0538 (.02365)	-.1795 (.01840)
	σ	.0340 (.00052)	.0218 (.00032)	.0501 (.00071)
ML	ϕ_1	-.9844 (.00252)	-.9815 (.00266)	-.9733 (.00239)
estimation	ζ_1	-.2909 (.01555)	.0601 (.01630)	-.1830 (.01539)
	σ	.0330 (.00035)	.0216 (.00024)	.0489 (.00051)

We plot the log of the periodograms and the log of the corresponding spectrum estimates minus 10 considering the best fitted models $ARMA(5,7)$ for the Halifax data, $ARMA(8,3)$ for the Sydney data and $ARMA(6,5)$ for the Yarmouth data. For comparison, we estimate the power spectrums using both the M estimation and the ML estimation methods. The plots are shown in Figure 2.6. From the plots we observe a significant difference in the power spectrums estimated by the M and ML methods. For the ML method, we see that the spectrum estimates give a large peak at frequencies near zero which is due to the huge spikes in the periodograms. On
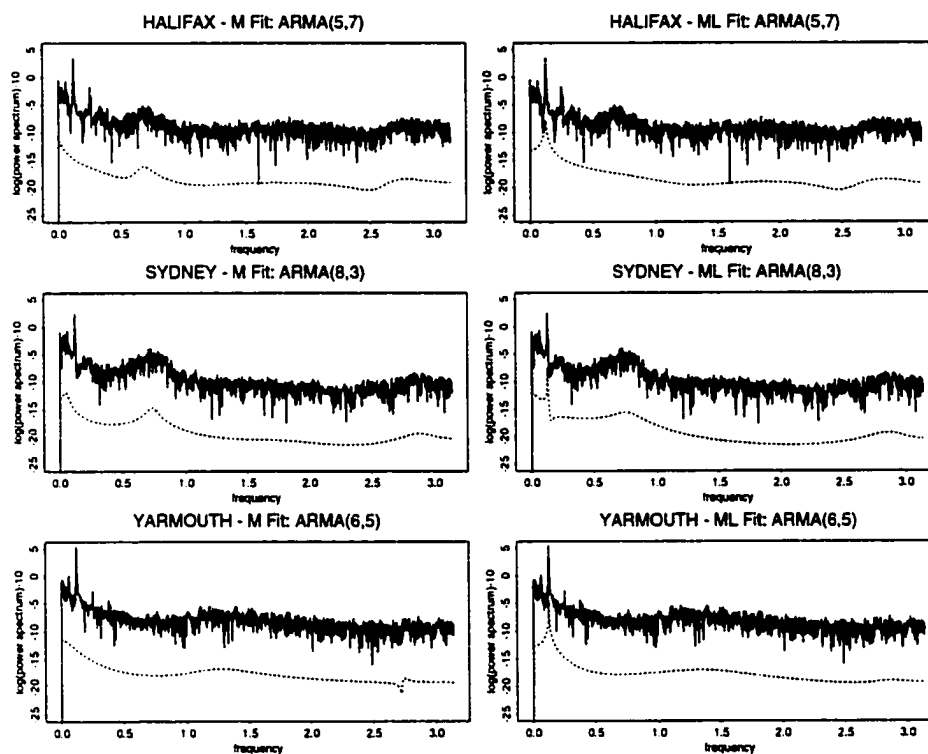
Figure 2.6: Plots of the log of the estimated power spectrum minus 10 based on M estimation and ML estimation methods

the other hand, those peaks in the spectrum estimates disappear when we estimate them using the M method. The fact is that the M method estimates the spectrum by downweighting the huge spikes in the periodograms.

2.6 Discussion

In this work, we mainly focus on the robust model selection and robust spectrum estimation in an $ARMA$ process contaminated with harmonic components. Given orders p and q of the $ARMA(p,q)$ process, we can estimate the corresponding parameters of the process by minimizing the objective function $H_1(\theta)$ in equation (2.13) or, equivalently, by solving the M estimating equation (2.16) iteratively. The proposed M estimators are Fisher consistent and have bounded influence functions. The M estimators can be related to the classical maximum likelihood type estimators. The choice of the bound $k = \infty$ for the Huber's ψ function gives the corresponding classical estimators of an $ARMA$ process. But as pointed out earlier, these classical estimators have unbounded influence functions. The influence function of an estimator may be used to find its asymptotic variance. This asymptotic variance can be approximated by replacing the true distribution F with its empirical distribution function F_n. For choosing a suitable model, we minimize the robust model selection criterion $RAIC_1$ defined in equation (2.24) with respect to the orders p and q of the $ARMA(p,q)$ process. For each p and q, we estimate the corresponding $ARMA$ components and hence the $RAIC_1$ of the process. The model with minimum $RAIC_1$ is considered to be the best model.

The proposed M estimation and robust model selection criterion have been applied to three data sets on sea level records, where we observe harmonic contamination in the process along with the original $ARMA$ process. The classical method is severely affected by this harmonic contamination when we try to estimate the power spectrum assuming an $ARMA(p,q)$ process. It is found that the proposed M estimation is useful in estimating the $ARMA$ components as well as the power spectrum of the process robustly by avoiding the presence of harmonic (or tidal) components in the process.

Halifax and Sydney include broad band spectral features at frequency near .7. There also appear to be broad band features at frequency near 2.8, which corresponds to the 5th harmonic of the feature at frequency .7, folded back into the interval $[0, \pi]$ (e.g. $2\pi - 5 \times .7$). These features were identified by Smith and Miyaoka (1999) as

being indicative of harbor seiches, which should, in theory, have spectral features at odd harmonics of a fundamental frequency determined by harbor and forcing characteristics.

It is interesting to note that the maximum likelihood fit at Halifax nearly fails to identify the feature near frequency .7. The influence of the tide on the spectral estimate is very apparent in the Yarmouth record. This is perhaps due to the high concentration of energy in the M2 cycle at Yarmouth, which is located near the mouth of the Bay of Funday, which has the world's highest tides because its natural resonant period is close to 12.5 hours (the M2 period). The Yarmouth record also includes a broad band feature near frequency 1.3 which may represent a seiche, and there is an unusual looking feature (in the M estimate only) near frequency 2.7. The latter feature requires further investigation to determine whether it is a real feature, or an artifact of the M estimation method.

Chapter 3

Robustness Aspects of Model Selection in Autoregressive Processes

3.1 Introduction

In chapter 2, we introduced the robust model selection criterion $RAIC_1$ in the frequency domain for choosing the appropriate orders of an $ARMA$ process. The $RAIC_1$ was based on the approximate likelihood function of the discrete Fourier transforms of a time series process, where we assumed that the process consists of the sum of two parts, the tide and the surge. We developed a robust method to estimate the parameters of the surge spectrum, where the surge was assumed to follow an $ARMA(p,q)$ process. This robust method was found to be useful in estimating the parameters of the $ARMA(p,q)$ process by downweighting the influence of tidal (or harmonic) components on the data. The resulting robust estimates of the $ARMA$ components were used to calculate the $RAIC_1$ for choosing the correct order of an $ARMA$ process. Note that this technique of robust estimation and model selection is useful in the case where a time series $ARMA$ process is mixed with some tidal or harmonic components. In the frequency domain, we can identify and downweight the tidal components using the proposed M estimation method. However, there may be certain situations where a time series process does not contain any harmonic or tidal components. Instead,

38

the process may be contaminated by some random outliers. The outliers commonly encountered in time series process are termed as "innovation" or "additive" outliers which will be discussed later. In this chapter, we introduce modifications of the usual robust model selection criterion in the time domain for choosing the appropriate order of an autoregressive process in the presence of outliers.

Robust model selection in an autoregressive process has received attention from a number of authors (see Martin, 1980; Hampel et al, 1986; Behrens, 1990; Ronchetti, 1997). Ronchetti (1997) discusses some robust procedures for choosing order q of an $AR(q)$ process. For an autoregressive process of order q of the form

$$y_t = \sum_{j=1}^{q} \phi_j y_{t-j} + u_t = \mu_t + u_t \tag{3.1}$$

where $\mu_t = \sum_{j=1}^{q} \phi_j y_{t-j}$, u_t are i.i.d. $N(0, \sigma^2)$, $(t = q+1, \ldots, n)$, Ronchetti (1997) suggests minimizing the robust AIC

$$RAIC^*(q) = \sum_{t=q+1}^{n} w(y_{t-1}^{t-q}) \, \rho \left(\frac{y_t - \hat{\mu}_t}{\sigma} \right) + \alpha_q \tag{3.2}$$

with respect to order q, for $y_{t-1}^{t-q} = (y_{t-1}, \ldots, y_{t-q})^T$, $\hat{\mu}_t = \sum_{j=1}^{q} \hat{\phi}_j y_{t-j} = \hat{\phi}^T y_{t-1}^{t-q}$ and $\hat{\phi} = (\hat{\phi}_1, \ldots, \hat{\phi}_q)^T$. For simplicity, σ is assumed to be known. A robust estimate, $\hat{\phi}$ is chosen as the generalized M (GM) estimate of the autoregressive parameter ϕ. A possible choice of the ρ function is of the form

$$\rho(x) = \begin{cases} \frac{1}{2}x^2 & \text{for } |x| \leq k \\ k|x| - \frac{1}{2}k^2 & \text{otherwise} \end{cases} \tag{3.3}$$

which corresponds to the Huber's least favorable distribution for location. The commonly used value of the constant k is 1.345. The term α_q (as proposed by Behrens, 1991) can be defined as $\alpha_q = \text{tr}(M^{-1}Q)$ with $M = E\left[w\left(y_{t-1}^{t-q}\right) \psi'(u_t/\sigma) \left(y_{t-1}^{t-q}\right) \left(y_{t-1}^{t-q}\right)^T \right]$, $Q = E\left[w^2 \left(y_{t-1}^{t-q}\right) \psi^2(u_t/\sigma) \left(y_{t-1}^{t-q}\right) \left(y_{t-1}^{t-q}\right)^T \right]$. ψ is the derivative of the function ρ. Here $w(x)$ is a weight function which downweights any outlying point x. For a good point

x, $w(x)$ is expected to be close to 1. α_q can be estimated by its sample counterpart as $\hat{\alpha}_q = \text{tr}(\hat{M}^{-1}\hat{Q})$, where $\hat{M} = \sum_{t=q+1}^{n} \left[w\left(y_{t-1}^{t-q}\right) \psi'((y_t - \hat{\mu}_t)/\sigma) \left(y_{t-1}^{t-q}\right) \left(y_{t-1}^{t-q}\right)^T \right]$ and $\hat{Q} = \sum_{t=q+1}^{n} \left[w^2\left(y_{t-1}^{t-q}\right) \psi^2((y_t - \hat{\mu}_t)/\sigma) \left(y_{t-1}^{t-q}\right) \left(y_{t-1}^{t-q}\right)^T \right]$.

Note that for the choice $w(x) = 1$ and $\rho(x) = \frac{1}{2}x^2$, the robust model selection criterion $RAIC^*(q)$ reduces to the classical Akaike Information Criterion, $AIC^*(q)$, defined by

$$AIC^*(q) = \frac{1}{2} \sum_{t=q+1}^{n} \left(\frac{y_t - \hat{\mu}_t}{\sigma} \right)^2 + q. \tag{3.4}$$

In the following section, we point out some limitations of the robust criterion $RAIC^*$ and suggest a modified version of this criterion.

3.2 A Modified Robust Model Selection Criterion

The $RAIC^*(q)$ in equation (3.2) seems to have the following drawbacks. The first term in the expression for $RAIC^*(q)$ is the sum of $(n - q)$ observations which clearly depends on the order q. So the number of observations to be summed up changes when we change the order q of the process. In practice, we need to keep the number of residuals fixed for comparing one model with another. As a remedy, we can vary t in equation (3.2) from $q^* + 1$ to n, where q^* is the maximum possible value of q under consideration for the order of the autoregressive process.

Table 3.1: Weight functions for $AR(1)$ and $AR(2)$ models for the setup in equation (3.6)

Index	y_t	$AR(1)$ Model			$AR(2)$ Model		
		y_{t-1}	$w(y_{t-1})$		y_{t-1}	y_{t-2}	$w(y_{t-1}, y_{t-2})$
1	y_3^*	y_2	1		y_2	y_1	1
2	y_4	y_3^*	0		y_3^*	y_2	0
3	y_5^*	y_4	1		y_4	y_3^*	0
4	y_6	y_5^*	0		y_5^*	y_4	0
5	y_7^*	y_6	1		y_6	y_5^*	0
6	y_8	y_7^*	0		y_7^*	y_6	0
7	y_9	y_8	1		y_8	y_7^*	0
8	y_{10}	y_9	1		y_9	y_8	1

Another important issue concerns with the weight functions. Note that the first term in the expression for $RAIC^*(q)$ in equation (3.2) is given by

$$\sum_t w(y_{t-1}^{t-q}) \rho \left(\frac{y_t - \hat{\mu}_t}{\sigma} \right). \tag{3.5}$$

The weight function $w(y_{t-1}^{t-q})$ in the above equation depends on the order q of the process. As a typical example, we may consider the following setup of the process

$\{y_t\}$. Suppose, the first 10 values of $\{y_t\}$ are obtained as

$$(y_1, y_2, y_3^*, y_4, y_5^*, y_6, y_7^*, y_8, y_9, y_{10}), \tag{3.6}$$

where y_3^*, y_5^* and y_7^* are considered to be extreme outliers. Now, if one wants to compare an $AR(2)$ model with an $AR(1)$ model, the regression setup and the corresponding weight functions for the first 10 points may be defined as in Table 3.1. In the table, the weights $w(y_{t-1})$ for the $AR(1)$ process corresponding to the indices 2, 4 and 6 can be considered as zero as we assume that the values y_3^*, y_5^* and y_7^* are extreme outliers. For the same reason, the weights $w(y_{t-1}, y_{t-2})$ for the $AR(2)$ process corresponding to the indices 2, ... , 7 can be considered as zero. In such cases, even if the $\rho((y_t - \hat{\mu}_t)/\sigma)$ in equation (3.5) are the same for both $AR(1)$ and $AR(2)$ processes, the weight function $w(y_{t-1})$ for the $AR(1)$ process differs significantly from the weight function $w(y_{t-1}, y_{t-2})$ for the $AR(2)$ process. As a consequence, the decision based on the $RAIC^*(q)$ in equation (3.2) may be quite misleading. In order to overcome these problems, we suggest normalizing the $RAIC^*(q)$ by dividing it with the sum of their corresponding weight functions. So as modified form of the robust model selection criterion, we propose minimizing the objective function

$$RAIC(q) = \left(1 \bigg/ \sum_{t=q^*+1}^{n} w\left(y_{t-1}^{t-q}\right) \right) \left[\sum_{t=q^*+1}^{n} w\left(y_{t-1}^{t-q}\right) \rho\left(\frac{y_t - \hat{\mu}_t}{\sigma}\right) + c_q\, \alpha_q \right] \tag{3.7}$$

with respect to q, $q = 1, \ldots, q^*$, q^* is the maximum possible value of q, c_q is a tuning constant to be chosen as $c_q = \kappa\,(n - q^*)/\sum_{t=q^*+1}^{n} w\left(y_{t-1}^{t-q}\right)$ for a constant κ, and $\alpha_q = \text{tr}(M^{-1}Q)$. M and Q can be estimated as

$$\hat{M} = \sum_{t=q^*+1}^{n} w\left(y_{t-1}^{t-q}\right) \psi'\left(\frac{y_t - \hat{\mu}_t}{\sigma}\right) \left(y_{t-1}^{t-q}\right) \left(y_{t-1}^{t-q}\right)^T$$

$$\hat{Q} = \sum_{t=q^*+1}^{q} w^2\left(y_{t-1}^{t-q}\right) \psi^2\left(\frac{y_t - \hat{\mu}_t}{\sigma}\right) \left(y_{t-1}^{t-q}\right) \left(y_{t-1}^{t-q}\right)^T.$$

Note that for the choice $w(x) = 1$ and $\rho(x) = \frac{1}{2}x^2$, the $RAIC(q)$ in equation (3.7) reduces to the corresponding classical Akaike Information Criterion. The value of

the constant κ is chosen as 2 in our simulation study as this value was found to be reasonable in determining the correct order of a contamination process. To calculate the $RAIC(q)$, we need to estimate the corresponding autoregressive parameters using a suitable robust method. Martin (1980) discusses GM estimation of the autoregressive parameters in an $AR(q)$ process. Here we use this GM estimation method with some modifications of the weight functions. Once we have the parameter estimates, we can apply the proposed $RAIC(q)$ criterion for choosing the appropriate order of an autoregressive process.

A simulation has been carried out to study the performance of the robust criterion $RAIC$ based on samples of size $n = 200$. Before we study the simulation results, we discuss the robust estimation of the autoregressive parameters in the next section. In section 3.4, we present the simulation results comparing the performance of the proposed robust model selection criterion with the classical AIC criterion in the presence of innovation and additive outliers. Section 3.5 gives the conclusions of the chapter.

3.3 GM Estimation of the Autoregressive Parameters

Let us recall the autoregressive model of order q defined in equation (3.1). Following Martin (1980), the autoregressive parameter ϕ can be estimated by minimizing an objective function of the form

$$H(\phi) = \sum_{t=q+1}^{n} w(y_{t-1}^{t-q}) \, \rho \left(\frac{y_t - \phi^T y_{t-1}^{t-q}}{\sigma} \right) \tag{3.8}$$

or, equivalently, by solving the GM estimating equation

$$\sum_{t=q+1}^{n} w(y_{t-1}^{t-q}) \, \psi \left(\frac{y_t - \phi^T y_{t-1}^{t-q}}{\sigma} \right) (y_{t-1}^{t-q}) = 0 \tag{3.9}$$

iteratively. Here ψ is the derivative of the function ρ. Note that for the choice $\psi(x) = x$ and $w(x) = 1$ for each x, equation (3.9) gives the conditional LS estimate of the autoregressive parameter ϕ. Starting from a set of initial estimates ϕ_0 and σ_0 of ϕ and σ, respectively, the above GM estimating equation can be solved iteratively as $\phi = \phi_0 + M_0^{-1} q_0$, where $M_0 = (1/\sigma_0) \sum_{t=q+1}^{n} w\left(y_{t-1}^{t-q}\right) \psi'((y_t - \phi_0^T y_{t-1}^{t-q})/\sigma_0)\left(y_{t-1}^{t-q}\right)\left(y_{t-1}^{t-q}\right)^T$ and $q_0 = \sum_{t=q+1}^{n} w\left(y_{t-1}^{t-q}\right) \psi((y_t - \phi_0^T y_{t-1}^{t-q})/\sigma_0)\left(y_{t-1}^{t-q}\right)$. For the initial estimate ϕ_0, the initial σ_0 can be obtained as $\sigma_0 = \text{med} |y_t - \phi_0^T y_{t-1}^{t-q}|/.6745$. For each updated estimate of ϕ, the estimate of σ should be updated accordingly. The weights $w(\mathbf{x})$ are chosen as

$$w(\mathbf{x}) = w(\mathbf{x}, S_x) = \min\left\{1, \left(\frac{b}{\mathbf{x}^T S_x^{-1} \mathbf{x}}\right)^{\frac{\gamma}{2}}\right\} \qquad (3.10)$$

where S_x is the minimum volume ellipsoid (MVE) estimator (developed by Rousseeuw and van Zomeren, 1990) of the scale of $\mathbf{x}$. γ is a positive integer to be chosen as greater than or equal to 1. b is chosen as the $(1 - \alpha_0)$ quantile of the chi-squared distribution with degrees of freedom equal to the dimension of $\mathbf{x}$. In the simulation study, $\gamma = 2$ and $\alpha_0 = .05$ have been used. Note that the influence function of the GM estimates is bounded. In Chapter 4, we discuss the infinitesimal behavior of the GM estimates based on the influence function for a general model which includes nonlinear regression with autoregressive errors.

3.4 Simulation Study

To study the performance of the robust Akaike Information Criterion $RAIC$, a number of simulations have been carried out. Each simulation is based on 500 random samples each with sample size $n = 200$. Samples have been drawn from the processes $AR(1)$ (with parameter $\phi_1 = .6$) and $AR(2)$ (with parameters $\phi_1 = .6$ and $\phi_2 = .3$) assuming that $(1 - \epsilon)$ proportion of the innovation process u_t in equation (3.1) comes from $N(0, 1)$ population and ϵ proportion of u_t from $N(0, \tau^2)$, with $\tau^2 = 9, 25, 100$. This type of contamination is due to innovation outliers. We are interested in finding the proportion of times that the $RAIC$ is capable of choosing the true model. For each data set, $RAIC(1)$, $RAIC(2)$ and $RAIC(3)$ have been computed and the

observed correct model is that which minimizes $RAIC(q)$; $q = 1, 2, 3$. If the true model is $AR(1)$, we find the proportion of times the observed correct model chooses the true model $AR(1)$. Similarly, for the true $AR(2)$ model, we find the proportion of times the observed correct model chooses the true model $AR(2)$. We also study the performance of the classical model selection criterion

$$AIC(q) = \frac{1}{2} \sum_{t=q*+1}^{n} \left(\frac{y_t - \hat{\mu}_t}{\sigma} \right)^2 + q \qquad (3.11)$$

in the presence of outliers. The classical Akaike's Criterion amounts to choosing the model that minimizes $AIC(q)$.

Table 3.2 presents the simulation results based on 500 random samples each of size $n = 200$ drawn from the $AR(1)$ process with parameter $\phi_1 = .6$ in the presence of innovation outliers. For a set of values of ϵ (proportion of contamination) and τ^2 (variance of the contamination distribution), we find the success rates (the percentages of choosing the true model) based on both the robust model selection criterion $RAIC$ and the classical Akaike's Criterion AIC. It is evident that for uncontaminated data, the performances of both the $RAIC$ and AIC in choosing the true model are quite satisfactory. When the proportion of contamination increases, we observe poor performance in the classical AIC. On the other hand, the robust criterion $RAIC$ is more stable as compared to the classical AIC in selecting the correct model. However, we observe a decreasing tendency in the success rates of $RAIC$ when the proportion of contamination ϵ increases. The success rate also decreases when the variance of the contamination distribution τ^2 increases.

Table 3.2: Performances of *RAIC* and *AIC* in presence
of innovation outliers. The values are the success rates
(in percentages) of choosing the true model.

