INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

EMPIRICAL CHARACTERISTIC FUNCTION IN TIME SERIES ESTIMATION AND A TEST STATISTIC IN FINANCIAL MODELLING

by

Jun Yu

Graduate Program
in
Economics

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

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
June 1998

0-612-31169-4

THE UNIVERSITY OF WESTERN ONTARIO
FACULTY OF GRADUATE STUDIES

CERTIFICATE OF EXAMINATION

Chief Advisor

Advisory Committee

Examining Board

R. J. Kulperger

The thesis by

Jun Yu

entitled

EMPIRICAL CHARACTERISTIC FUNCTION IN TIME SERIES ESTIMATION AND A TEST STATISTIC IN FINANCIAL MODELLING

is accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Date 18 September 1998

Chair of Examining Board

ABSTRACT

This thesis consists of five chapters. The first four chapters deal with an estimation method for time series models via the empirical characteristic function (ECF). The fifth chapter proposes a test statistic which is useful in financial modelling.

In the first chapter we first review what has been done for the ECF method in both the independently and identically distributed (i.i.d.) case and the dependent case. It appears that the ECF method has received little attention in the dependent case. Consequently we propose an ECF method to estimate stationary time series models and obtain the asymptotic properties for the resulting estimator. The conditions which ensure the asymptotic properties appear regular. The ECF method is useful because many examples have been found in economics where the traditional estimation methods, such as the maximum likelihood (ML) estimation method, do not work or are not easy to work with. In some cases the likelihood has no closed form and in some other cases the likelihood function is not bounded over the parameter space. In some cases the likelihood has a closed form and is bounded over the parameter space, but it is too difficult to numerically approximate. The intuition why the ECF method can work well is that the characteristic function (CF) has a one-to-one correspondence with the distribution function and hence the ECF contains all the information in the data. Also the ECF is uniformly bounded and absolutely continuous.

In Chapter 2 we apply the ECF method to estimate the linear ARMA models. To use the ECF method efficiently, the optimal weight functions and the estimating equations are derived for the ARMA models. The Monte Carlo studies confirm the validity of the ECF method in terms of finite sample properties. For example, for the MA(1) and AR(1) models, the ECF method can work as well as the ML method. For

the ARMA(1, 1) model, the ECF method outperforms the conditional ML method.

In the next two chapters we apply the ECF method to estimate the non-linear time series models. In Chapter 3, we focus on the estimation of the stochastic volatility (SV) model. The SV model is considered because this model has attracted a great deal of attention in both finance and macroeconomics. However, the estimation of the model is hard because the likelihood function for the model has no closed form. Consequently, the characteristic function for a SV model is derived and an estimation method based on the ECF is used. The Monte Carlo study shows that the ECF method is a viable alternative. An empirical application is considered in this chapter and suggests different empirical results from what are normally found. In Chapter 4, we estimate a diffusion jump process by using the ECF method, where the intensity parameter in the Poisson jump process is self-exciting. This model is considered because it is an alternative way to the ARCH-type models to describe the movement of stock prices but it allows for the dis-continuity in the sample path. However, the estimation of the model is hard because the likelihood function for the model has no closed form. Consequently, the characteristic function for this model is obtained and hence we estimate the model using the ECF method. The Monte Carlo study shows that the ECF method outperforms the generalized method of moments (GMM). An empirical application is considered and we find some interesting empirical results.

In Chapter 5 we propose a test statistic to distinguish finite-variance models and infinite-variance models for daily stock returns. The distribution form of stock returns is important for both theoretical and empirical analysis in finance. Most of recent literatures suggests that finite-variance models have greater descriptive power than infinite-variance Stable distribution for daily stock returns. When we apply our test statistic to S&P 500 daily returns, we find that most finite-variance models can not be rejected when the crash days are excluded. However, all the existing finite variance models have been rejected when the crash days are included.

Keywords: Empirical Characteristic Function, Stationary Process, Consistency, Asymptotic Normality, Efficiency, ARMA Models, Stochastic Volatility Models, Diffusion Jump Processes, Test Statistic, Finite-Variance Distributions, Infinite-Variance Distributions, Interquartile Range

Dedicated To My Family

ACKNOWLEDGEMENTS

I would like to thank all the people who have helped me make this possible. I should apologize to those who should be mentioned but are not.

I acknowledge with the most sincere thanks, Professor John Knight, for the supervision of most parts of this thesis. He introduced a fascinating and challenging topic to me. During the past two years, he unceasingly encouraged me to pursue this subject. I have benefited from his insights, his approachability, his patience, his financial support and many other aspects. I suspect, I will never be able to repay him for what he has done for me.

My gratitude towards to Dr. Hao Yu is also infinite, not only because he supervised one chapter of my thesis, but also because his encouragement, his advice, his financial support and his personal friendship.

I also would like to thank Professor Chris Robinson for his financial support and persistent encouragements since I came to Western, Professor Victor Aguirregabiria for many conversations, Professor Audra Bowlus, Professor Robin Carter, Professor Reg Kulperger, Professor Brendan McCabe, Dr. Steve Satchell, Professor Jeff Smith, Dr. Jim Talman, Professor Jian Yang for useful discussion and comments. Thanks also go to Professor Al Slivinsky and Professor Ig Horstmann and Professor Arthur Robson for their encouragement and help when I was on the job market.

It is a pleasure to thank the NSERC of Canada for their NSERC Scholarship, Faculty of Graduate Studies at the UWO for their Special University Scholarship, Department of Economics for arranging Teaching Assistantships in the past five years and summer teaching in the past three summers. Thanks also go to Sue Brown, Melissa Broadfoot, Jane McAndrew and Yvonne Adams for secretarial and administrative assistance.

The UWO Department of Statistical and Actuarial Sciences is gratefully acknowledged for providing the computer facilities.

Finally, I express my deep appreciation to my family. I thank my parents, who have sacrificed so much for my education and well-being since my childhood and supported me to come to Canada to study. I also thank my wife Junying Huang for her love, patience and emotional support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

EMPIRICAL CHARACTERISTIC FUNCTION IN TIME SERIES ESTIMATION

1.1 Introduction

Maximum likelihood estimation under parametric assumptions is one of the most widely used estimation methods. One reason is that it results in estimators which are consistent, asymptotically normal and asymptotically efficient under appropriate regularity conditions. To implement the maximum likelihood method, however, the likelihood function must be of a tractable form and is sometimes required to be bounded in parameter space. Unfortunately, there are many processes in economics where the maximum likelihood approach is difficult to implement, both in the independently, identically distributed (i.i.d.) case and the dependent case. In the i.i.d. case, the processes sometimes have an unbounded likelihood function in parameter space. Examples include mixture of normals and switching regression models (Titterington *et al* (1985) and references therein). In the economic context, the mixture of normals and switching regression models can be viewed as contaminated data or structural change problems (Granger and Orr (1972)). For example, in a firm's monthly production series, the contamination may be due to a sudden strike, a sale promotion, or an annual vacation shutdown. In daily interest rate changes, the result of a governmental policy action can be viewed as structural change. In stock markets, the disclosure of information signals may lead to a parameter shift (Kon (1984)).

In other examples, however, the density functions of the processes can not be written in a closed form, such as the Stable distributions and compound normal and log-normal models. Both models are proposed to describe the behavior of stock returns (Mandelbrot (1963); Fama (1965); Clark (1973)). This problem also arises in the dependent case. Such examples include stochastic volatility (SV) models (Ghysels *et al* (1996)), ARCH-type models (Bollerslev *et al* (1992)), and processes which are compound Poisson-Normal and where the Poisson intensity is random, possibly dependent on past information in the series (Knight *et al* (1993)). Some of these models have found wide use in macroeconomics and finance.

Although some processes have a known density, evaluation of the exact likelihood can be extremely difficult for various reasons. For instance, in order to calculate the exact likelihood function of stationary ARMA models, one needs to deal with the determinant and inverse of the covariance matrix (Zinde-Walsh (1988)). However, such calculations can be computationally very expensive or even infeasible for a large number of observations.

The usual response to such difficulties arising from the likelihood is to use alternative methods[1]. For example, one might use some variant of the method of moments (Hansen (1982)), the conditional maximum likelihood (Bollerslev *et al* (1992)), the quasi-maximum likelihood (QML) (White (1982)), or the simulation based method (Danielsson (1994b); Duffie and Singleton (1993)). Although all methods are consistent under regular conditions, some of them are not asymptotically efficient. Fur-

[1] In estimating ARMA models, the exact maximum likelihood method is available by use of the Kalman Filter (see Hamilton (1994)).

thermore, the small sample properties of these methods may be unsatisfactory. This thesis discusses another alternative, a method that uses the empirical characteristic function (ECF).

Initiated by Parzen (1962) , the ECF has been used in many areas of inference such as testing for goodness of fit (Fan (1996)), testing for independence (Feuerverger (1990)), testing for symmetry (Feuerverger and Mureika (1977)), and parameter estimation.[2] The main justification for the ECF method is that the characteristic function (CF) has a one-to-one correspondence with the distribution function, and hence the ECF retains all the information present in the sample. Theoretically, therefore, the inference based on the ECF should work as well as that based on the empirical distribution function. The advantage of using the CF is that it is uniformly bounded and thus, should result in more robustness and greater numerical stability. The theory for the ECF method in the i.i.d. case is complete (Feuerverger and Mureika (1977); Csörgő (1981)). Surprisingly, however, the dependent case has received little attention and consequently there is great scope for research.

The purpose of this chapter is to discuss the ECF estimation method for stationary stochastic processes and establish the asymptotic properties of the ECF estimators for general stationary processes.

Section 1.2 conducts a literature review of the ECF method in both the i.i.d. case and dependent case. Section 1.3 proposes the ECF method in a general framework, specifies the assumptions and obtains the asymptotic properties of the ECF estimator for stationary processes. It turns out that the conditions which ensure the asymptotic properties are regular. Theoretically, therefore, the proposed ECF method is justified.

[2]The references for parameter estimation include Feuerverger and McDunnough (1981a,1981b); Feuerverger (1990) , Heathcote (1977), Knight and Satchell (1996, 1997), Press (1972), Quandt and Ramsey (1978); Schmidt (1982), Paulson *et al* (1975) and Yao and Morgan (1996).

1.2 Literature Review

In this section, a literature review is performed on the ECF method in the i.i.d. case and dependent case. The ECF method in both cases shares the same spirit, that is, to match some distance measure between the ECF and the CF. While the literature in the i.i.d. case is extensive, very little research has been reported in the dependent case. I will present them in detail.

1.2.1 The I.I.D. Case

Suppose $\{X_i\}_{i=-\infty}^{\infty}$ is a sequence of independent univariate random variables with common distribution function $F_{\boldsymbol{\theta}}(x)$ which depends on a vector of unknown parameters $\boldsymbol{\theta}$. We have a finite realization $\{x_1, x_2, \ldots, x_n\}$ and we wish to estimate $\boldsymbol{\theta}$. The CF is defined as $c(r;\boldsymbol{\theta}) = \int \exp(irx)\, dF_{\boldsymbol{\theta}}(x)$, and the ECF is defined as $c_n(r) = \frac{1}{n}\sum_{j=1}^{n} \exp(irx_j)$, where r is the transformation variable. Hence the ECF is the sample counter part of the CF and contains the information of the data, while the CF contains the information of the parameters.

By the S.L.L.N. $c_n(r) \stackrel{a.s.}{\to} c(r;\boldsymbol{\theta})$ for any fixed $r < +\infty$. If $E|X|^{1+\delta} < +\infty$ for some positive δ, Feuerverger and Mureika (1977) provide a weak convergence result for $Y_n(r;\boldsymbol{\theta}) = \sqrt{n}(c_n(r) - c(r;\boldsymbol{\theta}))$. That is, $Y_n(r;\boldsymbol{\theta}) \Rightarrow Y(r;\boldsymbol{\theta})$ in every finite interval, where $Y(r;\boldsymbol{\theta})$ is a complex valued Gaussian process.[3] Moreover, Csörgő (1981) obtains the strong approximation for $Y_n(r;\boldsymbol{\theta})$. Both the weak convergence and strong approximation justify the estimation method via the ECF.

An estimation procedure using the ECF is to choose $\tilde{\boldsymbol{\theta}}_n$ to minimize[4]

$$I_n(\boldsymbol{\theta}) = \int |c_n(r) - c(r;\boldsymbol{\theta})|^2\, dG(r), \tag{2.1}$$

[3] "$\Rightarrow$" is defined as weak convergence (see Billingsley (1986)).

[4] $\int$ means $\int_{-\infty}^{+\infty}$ in the thesis unless specified.

or

$$I_n(\boldsymbol{\theta}) = \int |c_n(r) - c(r;\boldsymbol{\theta})|^2 g(r)\, dr, \tag{2.2}$$

or to solve the estimating equation

$$\int w_{\boldsymbol{\theta}}(r)(c_n(r) - c(r;\boldsymbol{\theta}))\, dr = 0, \tag{2.3}$$

where $G(r), g(r)$ and $w_{\boldsymbol{\theta}}(r)$ are weight functions. Under suitable regularity assumptions Heathcote (1977) establishes the strong consistency and asymptotic normality for $\bar{\boldsymbol{\theta}}_n$.

A simple estimating procedure is to choose $G(r)$ to be a step function with a finite number of jumps. This discrete type procedure consists of two steps. Firstly, we choose q (no less than the number of the unknown parameters), and $r_1, r_2, ..., r_q$. Secondly,

$$\min_{\boldsymbol{\theta}\in\Theta} \sum_{j=1}^{q} |c_n(r_j) - c(r_j;\boldsymbol{\theta})|^2. \tag{2.4}$$

Basically this is the procedure proposed by Quandt and Ramsey (1978) and is essentially the ordinary least square (OLS) procedure.[5]

Since $c_n(r_1), \cdots, c_n(r_q)$ are not i.i.d., the OLS procedure results in inefficient estimators. Consequently, Schmidt (1982) proposes a more efficient estimator of $\boldsymbol{\theta}$ by performing the generalized least square (GLS) procedure,[6,7]

$$\min_{\boldsymbol{\theta}\in\Theta} \sum_{j=1}^{q}\sum_{k=1}^{q}(c_n(r_j) - c(r_j;\boldsymbol{\theta}))\hat{\Omega}_{(j,k)}(\bar{c}_n(r_k) - \bar{c}(r_k;\boldsymbol{\theta})), \tag{2.5}$$

where $\hat{\Omega} = (\hat{\Omega}_{(j,k)})$ is the inverse of a consistent estimate of the variance-covariance matrix of $c_n(r_1), \cdots, c_n(r_q)$, and $\bar{c}$ is the conjugate of c. For both the OLS and GLS,

---

[5]They use the moment generating function instead of the CF.

[6]He also uses the moment generating function.

[7]Basically the procedure is a feasible generalized least square FGLS which results in an estimator asymptotically equivalent to the GLS estimator.

the resulting estimators as well as the asymptotic variance depend on the transformation variables, the r's. Not surprisingly the efficiency also depends on these r's. Feuerverger and McDunnough (1981b) show that if the r's are sufficiently fine and extended the Cramér-Rao lower bound can be reached. However, at least two practical difficulties arise when the optimal r's are to be chosen. First, how many r's should be used (i.e., the value of q)? Second, how to make them optimal? Regarding the second question, Schmidt (1982) proposes to choose r's to minimize the determinant of the asymptotic covariance matrix of the estimator for a given q. This makes sense in the single parameter case. However, there is no clear procedure in the multi-parameter case. Yao and Morgan (1996) advocate minimizing the mean square error rather than the asymptotic variance.[8] To make the minimization numerically more efficient, Feuerverger and McDunnough (1981b) suggest using equally spaced points, i.e., $r_j = \tau \cdot j$, $j = 1, 2, \ldots, q$. The greater difficulty involves the first question, the choice of q. Schmidt (1982) points out that the choice of q would have to await the discovery of the sample property of the estimates, which is still an open question.