Proportion of Contamination	True Model: $AR(1)$					
	$\tau^2 = 9$		$\tau^2 = 25$		$\tau^2 = 100$	
ϵ	RAIC	AIC	RAIC	AIC	RAIC	AIC
0.00	91.0	80.6	90.8	77.2	90.2	76.2
0.03	88.6	70.4	86.8	64.2	74.0	51.6
0.05	85.8	69.8	79.6	55.8	73.4	40.2
0.10	83.6	62.4	70.4	42.6	61.4	21.6

We repeat the above simulation for samples drawn from the $AR(2)$ process with innovation outliers. The parameters of the process were chosen as $\phi_1 = .6$ and $\phi_2 = .3$. The simulation results are presented in Table 3.3. We observe similar behavior to Table 3.2 for the *RAIC* in choosing the true model. However, the performance of the classical *AIC* seems to be better for the $AR(2)$ process than for the $AR(1)$ process. This is probably due to the fact that the classical *AIC* has a tendency to choose a model always in higher orders. On the other hand, *RAIC* chooses the correct model more frequently for all cases.

Table 3.3: Performances of *RAIC* and *AIC* in presence
of innovation outliers. The values are the success rates
(in percentages) of choosing the true model.

Proportion of Contamination	True Model: $AR(2)$					
	$\tau^2 = 9$		$\tau^2 = 25$		$\tau^2 = 100$	
ϵ	RAIC	AIC	RAIC	AIC	RAIC	AIC
0.00	91.8	84.8	89.6	86.0	89.8	84.6
0.03	88.0	78.6	85.6	73.2	82.6	66.4
0.05	87.2	78.8	82.8	68.2	77.0	52.8
0.10	84.6	72.6	76.8	55.4	67.2	38.2

We also study the performance of *RAIC* in the presence of additive outliers. For additive outliers, a time series process y_t is not itself an $AR(q)$ process. Instead, we have $y_t = x_t + z_t$, where x_t is an $AR(q)$ process satisfying $x_t = \phi^T y_{t-1}^{t-q} + u_t$ with $u_t \sim N(0, \sigma^2)$ and z_t is an independent process. The process z_t has distribution G^* given by $G^* = (1 - \epsilon)\delta_0 + \epsilon G$, where δ_0 is the distribution that assigns probability 1 to zero and G is an arbitrary distribution. Thus with probability $(1 - \epsilon)$, the $AR(q)$ process x_t itself is observed, and with probability ϵ, y_t is the $AR(q)$ process x_t plus an error with distribution G. Here we consider G as $N(0, \tau^2)$, with $\tau^2 = 9, 25, 100$. u_t is considered to be distributed as $N(0,1)$, as before. Table 3.4 provides the simulation results for samples drawn from the $AR(1)$ process with additive outliers. We repeat this simulation for the $AR(2)$ process with the same additive outliers. The results have been presented in Table 3.5. From Tables 3.4 and 3.5, it is clear that the *RAIC* outperforms *AIC*, often quite dramatically. If we compare the simulation results for the innovation outliers with those for the additive outliers, we observe that the classical *AIC* is more affected by the additive outliers than by the innovation outliers, while *RAIC* is relatively stable in both cases.

Table 3.4: Performances of *RAIC* and *AIC* in presence of additive outliers. The values are the success rates (in percentages) of choosing the true model.

Proportion of	True Model: $AR(1)$					
Contamination	$\tau^2 = 9$		$\tau^2 = 25$		$\tau^2 = 100$	
ϵ	RAIC	AIC	RAIC	AIC	RAIC	AIC
0.00	92.4	77.8	91.6	80.4	92.2	77.8
0.03	80.0	55.2	74.2	40.4	73.2	25.0
0.05	71.8	49.0	65.6	30.2	62.8	18.2
0.10	65.0	30.0	50.0	19.2	50.6	12.8

Table 3.5: Performances of *RAIC* and *AIC* in presence of additive outliers. The values are the success rates (in percentages) of choosing the true model.

Proportion of Contamination ϵ	True Model: $AR(2)$					
	$\tau^2 = 9$		$\tau^2 = 25$		$\tau^2 = 100$	
	RAIC	AIC	RAIC	AIC	RAIC	AIC
0.00	91.2	85.6	91.4	84.2	89.2	84.2
0.03	89.4	63.0	85.0	32.6	82.0	7.2
0.05	82.6	45.6	84.2	20.0	75.4	3.6
0.10	77.2	16.4	74.4	6.2	66.6	6.2

3.5 Discussion

From the simulation study, we observe that the overall performance of the robust Akaike Information Criterion $RAIC$ is better than that of the classical AIC in choosing the appropriate order of an autoregressive process. In the case of a contaminated time series process, the $RAIC$ gives better results when the proportion of contamination ϵ and the variance of the contamination process τ^2 are relatively small. Furthermore, the performance of the $RAIC$ for a process with innovation outliers is found to be better relative to the performance of the $RAIC$ for a process with additive outliers. This is perhaps reasonable, because the affect of additive outliers is relatively large as compared to the affect of innovation outliers on the GM estimates of time series parameters. A careful study of the breakdown properties of the GM estimates in a time series setting is needed in order to improve the performance of the $RAIC$ further. Our simulation study indicates that the proportion of contamination ϵ should be reasonably small for a better performance of the $RAIC$.

In this chapter, we focused on the model selection criterion for autoregressive processes. It seems important to develop a robust criterion for general $ARMA$ processes. Note that the parameters of an $ARMA$ process are very hard to estimate robustly (see Ronchetti, 1997). When moving-average terms are present in a time series process, the GM estimates result in unbounded influence function. As a consequence, the GM estimates and the model selection criterion based on these estimates are not robust. So in order to develop a robust criterion for $ARMA$ processes, it is important to find a robust technique for estimating the corresponding $ARMA$ components. Much work remains to be done in this area.

Chapter 4

Robust Estimation of Nonlinear Regression With Autoregressive Errors

4.1 Introduction

Nonlinear regression models play an important role in many fields. The classical estimates of nonlinear regression are often very sensitive to outliers in the data. As a remedy, robust versions of these classical estimates have been studied by a number of authors. Some of the robust techniques of nonlinear regression are discussed in Fraiman (1983), Stromberg and Ruppert (1992), Stromberg (1993). Fraiman (1983) proposed an optimal M estimate with bounded gross error sensitivity in the case of a nonlinear regression setting with independent errors. Stromberg and Ruppert (1992) discuss the high breakdown point robust estimation in nonlinear regression. They extended the idea of the finite-sample breakdown analysis from linear to nonlinear regression. In a linear regression setting, the finite-sample breakdown point of an estimator is defined as the smallest proportion of data that must be changed to cause an infinite perturbation of an estimator. In nonlinear regression, Stromberg and Ruppert (1992) introduce the concepts of upper and lower breakdown functions to give a new definition of the finite-sample breakdown point. Instead of defining breakdown in terms of the estimated parameter, the breakdown point in nonlinear regression is

defined in terms of the estimated regression function. Stromberg and Ruppert (1992) pointed out that the least median squares (LMS), proposed by Rousseeuw (1984), and the least trimmed squares (LTS) estimators in a nonlinear regression setting have high breakdown point at most 1/2. The LMS and the LTS estimators are defined by the minimization of the median or the trimmed mean of the squares of the residuals, respectively. Although the LMS and the LTS estimators have high breakdown points, they are not efficient (see Rousseeuw and Leroy, 1987). To find an efficient as well as a high breakdown point estimator in a nonlinear regression problem, Stromberg (1993) developed an algorithm which uses Yohai's MM estimate (see Yohai, 1987) starting from a high breakdown point estimate such as the LMS estimate. Yohai (1987) developed the MM estimate using a redescending ψ function such as the bisquare ψ function. However, the MM estimate has unbounded influence function as its influence function depends on the vector of covariates x, which is unbounded. Yohai and Zamar (1988) introduced a new class of robust estimates called τ estimates. The τ estimates are defined by minimizing a scale estimate τ applied to the residuals. Here also redescending ψ functions are used to find the τ estimates. Like MM estimates, the τ estimates have unbounded influence functions in the x's. Recently, Ferri, Kelmansky, Yohai and Zamar (1999) proposed a class of generalized τ estimates which have bounded influence function in the x's. These estimates are defined by minimizing the τ scale of the weighted residuals, with weights that penalize high-leverage observations. Like MM and τ estimates, redescending ψ functions are used to obtain the generalized τ estimates. Redescending ψ functions vanish outside some central region which are able to reject extreme outliers entirely. However, the use of these ψ functions may not be efficient in certain situations. Huber (1981) commented that the redescending ψ functions are certainly beneficial if there are extreme outliers, but the improvement is relatively minor with respect to the asymptotic variance and is counterbalanced by an increase of the minimax risk. It may be shown that with a small proportion of moderate outliers, the generalized τ estimates based on the redescending ψ functions are not efficient as compared to the generalized M (GM) estimates based on the Huber's monotone ψ function. In this thesis, Huber's ψ function has been used in the GM estimation of nonlinear regression with autoregressive errors. A method has been developed for the joint estimation of

the regression parameters and the autoregressive parameters of a nonlinear regression model with autoregressive errors.

Situations in which data are collected sequentially over time may result in substantial serial correlations in the errors. It often occurs with economic data where the response y measures the state of a market at a particular time. Another typical example where serially correlated errors often arise is in the modeling of growth curves to data on a single animal over a certain period. In many cases, we observe a nonlinear relationship between the response variable y and the vector of covariates x. We develop a generalized M (GM) estimation method for the joint estimation of the regression parameters and the autoregressive parameters of a nonlinear regression model with autoregressive errors. In section 4.2, we introduce the proposed GM estimation method. The resulting estimates are similar to Mallows-type GM estimates. For an $AR(q)$ process of the errors, we minimize an objective function which is based on the robustified version of the conditional negative log-likelihood function of $y_{q+1}, \ldots, y_n$ for the given values $y_1, \ldots, y_q$.

We then study the infinitesimal behavior of the robust estimates based on the time series analogue of Hampel's influence function in section 4.3. It is shown that the resulting GM estimates have bounded influence functions under some regularity conditions. The influence functions are derived from the time series extension of Hampel's influence function as described in Martin and Yohai (1986). In section 4.4, we study the asymptotic properties of the robust estimates of nonlinear regression in some detail. We show that under some suitable assumptions, the proposed GM estimates of the regression parameters and the autoregressive parameters are asymptotically normally distributed with a certain mean vector and a covariance matrix. To choose an appropriate order for the autoregressive error process in a nonlinear regression setting, we extend the robust Akaike Information Criterion introduced in chapter 3 to the case of a nonlinear regression with autoregressive errors. This extended robust criterion is presented in section 4.5.

As an application of the proposed robust method, in section 4.6 we analyze a

ground level ozone data set which appears to have nonlinear relationship with some meteorological variables as covariates. As the data are collected over equally spaced time, there appears to be a significant serial correlation in the errors. Moreover, the error process is found to be heteroscedastic with respect to the covariates. We develop a robust technique to model the variance of the heteroscedastic errors as a function of the covariates. Section 4.7 gives the conclusions of the chapter.

4.2 GM Estimation of Nonlinear Regression With Autoregressive Errors

Let us consider a nonlinear regression model of the form

$$y_t = h(\mathbf{x}_t, \boldsymbol{\beta}) + \epsilon_t \tag{4.1}$$

where $y_t \in \mathbb{R}$, $\mathbf{x}_t \in \mathbb{R}^p$, $\boldsymbol{\beta} \in \mathbb{R}^{p_1}$ and h is a deterministic model function that is assumed to be continuous in $\boldsymbol{\beta}$ for each $\mathbf{x}_t$. It is also assumed that the error process ϵ_t follows an autoregressive process of order q defined by

$$\epsilon_t = \phi_1 \epsilon_{t-1} + \ldots + \phi_q \epsilon_{t-q} + u_t \tag{4.2}$$

where u_t are assumed to be independent $N(0, \sigma_t^2)$ and $\phi_1, \ldots, \phi_q$ are the autoregressive parameters. Here $\text{var}(u_t) = \sigma_t^2$ is considered to accommodate the heteroscedasticity among the residuals. Using equation (4.2), (4.1) can be reexpressed as

$$y_t = h_t + \phi_1(y_{t-1} - h_{t-1}) + \ldots + \phi_q(y_{t-q} - h_{t-q}) + u_t \tag{4.3}$$

where we use h_t in place of $h(\mathbf{x}_t; \boldsymbol{\beta})$ for notational convenience. We can rewrite model (4.3) in the form

$$y_t = \mu_t + u_t \quad (t = q + 1, \ldots, n) \tag{4.4}$$

where

$$\mu_t = h_t + (y_{t-1}^{t-q} - h_{t-1}^{t-q})^T \phi \tag{4.5}$$

and $y_{t-1}^{t-q} = (y_{t-1}, \dots, y_{t-q})^T$, $h_{t-1}^{t-q} = (h_{t-1}, \dots, h_{t-q})^T$, $\phi = (\phi_1, \dots, \phi_q)^T$. Model (4.3) can be fitted by the method of least squares. Alternatively, under normality assumptions, we can maximize the conditional likelihood function of $y_{q+1}, \dots, y_n$ given $y_1, \dots, y_q$. The resulting estimates are referred to as conditional ML or conditional LS estimates. For now, we assume that σ_t is known. Then the conditional negative log-likelihood function of $y_{q+1}, \dots, y_n$ for the given $y_1, \dots, y_q$ can be expressed, up to a constant, in the form

$$-\log L(\theta; s) = \frac{1}{n} \sum_{t=q+1}^{n} \frac{1}{2} \left(\frac{y_t - \mu_t}{\sigma_t} \right)^2 \tag{4.6}$$

where $\theta^T = (\beta^T, \phi^T)$ is the vector of parameters to be estimated. Here s may be considered as a random variable containing all the elements y_t and x_t for $t = 1, \dots, n$. The maximum likelihood (ML) estimate of θ can be obtained by minimizing this negative log-likelihood function with respect to θ. This ML estimate of θ is not robust in the sense that arbitrarily small departures of the underlying distribution from normality may result in arbitrarily large asymptotic variances and biases of the estimator (Maronna, Bustos and Yohai, 1979). For a robust estimate of θ, Huber (1981) suggested minimizing a robust version of the negative log-likelihood function in the form

$$H_n^*(s, \theta) = \frac{1}{n} \sum_{t=q+1}^{n} \rho \left(\frac{y_t - \mu_t}{\sigma_t} \right) \tag{4.7}$$

for an arbitrary continuous even function ρ. The commonly used forms of ρ function include $\rho(x) = \frac{1}{2}x^2$, which corresponds to the normal density and

$$\rho(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq k \\ k|x| - \frac{1}{2}k^2 & \text{otherwise} \end{cases} \tag{4.8}$$

which corresponds to the Huber's least favorable distribution for location (see Huber, 1981). The most common choice of the constant k is 1.345. For the choice $k = 1.345$, the corresponding location estimator has 95% efficiency at the normal model. Let ψ be the derivative of ρ. Then minimization of the objective function $H_n^*(\mathbf{s}, \boldsymbol{\theta})$ in equation (4.7) is equivalent to the solution of the M estimating equation

$$\sum_{t=q+1}^{n} \psi\left(\frac{y_t - \mu_t}{\sigma_t}\right) \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} = \mathbf{0}. \tag{4.9}$$

The estimates obtained by minimizing the objective function $H_n^*(\mathbf{s}, \boldsymbol{\theta})$ or, equivalently, by solving the estimating equation (4.9) iteratively are termed as M estimates. Note that the M estimators bound the influence of the residuals but not the influence of the position in the design space. It can be shown that the influence function of an M estimator is proportional to $\psi((y_t - \mu_t)/\sigma_t)(\partial \mu_t/\partial \boldsymbol{\theta})$, which may be unbounded with respect to the covariates. As a consequence, the M estimates may become nonrobust with respect to outliers in the x-direction. To bound the joint influence of the residuals and the position in the design space, we consider using GM estimates. The GM estimates are obtained by minimizing an objective function of the form

$$H_n^{**}(\mathbf{s}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{t=q+1}^{n} \left(\frac{w_t}{v_t}\right) \rho\left(\frac{y_t - \mu_t}{\sigma_t} v_t\right) \tag{4.10}$$

with respect to $\boldsymbol{\theta}$, where $w_t = w\left(\mathbf{x}_t, \mathbf{x}_{t-1}^{t-q}, y_{t-1}^{t-q}\right)$ and $v_t = v\left(\mathbf{x}_t, \mathbf{x}_{t-1}^{t-q}, y_{t-1}^{t-q}\right)$ are termed as weight functions. A number of choices of w_t and v_t are available (see Hampel et al, 1986, p.347). The commonly used weights are the classical ($w_t = v_t = 1$) and the Mallows ($v_t = 1$) weights. Note that the choice $w_t = v_t = 1$ gives the M estimate of $\boldsymbol{\theta}$. Here we consider the commonly used Mallows-type GM estimation which uses the weight function w_t and $v_t = 1$. Thus for the Mallows-type GM estimates, the objective function $H_n^{**}(\mathbf{s}, \boldsymbol{\theta})$ in equation (4.10) reduces to

$$H_n(\mathbf{s}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{t=q+1}^{n} w_t \, \rho\left(\frac{y_t - \mu_t}{\sigma_t}\right). \tag{4.11}$$

We minimize this objective function for the GM estimate of $\boldsymbol{\theta}$ based on the Huber's ρ function defined in equation (4.8). The weight function $w_t = w(\mathbf{z}_t)$ downweights the leverage points in the design space $\mathbf{z}_t = (\mathbf{x}_t, \mathbf{x}_{t-1}^{t-q}, y_{t-1}^{t-q})$. We use the weight function

$$w(\mathbf{x}) = w(\mathbf{x}, \mathbf{m}_x, \mathbf{S}_x) = \min \left[1, \left\{ \frac{b}{(\mathbf{x} - \mathbf{m}_x)^T \mathbf{S}_x^{-1} (\mathbf{x} - \mathbf{m}_x)} \right\}^{\frac{\gamma}{2}} \right] \qquad (4.12)$$

as suggested by Simpson, Ruppert and Carroll (1992). Here $\mathbf{m}_x$ and $\mathbf{S}_x$ are considered as the minimum volume ellipsoid (MVE) estimators of location and scale of $\mathbf{x}$, respectively. The MVE estimators are defined as the location and scale of the smallest ellipsoid containing at least $[(n + p + 1)/2]$ points, where p is the dimension of $\mathbf{x}$. (see Rousseeuw and van Zomeren, 1990). These MVE estimators of location and scale are highly robust with a breakdown point of almost $1/2$. The constant b is chosen as the $(1 - \alpha_0)$ quantile of the chi-squared distribution with degrees of freedom equal to the dimension of $\mathbf{x}$. The common choice of α_0 is $.05$. γ is chosen as a positive integer greater than or equal to 1. Note that larger value of γ results in smaller value of the weight function corresponding to a leverage point. In the application section of this chapter we analyze a data set, where we use $\gamma = 2$ for the weight functions.

For the GM estimate of $\boldsymbol{\theta}$, we can minimize the objective function $H_n(\mathbf{s}, \boldsymbol{\theta})$ directly or, equivalently, we can solve the GM estimating equations obtained by taking the derivative of $H_n(\mathbf{s}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and equating to zero. Differentiating the objective function $H_n(\mathbf{s}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}^T = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T)$ (which is a vector of the regression parameter $\boldsymbol{\beta}$ and the autoregressive parameter $\boldsymbol{\phi}$) gives the GM estimating equations

$$\sum_{t=q+1}^{n} \frac{1}{\sigma_t} \psi \left(\frac{y_t - \mu_t}{\sigma_t} \right) w_t \frac{\partial \mu_t}{\partial \boldsymbol{\beta}} = 0 \qquad (\text{for } \boldsymbol{\beta}) \qquad (4.13)$$

and

$$\sum_{t=q+1}^{n} \frac{1}{\sigma_t} \psi \left(\frac{y_t - \mu_t}{\sigma_t} \right) w_t \frac{\partial \mu_t}{\partial \boldsymbol{\phi}} = 0 \qquad (\text{for } \boldsymbol{\phi}) \qquad (4.14)$$

where $\frac{\partial \mu_t}{\partial \beta} = \frac{\partial h_t}{\partial \beta} - \phi_1 \frac{\partial h_{t-1}}{\partial \beta} - \ldots - \phi_q \frac{\partial h_{t-q}}{\partial \beta}$ and $\frac{\partial \mu_t}{\partial \phi} = (y_{t-1}^{t-q} - h_{t-1}^{t-q})$. These equations can be solved iteratively for the GM estimate of $\boldsymbol{\theta}^T = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T)$. The above two GM estimating equations can be expressed in the form

$$\sum_{t=q+1}^{n} \Psi(\mathbf{y}_t^1, \mathbf{X}_t^1; \boldsymbol{\theta}) = 0 \qquad (4.15)$$

where $\mathbf{y}_t^1 = (y_t, y_{t-1}, \ldots, y_1)$, $\mathbf{X}_t^1 = (\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1)$, $\Psi(\mathbf{y}_t^1, \mathbf{X}_t^1; \boldsymbol{\theta}) = (1/\sigma_t) \times w(\mathbf{z}_t)\psi((y_t - \mu_t(\mathbf{z}_t; \boldsymbol{\theta}))/\sigma_t)(\partial \mu_t/\partial \boldsymbol{\theta})$ and $\mathbf{z}_t = (\mathbf{x}_t, \mathbf{x}_{t-1}^{t-q}, y_{t-1}^{t-q})$. Note that so far we have assumed that σ_t (the standard error of u_t) is known. In practice, we need to estimate σ_t by using a suitable technique which is usually based on the pattern of the estimated residuals. For the choice $\sigma_t = \sigma$, we can use the median absolute deviation (multiplied by a tuning constant) $\text{med}_t |y_t - \hat{\mu}_t|/.6745$ ($t = q+1, \ldots, n$) as a robust estimate of σ. However, for a general case, estimation of σ_t becomes harder as modeling of σ_t depends on the characteristics of the data. In the application section of this chapter, we develop a model for estimating σ_t which mainly depends on the pattern of the residuals with respect to the covariates. To solve the GM estimating equations (4.13) and (4.14), we use the Newton-Raphson method of iteration. Using the first order Taylor's series approximation, the left hand side of equation (4.13) can be approximated as

$$\sum_{t=q+1}^{n} \frac{1}{\sigma_t} \psi \left(\frac{y_t - \mu_t}{\sigma_t} \right) w_t \frac{\partial \mu_t}{\partial \beta}$$

$$= \sum_{t=q+1}^{n} \frac{1}{\sigma_{t0}} \psi \left(\frac{y_t - \mu_{t0}}{\sigma_{t0}} \right) w_t \frac{\partial \mu_t}{\partial \beta} \bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0}$$

$$- (\beta - \beta_0) \times \left\{ \frac{1}{\sigma_{t0}^2} \sum_{t=q+1}^{n} \psi' \left(\frac{y_t - \mu_{t0}}{\sigma_{t0}} \right) w_t \frac{\partial \mu_t}{\partial \beta} \frac{\partial \mu_t}{\partial \beta}^T \bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} \right.$$

$$\left. - \sum_{t=q+1}^{n} \frac{1}{\sigma_{t0}} \psi \left(\frac{y_t - \mu_{t0}}{\sigma_{t0}} \right) w_t \frac{\partial^2 \mu_t}{\partial \beta \partial \beta^T} \bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} \right\}$$

where $\boldsymbol{\theta}_0^T = (\beta_0^T, \phi_0^T)$ and σ_{t0} are some initial values of $\boldsymbol{\theta}^T = (\beta^T, \phi^T)$ and σ_t, respectively. Replacing the last term in the above expression by its expectation

(which is zero), we can write

$$\beta = \beta_0 + M_{10}^{-1} q_{10} \tag{4.16}$$

where

$$M_{10} = \sum_{t=q+1}^{n} \frac{1}{\sigma_{t0}^2} \psi' \left(\frac{y_t - \mu_{t0}}{\sigma_{t0}} \right) w_t \frac{\partial \mu_t}{\partial \beta} \frac{\partial \mu_t}{\partial \beta}^T \Bigg|_{\theta=\theta_0}$$

and

$$q_{10} = \sum_{t=q+1}^{n} \frac{1}{\sigma_{t0}} \psi \left(\frac{y_t - \mu_{t0}}{\sigma_{t0}} \right) w_t \frac{\partial \mu_t}{\partial \beta} \Bigg|_{\theta=\theta_0} .$$

Applying similar technique, the GM estimating equation (4.14) gives

$$\phi = \phi_0 + M_{20}^{-1} q_{20} \tag{4.17}$$

where

$$M_{20} = \sum_{t=q+1}^{n} \frac{1}{\sigma_{t0}^2} \psi' \left(\frac{y_t - \mu_{t0}}{\sigma_{t0}} \right) w_t \frac{\partial \mu_t}{\partial \phi} \frac{\partial \mu_t}{\partial \phi}^T \Bigg|_{\theta=\theta_0}$$

and

$$q_{20} = \sum_{t=q+1}^{n} \frac{1}{\sigma_{t0}} \psi \left(\frac{y_t - \mu_{t0}}{\sigma_{t0}} \right) w_t \frac{\partial \mu_t}{\partial \phi} \Bigg|_{\theta=\theta_0} .$$

The equations (4.16) and (4.17) can be solved iteratively for the GM estimates of the regression parameter β and the autoregressive parameter ϕ, respectively. The usual choice of the initial estimates of β and ϕ are the classical least squares (LS) estimates. However, as the LS estimates are not robust, some authors suggest using robust estimates such as the least median squares (LMS) estimates as the initial values of β and ϕ. The estimation of σ_t is discussed in the application section of this chapter where we model σ_t as a function of the covariates. We update estimate of σ_t

for each updated estimate of $\boldsymbol{\theta}^T = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T)$.

In the next section, we discuss the infinitesimal behavior of the GM estimates in terms of their corresponding influence functions.

4.3   Influence Function of the GM Estimators

The influence function introduced by Hampel (1974) is considered as a technique of studying the local stability of an estimator under a point-mass contamination of the data or the underlying distribution. Hampel (1974) defines the influence function of an estimator based on a set of independent observations $y_1, \ldots, y_n$ with common distribution function F. It is assumed that the estimator $T_n = T_n(y_1, \ldots, y_n)$ may be obtained from a functional $T = T(F)$ defined on a space of distributions of y by evaluating T at the empirical distribution function F_n : $T_n = T(F_n)$. Let $F_\epsilon = (1 - \epsilon)F + \epsilon \Delta_y$ be a point-mass contamination of the true distribution F, where Δ_y is a delta function which has point-mass 1 at y and 0 elsewhere. Then the influence function is the directional derivative of the functional T at F in the direction determined by Δ_y :

$$IF(y; T, F) = \lim_{\epsilon \to 0} \frac{T(F_\epsilon) - T(F)}{\epsilon}. \tag{4.18}$$

Note that this approach is certainly useful in the case of an i.i.d. setting as estimators with i.i.d. data are invariant under permutations of the data, and may be obtained from functionals $T(F)$ of the marginal distribution F by evaluating T at the empirical distribution function F_n. When the estimators are based on a sequence of time series data, the time configuration of the contaminating points will be important in defining the influence function of a time series parameter estimate. Martin and Yohai (1986) discuss Hampel's influence function in a time series parameter estimation setting. Assume that the observations y_i are realizations of a stationary and ergodic process on $\mathbb{R}^{-\infty, \infty}$, with probability space $(\mathbb{R}^{-\infty, \infty}, \mathcal{B}, P)$, $\mathcal{B}$ being the family of Borel sets in $\mathbb{R}^{-\infty, \infty}$, with P in the set $\mathcal{P}_{SE}$ of all stationary and ergodic measures on $(\mathbb{R}^{-\infty, \infty}, \mathcal{B})$. Here $\mathbb{R}^{-\infty, \infty}$ represents the space of all doubly infinite sequences such as

$(\ldots, y_{-1}, y_0, y_1, y_2, \ldots)$. The influence function of a time series parameter estimate can be defined for general functionals $T(P)$. Suppose, the univariate contamination distribution F_{ϵ} is replaced by the process contamination measure $P_{\epsilon} = (1-\epsilon)P + \epsilon \Delta_y$, where in general $\mathbf{y} = (\ldots, y_{-1}, y_0, y_1, \ldots) \in \mathbb{R}^{-\infty,\infty}$, Δ_y has unit mass at $\mathbf{y}$, and P is a measure in $\mathcal{P}_{SE}$. Assume that the time-series parameter estimate is defined for not only stationary and ergodic measure P, but also for the contamination measure $P_{\epsilon} = (1-\epsilon)P + \epsilon \Delta_y$, $0 \leq \epsilon \leq 1$. Then Martin and Yohai (1986) define the time-series analogue of Hampel's influence function in the form

$$IF(\mathbf{y}; T, P) = \lim_{\epsilon \to 0} \frac{T(P_{\epsilon}) - T(P)}{\epsilon} \tag{4.19}$$

provided the limit exists.

We shall use this definition to find the influence functions of the GM estimates of nonlinear regression with autoregressive errors.

Let the observations $\{y_i, \mathbf{x}_i\}$ be co-ordinates of a vector-valued stationary and ergodic process $\{\mathbf{y}, \mathbf{X}\}$ with the measure P in $\mathcal{P}_{SE}$, where $\mathbf{y} = (\ldots, y_{-1}, y_0, y_1, \ldots)$ and $\mathbf{X} = (\ldots, \mathbf{x}_{-1}, \mathbf{x}_0, \mathbf{x}_1, \ldots)$. Note that instead of the general functional $T(P)$, all ensuing results are for a special class of functionals T associated with time-series parameter estimates T_n which can be computed as a solution to the GM estimating equation

$$\sum_{i=1}^{n} \Psi_i(\mathbf{y}_i^1, \mathbf{X}_i^1; T_n) = 0 \tag{4.20}$$

where $\mathbf{y}_i^1 = (y_i, y_{i-1}, \ldots, y_1)$ and $\mathbf{X}_i^1 = (\mathbf{x}_i, \mathbf{x}_{i-1}, \ldots, \mathbf{x}_1)$. The functionals Ψ_i and T_n are both vector-valued. The subscript i on Ψ_i accounts for "end-effects" which vanish after a finite number of observations for autoregressive models. The asymptotic value $T = T(P)$ of T_n can be determined by using a fixed psi function which satisfies

$$\Psi_i(\mathbf{y}_i^1, \mathbf{X}_i^1; \mathbf{t}) = \Psi(y_i, \mathbf{X}_i, ; \mathbf{t}) \tag{4.21}$$

for each $\mathbf{t}$ and $i \geq l$ with $\mathbf{y}_i = (y_i, y_{i-1}, \ldots) \in \mathbb{R}^{\infty}$ and $\mathbf{X}_i = (\mathbf{x}_i, \mathbf{x}_{i-1}, \ldots) \in \mathbb{R}^{\infty}$.

For an $AR(q)$ process, we can choose $l = q + 1$. Under suitable regularity conditions including ergodicity, we may have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Psi_i(\mathbf{y}_i^1, \mathbf{X}_i^1; \mathbf{t}) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Psi(\mathbf{y}_i, \mathbf{X}_i; \mathbf{t}) = E\Psi(\mathbf{y}_1, \mathbf{X}_1; \mathbf{t}). \qquad (4.22)$$

So we assume that the asymptotic value $T = T(P)$ of the time-series estimate T_n is defined by

$$\int \Psi(\mathbf{y}_1, \mathbf{X}_1; T) dP(\mathbf{y}_1, \mathbf{X}_1) = 0. \qquad (4.23)$$

Here T is defined on $\mathcal{P}_0$ consisting of all P in $\mathcal{P}_{SE}$ for which the integral (4.23) exists and is finite. We assume that the equation (4.23) has a unique root $\boldsymbol{\theta}_0 = T(P)$.

In the specific case of $H_n(\mathbf{s}, \boldsymbol{\theta})$ defined by (4.11), for a unique root $\boldsymbol{\theta}_0$, the corresponding objective function $E[H_n(\mathbf{s}, \boldsymbol{\theta})]$ must be convex. In fact, $E[H_n(\mathbf{s}, \boldsymbol{\theta})]$ based on Huber's ρ function is convex and ensures a unique root $\boldsymbol{\theta}_0$. Note that for the GM estimating equations (4.13) and (4.14), the Ψ_i function in equation (4.20) can be defined as (assuming var(u_i)=1, for simplicity)

$$\Psi_i(\mathbf{y}_i^1, \mathbf{X}_i^1; \boldsymbol{\theta}) = w(\mathbf{z}_i)\psi(y_i - \mu_i(\mathbf{z}_i; \boldsymbol{\theta}))\frac{\partial\mu_i}{\partial\boldsymbol{\theta}} \qquad (4.24)$$

where $\mathbf{z}_i = (\mathbf{x}_i, \mathbf{x}_{i-1}^{i-q}, y_{i-1}^{i-q})$ for $i \geq q + 1$, and $w(\mathbf{z}_i) = w(\mathbf{z}_i, \mathbf{m}_z, \mathbf{S}_z)$ is defined as in (4.12). Correspondingly, the limit Ψ function can be defined as

$$\Psi(\mathbf{y}_1, \mathbf{X}_1; \boldsymbol{\theta}) = w(\mathbf{z}_1)\psi(y_1 - \mu_1(\mathbf{z}_1; \boldsymbol{\theta}))\frac{\partial\mu_1}{\partial\boldsymbol{\theta}} \qquad (4.25)$$

where $w(\mathbf{z}_1)$ may be considered as a functional $w(\mathbf{z}_1, \mathbf{m}(P), \mathbf{S}(P))$ with respect to the probability measure P. Thus under suitable regularity conditions, equation (4.19) gives the influence function of the functional $T(P)$ in the form

$$IF(\mathbf{y}_1, \mathbf{X}_1; T, P) = \lim_{\epsilon\to 0} \frac{T((1 - \epsilon)P + \epsilon\Delta_{\mathbf{y}_1, \mathbf{x}_1}) - T(P)}{\epsilon} \qquad (4.26)$$

where $\Delta_{\mathbf{y}_1,\mathbf{x}_1}$ has point-mass 1 at $\{\mathbf{y}_1, \mathbf{X}_1\}$ and 0 elsewhere. Therefore, using equation (4.23), the influence function of the functional $T(P)$ is obtained as

$$IF(\mathbf{y}_1, \mathbf{X}_1; T, P) = \left[- E \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \Psi(\mathbf{y}_1, \mathbf{X}_1; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=T} \right\} \right]^{-1} \Psi(\mathbf{y}_1, \mathbf{X}_1; T). \qquad (4.27)$$

Note that this influence function is bounded in both $\mathbf{y}_1$ and $\mathbf{X}_1$ when $w(\mathbf{z}_1) \times \frac{\partial}{\partial \boldsymbol{\theta}} \mu_1(\mathbf{z}_1; \boldsymbol{\theta})$ is bounded. For some nonlinear regression functions, the weight function w defined by (4.12) may result in unbounded influence functions of the GM estimates in terms of unbounded values of the product $w(\mathbf{z}_1) \times \frac{\partial}{\partial \boldsymbol{\theta}} \mu_1(\mathbf{z}_1; \boldsymbol{\theta})$. We address this issue in the application section of this chapter and suggest a modified form of the weight function w in the case of an unbounded influence function.

In the next section, we discuss the asymptotic normality of the GM estimates in the case of a nonlinear regression setting with dependent errors.

4.4 Asymptotic Normality of the GM Estimators

To develop the asymptotic normality of the GM estimates, we use the approach of mean value theorem following Domowitz and White (1982). Recall that the GM estimate of $\boldsymbol{\theta}$ can be obtained by minimizing an objective function of the form

$$H_n(\mathbf{s}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{t=q+1}^{n} \nu(y_t, \mathbf{z}_t, \boldsymbol{\theta}) \qquad (4.28)$$

where $\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta}) = w(\mathbf{z}_t) \rho(y_t - \mu_t(\mathbf{z}_t, \boldsymbol{\theta}))$, with a weight function w and a suitable ρ function. Here $\mathbf{s}$ can be considered as a random variable containing all the elements $\mathbf{z}_t$ and y_t. We assume that $\mathrm{var}(y_t) = \sigma_t^2 = 1$, for simplicity. Under stationarity conditions, we have

$$\begin{aligned} \bar{H}_n(\boldsymbol{\theta}) &= E[H_n(\mathbf{s}, \boldsymbol{\theta})] \\ &= \frac{1}{n} \sum_{t=q+1}^{n} E[\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})] \\ &\rightarrow E[\nu(y_1, \mathbf{z}_1, \boldsymbol{\theta})] \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Let us consider the functional

$$\bar{H}(\boldsymbol{\theta}) = \lim_{n \to \infty} \bar{H}_n(\boldsymbol{\theta}) = E[\nu(y_1, \mathbf{z}_1, \boldsymbol{\theta})]. \tag{4.29}$$

We use the following general assumptions to establish the asymptotic results for the GM estimates.

(a1) *The function $\rho(x)$ is continuous and even in x.*

(a2) *The function $\psi(x)$, derivative of $\rho(x)$, is continuous, odd and bounded in x.*

(a3) *The weight function $w(x)$ is continuous in x.*

(a4) *The function $\bar{H}(\boldsymbol{\theta})$ has an identifiably unique minimizer $\boldsymbol{\theta}_0$, where $\boldsymbol{\theta}$ is an element of Θ, a compact subset of an Euclidean space.*

(a5) $E\left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \Psi(\mathbf{y}_1, \mathbf{X}_1; \boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right\}$ *is non-singular, where $\Psi(\mathbf{y}_1, \mathbf{X}_1; \boldsymbol{\theta})$ is defined in* (4.25).

(a6) *The sequence $\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})$ is continuous on Θ, and is a measurable function of $(y_t, \mathbf{z}_t)$.*

(a7) *The sequence $\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})$ is continuously differentiable of order 2 in $\boldsymbol{\theta}$.*

Assumptions (a1)-(a3) are "very mild" assumptions in our situation. Huber's ρ and ψ functions are both well-behaved continuous functions. Also the weight function $w(x)$, which is considered as a function of the Mahalanobis distance as defined in equation (4.12), is continuous in x. In assumption (a4), for the minimizer $\boldsymbol{\theta}_0$ to be identifiably unique, the function $\bar{H}(\boldsymbol{\theta})$ must be locally uniformly strictly convex. Note that $\bar{H}(\boldsymbol{\theta})$ based on Huber's ρ function is convex and gives a minimizer which is identifiably unique (see Huber, 1981 for details). Assumption (a5) is required for the existence of the variance as well as the influence function of a GM estimator.

Assumption (a7) requires that the sequence $\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})$ is continuously differentiable of order 2 on Θ. Note that Huber's ρ function defined in equation (4.8) does

not have a continuous second derivative, which clearly violates this assumption. As a remedy, we can consider a modified form of the Huber's ρ function, which provides a continuous second derivative. We define the modified Huber's ρ function in the form

$$
\rho_{mh}(x) = \begin{cases}
-kx - \frac{13}{24}k^2 & \text{for } x \leq -\frac{3}{2}k \\
\frac{1}{6k}\left(x + \frac{3k}{2}\right)^3 - kx - \frac{13}{24}k^2 & \text{for } -\frac{3}{2}k < x < -\frac{1}{2}k \\
\frac{1}{2}x^2 & \text{for } -\frac{1}{2}k \leq x \leq \frac{1}{2}k \\
-\frac{1}{6k}\left(x - \frac{3k}{2}\right)^3 + kx - \frac{13}{24}k^2 & \text{for } \frac{1}{2}k < x < \frac{3}{2}k \\
kx - \frac{13}{24}k^2 & \text{for } x \geq \frac{3}{2}k.
\end{cases}
\tag{4.30}
$$

The corresponding modified Huber's ψ function can be expressed as

$$
\psi_{mh}(x) = \begin{cases}
-k & \text{for } x \leq -\frac{3}{2}k \\
\frac{1}{2k}\left(x + \frac{3k}{2}\right)^2 - k & \text{for } -\frac{3}{2}k < x < -\frac{1}{2}k \\
x & \text{for } -\frac{1}{2}k \leq x \leq \frac{1}{2}k \\
-\frac{1}{2k}\left(x - \frac{3k}{2}\right)^2 + k & \text{for } \frac{1}{2}k < x < \frac{3}{2}k \\
k & \text{for } x \geq \frac{3}{2}k.
\end{cases}
\tag{4.31}
$$

In Figure 4.1, the modified Huber's ρ and ψ functions are shown with their corresponding original Huber's ρ and ψ functions. Here $k = 1.345$ has been used for the ρ and ψ functions. As both ρ functions are close to each other, we expect similar results from the two ρ functions.

Note that to derive the asymptotic results for the GM estimates, we use a uniform law of large numbers and a central limit theorem valid for time dependent data. The key concept employed is uniform integrability to ensure the uniform convergence of a sequence. Hoadley (1971) uses the assumption of uniform integrability to establish the asymptotic properties of an estimator in the case where the observations are independent. White and Domowitz (1981) extend the results of Hoadley to establish asymptotic results for time dependent data. White and Domowitz impose some mixing conditions, which restrict the memory of a process in a fashion analogous to the role of ergodicity for a stationary stochastic process. The concepts of mixing and
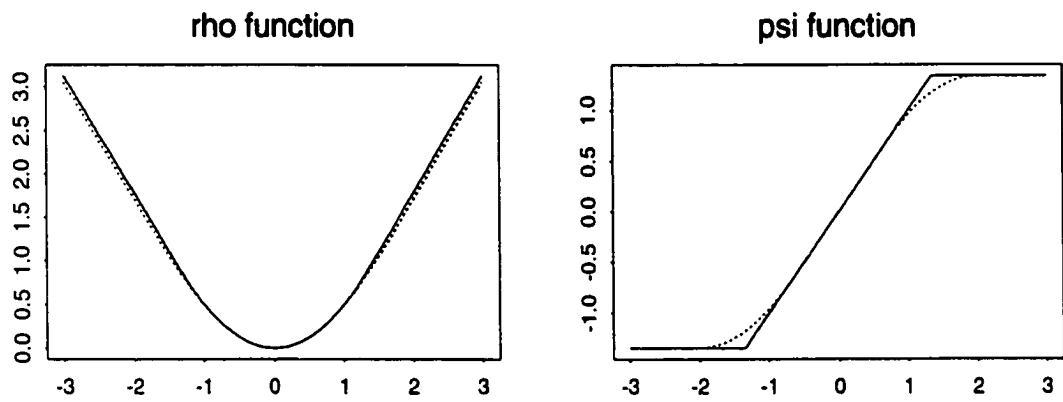
Figure 4.1: Modified Huber's ρ and ψ functions (dotted lines) and original Huber's ρ and ψ functions (solid lines)

uniform integrability of a sequence are discussed below with some definitions.

Let us define two measures of dependence between σ-algebras. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two σ-algebras and define

$$\phi(\mathcal{B}_1, \mathcal{B}_2) = \sup_{\{B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2, P(B_1) > 0\}} |P(B_2 \mid B_1) - P(B_2)|$$

$$\alpha(\mathcal{B}_1, \mathcal{B}_2) = \sup_{\{B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2\}} |P(B_1 B_2) - P(B_1)P(B_2)|$$

where P is a probability measure on a probability space $(\Omega, \mathcal{B}, P)$. Intuitively, the coefficients ϕ and α measure the dependence of the events in $\mathcal{B}_2$ on those in $\mathcal{B}_1$ in terms of how much the probability of the joint occurrence of an event in each σ-algebra differs from the product of the probabilities of each event occurring. The events in $\mathcal{B}_1$ and $\mathcal{B}_2$ are independent if and only if ϕ and α are zero. The function α provides an absolute measure of dependence, while ϕ measures dependence relative to $P(\mathcal{B}_1)$.

Definition 4.4.1 Mixing: For a sequence of random vectors $\{Y_t\}$ defined on the probability space $(\Omega, \mathcal{B}, P)$, let $\mathcal{B}_a^b$ be the Borel σ-algebra of events generated by $\{Y_a, Y_{a+1}, \ldots, Y_b\}$. Define the mixing coefficients

$$\phi(m) = \sup_n \phi(\mathcal{B}_{-\infty}^n, \mathcal{B}_{n+m}^\infty) \quad \text{and} \quad \alpha(m) = \sup_n \alpha(\mathcal{B}_{-\infty}^n, \mathcal{B}_{n+m}^\infty).$$

A sequence for which $\phi(m) \to 0$ as $m \to \infty$ is termed as uniform or ϕ-mixing and a sequence for which $\alpha(m) \to 0$ as $m \to \infty$ is termed as strong or α-mixing.

The coefficients $\phi(m)$ and $\alpha(m)$ measure the dependence between events separated by at least m time periods. Thus if $\phi(m) = 0$ or $\alpha(m) = 0$ for some m, events m periods apart are independent. By allowing $\phi(m)$ or $\alpha(m)$ to approach zero as $m \to \infty$, we allow considerations of situations where events are asymptotically independent. Note that as $\phi(m) \geq \alpha(m)$, ϕ-mixing implies α-mixing. For a real number r, $1 \leq r \leq \infty$,

(i) if $\phi(m) = O(m^{-\tau})$ for $\tau > r/(2r - 1)$, we say that $\phi(m)$ is of size $r/(2r - 1)$ and

(ii) if $\alpha(m) = O(m^{-\tau})$ for $\tau > r/(r - 1)$, $r > 1$, we say that $\alpha(m)$ is of size $r/(r - 1)$.

This definition gives precise idea about the memory of a random sequence that can be

related to moment conditions expressed in terms of r. If $r \to \infty$, a sequence exhibits more and more dependence, and if $r \to 1$, a sequence exhibits less dependence. If any sequence $\{Y_t\}$ is independent $N(0, \sigma_t^2)$, then $\{Y_t\}$ has $\phi(m)$ of size 1. If the sequence $\{Y_t\}$ is a Gaussian $AR(1)$ process, then $\{Y_t\}$ has $\alpha(m)$ of size $r/(r-1)$ for any $r > 1$. In fact, $\alpha(m)$ decreases exponentially with m for the $AR(1)$ process. Under general conditions, $ARMA$ processes have exponentially decaying memories. The result for the $AR(1)$ process also extends to an $ARMA$ process (White, 1984).