Alternatively one can choose the transformation variables continuously. The technique can be accomplished by using a continuous weight. For example, Paulson *et al* (1975) choose $g(r)$ to be $\exp(-r^2)$. However, the resulting estimators for an arbitrary weight are inefficient.

To find an optimal weight function, we focus on Equation (2.3). Consider the first order condition of maximizing the log likelihood function,

$$\int \frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial \boldsymbol{\theta}}\, d(F_n(x) - F_{\boldsymbol{\theta}}(x)) = 0,$$

where $F_n(x)$ is the empirical distribution function defined in Billingsley (1968). Ap-

[8]They detail the application using the Laplace transformation.

plying the Parseval Theorem to the above equation, we obtain

$$\int \frac{1}{2\pi} \int \frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial \boldsymbol{\theta}} e^{-irx}\, dx\, (c_n(r) - c(r;\boldsymbol{\theta}))\, dr = 0.$$

Therefore, the resulting estimator of (2.3) is equivalent to the maximum likelihood estimator (MLE) if the weight function $w_{\boldsymbol{\theta}}$ is proportional to the function below,

$$\frac{1}{2\pi} \int \frac{\partial \log f(x)}{\partial \boldsymbol{\theta}} e^{-irx}\, dx. \tag{2.6}$$

An example is provided by the normal distribution. Let $\{x_1, x_2, \ldots, x_T\}$ be independent and identically distributed $N(0, \sigma^2)$ random variables, where σ^2 is the only unknown parameter. Thus

$$\log f(x) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log \sigma^2 - \frac{x^2}{2\sigma^2},$$

and

$$\frac{\partial \log f(x)}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{x^2}{2\sigma^4}.$$

From (2.6), the optimal weight function can chosen as

$$\begin{aligned} w_{\sigma^2}(r) &= \frac{1}{2\pi}\int (-\frac{1}{2\sigma^2} + \frac{x^2}{2\sigma^4}) \exp(-irx)\, dx \\ &= -\frac{1}{2\sigma^2}\delta(r) - \frac{1}{2\sigma^4}\delta''(r), \end{aligned} \tag{2.7}$$

where $\delta(\cdot)$ is the Dirac delta function and defined as,

$$\delta(x) = \begin{cases} 0 & \text{if } x \neq 0 \\ +\infty & \text{if } x = 0, \end{cases}$$

For the rigirious definition on $\delta(x)$ and the discussion of its properties, see Gel'Fan (1964). Substituting the optimal weight function into the estimating equation (2.3) we have

$$\begin{aligned} 0 &= \int \{-\frac{1}{2\sigma^2}\delta(r) - \frac{1}{2\sigma^4}\delta''(r)\} c_n(r)\, dr \\ &= -\frac{1}{2\sigma^2}\int \delta(r) c_n(r)\, dr - \frac{1}{2\sigma^4}\int \delta''(r) c_n(r)\, dr \\ &= -\frac{1}{2\sigma^2} c_n(0) - \frac{1}{2\sigma^4}\frac{\partial^2 c_n(r)}{\partial r^2}\Big|_{r=0}, \end{aligned}$$

This implies

$$\tilde{\sigma}^2{}_n = -\frac{\partial^2 c_n(r)}{\partial r^2}|_{r=0} = \frac{1}{n}\sum_{j=1}^{T} x_j^2,$$

which is the same as the MLE of σ^2. Note that Paulson *et al* (1975) choose the weight function to be $\exp(-r^2)$ and Heathcote (1977) chooses $\delta(x)$ as the weight. Both $\exp(-r^2)$ and $\delta(x)$ together with our weight function assign more weight on an interval around the origin and hence are consistent with the theory that the CF contains the most information. See Theorem 7.2.1 of Lukacs (1970).[9] However, by including the second order derivative of $\delta(x)$, our weight function puts even larger weight on the origin and is expected to result in the most efficient estimator.

1.2.2 The Dependent Case

Let $\{y_j\}_{j=-\infty}^{\infty}$ be a univariate, stationary time series whose distribution depends upon a vector of unknown parameters, $\boldsymbol{\theta}$. We wish to estimate $\boldsymbol{\theta}$ from a finite realization $\{y_1, y_2, ..., y_T\}$. The overlapping blocks for $y_1, y_2, ..., y_T$ are defined as,

$$\boldsymbol{x}_j = (y_j, ..., y_{j+p})',\ j = 1, ..., T-p.$$

Hence each block has p observations overlapping with the adjacent blocks. The CF of each block is defined as

$$c(\boldsymbol{r}; \boldsymbol{\theta}) = E(\exp(i\boldsymbol{r}'\boldsymbol{x}_j)),$$

where $\boldsymbol{r} = (r^1, ..., r^{p+1})'$. The ECF is defined as

$$c_n(\boldsymbol{r}) = \frac{1}{n}\sum_{j=1}^{n} \exp(i\boldsymbol{r}'\boldsymbol{x}_j),$$

where $n = T - p$.

[9] For a unbounded random variable which has moment generating function discussed in this thesis, the uniqueness theorem still applies.

The estimation procedure is similar to the i.i.d. case, i.e., to match some distance between the ECF and the CF. Unfortunately, only two papers are related to the dependent case. The first work is published by Feuerverger (1990) and the second one is reported by Knight and Satchell (1996). Both papers propose to match the ECF and the CF over a grid of finite points and hence we call the procedure "the discrete ECF method (DECF)". Feuerverger (1990) proves that under some regularity conditions, the resulting estimators can achieve the Cramér-Rao lower bound if p is sufficiently large and the grid of points is sufficiently fine and extended. However, he has not applied the procedure to estimate any time-series model. Knight and Satchell (1996) detail the application of the DECF method to stationary stochastic processes and give a multi-step procedure. We review their procedure in detail. Firstly, choose q and an arbitrary set of vectors $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_q)$, where each vector is of length $p+1$. That is, choose $G(\boldsymbol{r})$ to be a step function with q jumping points. Secondly, define $V = (Re\, c(\boldsymbol{r}_1;\boldsymbol{\theta}), ..., Re\, c(\boldsymbol{r}_q;\boldsymbol{\theta}), Im\, c(\boldsymbol{r}_1;\boldsymbol{\theta}), ..., Im\, c(\boldsymbol{r}_q;\boldsymbol{\theta}))'$, and $V_n = (Re c_n(\boldsymbol{r}_1), ..., Re\, c_n(\boldsymbol{r}_q), Im\, c_n(\boldsymbol{r}_1), ..., Im\, c_n(\boldsymbol{r}_q))'$, and choose $\boldsymbol{r}$ to minimize $(V_n - V)'(V_n - V)$ to obtain a consistent estimate for Ω, $\hat{\Omega}$, where $\Omega = \begin{bmatrix} \Omega_{RR} & \Omega_{RI} \\ \Omega_{IR} & \Omega_{II} \end{bmatrix}$ is the covariance matrix of V_n. Thirdly, choose $\boldsymbol{r}$ to minimize some measure of the asymptotic covariance matrix of the GLS estimators. Fourthly, based on the optimal $\boldsymbol{r}$ obtained at step 3, repeat step 1 to obtain another consistent estimate for Ω, say $\hat{\hat{\Omega}}$. Finally choose $\boldsymbol{\theta}$ to minimize $(V_n - V)'\hat{\hat{\Omega}}^{-1}(V_n - V)$. The explicit form for the covariance matrix Ω is given by Knight and Satchell (1996). Choosing $p = 2$, $q = 5$ and an arbitrary $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_q)$, Knight and Satchell (1996) use the DECF method to estimate an MA(1) model and perform a Monte Carlo simulation. They find that the DECF method is a viable alternative method, however, the performance of the DECF method is strictly dominated by that of the MLE.