We state the following assumption on the mixing conditions.

(a8) *The sequence $\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})$ is either ϕ-mixing, with $\phi(m)$ of size $r/(2r-1)$, $r \geq 1$, or α-mixing, with $\alpha(m)$ of size $r/(r-1)$, $r > 1$.*

Definition 4.4.2 Uniform Integrability: A family $\{X_t : t \in I\}$ of integrable random variables is said to be uniformly integrable if

$$\limsup \left\{ \int_{|X_t| > K} |X_t| dP : t \in I \right\} = 0 \quad \text{as } K \to \infty.$$

A sufficient condition for $\{X_t : t \in I\}$ to be uniformly integrable is that $E|X_t|^{1+\delta} \leq \Delta < \infty$ for some positive constants Δ and δ (see Hoadley, 1971). Moreover, if $E|X_t|^{r+\delta} \leq \Delta < \infty$ for some $r \geq 1$ and $0 < \delta \leq r$, then the integrable function $\{X_t : t \in I\}$ is said to be uniformly $(r+\delta)$-integrable (see Domowitz and White, 1982). The concept of uniform integrability is important to the approach adopted here to prove the uniform convergence of $H_n(\mathbf{s}, \boldsymbol{\theta})$ and $H_n''(\mathbf{s}, \boldsymbol{\theta})$ to $\bar{H}(\boldsymbol{\theta})$ and $\bar{H}''(\boldsymbol{\theta})$, respectively. The following assumptions of uniform integrability of a sequence $\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})$ will be used to establish the asymptotic results of consistency and normality of a GM estimator.

(a9) *For the measurable sequence $\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})$, which is continuous on Θ,*
$E|\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})|^{r+\delta} \leq \Delta < \infty$ for all $\boldsymbol{\theta} \in \Theta$, $r \geq 1$, $0 < \delta \leq r$; that is, the sequence
$\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})$ is uniformly $(r + \delta)$-integrable.

(a10) *For $\theta = (\theta_1, \ldots, \theta_p)^T$, $\{\partial \nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})/\partial \theta_i\}$, $i = 1, \ldots, p$, are uniformly $2r$-integrable for all $\boldsymbol{\theta} \in \Theta$, and for some $r > 1$.*

(a11) *For $\theta = (\theta_1, \ldots, \theta_p)^T$, $\{\partial^2 \nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})/\partial \theta_i \partial \theta_j\}$, $i, j = 1, \ldots, p$, are uniformly $(r + \delta)$-integrable for some $r \geq 1$, $0 < \delta \leq r$ all $\boldsymbol{\theta} \in \Theta$.*

Assumption (a9) can be established for some specific choices of r and δ. Note that the assumption of uniform $(r + \delta)$-integrability of the measurable sequence $\{\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})\}$ requires that

$$\sup_{\boldsymbol{\theta} \in \Theta} E |\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})|^{r+\delta} \leq \Delta < \infty$$

for some $r \geq 1$, $0 < \delta \leq r$. Since $\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta}) = w(\mathbf{z}_t)\rho(y_t - \mu_t(\mathbf{z}_t, \boldsymbol{\theta})) \leq \rho(y_t - \mu_t(\mathbf{z}_t, \boldsymbol{\theta})) \leq (y_t - \mu_t(\mathbf{z}_t, \boldsymbol{\theta}))^2$, it is sufficient to show that

$$\sup_{\boldsymbol{\theta} \in \Theta} E |y_t - \mu_t(\mathbf{z}_t, \boldsymbol{\theta})|^{2(r+\delta)} < \infty$$

for some $r \geq 1$, $0 < \delta \leq r$. As Θ is assumed to be a compact subset of an Euclidean space, the above expectation may be considered as finite for some choices of r and δ. As a simple illustration, let us consider $r + \delta = 1$. Then

$$\sup_{\boldsymbol{\theta} \in \Theta} E |y_t - \mu_t(\mathbf{z}_t, \boldsymbol{\theta})|^2$$
$$= \sup_{\boldsymbol{\theta} \in \Theta} \int [(y_t - E(y_t)) + (E(y_t) - \mu_t(\mathbf{z}_t, \boldsymbol{\theta}))]^2 dP$$
$$= \text{var}(y_t) + \sup_{\boldsymbol{\theta} \in \Theta} \int (E(y_t) - \mu_t(\mathbf{z}_t, \boldsymbol{\theta}))^2 dP.$$

Here $\text{var}(y_t)$ is finite. As Θ is compact, the second term in the above equation is also finite. Similarly, it may be shown that for some choices of $r \geq 1$, $0 < \delta \leq r$, $\sup_{\boldsymbol{\theta} \in \Theta} E |y_t - \mu_t(\mathbf{z}_t, \boldsymbol{\theta})|^{2(r+\delta)}$ is finite. The value of r depends on the allowable dependence in a stationary sequence.

By a similar argument, it may be shown that (a10) and (a11) are also reasonable assumptions in our situation. It will be shown that under assumptions (a7),

(a8) and (a11), $\bar{H}''(\boldsymbol{\theta})$ is continuous and $\left| H_n''(\mathbf{s}, \boldsymbol{\theta}) - \bar{H}''(\boldsymbol{\theta}) \right| \to 0$ a.s. uniformly for all $\boldsymbol{\theta} \in \Theta$, where $\bar{H}''(\boldsymbol{\theta}) = \lim_{n \to \infty} \bar{H}_n''(\boldsymbol{\theta}) = \lim_{n \to \infty} E\left[H_n''(\mathbf{s}, \boldsymbol{\theta}) \right]$ and $H_n''(\mathbf{s}, \boldsymbol{\theta}) = (1/n) \sum_{t=q+1}^{n} [\nu''(y_t, \mathbf{z}_t, \boldsymbol{\theta})]$, $\nu''(y_t, \mathbf{z}_t, \boldsymbol{\theta})$ is the second derivative of $\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

Now, we state the following lemmas which lead to the consistency and asymptotic normality results of the *GM* estimators.

Lemma 4.4.1 Let $\{Y_t\}$ be a scalar sequence with $\phi(m)$ of size $r/(2r-1)$ or $\alpha(m)$ of size $r/(r-1), r > 1$, and let $E(Y_t) = \mu_t$. If for $0 < \delta \leq r$, $\sum_{t=1}^{\infty} \left(E|Y_t - \mu_t|^{r+\delta}/t^{r+\delta} \right)^{1/r} < \infty$, then $\left| \frac{1}{n} \sum_{t=1}^{n} (Y_t - \mu_t) \right| \to 0$ a.s. as $n \to \infty$. (See McLeish, 1975 for proof).

Using this lemma, we obtain the following corollary.

Corollary 4.4.2 Let $\{Y_t\}$ be a sequence with $\phi(m)$ of size $r/(2r-1)$ or $\alpha(m)$ of size $r/(r-1), r > 1$, such that $E|Y_t|^{r+\delta} < \Delta < \infty$ for $0 < \delta \leq r$, and all t. Then $\left| \frac{1}{n} \sum_{t=1}^{n} (Y_t - \mu_t) \right| \to 0$ a.s. as $n \to \infty$.

Proof: Let Y and Z be two random variables with $E|Y|^r < \infty$ and $E|Z|^r < \infty$ for some $r > 0$. Then the "c_r inequality" gives $E|Y+Z|^r \leq c_r \left(E|Y|^r + E|Z|^r \right)$, where $c_r = 1$ if $r \leq 1$ and $c_r = 2^{r-1}$ if $r > 1$ (White, 1984). Using this inequality, we have

$$E|Y_t - \mu_t|^{r+\delta} \leq 2^{r+\delta-1} \left(E|Y_t|^{r+\delta} + E|\mu_t|^{r+\delta} \right).$$

By assumption, $E|Y_t|^{r+\delta} < \Delta$ and using Jensen's inequality, we have $|\mu_t| \leq E|Y_t| \leq \left(E|Y_t|^{r+\delta} \right)^{1/(r+\delta)}$. This implies that for all t, $|\mu_t|^{r+\delta} < \Delta$. Hence for all t, $E|Y_t - \mu_t|^{r+\delta} < 2^{r+\delta-1}(\Delta + \Delta) = 2^{r+\delta}\Delta$. So

$$\sum_{t=1}^{\infty} \left(E|Y_t - \mu_t|^{r+\delta}/t^{r+\delta} \right)^{1/r} < 2^{r+\delta}\Delta \sum_{t=1}^{\infty} t^{-(1+\delta/r)} < \infty$$

since $\sum_{t=1}^{\infty} t^{-(1+\delta/r)} < \infty$ for any $(\delta/r) > 0$. Thus the conditions of Lemma 4.4.1 are satisfied and the result follows. ∎

Note that for sequences with longer memories, r is greater, and the moment restrictions increase accordingly. So there is a trade-off between the amount of allowable dependence and the sufficient moment restrictions. Corollary 4.4.2 can be extended to the lemma below which ensures the almost sure uniform convergence of $H_n(\mathbf{s}, \boldsymbol{\theta})$ to $\bar{H}(\boldsymbol{\theta})$, which will be required to establish the asymptotic results.

Lemma 4.4.3 Let the sequence $\{q_t(\boldsymbol{\theta})\}$ be measurable for each $\boldsymbol{\theta}$ in Θ, and continuous on Θ, a compact subset of an Euclidean space, uniformly in t, a.s. with respect to a probability measure P. If $\phi(m)$ is of size $r/(2r-1)$, $r \geq 1$, or $\alpha(m)$ is of size $r/(r-1)$, $r > 1$, and for all $\boldsymbol{\theta} \in \Theta$, if $E|q_t(\boldsymbol{\theta})|^{r+\delta} \leq \Delta < \infty$ for $0 < \delta \leq r$, and all t, then

(a) $E[q_t(\boldsymbol{\theta})]$ is continuous on Θ uniformly in t, and

(b) $\left| \frac{1}{n} \sum_{t=1}^{n} [q_t(\boldsymbol{\theta}) - E(q_t(\boldsymbol{\theta}))] \right| \to 0$ a.s. as $n \to \infty$, uniformly in $\boldsymbol{\theta}$.

Proof: (a) Since $q_t(\boldsymbol{\theta})$ is continuous on Θ, uniformly in t, a.s., for each $\boldsymbol{\theta}_0 \in \Theta$, $\lim q_t(\boldsymbol{\theta}) = q_t(\boldsymbol{\theta}_0)$ as $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$ uniformly in t, a.s. Moreover, for each $\boldsymbol{\theta} \in \Theta$ and all t, $q_t(\boldsymbol{\theta})$ is uniformly integrable under the given assumptions. Therefore, using Theorem A.3 (ii) of Hoadley (1971), we have $\lim E|q_t(\boldsymbol{\theta}) - q_t(\boldsymbol{\theta}_0)| = 0$ as $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$ uniformly in t. Hence we have $\lim E[q_t(\boldsymbol{\theta})] = E[q_t(\boldsymbol{\theta}_0)]$ as $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$ uniformly in t. This completes the proof.

(b) Since under the mixing conditions, $q_t(\boldsymbol{\theta})$ is assumed to be $(r + \delta)$-integrable uniformly in t, and all $\boldsymbol{\theta} \in \Theta$, the result follows from Corollary 4.4.2. ∎

This lemma will be applied to the sequence $\{\nu(y_t, \mathbf{z}_t, \boldsymbol{\theta})\}$ in Theorem 4.4.4 and to the sequence $\{\nu''(y_t, \mathbf{z}_t, \boldsymbol{\theta})\}$ in Theorem 4.4.7 to establish the uniform convergence of the sequences $H_n(\mathbf{s}, \boldsymbol{\theta})$ and $H_n''(\mathbf{s}, \boldsymbol{\theta})$, respectively.

Now, we state the following theorem to obtain a consistency result of the *GM* estimates.

Theorem 4.4.4 (Consistency) Under assumptions (a4), (a6), (a8) and (a9), for the GM estimator $\hat{\theta}_n$, obtained by minimizing the objective function $H_n(\mathbf{s}, \theta)$, $\hat{\theta}_n \to \theta_0$ a.s. as $n \to \infty$.

Proof: Under assumptions (a6), (a8) and (a9), we can apply Lemma 4.4.3 to the sequence $\{\nu(y_t, \mathbf{z}_t, \theta)\}$ to show that $\bar{H}(\theta) = E[\nu(y_1, \mathbf{z}_1, \theta)]$ is continuous on Θ and $|H_n(\mathbf{s}, \theta) - \bar{H}(\theta)| \to 0$ a.s. uniformly for all $\theta \in \Theta$. Under assumption (a4), $\bar{H}(\theta)$ has a unique minimizer θ_0 and the result then follows from Theorem 2.2 of Domowitz and White (1982). ∎

The asymptotic normality of an estimator can be obtained by taking a mean value expansion of the first-order conditions for a minimum of $H_n(\mathbf{s}, \theta)$. The mean value expansion gives

$$H_n'(\mathbf{s}, \hat{\theta}_n) = H_n'(\mathbf{s}, \theta_0) + H_n''(\mathbf{s}, \bar{\theta}_n)(\hat{\theta}_n - \theta_0) \tag{4.32}$$

where $\bar{\theta}_n$ is lying on the segment connecting $\hat{\theta}_n$ and θ_0, and H_n' and H_n'' are the corresponding first and second derivatives of H_n with respect to θ. The function H_n'' gives the Hessian of H_n. From (4.32), we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \left[-H_n''(\mathbf{s}, \bar{\theta}_n)\right]^{-1} \left(\sqrt{n} H_n'(\mathbf{s}, \theta_0)\right) \tag{4.33}$$

since $H_n'(\mathbf{s}, \hat{\theta}_n) = 0$ for the minimizer $\hat{\theta}_n$. Now, the asymptotic normality of $\hat{\theta}_n$ follows if the Hessian is properly behaved, that is, if it converges appropriately and if $\sqrt{n} H_n'(\mathbf{s}, \theta_0)$ in (4.33) has the central limit property.

Using Theorem 2.6 of Domowitz and White (1982), we state the following lemma to establish the central limit property of $\sqrt{n} H_n'(\mathbf{s}, \theta)$.

Lemma 4.4.5 Let $S_n = \sum_{t=1}^{n} Y_t$ and $T_a(n) = n^{-\frac{1}{2}} \sum_{t=a+1}^{a+n} Y_t$ for a sequence $\{Y_t\}$ of random variables satisfying (a) $E(Y_t) = 0$; (b) there exists a finite and non-zero Q such that $E((T_a(n))^2) \to Q$ as $n \to \infty$, uniformly in a; and (c) $E|Y_t|^{2r} \leq \Delta < \infty$ for all t and some $r > 1$. If either $\phi(m)$ or $\alpha(m)$ is of size $r/(r-1)$, then

$n^{-\frac{1}{2}}Q^{-\frac{1}{2}}S_n \stackrel{\cdot}{\sim} N(0,1)$. (For proof, see White and Domowitz, 1981).

Recall that the derivative of the objective function $H_n(\mathbf{s}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ gives

$$\sqrt{n}H_n'(\mathbf{s}, \boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{t=q+1}^{n} w(\mathbf{z}_t)\psi(y_t - \mu_t(\mathbf{z}_t, \boldsymbol{\theta}))\frac{\partial\mu_t}{\partial\boldsymbol{\theta}} \qquad (4.34)$$

where ψ is the derivative of the function ρ. Under assumptions (a8) and (a10), the sequence $Y_t = w(\mathbf{z}_t)\psi(y_t - \mu_t(\mathbf{z}_t, \boldsymbol{\theta}))(\partial\mu_t/\partial\boldsymbol{\theta})$ satisfies the conditions of Lemma 4.4.5. For $S_n = \sum_{t=1}^{n} Y_t$ and $T_a(n) = n^{-\frac{1}{2}}\sum_{t=a+1}^{a+n} Y_t$, we have

$$
\begin{aligned}
E(Y_t) &= E[w(\mathbf{z}_t)\psi(y_t - \mu_t(\mathbf{z}_t, \boldsymbol{\theta}))(\partial\mu_t/\partial\boldsymbol{\theta})] \\
&= -(\partial/\partial\boldsymbol{\theta})E[w(\mathbf{z}_t)\rho(y_t - \mu_t(\mathbf{z}_t, \boldsymbol{\theta}))] \\
&= -(\partial/\partial\boldsymbol{\theta})E[w(\mathbf{z}_1)\rho(y_1 - \mu_1(\mathbf{z}_1, \boldsymbol{\theta}))], \text{ under stationarity} \\
&= \mathbf{0} \quad \text{at } \boldsymbol{\theta} = \boldsymbol{\theta}_0.
\end{aligned}
$$

Also at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, we have

$$
\begin{aligned}
&E\left[(T_a(n))(T_a(n))^T\right] \\
&= n^{-1}E\left[\left(\sum_{t=a+1}^{a+n} Y_t\right)\left(\sum_{t=a+1}^{a+n} Y_t\right)^T\right] \\
&= n^{-1}\left[\sum_{t=a+1}^{a+n} E\left(Y_t Y_t^T\right) + \sum_{t=a+1}^{a+n}\sum_{t'\neq t=a+1}^{a+n} E\left(Y_t Y_{t'}^T\right)\right] \\
&= n^{-1}\sum_{t=a+1}^{a+n} E\left(Y_t Y_t^T\right)
\end{aligned}
$$

[the cross-product term is zero since $E\left(Y_t Y_{t+1}^T\right) = E\left(Y_t E^T\left(Y_{t+1} \mid \mathbf{z}_{t+1}\right)\right)$ and $E\left(Y_{t+1} \mid \mathbf{z}_{t+1}\right)$ is zero, and so on]

$$
\begin{aligned}
&\longrightarrow \text{var}(Y_1) \quad \text{as } n \to \infty \\
&= E\left[w^2(\mathbf{z}_1)\psi^2(y_1 - \mu_1(\mathbf{z}_1, \boldsymbol{\theta}))\frac{\partial\mu_1}{\partial\boldsymbol{\theta}}\frac{\partial\mu_1}{\partial\boldsymbol{\theta}}^T\right] \\
&= Q \quad \text{(say)}.
\end{aligned}
$$

Thus under the given assumptions, the central limit theorem applies to $(1/\sqrt{n})S_n = \sqrt{n}H'_n(\mathbf{s}, \boldsymbol{\theta})$, that is, $\sqrt{n}H'_n(\mathbf{s}, \boldsymbol{\theta}_0) \overset{\cdot}{\sim} N(0, Q)$.

Now, we state the following lemma which will ensure that the Hessian converges properly.

Lemma 4.4.6 Suppose, $G_n(\mathbf{s}, \boldsymbol{\theta})$ is a measurable function on a measurable space Ω, and for each $\mathbf{s} \in \Omega$, a continuous function on Θ, a compact subset of an Euclidean space. If the function $\bar{G}(\boldsymbol{\theta}) = E[G_n(\mathbf{s}, \boldsymbol{\theta})]$ is continuous on Θ with $\left| G_n(\mathbf{s}, \boldsymbol{\theta}) - \bar{G}(\boldsymbol{\theta}) \right| \to 0$ a.s. uniformly for all $\boldsymbol{\theta} \in \Theta$, and if $\hat{\boldsymbol{\theta}}_n(\mathbf{s}) - \boldsymbol{\theta}_0 \to 0$ a.s., then $\left| G_n(\mathbf{s}, \hat{\boldsymbol{\theta}}_n(\mathbf{s})) - \bar{G}(\boldsymbol{\theta}_0) \right| \to 0$ a.s. (The proof is identical to that of Theorem 2.3 of Domowitz and White, 1982).

The central limit property follows from the next theorem. In this theorem we address the convergence of the Hessian as well as the asymptotic normality of the GM estimates.

Theorem 4.4.7 (Asymptotic Normality) Under assumptions (a1)-(a11), the central limit property of the GM estimator $\hat{\boldsymbol{\theta}}_n$ gives

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \overset{\cdot}{\sim} N(0, V) \tag{4.35}$$

where $V = \left[-\bar{H}''(\boldsymbol{\theta}_0) \right]^{-1} Q \left[-\bar{H}''(\boldsymbol{\theta}_0) \right]^{-1}$ provides the asymptotic variance of the GM estimator $\hat{\boldsymbol{\theta}}_n$.

Proof: Recall equation (4.33) of the mean value expansion. Under assumptions (a4), (a6), (a8) and (a9), $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s. as $n \to \infty$ (see Theorem 4.4.4). Under assumptions (a7), (a8) and (a11), we can apply Lemma 4.4.3 to the sequence $\{\nu''(y_t, \mathbf{z}_t, \boldsymbol{\theta})\}$ to show that $\bar{H}''(\boldsymbol{\theta})$ is continuous on Θ and $\left| H''_n(\mathbf{s}, \boldsymbol{\theta}) - \bar{H}''(\boldsymbol{\theta}) \right| \to 0$ a.s. uniformly for all $\boldsymbol{\theta} \in \Theta$. Now, since $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s. and $\bar{\boldsymbol{\theta}}_n$ is a suitable mean value, and since $\left| H''_n(\mathbf{s}, \boldsymbol{\theta}) - \bar{H}''(\boldsymbol{\theta}) \right| \to 0$ a.s. uniformly for all $\boldsymbol{\theta} \in \Theta$, using Lemma 4.4.6, we can show that $\left| H''_n(\mathbf{s}, \bar{\boldsymbol{\theta}}_n) - \bar{H}''(\boldsymbol{\theta}_0) \right| \to 0$ a.s. Also under assumptions (a8) and (a10), we can apply Lemma 4.4.5 to show that the central limit theorem applies to $\sqrt{n}H'_n(\mathbf{s}, \boldsymbol{\theta}_0)$,

that is, $\sqrt{n}H'_n(\mathbf{s}, \boldsymbol{\theta}_0) \sim N(0, Q)$. Then the theorem follows from equation (4.33). ∎

It can be shown that the function $-\bar{H}''(\boldsymbol{\theta}) = \lim_{n\to\infty} E[-H''_n(\mathbf{s}, \boldsymbol{\theta})] = M$, where $M = E\left[w(\mathbf{z}_1)\psi'(y_1 - \mu_1(\mathbf{z}_1, \boldsymbol{\theta}))\frac{\partial\mu_1}{\partial\boldsymbol{\theta}}\frac{\partial\mu_1}{\partial\boldsymbol{\theta}}^T\right]$. It is important to note that the variance function V can also be obtained from the influence function as

$$V = E\left[IF(\mathbf{y}_1, \mathbf{X}_1; T, P)\ IF(\mathbf{y}_1, \mathbf{X}_1; T, P)^T\right] \tag{4.36}$$

where the influence function $IF(\mathbf{y}_1, \mathbf{X}_1; T, P)$ is defined in equation (4.27) for the functional $T(P) = \boldsymbol{\theta}_0$. Note that since $\boldsymbol{\theta}^T = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T)$, the function M can be expressed as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

where

$$M_{11} = E\left[w(\mathbf{z}_1)\psi'(y_1 - \mu_1(\mathbf{z}_1, \boldsymbol{\theta}))\frac{\partial\mu_1}{\partial\boldsymbol{\beta}}\frac{\partial\mu_1}{\partial\boldsymbol{\beta}}^T\right]$$

$$M_{22} = E\left[w(\mathbf{z}_1)\psi'(y_1 - \mu_1(\mathbf{z}_1, \boldsymbol{\theta}))\frac{\partial\mu_1}{\partial\boldsymbol{\phi}}\frac{\partial\mu_1}{\partial\boldsymbol{\phi}}^T\right]$$

$$M_{12} = M_{21}^T = E\left[w(\mathbf{z}_1)\psi'(y_1 - \mu_1(\mathbf{z}_1, \boldsymbol{\theta}))\frac{\partial\mu_1}{\partial\boldsymbol{\beta}}\frac{\partial\mu_1}{\partial\boldsymbol{\phi}}^T\right],$$

with each term evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Similarly, the function Q can be expressed as

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$$

where

$$Q_{11} = E\left[w^2(z_1)\psi^2(y_1 - \mu_1(z_1,\boldsymbol{\theta}))\frac{\partial\mu_1}{\partial\beta}\frac{\partial\mu_1}{\partial\beta}^T\right]$$

$$Q_{22} = E\left[w^2(z_1)\psi^2(y_1 - \mu_1(z_1,\boldsymbol{\theta}))\frac{\partial\mu_1}{\partial\phi}\frac{\partial\mu_1}{\partial\phi}^T\right]$$

$$Q_{12} = Q_{21}^T = E\left[w^2(z_1)\psi^2(y_1 - \mu_1(z_1,\boldsymbol{\theta}))\frac{\partial\mu_1}{\partial\beta}\frac{\partial\mu_1}{\partial\phi}^T\right],$$

with each term evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

The asymptotic variance $V = M^{-1} Q M^{-1}$ can be estimated by its sample counterpart as

$$\hat{V} = \hat{M}^{-1} \hat{Q} \hat{M}^{-1}$$

$$= \begin{bmatrix} \hat{M}_{11} & \hat{M}_{12} \\ \hat{M}_{21} & \hat{M}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \hat{Q}_{11} & \hat{Q}_{12} \\ \hat{Q}_{21} & \hat{Q}_{22} \end{bmatrix} \begin{bmatrix} \hat{M}_{11} & \hat{M}_{12} \\ \hat{M}_{21} & \hat{M}_{22} \end{bmatrix}^{-1}$$

where

$$\hat{M}_{11} = \frac{1}{n}\sum_{t=q+1}^{n} w(z_t)\psi'(y_t - \mu_t(z_t,\boldsymbol{\theta}))\frac{\partial\mu_t}{\partial\beta}\frac{\partial\mu_t}{\partial\beta}^T$$

$$\hat{Q}_{11} = \frac{1}{n}\sum_{t=q+1}^{n} w^2(z_t)\psi^2(y_t - \mu_t(z_t,\boldsymbol{\theta}))\frac{\partial\mu_t}{\partial\beta}\frac{\partial\mu_t}{\partial\beta}^T$$

and so on, with each term evaluated at the *GM* estimate $\hat{\boldsymbol{\theta}}^T = (\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\phi}}^T)$ of the parameter $\boldsymbol{\theta}^T = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T)$. We use $\hat{\boldsymbol{\theta}}$ in place of the *GM* estimate $\hat{\boldsymbol{\theta}}_n$ for notational convenience. Note that throughout this section, we assume that $\text{var}(u_t) = \sigma_t^2 = 1$, for simplicity. In practice, σ_t is estimated by a suitable method which we discuss in the application section of this chapter. For estimated $\hat{\sigma}_t$, the variance function V should be estimated accordingly.

4.5 Choosing Order of the Autoregressive Error Process in Nonlinear Regression

We use the robust version of the Akaiki Information Criterion ($RAIC$) for choosing the appropriate order of an autoregressive error process. In Chapter 3, we discussed the $RAIC$ in order to choose the correct order of an autoregressive process. In this section, we extend this $RAIC$ to the nonlinear regression with autoregressive errors. For the nonlinear regression model (4.1), the order q of the $AR(q)$ process $\{\epsilon_t\}$ can be determined by minimizing the robust version of the Information Criterion

$$RAIC(q) = \left(1 \Big/ \sum_{t=q^*+1}^{n} w(\mathbf{z}_t)\right) \left[\sum_{t=q^*+1}^{n} w(\mathbf{z}_t)\rho\left(\frac{y_t - \mu_t(\mathbf{z}_t, \hat{\boldsymbol{\theta}})}{\hat{\sigma}_t}\right) + c_q \hat{\alpha}_q\right] \quad (4.37)$$

where $\mathbf{z}_t = (\mathbf{x}_t, \mathbf{x}_{t-1}^{t-q}, y_{t-1}^{t-q})$, the tuning constant $c_q = 2(n - q)/\sum_{t=q^*+1}^{n} w(\mathbf{z}_t)$, $\hat{\alpha}_q = \text{tr}(\hat{M}^{-1}\hat{Q})$ with

$$\hat{M} = \sum_{t=q^*+1}^{n} \frac{1}{\hat{\sigma}_t^2} w(\mathbf{z}_t)\psi'\left(\frac{y_t - \mu_t(\mathbf{z}_t, \hat{\boldsymbol{\theta}})}{\hat{\sigma}_t}\right) \frac{\partial\mu_t}{\partial\boldsymbol{\theta}} \frac{\partial\mu_t}{\partial\boldsymbol{\theta}}^T \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

$$\hat{Q} = \sum_{t=q^*+1}^{n} \frac{1}{\hat{\sigma}_t^2} w^2(\mathbf{z}_t)\psi^2\left(\frac{y_t - \mu_t(\mathbf{z}_t, \hat{\boldsymbol{\theta}})}{\hat{\sigma}_t}\right) \frac{\partial\mu_t}{\partial\boldsymbol{\theta}} \frac{\partial\mu_t}{\partial\boldsymbol{\theta}}^T \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

and q^* is the maximum possible value of q under consideration for the order of the autoregressive error process. We choose that value of q which minimizes $RAIC(q)$ for the best autoregressive model.

4.6 Application: Analysis of Ground Level Ozone Data

Ozone (O_3) is a trace gas in the atmosphere. The highest concentration of ozone is in the stratosphere, where it shields the earth's surface from harmful ultraviolet radiation. At the surface, however, ozone itself is harmful, with serious impacts on public

health and environment. Excessive tropospheric ozone concentrations, that arise as a consequence of changes in precursor emissions, are of main concern in air pollution. Ozone concentrations have been high enough in certain areas to be of concern for several decades. The United States Environmental Protection Agency (USEPA) has been trying to improve the environmental risks by developing an appropriate strategy to control emission sources, which is found to be an effective way of reducing the ozone level concentrations. The current US air pollution standard stipulates that the daily maximum ozone concentration can exceed 120 parts per billion (ppb) only three times in a three year period (Niu, 1996).

As ozone is a secondary pollutant, it is difficult to control the level of ozone concentrations. Ozone is an outcome from photochemical reactions involving precursor pollutants. The precursors include a variety of volatile organic compounds, comprised mainly of nonmethane hydrocarbons, nitric oxide (NO) and nitrogen dioxide (NO_2). Transportation and industrial processes contribute to the emission of both nonmethane hydrocarbons and nitrogen oxides. On the other hand, diverse sources such as automobiles, chemical manufacturers and other facilities using chemical solvents contribute to the emission of volatile organic compounds. Along with the sources of precursor emissions, variations in meteorological conditions also contribute to the fluctuations of ozone level concentrations. For instance, high temperatures coupled with lower wind speeds usually result in high ozone levels. A major issue in the analysis of ozone data is how to separate trends due to the precursor emissions from the effects of meteorological variability. Many statistical contributions are focused on determining the relationship between ozone concentrations and meteorology. In many cases, nonlinear regression procedures have been developed to relate the ozone concentrations with meteorology. Niu (1996) analyzed the daily maxima of ground level ozone concentrations in the Chicago area of the United States . He introduced a class of nonlinear additive models for environmental time series, in which both mean levels and variances of the series are nonlinear functions of relevant meteorological variables.

In this section, we analyze a ozone data set observed from a meteorological site in the Bexar County, Texas, USA. The data were collected by the Texas Natural

Resource Conservation Commission (TNRCC) and local monitoring networks. We take the daily maximum of hourly surface ozone concentrations (OZONE) in parts per billion (ppb) over the ten year period 1986-1995. The corresponding five meteorological variables selected are daily maximum surface temperature (TEMP) in degrees Fahrenheit, morning average wind speed resultant (MWS) in miles per hour, afternoon average wind speed resultant (AWS), morning average wind direction resultant (MWD) in degrees clockwise from true north and afternoon average wind direction resultant (AWD).

Figure 4.2 presents the plots of daily maximum ozone concentrations against time as well as three meteorological variables. From the time series plot of the ozone data, we observe some seasonal pattern in the ozone concentrations. Ozone level appears to be positively associated with temperature, and the relationship is nonlinear. Both morning and afternoon wind speeds have negative impacts on ozone concentrations. In fact, high temperatures along with lower wind speeds usually result in high ozone level concentrations.

4.6.1 Model Selection

To relate the ozone concentrations with meteorology, we consider a nonlinear regression model with autoregressive errors in the form

$$y_t = f(\mathbf{x}_t; \boldsymbol{\beta}) + g(t; \boldsymbol{\gamma}) + \epsilon_t \tag{4.38}$$

where $f(\mathbf{x}_t; \boldsymbol{\beta})$ is a nonlinear function of the meteorological variables, $g(t; \boldsymbol{\gamma})$ represents the trend and seasonal effects and ϵ_t is an autoregressive process of order q of the form

$$\epsilon_t = \phi_1 \epsilon_{t-1} + \ldots + \phi_q \epsilon_{t-q} + u_t. \tag{4.39}$$

As the daily maximum of ozone series $\{y_t\}$ is usually nonstationary, we assume the u_t's are independent $N(0, \sigma_t^2)$, where σ_t can be modeled as a function of the meteorological variables. Niu (1996) used this type of model except for the fact that instead of using
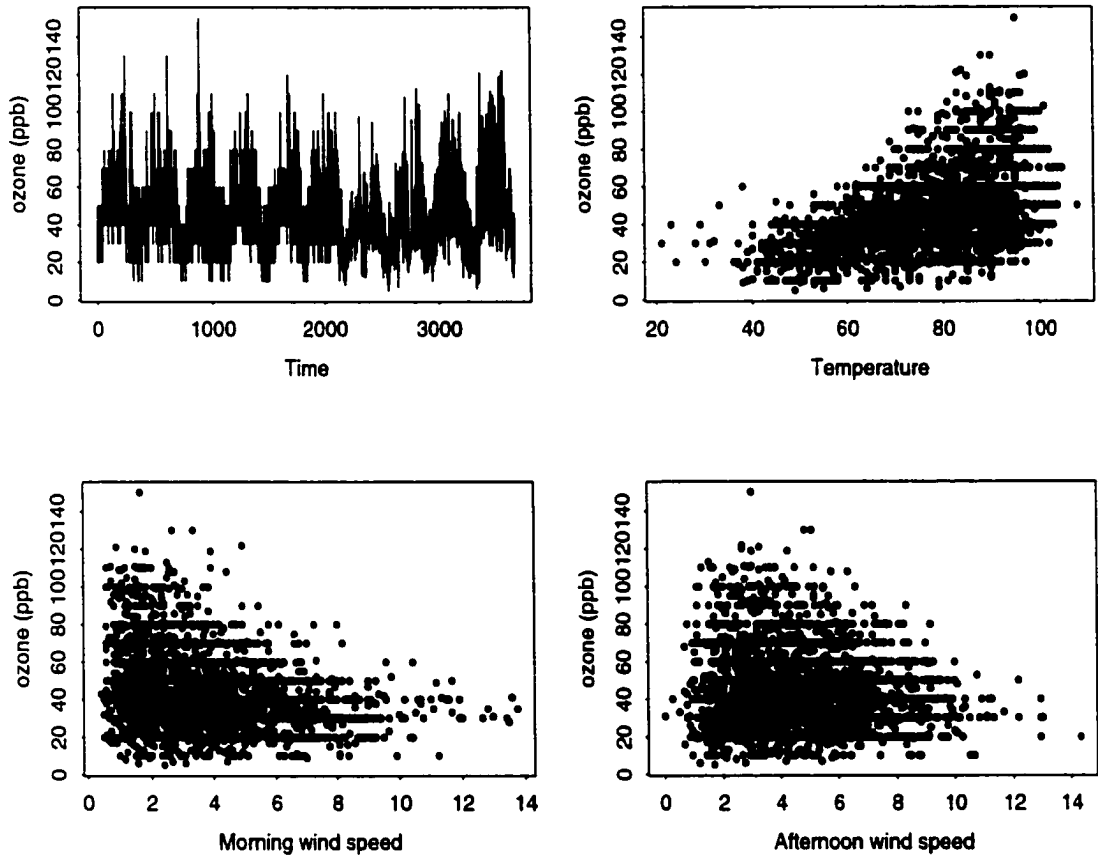
Figure 4.2: Plots of ozone concentrations against time as well as some meteorological variables

a nonlinear parametric model $f(\mathbf{x}_t; \boldsymbol{\beta})$, he used a nonlinear additive model for the meteorological variables. By investigating the trends of ozone levels, we consider that the meteorological variables are exponentially related to the mean ozone level concentrations:

$$f(\mathbf{x}_t; \boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t}) \qquad (4.40)$$

where x_{1t} represents the daily maximum surface temperature (TEMP), x_{2t} the morning average wind speed resultant (MWS), x_{3t} the afternoon average wind speed resultant (MWS), x_{4t} the morning average wind direction resultant (MWD) and x_{5t} the afternoon average wind direction resultant (AWD) at time t. The function $g(t; \boldsymbol{\gamma})$ includes the seasonal and trend terms which is assumed to be of the form

$$
\begin{aligned}
g(t; \boldsymbol{\gamma}) &= \gamma_1(t/365) + \gamma_2 \sin(2\pi t/365) + \gamma_3 \cos(2\pi t/365) \\
&\quad + \gamma_4 \sin(4\pi t/365) + \gamma_5 \cos(4\pi t/365)
\end{aligned} \qquad (4.41)
$$

where γ_1 is a trend parameter which is of great interest in the analysis of ozone concentrations. The parameters γ_i ($i = 2, \ldots, 5$) represent the yearly and half-yearly seasonal terms. Defining $h_t = f(\mathbf{x}_t; \boldsymbol{\beta}) + g(t; \boldsymbol{\gamma})$, the model (4.38) can be rewritten as $y_t = \mu_t + u_t$ ($t = q + 1, \ldots, n$), where $\mu_t = h_t + (y_{t-1}^{t-q} - h_{t-1}^{t-q})^T \boldsymbol{\phi}$ with $\boldsymbol{\phi}^T = (\phi_1, \ldots, \phi_q)$, and u_t's are independent $N(0, \sigma_t^2)$.

Remark: Note that for the GM estimate of $\boldsymbol{\theta}^T = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T, \boldsymbol{\phi}^T)$ of models (4.38) and (4.39), the corresponding influence function is unbounded as the function $w(\mathbf{z}_1) \times \frac{\partial}{\partial \boldsymbol{\theta}} \mu_1(\mathbf{z}_1, \boldsymbol{\theta})$ is unbounded (see equation (4.27)). This is due to the fact that when $\mathbf{z}_1$ tends to ∞, the weight function $w(\mathbf{z}_1)$ tends to zero at a slower rate as compared to the function $\frac{\partial}{\partial \boldsymbol{\theta}} \mu_1(\mathbf{z}_1, \boldsymbol{\theta})$, which tends to ∞. For a bounded influence function of the GM estimates, one may choose an exponentially decaying weight function once $w(\mathbf{z}_1)$ is small. We may consider $w(\mathbf{z}_1)$ as small when it is less than .05.

Now, the next step is to find a suitable model for estimating σ_t. Initially, we assume that $\sigma_t = \sigma$, for each t. Then σ can be estimated as $\hat{\sigma} = \text{med}_t |y_t - \hat{\mu}_t| / .6745$, where $\hat{\mu}_t$ is obtained by evaluating μ_t at the GM estimate $\hat{\boldsymbol{\theta}}$. For a fixed
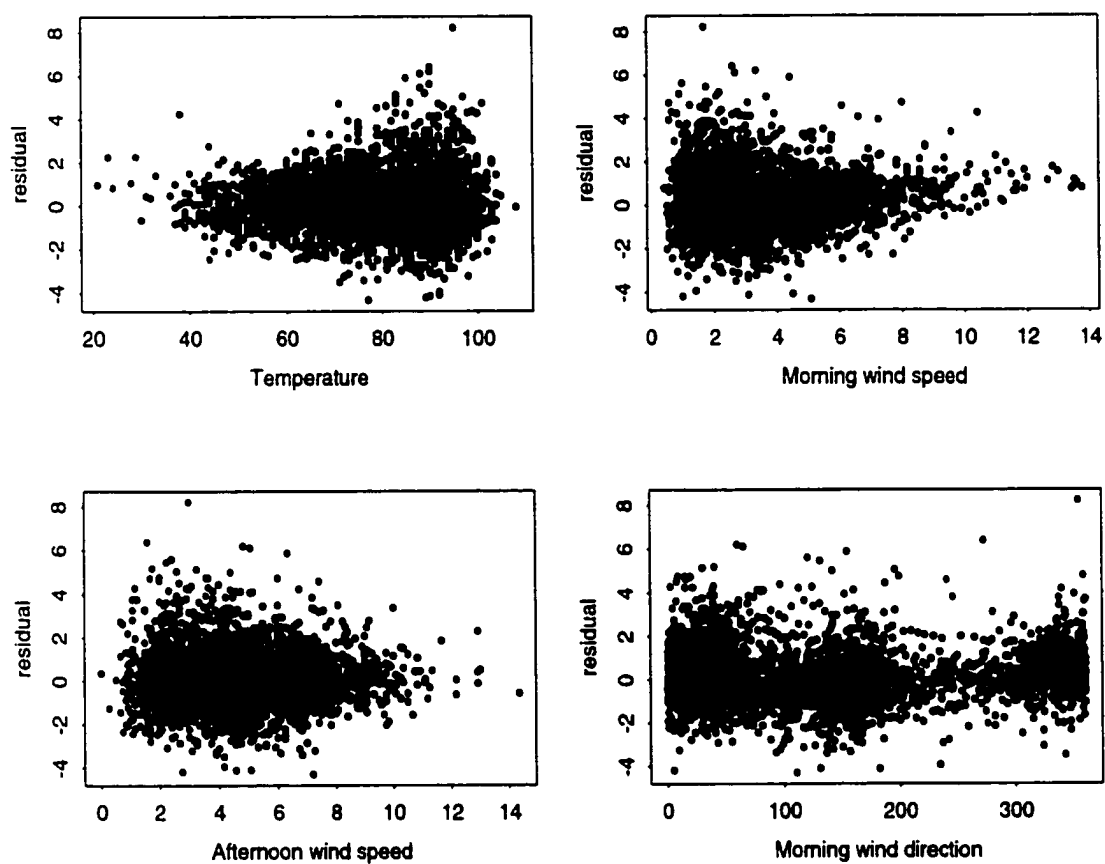
Figure 4.3: Scatterplots of $(y_t - \hat{\mu}_t)/\hat{\sigma}$ versus the predictors

σ, model (4.38) was fitted assuming that the error ϵ_t follows an $AR(1)$ process. For this fitted model, we plot the standardized residuals $(y_t - \hat{\mu}_t)/\hat{\sigma}$ against the meteorological variables TEMP, MWS, AWS and MWD. The scatterplots are shown in Figure 4.3. From the scatterplots, we certainly observe a pattern in the residuals corresponding to the three important meteorological variables TEMP, MWS and AWS. The other two meteorological variables MWD and AWD were found to be insignificant and no pattern in the residuals was observed corresponding to these variables. The scatterplot of the residuals against temperature shows that the variability in the residuals increases for increased temperature. On the other hand, this variability decreases when wind speed increases. In fact, high temperature coupled with lower wind speed cause increased variability in the residuals. We consider modeling σ_t as a function of the three meteorological variables TEMP, MWS and AWS. To define a suitable model for σ_t, we first classify the predictor variables into different categories. For each of the three predictor variables, six class intervals were defined in such a way that each class interval contains approximately equal number of residuals. For example, for the variable TEMP, the six class intervals were defined as $\{(0, 63], (63, 73], (73, 80], (80, 87], (87, 92], (92, 120]\}$. Then we define a new variable TEMP* with the values $(1, \ldots, 6)$, where TEMP* is 1 if the original TEMP lies in $(0, 63]$, 2 if TEMP lies in $(63, 73]$, and so on. Using similar technique, we define the new variables MWS* and AWS* with possible values $(1, \ldots, 6)$ corresponding to the original variables MWS and AWS. Now, for the three categorical variables, we define a $6 \times 6 \times 6$ cross-table with the corresponding residuals in each of the 216 cells. Then we find the MAD (median absolute deviation) of the residuals, denoted by MAD*, in each cell. Any cell with frequency less than or equal to 1 is ignored. Then assuming that σ_t is fixed for the residuals in each cell, we consider estimating σ_t as $(\widehat{MAD^*})_t$ from the model

$$MAD^* = \alpha_0 + \alpha_1 x_1^* + \alpha_2 x_2^* + \alpha_3 x_3^* + \alpha_4 (x_1^*)(x_2^*) + \alpha_5 (x_1^*)(x_3^*) + \text{error} \qquad (4.42)$$

where $x_1^* = \text{TEMP}^*$, $x_2^* = \text{MWS}^*$ and $x_3^* = \text{AWS}^*$. Note that in model (4.42), we include the interaction effects of TEMP* with MWS* and AWS* as these two effects were found to be very significant. Also note that the cell frequencies in each of the 216 cells are naturally different from each other. We downweight any point with

cell frequency less than 20 since the MAD of a small number of observations is less reliable. A weight function w^* is defined in the form

$$w_i^* = \begin{cases} 1 & \text{if } i\text{th cell frequency} \geq 20 \\ \left(\frac{i\text{th cell frequency}}{20}\right)^2 & \text{otherwise.} \end{cases} \tag{4.43}$$

Using these weight functions, we estimate the parameters $\alpha = (\alpha_0, \ldots, \alpha_5)^T$ of model (4.42) as the solution to a GM estimating equation of the form

$$\sum_i \psi\left(\frac{r_i^*(\alpha)}{\hat{\sigma}^*}\right) w_i^* x_i^* = 0 \tag{4.44}$$

where ψ is the Huber's ψ function defined earlier, x_i^* the i-th element of the vector of covariates TEMP*, MWS* and AWS*; $\hat{\sigma}^*$ is the MAD of the residuals $r_i^*(\alpha)$ from (4.42) and w_i^* is defined by (4.43). Thus σ_t's are estimated as the fitted values $\widehat{(MAD^*)}_t$ from model (4.42). Based on these estimated $\hat{\sigma}_t$'s, we fit our original model (4.38) for the estimates of the regression parameters β and γ, and the autoregressive parameter ϕ for an $AR(1)$ process using the GM estimation technique discussed in section 4.2. Based on the current residuals $(y_t - \hat{\mu}_t)$, we again estimate σ_t using the same technique as considered earlier. We continue the two steps until a convergence criterion is met. Note that the fitting of model (4.42) is found to be very significant as approximately 45% of the variation in the MAD*'s can be explained by this model. The GM estimate of $\alpha = (\alpha_0, \ldots, \alpha_5)^T$ is obtained as $\hat{\alpha} = (1.359, 3.508, 0.881, 0.770, -0.402, -0.367)^T$ with corresponding standard error $s.e.(\hat{\alpha}) = (1.127, 0.415, 0.240, 0.269, 0.076, 0.101)^T$. In Figure 4.4, we plot the standardized residuals $(y_t - \hat{\mu}_t)/\hat{\sigma}_t$ against the meteorological variables TEMP, MWS and AWS. Here a significant change in the pattern of the residuals is observed - the residuals have increased scatter throughout the plots as compared to Figure 4.3.