The finding is not surprising and can be explained intuitively. Since matching the ECF and the CF over a grid of finite points is equivalent to matching a finite number of moments, the DECF method is, in essence, equivalent to the GMM. Just like sometimes it is not obvious how many and which moments to choose for the GMM, the difficulties for the DECF method are how many and which $\boldsymbol{r}$'s we should use.

The difficulty also comes from the choice of p. For the blocks to capture all the information in the original series, the choice of p could be critical. However, there is a trade-off between a large p and a small p. If we choose a large p, we should expect the moving blocks contain all the information in the original series. By doing so we lose no information and the estimation via the ECF based on such blocks could be efficient when an optimal weight function is used. In fact, according to the inversion theorem, the joint density is the Fourier inversion of the CF of moving blocks. Provided the Fourier inversion can be implemented efficiently, the ECF estimator is asymptotically equivalent to the MLE. Unfortunately, such an inversion is high dimensional integral and sometimes cannot be simplified. Therefore, the procedure could be numerically infeasible. On the other hand, a smaller p could be chosen to achieve the feasibility for the procedure, however, the $n(= T - p)$ overlapping blocks may not retain all the important information of the series. Consequently, several issues arise here. The first issue is whether and under what conditions the ECF estimator has desirable asymptotic properties for general p and general weight. The second one is how to choose p and the weight in order to construct an asymptotically efficient estimator via the ECF method. The last issue that needs to be addressed is the finite sample properties of the ECF estimator. In the next section we will propose the ECF method for stationary stochastic processes in more general framework as well as obtain the asymptotic properties of the resulting estimator.

1.3 ECF Method and Asymptotic Properties

Observing the difficulties involved in the DECF method, for stationary stochastic processes, we propose the ECF method which minimizes the integral

$$I_n(\boldsymbol{\theta}) = \int \cdots \int |c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta})|^2 \, dG(\boldsymbol{r}), \tag{3.1}$$

or

$$I_n(\boldsymbol{\theta}) = \int \cdots \int |c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta})|^2 g(\boldsymbol{r}) \, dr^1 \cdots dr^{p+1}, \tag{3.2}$$

or solves the following estimating equation

$$\int \cdots \int w_{\boldsymbol{\theta}}(\boldsymbol{r})(c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta})) \, d\boldsymbol{r} = 0. \tag{3.3}$$

If the weight function $G(\boldsymbol{r})$ is chosen to be a step function, the procedure is indeed the one proposed by Feuerverger (1990) and Knight and Satchell (1996). Hence the procedure we propose includes the DECF method as a special case. We can certainly choose an alternative weight. If a continuous weight function is used, the procedure will basically match all the moments continuously, including integer moments, fraction moments and irrational moments. In this sense, the procedure exploits more information in the sample. Another advantage of using a continuous weight function is that one no longer needs to choose the transformation variables, $\boldsymbol{r}$'s, because they are simply integrated out. In this thesis we refer the ECF method with a continuous weight as "the continuous ECF method (CECF)". The simplest weight for the CECF method is probably $g(\boldsymbol{r}) = \mathbf{1}$ (or $G(\boldsymbol{r}) = \boldsymbol{r}$). This procedure is referred to "the OLS of the continuous ECF method (OLS-CECF))". We can also use "the weighted least square of the continuous ECF method (WLS-CECF)" by choosing $g(\boldsymbol{r})$ to be a non-equally weighted function such as an exponential function. Furthermore, by using the Parseval Theorem, we can obtain an optimal weight function, $w^*(\boldsymbol{r})$,

$$\left(\frac{1}{2\pi}\right)^{p+1} \int \cdots \int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j) \frac{\partial \log f(y_{j+p}|y_j, \ldots, y_{j+p-1})}{\partial \boldsymbol{\theta}} \, dy_j \ldots dy_{j+p}. \tag{3.4}$$

The weight is optimal in the sense that based on $w^*(\boldsymbol{r})$ and Equation (3.3), the resulting estimator can achieve the Cramér-Rao lower bound when p is sufficiently large (Feuerverger (1990)). The procedure associated with the optimal weight is referred to "the GLS of the continuous ECF method (GLS-CECF)".

For simplicity of notation we deal only with the single parameter case and the estimator resulting from (3.1) in the next two theorems. The multi-parameter case will be considered in Chapter 3 and Chapter 4. The equation (3.1) consists of minimizing a distance function and hence can be regarded as the Fourier version of the M-estimators first discussed in the i.i.d. case by Huber (1981) and extended to the dependent case by Martin and Yohai (1986) . The necessary regularity conditions will be outlined in detail. Based on the conditions the consistency and asymptotic normality are then reported and proved. Theoretically, therefore, we justify the proposed procedure.

Assumptions:

(A1) Let Θ be a compact set.

(A2) $I_n(\theta)$ can be differentiated under the integral sign with respect to θ.

(A3) $\tilde{\theta}_n = argmin_{\theta\in\Theta} I_n(\theta)$.

(A4) True value θ_0 lies in the interior of Θ.

(A5) The sequence $\{y_t\}$ is ergodic.

(A6) $c''(\boldsymbol{r};\theta)$ exists and is uniformly bounded by G-integrable function, where the prime denotes differentiation with respect to θ.

(A7) The sequence $\{y_t\}$ satisfies the ρ mixing condition. That is, if $\rho(q)$ is the maximum correlation possible between functions of the y_j's separated in time by a distance q, then

$$\sum_{q=1}^{\infty} \rho(q) < \infty.$$

(A8) $\int \cdots \int (c'(\boldsymbol{r};\theta))^2|_{\theta=\theta_0} dG(\boldsymbol{r}) \neq 0$.

Theorem 1.3.1 *For any fixed p, under assumptions A1-A5, the ECF estimator by minimizing* (3.1) *is strongly consistent, i.e.,*

$$\tilde{\theta}_n \stackrel{a.s.}{\to} \theta_0$$

Proof: Assumption (A5) implies that

$$c_n(\boldsymbol{r}) \stackrel{a.s.}{\to} c(\boldsymbol{r};\theta_0) \ \forall \boldsymbol{r} \in \Re^{p+1},$$

Hence, we have,

$$Re\, c_n(\boldsymbol{r}) \stackrel{a.s.}{\to} Re\, c(\boldsymbol{r};\theta_0),$$

and

$$Im\, c_n(\boldsymbol{r}) \stackrel{a.s.}{\to} Im\, c(\boldsymbol{r};\theta_0).$$

For any $\delta > 0$, consider

$$\begin{aligned} I_n(\theta_0 \pm \delta) - I_n(\theta_0) \;=\; & \int \cdots \int \Big\{ \Big[Re\, c(\boldsymbol{r};\theta_0 \pm \delta) - Re\, c(\boldsymbol{r};\theta_0) \Big] \\ & \Big[Re\, c(\boldsymbol{r};\theta_0 \pm \delta) - Re\, c(\boldsymbol{r};\theta_0) - 2Re\, c_n(\boldsymbol{r}) \Big] \\ & + \Big[Im\, c(\boldsymbol{r};\theta_0 \pm \delta) - Im\, c(\boldsymbol{r};\theta_0) \Big] \\ & \Big[Im\, c(\boldsymbol{r};\theta_0 \pm \delta) - Im\, c(\boldsymbol{r};\theta_0) - 2Im\, c_n(\boldsymbol{r}) \Big] \Big\}\, dG(\boldsymbol{r}). \end{aligned}$$

Taking expectation, we have

$$\begin{aligned} & E\{I_n(\theta_0 \pm \delta) - I_n(\theta_0)\} \\ = \; & \int \cdots \int \Big\{ \Big[Re\, c(\boldsymbol{r};\theta_0 \pm \delta) - Re\, c(\boldsymbol{r};\theta_0) \Big]^2 + \Big[Im\, c(\boldsymbol{r};\theta_0 \pm \delta) - Im\, c(\boldsymbol{r};\theta_0) \Big]^2 \Big\}\, dG(\boldsymbol{r}) \\ > \; & 0, \end{aligned}$$

and it follows from the strong law of large numbers that the inequality $I_n(\theta_0 \pm \delta) > I_n(\theta_0)$ holds almost surely. By Assumption (A3), we have

$$\tilde{\theta}_n \stackrel{a.s.}{\to} \theta_0.$$

This is the result of the strong consistency. ■

Theorem 1.3.2 *For any fixed p, if $\tilde{\theta}_n \xrightarrow{a.s.} \theta_0$, under assumptions A6-A8, the ECF estimator by minimizing* (3.1) *is asymptotically normal, i.e.,*

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \xrightarrow{D} N(0, W^2),$$

where the expression of W^2 is given in the proof.

Proof: Consider the first order condition of the minimization problem (3.1) and by Assumption (A2), we have,

$$\begin{aligned} I_n'(\theta) &= 2\int\cdots\int\Big\{\Big[Re\, c_n(\boldsymbol{r}) - Re\, c(\boldsymbol{r};\theta)\Big]Re\, c'(\boldsymbol{r};\theta) \\ &\quad +\Big[Im\, c_n(\boldsymbol{r}) - Im\, c(\boldsymbol{r};\theta)\Big]Im\, c'(\boldsymbol{r};\theta)\Big\}\, dG(\boldsymbol{r}) \\ &= -\frac{2}{n}\sum_{j=1}^{n}\int\cdots\int\Big\{\Big[cos(\boldsymbol{r}'\boldsymbol{x}_j) - Re\, c(\boldsymbol{r};\theta)\Big]Re\, c'(\boldsymbol{r};\theta) \\ &\quad +\Big[sin(\boldsymbol{r}'\boldsymbol{x}_j) - Im\, c(\boldsymbol{r};\theta)\Big]Im\, c'(\boldsymbol{r};\theta)\Big\}\, dG(\boldsymbol{r}) \\ &= 0. \end{aligned} \tag{3.5}$$

Define

$$\begin{aligned} K_j(\theta) &= \int\cdots\int\Big\{\Big[cos(\boldsymbol{r}'\boldsymbol{x}_j) - Re\, c(\boldsymbol{r};\theta)\Big]Re\, c'(\boldsymbol{r};\theta) \\ &\quad +\Big[sin(\boldsymbol{r}'\boldsymbol{x}_j) - Im\, c(\boldsymbol{r};\theta)\Big]Im\, c'(\boldsymbol{r};\theta)\Big\}\, dG(\boldsymbol{r}), \end{aligned} \tag{3.6}$$

hence $I_n'(\theta)$ is the partial sum of random sequence $\{K_1(\theta), K_2(\theta), \cdots\}$ multiplied by a constant. Of course the random sequence is not independently distributed since $\boldsymbol{x}_j's$ are correlated with each other. However, we may show that Assumption (A7) implies the ρ mixing for the sequence $\{K_1(\theta), K_2(\theta), \cdots\}$. Calculating the variance of the partial sum, we have

$$\frac{1}{n^2}Var(K_1(\theta) + \cdots + K_n(\theta)) \tag{3.7}$$

$$= \int \cdots \int \int \cdots \int \Big\{ Re\, c'(\boldsymbol{r};\theta) Re\, c'(\boldsymbol{s};\theta) \frac{1}{n^2} \sum_{j=1}^{n}\sum_{k=1}^{n} Cov\Big(cos(\boldsymbol{r}'\boldsymbol{x}_j), cos(\boldsymbol{s}'\boldsymbol{x}_k)\Big)$$

$$+Re\, c'(\boldsymbol{r};\theta) Im\, c'(\boldsymbol{s};\theta) \frac{2}{n^2} \sum_{j=1}^{n}\sum_{k=1}^{n} Cov\Big(cos(\boldsymbol{r}'\boldsymbol{x}_j), sin(\boldsymbol{s}'\boldsymbol{x}_k)\Big) +$$

$$Im\, c'(\boldsymbol{r};\theta) Im\, c'(\boldsymbol{s};\theta) \frac{1}{n^2} \sum_{j=1}^{n}\sum_{k=1}^{n} Cov\Big(sin(\boldsymbol{r}'\boldsymbol{x}_j), sin(\boldsymbol{s}'\boldsymbol{x}_k)\Big) \Big\} dG(\boldsymbol{r}) dG(\boldsymbol{s}).$$

If we define

$$\Psi_k(\boldsymbol{r},\boldsymbol{s}) = E[\exp(i\boldsymbol{r}'\boldsymbol{x}_1 + i\boldsymbol{s}'\boldsymbol{x}_{k+1})],$$

we can rewrite the covariance as,

$$\frac{1}{n^2} \sum_{j=1}^{n}\sum_{k=1}^{n} Cov\Big(cos(\boldsymbol{r}'\boldsymbol{x}_j), cos(\boldsymbol{s}'\boldsymbol{x}_k)\Big)$$

$$= \frac{1}{2n}(Re\, c(\boldsymbol{r}+\boldsymbol{s}) + Re\, c(\boldsymbol{r}-\boldsymbol{s})) - Re\, c(\boldsymbol{r}) Re\, c(\boldsymbol{s}) + \frac{1}{2n^2}\sum_{k=1}^{n-1}(n-k)(Re\,\Psi_k(\boldsymbol{r},\boldsymbol{s})$$

$$+Re\,\Psi_k(\boldsymbol{r},-\boldsymbol{s}) + Re\,\Psi_k(\boldsymbol{s},\boldsymbol{r}) + Re\,\Psi_k(\boldsymbol{s},-\boldsymbol{r})),$$

$$\frac{1}{n^2} \sum_{j=1}^{n}\sum_{k=1}^{n} Cov\Big(cos(\boldsymbol{r}'\boldsymbol{x}_j), sin(\boldsymbol{s}'\boldsymbol{x}_k)\Big)$$

$$= \frac{1}{2n}(Im\, c(\boldsymbol{r}-\boldsymbol{s}) + Im\, c(\boldsymbol{r}-\boldsymbol{s})) - Re\, c(\boldsymbol{r}) Re\, c(\boldsymbol{s}) + \frac{1}{2n^2}\sum_{k=1}^{n-1}(n-k)(Im\,\Psi_k(\boldsymbol{r},\boldsymbol{s})$$

$$-Im\,\Psi_k(\boldsymbol{r},-\boldsymbol{s}) + Im\,\Psi_k(\boldsymbol{s},\boldsymbol{r}) + Im\,\Psi_k(\boldsymbol{s},-\boldsymbol{r})),$$

$$\frac{1}{n^2} \sum_{j=1}^{n}\sum_{k=1}^{n} Cov\Big(sin(\boldsymbol{r}'\boldsymbol{x}_j), sin(\boldsymbol{s}'\boldsymbol{x}_k)\Big)$$

$$= \frac{1}{2n}(Re\, c(\boldsymbol{r}+\boldsymbol{s}) + Re\, c(\boldsymbol{r}-\boldsymbol{s})) - Im\, c(\boldsymbol{r}) Im\, c(\boldsymbol{s}) + \frac{1}{2n^2}\sum_{k=1}^{n-1}(n-k)(Re\,\Psi_k(\boldsymbol{r},-\boldsymbol{s})$$

$$-Re\,\Psi_k(\boldsymbol{r},\boldsymbol{s}) + Re\,\Psi_k(\boldsymbol{s},-\boldsymbol{r}) - Re\,\Psi_k(\boldsymbol{s},\boldsymbol{r})),$$

where $c(\boldsymbol{r}) = c(\boldsymbol{r};\theta)$. Assumption (A7) implies the convergence of

$$\sum_{k=1}^{\infty} Cov\Big(\exp(i\boldsymbol{r}'\boldsymbol{x}_1), \exp(i\boldsymbol{r}'\boldsymbol{x}_{k+1})\Big).$$