In the next step, we choose an appropriate order q for the $AR(q)$ process of the sequence $\{\epsilon_t\}$. Using the above procedure, we fit model (4.38) considering that the process $\{\epsilon_t\}$ follows an $AR(q)$ model for $q = 1, 2$ and 3. For the $AR(1)$ model, the corresponding ϕ estimate is obtained as $\hat{\phi} = 0.572$, whereas for $AR(2)$ and $AR(3)$ models
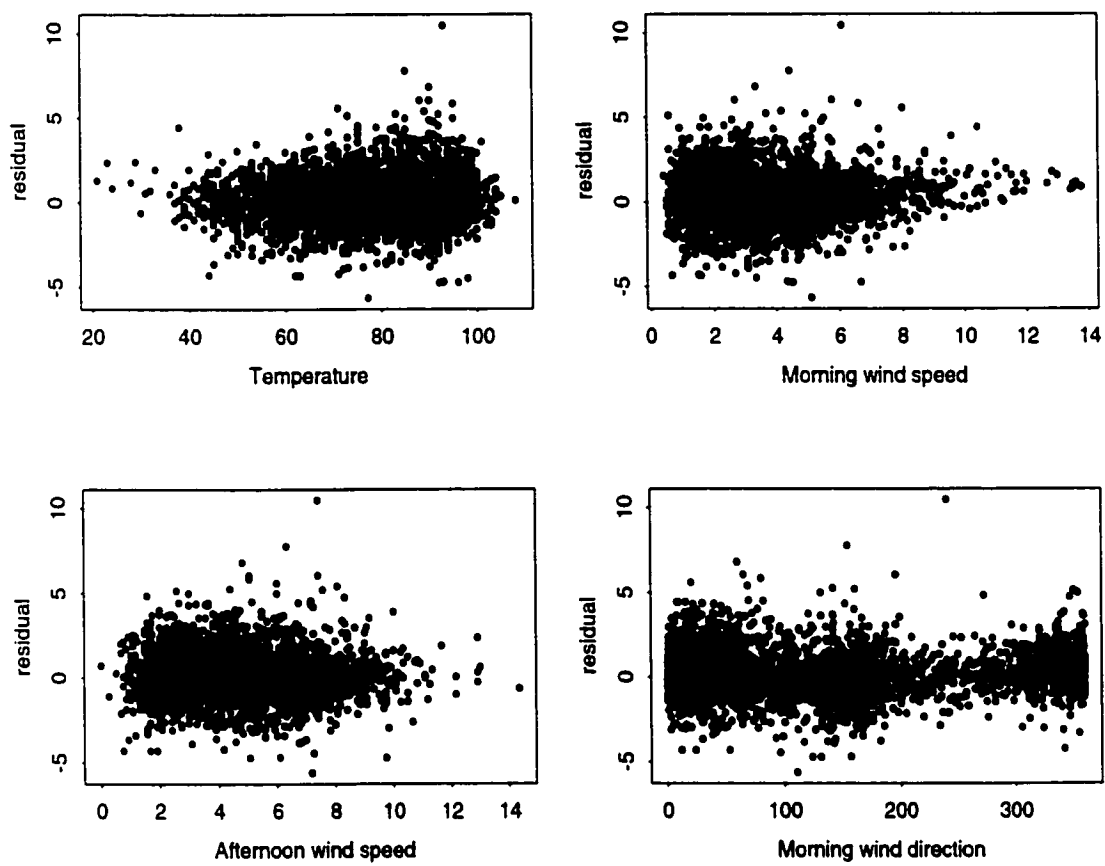
Figure 4.4: Scatterplots of $(y_t - \hat{\mu}_t)/\hat{\sigma}_t$ versus the predictors

the ϕ estimates are obtained as $\hat{\phi} = (0.535, 0.060)^T$ and $\hat{\phi} = (0.530, 0.039, 0.040)^T$, respectively. Then using equation (4.37), we find the corresponding robust Akaiki Information Criterion ($RAIC$) for each of the three autoregressive models in order to choose the correct model. The $RAIC$'s are obtained as $RAIC(1) = 0.6795$, $RAIC(2) = 0.6877$ and $RAIC(3) = 0.6885$. As $RAIC$ is minimum at $q = 1$, we choose the $AR(1)$ model for the process $\{\epsilon_t\}$.

Thus assuming that the error $\{\epsilon_t\}$ follows an $AR(1)$ process, the GM estimates of the regression parameters and the autoregressive parameter (with standard error in parenthesis) are obtained as in Table 4.1.

Table 4.1: Estimates of regression parameter and autoregressive parameter (with standard error in parenthesis) for an $AR(1)$ process of $\{\epsilon_t\}$

Parameter	GM estimate	LMS estimate	One-step GM estimate	LS estimate
β_0	2.96753 (.096809)	3.33369	2.92015 (.094216)	3.08943 (.089046)
β_1 (TEMP)	0.01166 (.001105)	0.00922	0.01268 (.001082)	0.01045 (.001015)
β_2 (MWS)	-0.02825 (.003397)	-0.04695	-0.03273 (.003688)	-0.01632 (.003067)
β_3 (AWS)	-0.01331 (.003120)	-0.02098	-0.01522 (.003016)	-0.01438 (.002902)
β_4 (MWD)	-0.00021 (.000053)	-0.00007	-0.00020 (.000048)	-0.00011 (.000047)
β_5 (AWD)	-0.00007 (.000061)	-0.00030	-0.00011 (.000058)	-0.00010 (.000055)
γ_1 (trend)	-0.06675 (.160930)	0.06790	0.06510 (.194627)	-0.10659 (.154025)
γ_2 (sin1)	3.38718 (.722034)	2.69173	3.19388 (.857380)	2.40193 (.716698)
γ_3 (cos1)	-0.14489 (.817089)	-2.07679	0.10743 (.928384)	-0.90521 (.867037)
γ_4 (sin2)	-1.31247 (.627188)	-0.52545	-0.85355 (.759221)	-0.92436 (.621174)
γ_5 (cos2)	-5.76445 (.622798)	-5.09578	-5.74749 (.766196)	-5.35951 (.622949)
ϕ_1	0.57157 (.016534)	0.63666	0.57594 (.015956)	0.53596 (.016518)

From the GM fit, we see that the parameter estimates corresponding to the meteorological variables TEMP, MWS and AWS are highly significant. Temperature has a significant positive association with the ozone level concentrations. On the other hand, both morning and afternoon wind speed have significant negative associations with the ozone levels. In fact, high temperature and lower wind speed are associated

with high ozone level concentrations. The variables MWD and AWD are negatively associated with the ozone concentrations. However, the association appears to be insignificant for AWD. The trend coefficient γ_1 also shows a negative effect on ozone levels, but the effect is found to be insignificant. The estimates of the seasonal parameters γ_2 and γ_5 are found to be highly significant. In fact, the variability of ozone concentrations is highly seasonal, with high variation in the summer and low variation in the other seasons. Also the autoregressive parameter ϕ_1 has a highly significant estimate with the value of .57, which justifies the use of autoregressive errors in modeling the ozone concentrations.

To have some idea about the potential influential observations, we consider finding the weights of the GM fit. Note that the GM estimating equations (4.13) and (4.14) can be expressed as $\sum_{t=q+1}^{n} v_t^*(1/\sigma_t)((y_t - \mu_t)/\sigma_t)(\partial\mu_t/\partial\theta) = 0$, where $v_t^* = w_t\psi((y_t-\mu_t)/\sigma_t)/((y_t-\mu_t)/\sigma_t)$ is the weight function at point t ($t = q+1,\ldots,n$). Also note that for the choice $\psi(x) = x$ and $w_t = 1$, we have $v_t^* = 1$, and the GM estimates become the classical ML estimates. In Figure 4.5, we plot the weights v_t^* of the GM estimates. From the plot, it is observed that some of the points are heavily downweighted with weights less than .4. However, a considerable number of points have larger weights with values greater than .6. The proportion of points having weights less than .6 is approximately .2. Note that the weight function w_t defined in (4.12) is based on the common assumption that the covariates follow a multivariate normal distribution. In our case, some of the covariates of the meteorological variables appear to have nonnormal structure. For possible improvement of the results, we analyzed the ozone data set based on weight functions for transformed covariates (for example, square root and log transformations). However, the results were found to be very similar to that obtained from the weight functions for the original covariates. So we choose the weight functions for the original covariates in the analysis of the ozone data.

We also study the one-step GM estimates of the regression parameters and the autoregressive parameters starting from a set of high breakdown point initial estimates such as the least median squares (LMS) estimates of Rousseeuw (1984). Simpson,
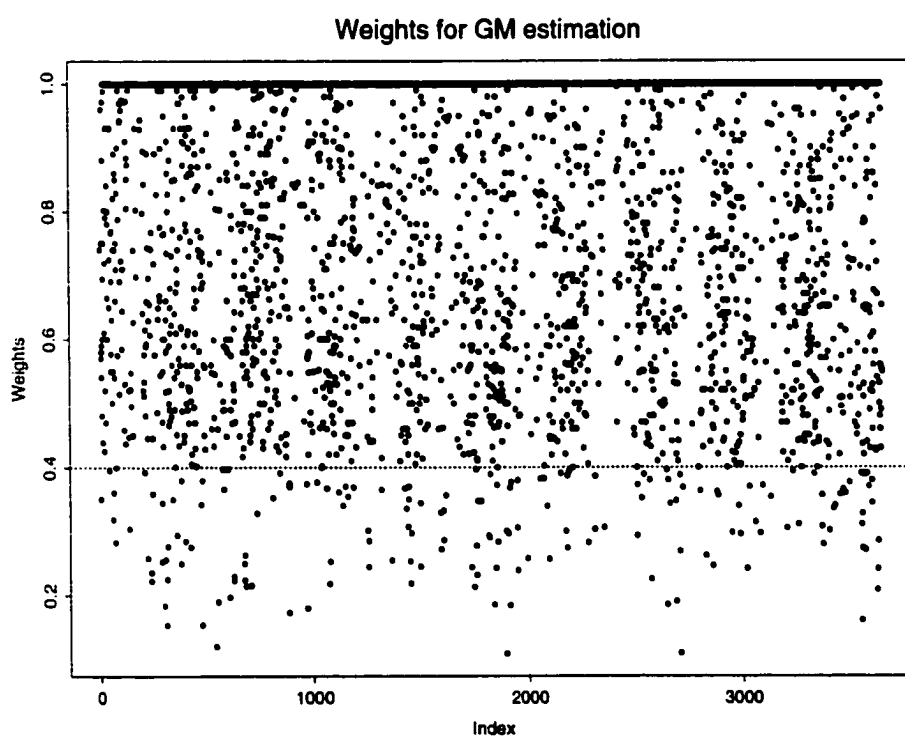
Figure 4.5: Weights for the *GM* estimation

Ruppert, and Carroll (1992) proposed the one-step GM estimation in a linear regression setting. We can extend this one-step GM estimation from linear to nonlinear regression. Suppose $\theta_0^T = (\beta_0^T, \phi_0^T)$ and σ_{t0} are the LMS estimates of $\theta^T = (\beta^T, \phi^T)$ and σ_t, respectively. Then the one-step GM estimates of β and ϕ can be obtained from equations (4.16) and (4.17) as $\beta_1 = \beta_0 + M_{10}^{-1} q_{10}$ and $\phi_1 = \phi_0 + M_{20}^{-1} q_{20}$, respectively. The initial LMS estimates are obtained by minimizing the median of the squares of the residuals. Here we adopt an algorithm which uses subsamples of size $[n/2]$ from a sample of size n. For each subsample, we find the LS estimates of the corresponding parameters. We use half of the data points in the subsamples based on the fact that the breakdown point of the LMS estimates is almost $1/2$ in most cases. When n is large, it is computationally expensive to use all possible subsamples of size $[n/2]$. So in our case, we choose $N = 2000$ subsamples in order to find the approximate LMS estimates. Suppose, $\hat{\theta}_1, \dots , \hat{\theta}_N$ are the corresponding LS estimates of θ for the N subsamples. For each of these $\hat{\theta}$'s, we find the corresponding median of the squared residuals for the original data. The LMS estimate is chosen as the $\hat{\theta}$ which minimizes the median of the squared residuals. Using similar technique, we find the LMS estimate σ_{t0} of σ_t from model (4.42). The LMS estimate and the corresponding one-step GM estimate of θ are displayed in Table 4.1. The one-step GM estimates are found to be very close to the fully iterated GM estimates except for the estimates of trend parameter γ_1 and the seasonal parameter γ_3. However, these parameters appear to be insignificant in both estimation procedures.

Finally, we compare the robust estimates with the corresponding classical LS estimates. Note that for the choice $\psi(x) = x$, $w(x) = 1$ and $MAD = $ standard deviation, the GM estimates of $\theta^T = (\beta^T, \phi^T)$ and σ_t become the classical LS estimates. The LS estimates of θ (with standard error in parenthesis) are shown in Table 4.1. The parameter estimates corresponding to the meteorological variables TEMP and MWS are found to be different in the LS method as compared to the GM and one-step GM methods. Also the estimates of the seasonal components γ_2, γ_3 and γ_4 are found to be slightly different in the LS method. As we observe some influential observations in the data from the weight functions in Figure 4.5, the LS estimates are naturally affected by those points and thus result in different values as compared to the GM

and one-step GM estimates. Ideally, robust fit is better than the classical LS fit as the robust procedure is based on downweighting the influential observations.

4.7 Discussion

In this chapter, we study the GM estimates of the regression parameters in a nonlinear regression setting with autoregressive errors. The asymptotic behavior of the estimates has been discussed in some detail which can accommodate a dependence structure among the residuals. To derive the asymptotic properties of the robust estimates, we use a uniform law of large numbers and a central limit theorem valid for time dependent data. The key concept employed is uniform integrability to ensure the uniform convergence of a sequence. We also impose some mixing conditions, which restrict the memory of a process in a fashion analogous to the role of ergodicity for a stationary stochastic process. We show that under some regularity conditions, the GM estimators are consistent and follow the asymptotic normal distribution with certain mean vectors and covariance matrices. We also study the infinitesimal behavior of the GM estimates based on the time series analogue of Hampel's influence function. We show that the variance function of a GM estimator can be expressed in terms of the corresponding influence function.

In the application section of this chapter, we develop a technique to model the scale parameter σ_t as a function of the covariates. This modeling of σ_t is found to be effective in accommodating the heteroscedasticity of the errors. To estimate σ_t, Niu (1996) proposed a model where σ_t is exponentially related to the covariates. Initially, we attempted to estimate σ_t using the model proposed by Niu (1996). However, this model was found to be ineffective in explaining the heteroscedasticity of the residuals. Later, we develop a new approach which was found to be more effective than the approach suggested by Niu (1996).

We apply the proposed GM estimation technique to the analysis of ground-level ozone concentration. Ozone, an important indicator of air pollution, is an outcome

from the emission of nonmethane hydrocarbons and nitrogen oxides into the atmosphere. Furthermore, ozone levels are also strongly affected by the variations in meteorological conditions. As expected, we found from the analysis of ozone data that ozone concentrations are positively related to daily maximum surface temperature but negatively related to wind speed. Also the variability of ozone levels is strongly seasonal, with high variation in the summer and low variation in the other seasons.

Recently, Ferri, Kelmansky, Yohai and Zamar (1999) introduce a new class of robust estimates in a linear regression setting. These estimates, termed as generalized τ (G_τ) estimates, are defined by minimizing the τ scale of the weighted residuals, where the weights penalize the influential observations. We can extend the G_τ estimates from linear to nonlinear regression. Suppose, the random sample $(y_t, \mathbf{x}_t)$ $(t = 1, \ldots, n)$ follow the nonlinear regression model $y_t = h(\mathbf{x}_t; \boldsymbol{\theta}) + u_t$ where u_t's are i.i.d. random variables, $\mathbf{x}_t$'s are p-dimensional vectors of covariates and h is a deterministic model which is continuous in $\boldsymbol{\theta}$ for each $\mathbf{x}_t$. Then for some weight function $w_t = w(\mathbf{x}_t)$, we can define the weighted residuals in the form $r_t(\boldsymbol{\theta}) = u_t(\boldsymbol{\theta})w_t$ $(t = 1, \ldots, n)$, where $u_t(\boldsymbol{\theta}) = y_t - h(\mathbf{x}_t; \boldsymbol{\theta})$. The weighted M scale of the residuals, $S_n(\boldsymbol{\theta})$, is defined by

$$\frac{1}{n} \sum_{t=1}^{n} \rho_1^*(r_t(\boldsymbol{\theta})/S_n(\boldsymbol{\theta})) = d \tag{4.45}$$

for some function ρ_1^* and a constant d, where $d < \sup_x \rho_1^*(x)$, and the loss function ρ_1^* is even, bounded, nondecreasing on $[0, \infty)$. The weighted τ scale, $\tau_n(\boldsymbol{\theta})$, is then defined by

$$\tau_n^\kappa(\boldsymbol{\theta}) = S_n^\kappa(\boldsymbol{\theta})\frac{1}{n} \sum_{t=1}^{n} \rho_2^*(r_t(\boldsymbol{\theta})/S_n(\boldsymbol{\theta})) \tag{4.46}$$

where $S_n(\boldsymbol{\theta})$ is defined by equation (4.45) and ρ_2^* is even, bounded and nondecreasing on $[0, \infty)$ and satisfies $0 < \lim_{u \to 0} \rho_2^*(u)/u^\kappa$ for some $\kappa > 0$. Finally, the G_τ estimate is defined by minimizing the weighted τ scale, $\hat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}} \tau_n(\boldsymbol{\theta})$. Note that for the choice $w_1 = \ldots = w_n = 1$, the G_τ estimate reduces to the τ estimate introduced by Yohai and Zamar (1988).

The G_τ estimates are based on the strict assumption that the errors u_t are i.i.d. $N(0,\sigma^2)$ for $t = 1,\ldots,n$. However, we are dealing with a situation where the errors are heteroscedastic, that is, we assume that the u_t are distributed as $N(0,\sigma_t^2)$ (see model (4.3)). In this case, we can modify the G_τ estimates for known σ_t. If σ_t are known, then we can replace $r_t(\boldsymbol{\theta})$ by the standardized weighted residuals $r_t^s(\boldsymbol{\theta}) = (u_t(\boldsymbol{\theta})/\sigma_t)w_t$ and then minimize $\tau_n(\boldsymbol{\theta})$ with respect to θ based on the $r_t^s(\boldsymbol{\theta})$'s. In practice, we need to estimate σ_t by using a suitable method. Estimation of σ_t needs to be investigated further. A similar technique as used in the analysis of the ozone data may be adopted to find the G_τ estimates of σ_t. As a future study, it would be interesting to see how the G_τ estimates behave as compared to the GM estimates in a nonlinear regression setting.

Chapter 5

Time Series Influence Function in the Frequency Domain

5.1 Introduction

In chapters 2 and 4, we discussed Hampel's influence function in order to investigate the infinitesimal behavior of a real-valued functional $T(F)$, where F represents the underlying distribution function. Hampel's influence function is considered as a tool of studying the local stability of an estimator in terms of the effect of point-mass contamination of the data or the underlying distribution. Hastings (personal communication) introduced a new class of influence functions which are referred to as frequency influence functions of a functional $g(F)$, where F is now considered as the cumulative (non-normalized) power spectrum of a time series process $\{x_t\}$. Frequency influence function is developed to investigate the behavior of the functional $g(F)$ in terms of the point-mass perturbation of the cumulative spectrum F. Throughout this chapter we will assume that F represents the (non-normalized) cumulative power spectrum and its derivative f represents the corresponding power spectrum of a stationary process $\{x_t\}$. In section 5.2, we give a formal definition of the frequency influence function. Like Hampel's influence function, it is illustrated with some examples that show that the asymptotic variance of an estimator can be obtained as a function of its frequency influence function. In section 5.3, we discuss some aspects of this new idea of time series influence functions.

5.2 Hastings' Frequency Influence Function

5.2.1 Power Spectrum of a Stationary Process

Let $\{x_t\}$ be a zero-mean stationary stochastic process. Then the (non-normalized) power spectrum of the process $\{x_t\}$ at frequency ω, denoted by $f(\omega)$, is defined as

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\tau\omega} R(\tau) d\tau \tag{5.1}$$

where $R(\tau)$ is the autocovariance function defined by $R(\tau) = E(x_t x_{t-\tau})$. In fact, $f(\omega)d\omega$ can be considered as the average (over all realizations) of the contribution to the total power from components in $\{x_t\}$ with frequencies between ω and $\omega + d\omega$.

The integrated (or cumulative) power spectrum, $F(\omega)$, defined by

$$F(\omega) = \int_{-\infty}^{\omega} f(\lambda) d\lambda \tag{5.2}$$

is the average contribution to the total power from all components with frequencies less than or equal to ω.

Note that if $\{x_t\}$ is observed at a discrete set of time points, say at $t = 0, \pm 1, \pm 2, \ldots$, then the power spectrum $f(\omega)$ in (5.1) can be expressed as

$$f(\omega) = \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} R(l) e^{-il\omega}, \quad -\pi \le \omega \le \pi \tag{5.3}$$

where $R(l) = E(x_t x_{t+l})$ (with $E(x_t) = 0$) is the autocovariance function of $\{x_t\}$. This function $R(l)$ can be reexpressed as

$$R(l) = \int_{-\pi}^{\pi} e^{il\omega} f(\omega) d\omega, \quad l = 0, \pm 1, \pm 2, \ldots \tag{5.4}$$

5.2.2 Frequency Influence Function

In chapter 2, we pointed out that many oceanographic measurements show the influence of discrete spectrum harmonic or tidal components and due to their relatively large size, these can obscure other signals (with continuous spectrum) in the data. Frequency influence function is developed to investigate the infinitesimal behavior of a continuous spectrum time series functional in terms of the point-mass contamination of the underlying spectral distribution. Suppose, an estimate of a time series functional $g = g(F)$ defined on a continuous spectral distribution F can be obtained by evaluating g at an estimate of F. Let $F(\omega) + \epsilon F_{\tilde{\omega}}(\omega)$ be a contamination spectral distribution, where $F_{\tilde{\omega}}(\omega)$ represents the discrete spectrum of a harmonic process, which has point-mass $1/2$ at frequencies $\pm\tilde{\omega}$, that is,

$$F_{\tilde{\omega}}(\omega) = \begin{cases} 0, & -\pi < \omega < -\tilde{\omega} \\ \frac{1}{2}, & -\tilde{\omega} \leq \omega < \tilde{\omega} \\ 1, & \tilde{\omega} \leq \omega \leq \pi. \end{cases} \tag{5.5}$$

Then the frequency influence function is defined as the directional derivative of the functional g at $\epsilon = 0$:

$$D_F(g(F) \mid \tilde{\omega}) = \lim_{\epsilon \to 0} \frac{g(F + \epsilon F_{\tilde{\omega}}) - g(F)}{\epsilon} \tag{5.6}$$

provided the limit exists. We can argue that this frequency influence function measures the effect of an infinitesimal contamination of the underlying spectral distribution F at frequencies $\pm\tilde{\omega}$ on the estimate of the time series functional $g(F)$, standardized by the mass of the contamination. Based on this frequency influence function, we study some properties of a number of time series functionals which are illustrated in the following examples.

Example 5.2.1 Consider a time series functional of the form

$$g(F) = \int_{-\pi}^{\pi} a(\lambda) dF(\lambda) \tag{5.7}$$

where $a(\lambda)$ is a function of λ. Then we have

$$
\begin{aligned}
g(F + \epsilon F_{\tilde{\omega}}) &= \int_{-\pi}^{\pi} a(\lambda) d(F + \epsilon F_{\tilde{\omega}})(\lambda) \\
&= g(F) + \epsilon \int_{-\pi}^{\pi} a(\lambda) dF_{\tilde{\omega}}(\lambda) \\
&= g(F) + \epsilon(1/2) \left(a(-\tilde{\omega}) + a(\tilde{\omega}) \right)
\end{aligned}
$$

since $dF_{\tilde{\omega}}(\lambda)$ is $1/2$ at $\pm\tilde{\omega}$ and 0 elsewhere. Thus from (5.6), the frequency influence function of $g(F)$ is obtained as

$$
D_F(g(F) \mid \tilde{\omega}) = \frac{1}{2} \left\{ a(-\tilde{\omega}) + a(\tilde{\omega}) \right\} . \tag{5.8}
$$

It is clear that $D_F(g(F) \mid \tilde{\omega}) = D_F(g(F) \mid -\tilde{\omega})$. Now, if
Case I. a is real and symmetric, then

$$
D_F(g(F) \mid \tilde{\omega}) = a(\tilde{\omega})
$$

Case II. $a(-\omega) = a^*(\omega)$, where $a^*(\omega)$ is the conjugate of $a(\omega)$, then

$$
D_F(g(F) \mid \tilde{\omega}) = \mathbb{R}[a(\tilde{\omega})]
$$

where $\mathbb{R}[a(\omega)]$ represents the real part of $a(\omega)$.

Note that an estimate of the functional $g(F)$ in (5.7) may be obtained by evaluating g at an estimate of F as

$$
\begin{aligned}
g_n(a) &= \int_{-\pi}^{\pi} a(\lambda) I_n(\lambda) d\lambda \\
&= \int_{-\pi}^{\pi} \frac{1}{2} \left\{ a(-\lambda) + a(\lambda) \right\} I_n(\lambda) d\lambda \\
&= \int_{-\pi}^{\pi} D_F(g(F) \mid \lambda) I_n(\lambda) d\lambda
\end{aligned} \tag{5.9}
$$

where $I_n(\lambda)$ is the periodogram of a time series process at frequency λ. The variance

of the estimate $g_n(a)$ can be obtained as

$$\text{var}\,(g_n(a)) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} D_F(g(F) \mid \lambda) D_F^*(g(F) \mid \omega) \text{cov}\,(I_n(\lambda), I_n(\omega))\, d\lambda d\omega \qquad (5.10)$$

where $D_F^*(g(F) \mid \lambda)$ is the conjugate of $D_F(g(F) \mid \lambda)$. Under some regularity conditions, it can be shown for a Gaussian process that

$$\text{var}\,(g_n(a)) \approx \frac{4\pi}{n} \int_{-\pi}^{\pi} D_F(g(F) \mid \lambda) D_F^*(g(F) \mid \lambda) f^2(\lambda) d\lambda \qquad (5.11)$$

(see Brillinger, 1981 p.168 for details). Now, if $D_F(g(F) \mid \lambda)$ is real, the approximate variance of $g_n(a)$ can be obtained as

$$V\,(g_n(a)) \quad \approx \quad \frac{4\pi}{n} \int_{-\pi}^{\pi} D_F^2(g(F) \mid \lambda) f^2(\lambda) d\lambda \qquad (5.12)$$

which indicates a close relationship between the frequency influence function and the variance of a time series functional.

Note that if we choose $a(\lambda) = e^{il\lambda}$, then equation (5.7) gives the autocovariance function

$$R(l) = \int_{-\pi}^{\pi} e^{il\lambda} dF(\lambda) \qquad (5.13)$$

The corresponding frequency influence function is obtained from equation (5.8) as

$$D_F(R(l) \mid \tilde{\omega}) = \cos l\tilde{\omega} \qquad (5.14)$$

So a point-mass perturbation of the cumulative spectrum F causes a sinusoidal perturbation of the covariance function $g(F) = R(l)$.

The estimate of $R(l)$ can be obtained from equation (5.9) as

$$\hat{R}(l) = \frac{1}{n} \sum_{t=1}^{n-|l|} x_t x_{t+|l|} \qquad (5.15)$$

(see Brillinger, 1981). If $\{x_t\}$ follows a Gaussian process, the asymptotic variance of $\hat{R}(l)$ can be obtained from equation (5.12) as

$$
\begin{aligned}
V(\hat{R}(l)) &\approx \frac{4\pi}{n} \int_{-\pi}^{\pi} D_F^2(R(l) \mid \lambda) f^2(\lambda) d\lambda \\
&= \frac{4\pi}{n} \int_{-\pi}^{\pi} \cos^2 l\lambda f^2(\lambda) d\lambda
\end{aligned}
\tag{5.16}
$$

Example 5.2.2 Recall the (non-normalized) power spectrum $f(\omega) = (1/2\pi) \times \sum_{l=-\infty}^{\infty} R(l)e^{-il\omega}$, $-\pi \leq \omega \leq \pi$. For a stationary process $\{x_t\}$ $(t = 1, \ldots, n)$, $f(\omega)$ can be estimated by

$$
\hat{f}_0(\omega) = \frac{1}{2\pi} \sum_{s=-m}^{m} \hat{R}(s) e^{-is\omega}
\tag{5.17}
$$

where m is an integer less than $(n-1)$. It can be shown that for a suitable choice of m (that is, make $m \to \infty$ as $n \to \infty$, but sufficiently slowly (relative to n) so that $(m/n) \to 0$ as $n \to \infty$), $\hat{f}_0(\omega)$ is a consistent estimate of the power spectrum $f(\omega)$ (see Priestley, 1981). The estimate $\hat{f}_0(\omega)$ can be regarded as a special case of the more general form of estimate

$$
\hat{f}(\omega) = \frac{1}{2\pi} \sum_{s=-(n-1)}^{(n-1)} \gamma(s) \hat{R}(s) e^{-is\omega}
\tag{5.18}
$$

where to recover $\hat{f}_0(\omega)$, $\gamma(s)$ can be defined as

$$
\gamma(s) = \begin{cases} 1, & \text{for } |s| \leq m \\ 0, & \text{otherwise.} \end{cases}
\tag{5.19}
$$

Note that we can also consider other estimates of $f(\omega)$ by choosing $\gamma(s)$ in such a way that it decreases gradually rather than having the discontinuous form (5.19). In fact, there are many different forms of $\gamma(s)$ that can be used, which lead to consistent estimates of $f(\omega)$ if the function $\gamma(s)$ decreases at the appropriate rate. $\gamma(s)$ is usually considered to be a real even function of s. The estimate $\hat{f}(\omega)$ can be rewritten as a

weighted integral of the periodogram $I_n(\omega)$, where $I_n(\omega)$ can be expressed as

$$I_n(\omega) = \frac{1}{2\pi} \sum_{s=-(n-1)}^{(n-1)} \hat{R}(s)e^{-is\omega} \tag{5.20}$$

which gives

$$\hat{R}(s) = \int_{-\pi}^{\pi} I_n(\lambda)e^{is\lambda}d\lambda \tag{5.21}$$

for $|s| \leq (n-1)$. Thus from (5.18), we now have

$$\hat{f}(\omega) = \int_{-\pi}^{\pi} I_n(\lambda) \left\{ \frac{1}{2\pi} \sum_{s=-(n-1)}^{(n-1)} \gamma(s)e^{-is(\omega-\lambda)} \right\} d\lambda$$

which may be reexpressed as

$$\hat{f}(\omega) = \int_{-\pi}^{\pi} W_n(\omega - \lambda)I_n(\lambda)d\lambda \tag{5.22}$$

where the weight function

$$W_n(\theta) = \frac{1}{2\pi} \sum_{s=-(n-1)}^{(n-1)} \gamma(s)e^{-is\theta} \tag{5.23}$$

is the (discrete) Fourier transform of the sequence $\{\gamma(s)\}$. For most of the commonly used sequences $\{\gamma(s)\}$, the function $W_n(\theta)$ typically is concentrated near $\theta = 0$. Note that if $\gamma(s)$ is considered as a real even function of s, then $W_n(\theta)$ is a real valued even function of θ. The estimate $\hat{f}(\omega)$ in (5.22) can be considered as a locally weighted average of periodograms in the neighborhood of the frequency ω. A number of choices of $\gamma(s)$ are available (see Priestley, 1981, p.437). It can be shown that

$$E[\hat{f}(\omega)] \approx \int_{-\pi}^{\pi} W_n(\omega - \lambda)f(\lambda)d\lambda. \tag{5.24}$$

Now, to find the frequency domain influence function of $\hat{f}(\omega)$, let us consider the

functional $g_{n,\omega}(F) = \int_{-\pi}^{\pi} W_n(\omega - \lambda)dF(\lambda)$. Then we have

$$
\begin{aligned}
g_{n,\omega}(F + \epsilon F_{\tilde{\omega}}) &= \int_{-\pi}^{\pi} W_n(\omega - \lambda)d(F + \epsilon F_{\tilde{\omega}})(\lambda) \\
&= g_{n,\omega}(F) + \epsilon \int_{-\pi}^{\pi} W_n(\omega - \lambda)dF_{\tilde{\omega}}(\lambda) \\
&= g_{n,\omega}(F) + \epsilon(1/2)(W_n(\omega - \tilde{\omega}) + W_n(\omega + \tilde{\omega}))
\end{aligned}
$$

Thus the frequency influence function is obtained as

$$
D_F(g_{n,\omega}(F) \mid \tilde{\omega}) = \frac{1}{2}\{W_n(\omega - \tilde{\omega}) + W_n(\omega + \tilde{\omega})\} \tag{5.25}
$$

When we estimate the power spectrum $f(\omega)$ by $\hat{f}(\omega)$, we can think of the frequency influence function $D_F(g_{n,\omega}(F) \mid \tilde{\omega})$ as a generalized spectral window. As the weight function W_n is a real valued even function, using equation (5.11) it is easy to show that for a Gaussian process, the asymptotic variance of the spectrum estimate $\hat{f}(\omega)$ can be expressed as a function of its frequency influence function as $V(\hat{f}(\omega)) \approx (4\pi/n) \int_{-\pi}^{\pi} D_F^2(g_{n,\omega}(F) \mid \lambda)f^2(\lambda)d\lambda$. This variance function can be simplified as

$$
\begin{aligned}
V(\hat{f}(\omega)) &\approx \frac{\pi}{n} \int_{-\pi}^{\pi} W_n^2(\omega - \lambda)f^2(\lambda)d\lambda + \frac{\pi}{n} \int_{-\pi}^{\pi} W_n^2(\omega + \lambda)f^2(\lambda)d\lambda \\
&\quad + \frac{2\pi}{n} \int_{-\pi}^{\pi} W_n(\omega - \lambda)W_n(\omega + \lambda)f^2(\lambda)d\lambda \\
&= \frac{2\pi}{n} \int_{-\pi}^{\pi} W_n^2(\omega - \lambda)f^2(\lambda)d\lambda + \frac{2\pi}{n} \int_{-\pi}^{\pi} W_n(\omega - \lambda)W_n(\omega + \lambda)f^2(\lambda)d\lambda
\end{aligned}
$$

Since $W_n(\theta) \to \delta(\theta)$ as $n \to \infty$, the term $\int_{-\pi}^{\pi} W_n(\omega - \lambda)W_n(\omega + \lambda)f^2(\lambda)d\lambda \to 0$ as $n \to \infty$ unless $\omega = 0$ or $\pm \pi$. Therefore, we may write $V(\hat{f}(\omega)) \approx (1 + \delta_{\omega,0,\pi})(2\pi/n) \int_{-\pi}^{\pi} W_n^2(\omega - \lambda)f^2(\lambda)d\lambda$, where $\delta_{\omega,0,\pi}$ is 1 at $\omega = 0$ or $\pm\pi$ and 0 elsewhere.

Note that the frequency domain influence function is directly related to the influence function of a functional in the time domain. To show this, let us recall the

general form of a time series functional $g(F)$ defined in equation (5.7) as

$$
\begin{aligned}
g(F) &= \int_{-\pi}^{\pi} a(\lambda) f(\lambda) d\lambda \\
&= \int_{-\pi}^{\pi} a(\lambda) \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\tau\lambda} R(\tau) d\tau \right\} d\lambda \\
&= \int_{-\pi}^{\pi} a(\lambda) \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\tau\lambda} E(x_t x_{t+\tau}) d\tau \right\} d\lambda
\end{aligned}
$$

Suppose, the time series process $\{x_t\}$ is contaminated by a harmonic process as $x_t' = x_t + A\cos(\tilde{\omega}t + \tilde{\varphi})$, where $\tilde{\varphi}$ is distributed as uniform$(-\pi, \pi)$ and is independent of $\{x_t\}$. Then we have

$$
E(x_t' x_{t+\tau}') = E(x_t x_{t+\tau} + A^2 \cos(\tilde{\omega}t + \tilde{\varphi}) \cos(\tilde{\omega}(t+\tau) + \tilde{\varphi})),
$$

the cross-product terms being zero, since $\{x_t\}$ and $\tilde{\varphi}$ are independent. Thus for the contaminated process $\{x_t'\}$, the above functional $g(F)$ becomes

$$
\begin{aligned}
g'(F) &= \int_{-\pi}^{\pi} a(\lambda) \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\tau\lambda} E(x_t' x_{t+\tau}') d\tau \right\} d\lambda \\
&= \int_{-\pi}^{\pi} a(\lambda) \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\tau\lambda} E(x_t x_{t+\tau}) d\tau \right\} d\lambda \\
&\quad + \int_{-\pi}^{\pi} a(\lambda) \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\tau\lambda} E(A^2 \cos(\tilde{\omega}t + \tilde{\varphi}) \cos(\tilde{\omega}(t+\tau) + \tilde{\varphi})) d\tau \right\} d\lambda \\
&= g(F) + \frac{A^2}{2} \int_{-\pi}^{\pi} a(\lambda) \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\tau\lambda} \cos\tilde{\omega}\tau d\tau \right\} d\lambda \\
&= g(F) + \frac{A^2}{2} \int_{-\pi}^{\pi} a(\lambda) \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \cos\tau\lambda \cos\tilde{\omega}\tau d\tau \right\} d\lambda \\
&= g(F) + \frac{A^2}{2} \int_{-\pi}^{\pi} a(\lambda) \frac{1}{2} [\delta(\lambda + \tilde{\omega}) + \delta(\lambda - \tilde{\omega})] d\lambda \\
&= g(F) + \frac{A^2}{2} \times \frac{1}{2} [a(-\tilde{\omega}) + a(\tilde{\omega})]
\end{aligned}
$$

Now, assuming $A^2/2 = \epsilon$ and letting $\epsilon \to 0$, the time domain influence function of

the functional $g(F)$ can be obtained as

$$IF(g(F)) = \lim_{\epsilon \to 0} \frac{g'(F) - g(F)}{\epsilon}$$

$$= \frac{1}{2} \{a(-\bar{\omega}) + a(\bar{\omega})\}$$

which coincides with the frequency influence function $D_F(g(F)|\bar{\omega})$ defined in (5.8). So the frequency influence function of a time series functional can be obtained by contaminating the original time series process or the underlying spectral distribution.

In the following example, we investigate the behavior of a time series functional for the case when the data follow an $AR(1)$ process.

Example 5.2.3 Let $\{x_t\}$ be a stationary stochastic process which follows the $AR(1)$ model: $x_t = \phi_1 x_{t-1} + u_t$, where u_t is a purely random process with mean 0 and variance σ^2. For this $AR(1)$ model, the (non-normalized) power spectrum can be defined as $f(\omega) = (\sigma^2/2\pi) \times (1/(1 - 2\phi_1 \cos\omega + \phi_1^2))$. The autoregressive parameter ϕ_1 is the autocorrelation function: $\phi_1 = R(1)/R(0)$, where $R(0) = \sigma_x^2$, the variance of x_t and $R(1) = E(x_t x_{t-1})$. From Example 5.2.1, we can find the frequency influence function of $R(0)$ as $D_F(R(0) \mid \bar{\omega}) = 1$. Let us consider the functionals $g_1(F) = \sigma_x^2 = R(0)$ and

$$g_2(F) = \phi_1 = \frac{R(1)}{R(0)} = \frac{\int_{-\pi}^{\pi} e^{-i\omega} dF(\omega)}{\int_{-\pi}^{\pi} dF(\omega)}. \tag{5.26}$$

Then we have

$$g_2(F + \epsilon F_{\bar{\omega}}) = \frac{\int_{-\pi}^{\pi} e^{-i\omega} d(F + \epsilon F_{\bar{\omega}})(\omega)}{\int_{-\pi}^{\pi} d(F + \epsilon F_{\bar{\omega}})(\omega)}$$

$$= \frac{R(1) + \epsilon \cos\bar{\omega}}{R(0) + \epsilon}$$

$$= \frac{R(1)}{R(0)} + \frac{R(1) + \epsilon \cos\bar{\omega}}{R(0) + \epsilon} - \frac{R(1)}{R(0)}$$

$$= g_2(F) + \frac{\epsilon(\cos\bar{\omega} - g_2(F))}{g_1(F) + \epsilon}$$

Thus the frequency influence function of ϕ_1 is obtained as $D_F(\phi_1 \mid \bar{\omega}) = (\cos\bar{\omega} - \phi_1)/\sigma_x^2$.

Now, we shall find the frequency influence function of the power spectrum $f(\omega)$ of an $AR(1)$ model defined above. The power spectrum $f(\omega)$ can be rewritten as

$$f(\omega) = \frac{\sigma_x^2}{2\pi} \frac{1 - \phi_1^2}{1 - 2\phi_1 \cos\omega + \phi_1^2} = g(F) \quad \text{(say)} \tag{5.27}$$

For the functionals $g_1(F) = \sigma_x^2 = R(0)$ and $g_2(F) = \phi_1 = R(1)/R(0)$, we already have

$$
\begin{aligned}
D_F(g_1(F) \mid \bar{\omega}) &= \left. \frac{\partial}{\partial \epsilon} g_1(F + \epsilon F_{\bar{\omega}}) \right|_{\epsilon=0} = 1 \\
D_F(g_2(F) \mid \bar{\omega}) &= \left. \frac{\partial}{\partial \epsilon} g_2(F + \epsilon F_{\bar{\omega}}) \right|_{\epsilon=0} = \frac{\cos\bar{\omega} - \phi_1}{\sigma_x^2}.
\end{aligned}
$$

Using these results, the frequency influence function of $g_2(F) = f(\omega)$ can be obtained as

$$
\begin{aligned}
D_F(g(F) \mid \bar{\omega}) &= \left. \frac{\partial}{\partial \epsilon} g(F + \epsilon F_{\bar{\omega}}) \right|_{\epsilon=0} \\
&= \frac{D_F(g_1(F) \mid \bar{\omega})}{2\pi} \left[\frac{1 - \phi_1^2}{1 - 2\phi_1 \cos\omega + \phi_1^2} \right] \\
&\quad + \frac{\sigma_x^2}{2\pi} \left[\frac{-2\phi_1 D_F(g_2(F) \mid \bar{\omega})}{1 - 2\phi_1 \cos\omega + \phi_1^2} + \frac{(1 - \phi_1^2)2(\cos\omega - \phi_1)D_F(g_2(F) \mid \bar{\omega})}{(1 - 2\phi_1 \cos\omega + \phi_1^2)^2} \right] \\
&= \frac{1}{2\pi(1 - 2\phi_1 \cos\omega + \phi_1^2)^2} \left[(1 - \phi_1^2)(1 - 2\phi_1 \cos\omega + \phi_1^2) \right. \\
&\quad \left. -2\phi_1(\cos\bar{\omega} - \phi_1)(1 - 2\phi_1 \cos\omega + \phi_1^2) + 2(1 - \phi_1^2)(\cos\bar{\omega} - \phi_1)(\cos\omega - \phi_1) \right] \\
&= \frac{(1 - 2\phi_1 \cos\omega + \phi_1^2)(1 - 2\phi_1 \cos\bar{\omega} + \phi_1^2) + 2(1 - \phi_1^2)(\cos\bar{\omega} - \phi_1)(\cos\omega - \phi_1)}{2\pi(1 - 2\phi_1 \cos\omega + \phi_1^2)^2}
\end{aligned}
$$

We have the $AR(1)$ spectral estimate if we replace σ_x^2 by $\hat{R}(0)$ and ϕ_1 by $\hat{R}(1)/\hat{R}(0)$. The spectral window is messy by comparison with that in Example 5.2.2, which is a function of $\omega - \bar{\omega}$.