Consequently, we can define Σ^2 as

$$\Sigma^2 = \lim_{n\to\infty} \frac{1}{n^2} Var(K_1(\theta_0) + \cdots + K_n(\theta_0)). \tag{3.8}$$

Using a central limit theorem for stationary processes (for example, Theorem 18.5.4 in Ibragimov and Linnik (1971)), we have

$$n^{1/2} I_n'(\theta_0) \xrightarrow{D} N(0, 4\Sigma^2). \tag{3.9}$$

Consider the second derivative of $I_n(\theta)$

$$\begin{aligned} I_n''(\theta) &= 2\int \cdots \int \Big\{ [Re\, c'(\boldsymbol{r};\theta)]^2 + [Im\, c'(\boldsymbol{r};\theta)]^2 - [Re\, c_n(\boldsymbol{r}) - Re\, c(\boldsymbol{r};\theta)] Re\, c''(\boldsymbol{r};\theta) \\ &\quad - [Im\, c_n(\boldsymbol{r}) - Im\, c(\boldsymbol{r};\theta)] Im\, c''(\boldsymbol{r};\theta) \Big\} \, dG(\boldsymbol{r}), \end{aligned} \tag{3.10}$$

and by the S.L.L.N., we have

$$I_n''(\theta_0) \xrightarrow{a.s.} E[I_n''(\theta_0)] = 2\int \cdots \int |c'(\boldsymbol{r};\theta_0)|^2 \, dG(\boldsymbol{r}).$$

The Taylor expansion

$$I_n'(\tilde{\theta}_n) = I_n'(\theta_0) + (\tilde{\theta}_n - \theta_0) + I_n''(\theta_0 + \epsilon(\tilde{\theta}_n - \theta_0))$$

implies

$$n^{1/2}(\tilde{\theta}_n - \theta_0) = -\frac{n^{1/2} I_n'(\theta_0)}{I_n''(\theta_0 + \epsilon(\tilde{\theta}_n - \theta_0))} \xrightarrow{D} N\Big(0, W^2\Big), \tag{3.11}$$

where

$$W^2 = \frac{\Sigma^2}{[\int \cdots \int |c'(\boldsymbol{r};\theta_0)|^2 \, dG(\boldsymbol{r})]^2}. \tag{3.12}$$

This is the result of the asymptotic normality. ■

Chapter 2

ESTIMATION OF LINEAR TIME SERIES MODELS VIA EMPIRICAL CHARACTERISTIC FUNCTION

2.1 Introduction

In this chapter, we will apply the ECF method proposed in Chapter 1 to estimate the Gaussian ARMA models, and examine the finite sample properties of the ECF method and compare it to the traditional estimation methods such as the maximum likelihood (ML) method and the conditional maximum likelihood (CML) method. The conclusion is that the ECF method can work as well as the maximum likelihood method and outperform the conditional maximum likelihood method. In Section 2.2 we review the traditional estimation methods for the Gaussian ARMA models. In Section 2.3 we use the ECF method to estimate the Gaussian ARMA models. To use the ECF method most efficiently, the optimal weights and estimating equations for the Gaussian ARMA model are derived in this section. Section 2.4 performs Monte Carlo studies and compares the finite sample properties of the ECF method with the those of the traditional estimation methods.

2.2 Review of Estimation of Gaussian ARMA Models

Consider the Gaussian ARMA(l, m) model,

$$Y_t = \rho_1 Y_{t-1} + \cdots + \rho_l Y_{t-1-l} + \varepsilon_t - \phi_1 \varepsilon_{t-1} - \cdots - \phi_m \varepsilon_{t-1-m}, \tag{2.1}$$

where $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. Let $\boldsymbol{\theta} = (\sigma^2, \phi_1, \cdots, \phi_l, \rho_1, \cdots, \rho_m)'$ denote the population parameters to be estimated. $\{y_1, ..., y_T\}$ is a realization. Three facts about the Gaussian ARMA(l, m) model are reviewed here. Firstly, some assumptions are needed for the stationary property of the model. For example, If $|\phi_1| < 1$, the ARMA(1,0) model (i.e., the AR(1) model) is stationary. In this chapter we assume all the series to be stationary. Secondly, the stationary Gaussian ARMA models satisfy the ρ mixing condition. Finally, the maximum likelihood approach provides the most efficient estimators, at least asymptotically. Denote the covariance matrix of $\boldsymbol{y} = (y_1, ..., y_T)$ by $\sigma^2\Phi_{T\times T}$. It is straightforward to obtain the log likelihood function,

$$L(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}log|\Phi| - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\boldsymbol{y}'\Phi^{-1}\boldsymbol{y}. \tag{2.2}$$

Maximizing the log likelihood function results in the MLE. Let $\hat{\boldsymbol{\theta}}_n$ be the MLE of $\boldsymbol{\theta}$, then $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \stackrel{D}{\rightarrow} N(0, I^{-1}(\boldsymbol{\theta}))$, where $I^{-1}(\boldsymbol{\theta})$ is the Cramér-Rao lower bound. Hence the MLE is asymptotically most efficient. Unfortunately, in practice the maximum likelihood approach by the use of inverting the covariance matrix is not always feasible, or is feasible but numerically not efficient. Consequently a state space representation of likelihood and then the Kalman Filter technique could be used. See Hamilton (1994) for more references. Alternatively, in this section we discuss three Gaussian ARMA models, only two of which can be easily estimated by the full maximum likelihood method via inversion the covariance matrix.

2.2.1 MA(1) Model

Consider the Gaussian MA(1) model

$$Y_t = \varepsilon_t - \phi\varepsilon_{t-1}, \tag{2.3}$$

where $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$, and $\boldsymbol{\theta} = (\sigma^2, \phi)$ is the population parameter. For the MA(1) model, we can write Φ as,

$$\begin{bmatrix} 1+\phi^2 & -\phi & 0 & \cdots & 0 \\ -\phi & 1+\phi^2 & -\phi & \cdots & 0 \\ 0 & -\phi & 1+\phi^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\phi^2 \end{bmatrix}. \tag{2.4}$$

Whittle (1983) has presented the exact inverse of Φ. The $(j,k)^{th}$ element of Φ^{-1} is given by

$$\frac{(\phi^{-j-1} - \phi^{j+1})(\phi^{-T+k} - \phi^{T-k})}{\phi(\phi^{-1} - \phi)(\phi^{-T-1} - \phi^{T+1})}, \tag{2.5}$$

where $k = 0, 1, ..., T-1$, and $j = 0, 1, ..., k-1$. Furthermore the determinant of Φ can be expressed as

$$|\Phi| = 1 + \phi^2 + \phi^4 + ... + \phi^{2T} = \frac{1 - \phi^{2(T+1)}}{1 - \phi^2}. \tag{2.6}$$

Even though an analytical solution for maximizing the exact likelihood function (2.2) is not readily found, it is straightforward to maximize it numerically given (2.4) and (2.5) and leads to the MLE. It should be noted, however, the exact inverse of Φ is complicated for high-order moving-average models. Because the exact likelihood function is computationally expensive to evaluate for high-order MA processes, in practice the conditional likelihood function is maximized instead. We review this approach for the MA(1) model in detail.

If the value of ε_{t-1} were known with certainty, then

$$Y_t | \varepsilon_{t-1} \sim N(\phi\varepsilon_{t-1}, \sigma^2). \tag{2.7}$$

If the initial value of ε_0 is set to be the expected value, i.e., $\varepsilon_0 = 0$, the conditional likelihood function can be obtained by

$$f_{Y_T,...,Y_1|\varepsilon_0=0}(y_t, \ldots, y_1) = f_{Y_1|\varepsilon_0=0}(y_1)\Pi_{t=2}^{T} f_{Y_t|\varepsilon_{t-1}}(y_t|\varepsilon_{t-1}), \tag{2.8}$$

where the sequence $\{\varepsilon_1, \cdots, \varepsilon_T\}$ can be calculated from $\{y_1, \cdots, y_T\}$ by the following iterative formula

$$\varepsilon_t = y_t - \phi y_{t-1} + \phi^2 y_{t-2} + \cdots + (-1)^{t-1}\phi^{t-1}y_1 + (-1)^t\phi^t\varepsilon_0. \tag{2.9}$$

In this thesis the resulting estimator by maximizing the conditional likelihood such as (2.8) is labeled "conditional maximum likelihood estimator (CMLE)".