Note that the asymptotic variance of an estimate of the functional $g(F)$ defined by (5.27) cannot be easily calculated. For such complicated functions, we may rely on the Taylor's series approximation of a time series functional. Consider the functional $g(F)$ and a perturbed cumulative spectrum F_ϵ : $F_\epsilon = F + \epsilon G^*$, where G^* is an arbitrary cumulative power spectrum. Let us define

$$U(\epsilon) = g(F_\epsilon) = g(F + \epsilon G^*) \tag{5.28}$$

Then using 1st order Taylor's series expansion, we have

$$U(1) = U(0) + U'(0) + \text{remainder} \tag{5.29}$$

Equivalently, we have

$$g(F + G^*) = g(F) + \left. \frac{\partial}{\partial \epsilon} g(F_\epsilon) \right|_{\epsilon=0} + \text{remainder} \tag{5.30}$$

Now, the second term on the right-hand side can be simplified under the assumption that the functional g is Gâteaux differentiable. Note that a functional g is said to be Gâteaux differentiable at F if there exists a linear functional L such that, for all $G \in \mathcal{G}$,

$$\lim_{\epsilon \to 0} \frac{g(F + \epsilon(G - F)) - g(F)}{\epsilon} = L(G - F) \tag{5.31}$$

(see Huber, 1981). For a suitable measurable function ψ_F, the Gâteaux derivative L is usually assumed to follow

$$L(G - F) = \int \psi_F \, d(G - F). \tag{5.32}$$

So we can write

$$\frac{\partial}{\partial\epsilon}g(F_\epsilon) = \lim_{h\to 0}\frac{g(F_{\epsilon+h}) - g(F_\epsilon)}{h}$$

$$= \lim_{h\to 0}\frac{g(F + (\epsilon + h)G^*) - g(F_\epsilon)}{h}$$

$$= \lim_{h\to 0}\frac{g(F_\epsilon + hG^*) - g(F_\epsilon)}{h}$$

$$= \int D_F(g(F_\epsilon) \mid \lambda)dG^*(\lambda)$$

provided the Gâteaux derivative exists. Thus equation (5.30) gives

$$g(F + G^*) = g(F) + \int_{-\pi}^{\pi} D_F(g(F) \mid \lambda)dG^*(\lambda) + \text{remainder} \qquad (5.33)$$

Replacing G^* by $G - F$ (G is also an arbitrary cumulative power spectrum), equation (5.33) can be rewritten as

$$g(G) = g(F) + \int_{-\pi}^{\pi} D_F(g(F) \mid \lambda)d(G - F)(\lambda) + \text{remainder} \qquad (5.34)$$

If G is near F, the remainder term will be negligible (see Huber, 1981 for details). Replacing G by $\hat{F}$, an estimate of the cumulative spectrum F, equation (5.34) gives

$$g(\hat{F}) = g(F) + \int_{-\pi}^{\pi} D_F(g(F) \mid \lambda)d(\hat{F} - F)(\lambda) + \text{remainder} \qquad (5.35)$$

The common choice of $\hat{F}$ is considered as $d\hat{F}(\lambda) = \hat{f}(\lambda)d\lambda = I_n(\lambda)d\lambda$, where $I_n(\lambda)$ is the periodogram of a time series process. For this choice of $\hat{F}$, we have from equation (5.35)

$$g(\hat{F}) = g(F) + \int_{-\pi}^{\pi} D_F(g(F) \mid \lambda)(I_n(\lambda) - f(\lambda))d\lambda + \text{remainder} \qquad (5.36)$$

Assuming that the remainder goes to zero for sufficiently large n, the asymptotic

variance of $g(\hat{F})$ can be obtained as

$$V(g(\hat{F})) = \text{var}\left\{ \int_{-\pi}^{\pi} D_F(g(F) \mid \lambda) I_n(\lambda) d\lambda \right\} \qquad (5.37)$$

Using equations (5.10) and (5.11), we can show that for a Gaussian process, the above variance function can be approximated as

$$V(g(\hat{F})) \approx \frac{4\pi}{n} \int_{-\pi}^{\pi} D_F(g(F) \mid \lambda) D_F^*(g(F) \mid \lambda) f^2(\lambda) d\lambda \qquad (5.38)$$

Note that this asymptotic variance is derived from the assumption that the remainder tends to zero for large n, which is not easy to prove. We will try to investigate the behavior of the remainder in future.

5.3 Discussion

Frequency influence function may be considered as a tool of studying the infinitesimal behavior of a time series functional in terms of the point-mass contamination of the underlying spectral distribution. The main difference between Hampel's influence function in a time series setting as discussed in Martin and Yohai (1986) and Hastings' frequency domain influence function is due to the fact that unlike Hampel's influence function, Hastings' frequency domain influence function is based on a particular type of perturbations of the data or the underlying spectral distribution. For example, frequency influence function is defined in terms of the harmonic contamination of the data or the point-mass contamination of the underlying spectral distribution. In Example 5.2.2, it was shown that the frequency domain influence functions can be derived as a particular case of the time domain influence functions. To be specific, time domain influence function reduces to frequency domain influence function in the case where the original process $\{x_t\}$ is contaminated by a harmonic process $\{y_t\}$ as $\{x_t + \sqrt{2\epsilon}\, y_t\}$, where $y_t = \cos(\tilde{\omega} t + \tilde{\varphi})$ and $\tilde{\varphi}$ is uniform$(-\pi, \pi)$, independent of $\{x_t\}$. Note that in time series analysis, it is reasonable to consider harmonic contamination of the data as we encounter this type of contaminated data in many oceanographic measurements.

Frequency influence function is found to be useful in finding the asymptotic variance of a time series functional. In Examples 5.2.1 and 5.2.2, it was shown explicitly that the asymptotic variance of a time series functional can be obtained as a function of its corresponding frequency influence function. In the case of a complicated time series functional $g(F)$ as illustrated in Example 5.2.3, we may find the corresponding asymptotic variance using the assumption of Gâteaux differentiability of the functional g. The asymptotic variance was established using the first-order Taylor's series expansion of the functional g. However, we found it hard to prove the required assumption that the remainder of the Taylor's series expansion is negligible for large samples. Further study is needed here to explore the asymptotic behavior of the remainder.

Chapter 6

Summary

6.1 Summary

In this section, we summarize the results obtained in this thesis.

In chapter 2, we focus on robust spectrum estimation and model selection in a time series process, which is presumed to follow an $ARMA$ process mixed with some discrete spectrum tidal harmonic components. We develop an M estimation method for estimating the power spectrum of the $ARMA$ process by downweighting the effect of possible harmonic components. The proposed M estimators are based on the robust likelihood function of the discrete Fourier transforms of a time series process. This robust likelihood function can be related to the Huber's least favorable distribution.

As pointed out earlier, many oceanographic measurements show the influence of harmonic tidal components, and due to their relatively large size, these can obscure other signals in the data. Our focus was on harbor seiches, which are resonant responses of harbors to external forcing. We analyze three data sets, which consist of sea level records at Sydney, Halifax and Yarmouth harbors of Nova Scotia, Canada. We presume that the observed sea level record consists of two parts, the tide and the surge, the latter being the residual obtained after removing the tide from the record. The tide consists of a number of harmonic components. Our interest is in

the structure of the surge, in which we look for characteristics of seiche motions. As a natural approach, one may try to remove the tidal component from the data using a certain tidal model and then study the residual. However, the resulting residual series often contains remnants of tidal components which are typically not distributed in a stationary fashion. We adopt an alternative approach which downweights the tidal components when estimating the surge. It is assumed that the surge data follow an $ARMA(p, q)$ model and our goal is to estimate the power spectrum of the $ARMA$ process. We develop an M estimation method which is found to be useful in estimating the power spectrum of the surge by downweighting the effect of the tidal component mixed with the surge. The proposed M estimators have good asymptotic properties. We show that the M estimators are Fisher consistent and have bounded influence functions. Under suitable regularity conditions, the M estimators are asymptotically normal with a certain mean vector and a covariance matrix. The variance of the M estimators can be expressed as a function of their corresponding influence functions. We also develop a robust model selection criterion for choosing the appropriate orders of an $ARMA$ process. The proposed robust model selection and M estimation technique may be applied to a wide class of oceanographic data, where we observe a strong presence of harmonic tidal components in the data.

In chapter 3, we introduce a robust model selection criterion in the time domain for choosing the appropriate order of an autoregressive process. We discuss aspects of robust estimation as well as model selection in an autoregressive process with possible time series outliers. The Akaike Information Criterion (AIC) is widely used as a classical model selection criterion. As this classical AIC is sensitive to outliers, robust versions of this criterion have been considered by a number of authors. We point out some drawbacks of the existing robust model selection criterion and propose a modified version of this criterion in order to choose the correct order of an autoregressive process.

Note that in chapter 3, we focus on the model selection criterion for autoregressive processes. It seems important to develop a robust criterion for general $ARMA$ processes. The parameters of an $ARMA$ process are very hard to estimate robustly.

When moving-average terms are present in a time series process, the *GM* estimates result in unbounded influence function. As a consequence, the *GM* estimates and the model selection criterion based on these estimates are not robust. So in order to develop a robust criterion for *ARMA* processes, it is important to find a robust technique for estimating the corresponding *ARMA* components. Much work remains to be done in this area.

In chapter 4, we develop a robust method for estimating the parameters in a nonlinear regression setting with autoregressive errors. Nonlinear regression plays an important role in many fields. Robust estimation in nonlinear regression has been considered by a number of authors. Most of the robust techniques are developed in the case of an i.i.d. setting of the residuals. But situations in which data are collected sequentially over time may result in substantial serial correlations in the errors. We develop a generalized M (*GM*) estimation method for the joint estimation of the regression parameters and the autoregressive parameters of a nonlinear regression model with autoregressive errors. To choose an appropriate order for the autoregressive error process, we extend the robust Akaike Information Criterion introduced in chapter 3 to the case of a nonlinear regression with autoregressive errors.

We study the infinitesimal behavior of the *GM* estimates based on the time series analogue of Hampel's influence function. The *GM* estimates have bounded influence function. We also study the consistency and asymptotic normality of the *GM* estimates in some detail. To derive the asymptotic results of the estimates, we use a uniform law of large numbers and a central limit theorem valid for time dependent data. The key concept employed is uniform integrability to ensure the uniform convergence of a sequence. We impose some mixing conditions, which restrict the memory of a process in a fashion analogous to the role of ergodicity for a stationary stochastic process. We show that under some regularity conditions, the *GM* estimators are consistent and follow the asymptotic normal distribution with certain mean vectors and covariance matrices. We also show that the variance function of a *GM* estimator can be expressed in terms of the corresponding influence function.

As an application of the proposed robust method, we analyze some ground level ozone data which appear to have nonlinear relationship with some meteorological variables as covariates. As the data are collected sequentially over time, there appears to be a significant serial correlation in the errors. Moreover, the error process is found to be heteroscedastic with respect to the covariates. We develop a robust technique to model the variance of the heteroscedastic errors as a function of the covariates. The analysis shows that ozone levels are strongly affected by the variations in meteorological conditions. It is found that ozone concentrations are positively related to daily maximum surface temperature but negatively related to wind speed.

In chapter 5, we discuss a new class of influence functions introduced by Hastings (personal communication), which are referred to as frequency influence functions of a functional $g(F)$, where F is considered as the cumulative (non-normalized) power spectrum of a time series process $\{x_t\}$. Frequency influence function investigates the behavior of the functional $g(F)$ in terms of the point-mass perturbation of the cumulative spectrum F. Like Hampel's influence function, it has been shown with some examples that the asymptotic variance of an estimator may be obtained as a function of its corresponding frequency influence function.

6.2 Further Research

The robust likelihood function of the discrete Fourier transforms, introduced in chapter 2 for estimating the power spectrum of a time series process, can be extended to the case of a nonlinear regression with a stationary noise process. Consider a model of the form

$$y_t = h(t \mid \boldsymbol{\theta}) + \epsilon_t, \quad t = 0, \ldots, n-1 \tag{6.1}$$

where $\boldsymbol{\theta}$ is a finite dimensional parameter, $h(t \mid \boldsymbol{\theta})$ a nonlinear function of $\boldsymbol{\theta}$ at time t and ϵ_t a stationary noise process. Then the discrete Fourier transform gives

$$d_y(\omega_j) = d_h(\omega_j \mid \boldsymbol{\theta}) + d_\epsilon(\omega_j) \tag{6.2}$$

where $\omega_j = 2\pi j/n$, $j = 1, \ldots, [(n-1)/2]$. The $d_\epsilon(\omega_j)$'s are approximately independent complex normals with mean 0 and variance $2\pi n f_{\epsilon\epsilon}(\omega_j)$, where $f_{\epsilon\epsilon}(\omega_j)$ represents the power spectrum of the noise process $\{\epsilon_t\}$ at frequency ω_j. The functional $d_h(\omega_j \mid \boldsymbol{\theta})$ being known up to $\boldsymbol{\theta}$, the problem can now be considered as one of nonlinear regression. Treating the $d_y(\omega_j)$ as exactly independent complex normals with mean $d_h(\omega_j \mid \boldsymbol{\theta})$ and variance $2\pi n f_{\epsilon\epsilon}(\omega_j)$, we can establish the robust likelihood function of the $d_y(\omega_j)$ following Huber's "Proposal 2" for a location parameter. This robust likelihood function would be able to identify and downweight any harmonic contamination of the stationary noise process.

In chapter 5, we introduce a new class of time series influence functions in the frequency domain. It is illustrated with some examples that the asymptotic variance of a time series functional can be expressed in terms of its corresponding frequency influence function. Frequency influence functions can be extended to the regression analysis as well. In a regression model as defined in equation (6.1), it may be possible to investigate the infinitesimal behavior of a time series functional corresponding to the regression parameter $\boldsymbol{\theta}$ in terms of the harmonic contamination of the data or the point-mass contamination of the underlying spectral distribution.

In chapter 4, we study the local stability of the GM estimators of nonlinear regression in terms of Hampel's influence function in a time series setting. The global stability of the GM estimators has not been investigated in this thesis. As a tool of studying the global stability of an estimator, its breakdown properties are investigated by many authors in the case of an i.i.d. setting. In a time series parameter estimation setting, the breakdown properties are yet to be explored. It is important to investigate the global stability in search of a high breakdown point robust estimator of a time series parameter. Much work needs to be done in this area.

Bibliography

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle", in *2nd International Symposium on Information Theory* (Edited by B. N. Petrov and F. Csaki), 267-281, Akademiai Kiado, Budapest.

Behrens, J. (1991), "Robust Ordnungswahl für autoregressive Prozesse", Ph. D. Thesis, University of Kaiserslautern, Germany.

Bhansali, R. J. and Downham, D. Y. (1977), "Some Properties of the Order of an Autoregressive Model Selected by a Generalization of Akaike's FPE Criterion", *Biometrika*, 64, 547-551.

Boos, D. D. and Serfling, R. J. (1980), "A Note on Differentials and the CLT and LIL for Statistical Functions, with Application to M-estimates", *The Annals of Statistics*, 8, 618-624.

Brillinger, D. R. (1981), *Time Series: Data Analysis and Theory* (Expanded Edition), Holden-Day, San Francisco.

Brillinger, D. R. (1985), "Fourier Inference: Some Methods for the Analysis of Array and Non-Gaussian Series Data", *Water Resources Bulletin: American Water Resources Association*, 3, 743-756.

Clarke, B. R. (1983), "Uniqueness and Frechet Differentiability of Functional Solutions to Maximum Likelihood Type Equations", *The Annals of Statistics*, 11, 1196-1205.

Clarke, B. R. (1986), "Nonsmooth Analysis and Frechet Differentiability of M-Functionals", *Probability Theory and Related Fields*, 73, 197-209.

Domowitz, I. and White, H. (1982), "Misspecified Models With Dependent Observations", *Journal of Econometrics*, 20, 35-58.

Ferretti, N., Kelmansky, D., Yohai, V. J. and Zamar, R. H. (1999), "A Class of Locally and Globally Robust Regression Estimates", *Journal of the American Statistical Association*, 94, 174-188.

Foreman, M. G. G. (1977), "Manual for Tidal Heights Analysis and Prediction", *Pac. Mar. Sci. Rep.*, Inst. of Ocean Sci, Patricia Bay, Sidney, B. C., Canada, 77-10, 101 pp.

Fraiman, R. (1983), "General M-Estimators and Applications to Bounded Influence Estimation for Non-linear Regression", *Commun. Statist.*, 12(22), 2617-2631.

Hampel. F. R. (1968), "Contributions to the Theory of Robust Estimation", Ph. D. Thesis, University of California, Berkeley.

Hampel, F. R. (1974), "The Influence Curve and Its Role in Robust Estimation", *Journal of the American Statistical Association*, 69, 383-393.

Hampel, F. R. (1978), "Optimally Bounding the Gross Error Sensitivity and the Influence of Position in Factor Space", Proceedings of the ASA Statistical Computing Section, ASA, Washington, D. C., pp. 59-64.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, New York, John Wiley.

Heritier, S. and Ronchetti, E. (1994), "Robust Bounded-Influence Tests in General Parametric Models", *Journal of the American Statistical Association*, 89, 897-904.

Hoadley, B. (1971), "Asymptotic Properties of Maximum Likelihood Estimators for the Independent not Identically Distributed Case", *Annals of Mathematical Statistics*, 42, 1977-1991.

Huber, P. J. (1964), "Robust Estimation of a Location Parameter", *Ann. Math. Statist.*, 35, 73-101.

Huber, P. J. (1973), "Robust Regression: Asymptotics, Conjectures, and Monte Carlo", *Annals of Statistics*, 1, 799-821.

Huber, P. J. (1977), "Robust Methods of Estimation of Regression Coefficients", *Math. Operationsforschung Statist. Ser. Statist.*, 8, 41-53.

Huber, P. J. (1967), "The Behavior of the Maximum Likelihood Estimates Under Nonstandard Conditions", *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 221-233.

Huber, P. J. (1981), *Robust Statistics*, Wiley, New York.

Kleiner, B., Martin, R. D. and Thompson, D. J. (1979), "Robust Estimation of Power Spectra", *Journal of the Royal Statistical Society*, Series B 41(3), 313-351.

Krasker, W. S. (1980), "Estimation in Linear Regression Models With Disparate Data Points", *Econometrica*, 48, 1333-1346.

Krasker, W. S. and Welsch, R. E. (1982), "Efficient Bounded Influence Regression Estimation", *Journal of the American Statistical Association*, 77, 595-604.

Mallows, C. L. (1975), "On Some Topics in Robustness", Technical Memorandum, Bell Telephone Laboratories, Murray Hill, N. J.

Maronna, R. A., Bustos, O. H. and Yohai, V. J. (1979), "Bias- and Efficiency-Robustness of General M-Estimators for Regression with Random Carriers", in *Smoothing Techniques for Curve Estimation*, eds. T. Gasser and M. Rosenblatt, New York, Springer-Verlag, 91-116.

Martin, R. D. (1980), "Robust Estimation of Autoregressive Models", in *Directions in Time Series*, (Edited by D. R. Brillinger and G. C. Tiao), 228-262, Institute of Mathematical Statistics, Hayward, CA.

Martin, R. D. (1981), "Robust Methods for Time Series II", in *Applied Time Series*, ed. D. F. Findley, Academic Press, New York.

Martin, R. D. (1983), "Robust-Resistant Spectral Analysis", in *Handbook of Statistics*, (Edited by D. R. Brillinger and P. R. Krishnaiah), 185-219, Elsevier Science Publishers B. V., Amsterdam.

Martin, R. D. and Yohai, V. J. (1986), "Influence Functionals for Time Series", *Annals of Statistics*, 14, 781-818.

McLeish, D. L. (1975), "A Maximal Inequality and Dependent Strong Laws", *Annals of Probability*, 3, 829-839.

Niu, Xu-Feng (1996), "Nonlinear Additive Models for Environmental Time Series, With Applications to Ground-Level Ozone Data Analysis", *Journal of the American Statistical Association*, 91, 1310-1321

Priestley, M. B. (1981), *Spectral Analysis and Time Series*, Academic Press, London.

Proudman, J. (1953), *Dynamical Oceanography*, Wiley, New York.

Ronchetti, E. (1997), "Robustness Aspects of Model Choice", *Statistica Sinica*, 7, 327-338.

Rousseeuw, P. J. (1984), "Least Median of Squares Regression," *Journal of the American Statistical Association*, 79, 871-880.

Rousseeuw, P. J. and Hubert, M. (1999), "Regression Depth", *Journal of the American Statistical Association*, 94, 388-402.

Rousseeuw, P. J. and Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, New York, John Wiley.

Rousseeuw, P. J. and van Zomeren, B. C. (1990), "Unmasking Multivariate Outliers and Leverage Points", *Journal of the American Statistical Association*, 85, 633-639.

Rousseeuw, P. J. and Yohai, V. J. (1984), "Robust Regression by Means of S Estimators", in *Robust and Nonlinear Time Series Analysis*, Lecture Notes in Statistics 26, eds. J. Franke, W. Hardle, and R. D. Martin, New York, Springer-Verlag, pp. 256-272.

Schuster, A. (1898), "On the Investigation of Hidden Periodicities With Application to a Supposed 26 Day Period of Meteorological Phenomena", *Terr. Magn.*, 3, 13-41.

Simpson, D. G., Ruppert, D. and Carroll, R. J. (1992), "On One-step *GM*-Estimates and Stability of Inferences in Linear Regression", *Journal of the American Statistical Association*, 87, 439-449.

Smith, B. and Miyaoka, E. (1999), "Frequency Domain Identification of Harbour Seiches", *Environmetrics*, 10, 575-587.

Stromberg, A. J. (1993), "Computation of High Breakdown Non-linear Regression Parameters", *Journal of the American Statistical Association*, 88, 237-244.

Stromberg, A. J. and Ruppert, D. (1992), "Breakdown in Non-linear Regression", *Journal of the American Statistical Association*, 87, 991-997.

Tukey, J. W. (1970), *Exploratory Data Analysis*, mimeographed preliminary edition.

White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press, Inc., London.

White, H. and Domowitz, I. (1984), "Nonlinear Regression With Dependent Observations", *Econometrica*, 52, 143-161.

White, H. and Domowitz, I. (1981), "Nonlinear Regression With Dependent Observations", Department of Economics discussion paper no. 81-32 (University of California, San Diego, CA).

Whittle, P. (1953), "The Analysis of Multiple Stationary Time Series", *Journal of the Royal Statistical Society*, Series B 15, 125-139.

Yohai, V. J. (1987), "High Breakdown Point and High Efficiency Robust Estimates for Regression", *The Annals of Statistics*, 15, 642-656.

Yohai, V. J. and Zamar, R. H. (1988), "High Breakdown Point Estimates of Regression by Means of the Minimization of an Efficient Scale", *Journal of the American Statistical Association*, 83, 406-413.