Note that the effect of imposing $\varepsilon_0 = 0$ will quickly die out when $|\phi|$ is substantially less than unity, thus equation (2.8) will give a good approximation to the unconditional likelihood for a reasonably large sample size.

2.2.2 AR(1) Model

Consider the Gaussian AR(1) model

$$Y_t = \rho Y_{t-1} - \varepsilon_t, \tag{2.10}$$

where $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$, and $\boldsymbol{\theta} = (\sigma^2, \rho)$ is the population parameter. For the AR(1) model, we can write Φ as,

$$\frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix}. \tag{2.11}$$

The exact inverse of Φ is

$$\begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}. \tag{2.12}$$

And the determinant of Φ is

$$|\Phi| = \frac{1}{1-\rho^2}. \tag{2.13}$$

Consider the first order conditions of maximizing (2.2) using (2.12) and (2.13). If we concentrate out σ^2, we have the following estimating equation for ρ,

$$\begin{aligned} 0 &= 2\rho(y_1^2 + y_T^2) + 2\rho(T+1+(1-T)\rho^2)\sum_{t=1}^{T-1} y_i^2 \\ &\quad -(4\rho^2 - 2T + 2T\rho^2)\sum_{t=1}^{T-1} y_t y_{t+1}. \end{aligned} \tag{2.14}$$

The real root of the above cubic equation leads to the MLE of ρ.

The CMLE of ρ for the AR(1) model can be obtained by

$$\max f_{Y_T,\ldots,Y_2|Y_1}(y_t,\ldots,y_2).$$

The procedure is equivalent to the OLS, i.e.,

$$\min_\rho \sum_{t=2}^{T}(y_t - \rho y_{t-1})^2. \tag{2.15}$$

Since the first observation in the OLS procedure is ignored, the finite sample properties of the MLE must be at least as good than those of the CMLE. However, if the sample size T is sufficiently large, the first observation makes a negligible contribution to the total likelihood. The MLE and the CMLE turn out to have the same large sample properties. Because the CMLE yields an analytical form for the estimators, it is used often in application when T is large.

2.2.3 Gaussian ARMA(1, 1) Model

Consider the Gaussian ARMA(1, 1) model

$$Y_t = \rho Y_{t-1} + \varepsilon_t - \phi\varepsilon_{t-1} \tag{2.16}$$

where $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$, and $\boldsymbol{\theta} = (\sigma^2, \rho, \phi)$ is the population parameter. For the Gaussian ARMA(1, 1) model, we can write Φ as,

$$\begin{bmatrix} \frac{1+\phi^2-2\rho\phi}{1-\rho^2} & \frac{\rho(1+\phi^2-2\rho\phi)}{1-\rho^2} - \phi & \cdots & \frac{\rho^{T-1}(1+\phi^2-2\rho\phi)}{1-\rho^2} - \rho^{T-2}\phi \\ \frac{\rho(1+\phi^2-2\rho\phi)}{1-\rho^2} - \phi & \frac{\rho(1+\phi^2-2\rho\phi)}{1-\rho^2} - \phi & \cdots & \frac{\rho^{T-2}(1+\phi^2-2\rho\phi)}{1-\rho^2} - \rho^{T-3}\phi \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\rho^{T-1}(1+\phi^2-2\rho\phi)}{1-\rho^2} - \rho^{T-2}\phi & \frac{\rho^{T-2}(1+\phi^2-2\rho\phi)}{1-\rho^2} - \rho^{T-3}\phi & \cdots & \frac{1+\phi^2-2\rho\phi}{1-\rho^2} \end{bmatrix}.$$

The AR(1) and the MA(1) are two special cases of the Gaussian ARMA(1, 1) with $\phi = 0$ and $\rho = 0$ respectively. However, for general (ρ, ϕ), there is no closed form for the inverse of Φ. Inverting this $T \times T$ matrix is time consuming when T is large, hence implementation of maximum likelihood method is time consuming for a large T. Because of the difficulties involved in calculating the exact likelihood function, in practice the conditional likelihood function is maximized instead. We present the method in detail.

It is common to obtain the conditional likelihood function conditions on both y and ε. One option is to set initial y and ε equal to their expected value. That is. $y_0 = 0$ and $\varepsilon_0 = 0$. The conditional likelihood function can be obtained by

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= f_{Y_T,\dots,Y_1|y_0=0,\varepsilon_0=0}(y_t, \dots, y_1|y_0 = 0, \varepsilon_0 = 0) \\ &= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \sum_{t=1}^{T}\frac{\varepsilon_t^2}{2\sigma^2}, \end{aligned} \tag{2.17}$$

where the sequence $\{\varepsilon_1, \cdots, \varepsilon_T\}$ can be calculated from the realization $\{y_1, \cdots, y_T\}$ by the following iterative formula

$$\varepsilon_t = y_t - \rho y_{t-1} + \phi\varepsilon_{t-1}. \tag{2.18}$$

Maximizing the conditional likelihood leads to the CMLE and the procedure is asymptotically equivalent to maximizing the exact likelihood because the initial conditions have negligible effect to the total likelihood when the sample size is very large.

2.3 Estimation of Gaussian ARMA Models Via ECF

As we point out in Chapter 1, to apply the ECF method to estimate a time-series model, both p and the weight function $G(\boldsymbol{r})$ need to be specified. Regarding the choice of p, we know that the moving blocks with larger p retain more information and hence the ECF method can work better. Regarding the weight function, we can choose it to be a step function, as Knight and Satchell (1996) proposed. Furthermore, we can choose any other weight function. As long as the regularity conditions in Chapter 1 hold, the resulting estimators are strongly consistent and asymptotically normally distributed. Consequently, three alternative weight functions are considered in this chapter to estimate the Gaussian ARMA models. The simplest one is to choose $g(\boldsymbol{r})$ to be 1 in Equation (3.2) of Chapter 1 and is indeed the OLS-CECF method. We use the WLS-CECF method by choosing $g(\boldsymbol{r}) = \exp(-a\boldsymbol{r}'\boldsymbol{r})$, where a is a non-negative constant. The exponential function is chosen for two reasons. First, it is a generalization of the weight proposed by Paulson *et al* (1975) and put more weight on the interval around the origin, consistent with the theory that the CF contains the most information around the origin. The second reason is for computational convenience. By choosing the exponential weight for the ECF method in Gaussian ARMA models, we can obtain the closed form expression for $I_n(\boldsymbol{\theta})$ given by Equation (3.2) in Chapter 1. In fact the OLS-CECF is a particular case of the WLS-CECF. We can show that for any exponential weight function $g(\boldsymbol{r})$ and any p, the assumptions in Chapter 1 hold for stationary Gaussian ARMA models. Therefore, the ECF estimators based on any exponential weight are strongly consistent and asymptotically normally distributed. The following theorem states that there is closed form expression for $I_n(\boldsymbol{\theta})$ for any stationary Gaussian ARMA model with any value of p.

Theorem 2.3.1 *If* $\{y_1, ..., y_T\}$ *is a finite realization of the Gaussian ARMA model* *defined by* (2.16) *with* $\boldsymbol{x}_j$, $c(\boldsymbol{r};\boldsymbol{\theta})$ *and* $c_n(\boldsymbol{r})$ *defined as before, there exits a closed form function to represent the following integral*

$$\int\cdots\int |c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta})|^2 \exp(-a\boldsymbol{r}'\boldsymbol{r})\, d\boldsymbol{r},$$

and the expression for this function is given in Appendix A.

Proof: See Appendix A.

We also choose $w_{\boldsymbol{\theta}}(\boldsymbol{r})$ in Equation (3.3) of Chapter 1 to be

$$(\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\frac{\partial \log f(y_{j+p}|y_j,\cdots,y_{j+p-1})}{\partial\boldsymbol{\theta}}\, dy_j\cdots dy_{j+P}. \qquad (3.1)$$

After solving the following estimating equation

$$\int c_n(\boldsymbol{r})w_{\boldsymbol{\theta}}(\boldsymbol{r})\, d\boldsymbol{r} = 0, \qquad (3.2)$$

we get the GLS-CECF estimator which is an efficient estimator provided p is large enough in the sense that the resulting estimator can achieve the Cramér-Rao lower bound; See Feuerverger and McDunnough (1981b). To obtain the more efficient ECF estimator, in Theorem 2.3.2 we derive the optimal weight functions and the estimating equations of the GLS-CECF method for a general Gaussian ARMA(l, m) model with any value of p.

Theorem 2.3.2 *Assume* $\{y_1, y_2, ..., y_T\}$ *to be a finite realization of the Gaussian ARMA model defined by* (2.16) *with* $\boldsymbol{\theta} = (\sigma^2, \boldsymbol{\rho}) = (\sigma^2, \rho_1, \cdots, \rho_l, \phi_1, \cdots, \phi_m)$. $\boldsymbol{x}_j$, $c(\boldsymbol{r};\boldsymbol{\theta})$ *and* $c_n(\boldsymbol{r})$ *are defined as before. Suppose the conditional density of* $y_{j+p}|y_j, \ldots, y_{j+p-1}$ *could be expressed as*

$$(y_{j+p}|y_j,\ldots,y_{j+p-1}) \sim N\left(f_1(\boldsymbol{\rho})y_j + \cdots + f_p(\boldsymbol{\rho})y_{j+p-1}, \sigma^2 g(\boldsymbol{\rho})\right),$$

where $g(\boldsymbol{\rho})$, $f_1(\boldsymbol{\rho})$, $\cdots$, $f_p(\boldsymbol{\rho}) \in C^1$. *Let*

$$A = \begin{pmatrix} f_1^2(\boldsymbol{\rho}) & f_1(\boldsymbol{\rho})f_2(\boldsymbol{\rho}) & \cdots & f_1(\boldsymbol{\rho})f_{p-1}(\boldsymbol{\rho}) & -f_1(\boldsymbol{\rho}) \\ f_2(\boldsymbol{\rho})f_1(\boldsymbol{\rho}) & f_2^2(\boldsymbol{\rho}) & \cdots & f_2(\boldsymbol{\rho})f_{p-1}(\boldsymbol{\rho}) & -f_2(\boldsymbol{\rho}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f_{p-1}(\boldsymbol{\rho})f_1(\boldsymbol{\rho}) & f_{p-1}(\boldsymbol{\rho})f_2(\boldsymbol{\rho}) & \cdots & f_{p-1}^2(\boldsymbol{\rho}) & -f_{p-1}(\boldsymbol{\rho}) \\ -f_1(\boldsymbol{\rho}) & -f_2(\boldsymbol{\rho}) & \cdots & -f_{p-1}(\boldsymbol{\rho}) & 1 \end{pmatrix}. \tag{3.3}$$

and

$$\begin{cases} B_1 & = \frac{\partial A}{\partial \phi_1} \\ \vdots & \\ B_{l+m} & = \frac{\partial A}{\partial \rho_m} \end{cases}$$

Furthermore, let $M\Lambda M', H_1\Lambda_1 H_1', \cdots, H_{l+m}\Lambda_{l+m}H_{l+m}'$ *be the eigenvalue decomposition of* $A, B_1, \cdots, B_{l+m}$ *respectively. Hence,* $M, H_1, \cdots, H_{l+m}$ *are orthonormal matrices and* $\Lambda, \Lambda_1, \cdots, \Lambda_{l+m}$ *are diagonal matrices with the eigenvalues of* $A. B_1, \cdots. B_{l+m}$ *respectively. Define*

$$\Lambda = \begin{pmatrix} \lambda^1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda^{p+1} \end{pmatrix},$$

and

$$\Lambda_k = \begin{pmatrix} \lambda_k^1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_k^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_k^{p+1} \end{pmatrix}.$$

for $k = 1, \cdots, l+m$. *Then the optimal weight functions are*

$$\begin{aligned} w_{\sigma^2}(\boldsymbol{r}) &= -\frac{1}{2\sigma^2}\delta(r^1)\cdots\delta(r^{p+1}) \\ &\quad -\frac{1}{2\sigma^4 g(\boldsymbol{\rho})}[\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1}) + \cdots \\ &\quad +\lambda^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1})], \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} w_{\boldsymbol{\rho}_k}(\boldsymbol{r}) &= -\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})}\delta(r^1)\cdots\delta(r^{p+1}) \\ &\quad -\frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}[\lambda_k^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1})+\cdots \\ &\quad +\lambda_k^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1})] \\ &\quad +\frac{1}{2\sigma^2 g(\boldsymbol{\rho})}[\lambda_k^1\delta''(t_k^1)\delta(t_k^2)\cdots\delta(t_k^{p+1})+\cdots \\ &\quad +\lambda_k^{p+1}\delta(t_k^1)\cdots\delta(t_k^p)\delta''(t_k^{p+1})], \end{aligned} \tag{3.5}$$

where $(s_1,\cdots,s_{p+1})' = M'\boldsymbol{r}$ *and* $(t_k^1,\cdots,t_k^{p+1})' = H_k'\boldsymbol{r}$ *with* $k = 1,\cdots,l+m$. *The delta function is the Dirac delta function defined as in Chapter 1 and* g_k *is the partial derivative of* $g(\boldsymbol{\rho})$ *with respect to* $\boldsymbol{\rho}_k$. *And the estimating equations are,*

$$\sigma^2 = \frac{\bar{P}(\boldsymbol{\rho})}{g(\boldsymbol{\rho})}, \tag{3.6}$$

$$\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})} - \frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}\bar{P}(\rho) + \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\bar{Q}_k(\rho) = 0. \tag{3.7}$$

where

$$\bar{P}(\rho) = \frac{1}{n}\sum_{j=1}^{n}[\lambda^1((M'\boldsymbol{x}_j)_{(1)})^2 + \cdots + \lambda^{p+1}((M'\boldsymbol{x}_j)_{(p+1)})^2],$$

and

$$\bar{Q}_k(\rho) = \frac{1}{n}\sum_{j=1}^{n}(\lambda_k^1((H_k'\boldsymbol{x}_j)_{(1)})^2 + \cdots + \lambda_k^{p+1}((H_k'\boldsymbol{x}_j)_{(p+1)})^2),$$

with $k = 1,\cdots,l+m$. *Combine* (3.6) *and* (3.7) *we have*

$$\begin{cases} \bar{Q}_1(\boldsymbol{\rho}) = 0 \\ \bar{Q}_2(\boldsymbol{\rho}) = 0 \\ \vdots \\ \bar{Q}_{l+m}(\boldsymbol{\rho}) = 0. \end{cases} \tag{3.8}$$

which determine the estimators of $\boldsymbol{\rho}$. *The estimator of* σ^2 *can be easily found by substituting the estimators of* $\boldsymbol{\rho}$ *back into the equation* (3.6).

Proof: See Appendix B.

Two examples are illustrated for the implementation of Theorem (2.3.2). For the MA(1) model defined by (2.3), if we choose $p = 1$, we have

$$y_{j+1}|y_j \sim N\left(-\frac{\phi}{1+\phi^2}y_j, \sigma^2\frac{1+\phi^2+\phi^4}{1+\phi^2}\right).$$

Hence,

$$A = \begin{pmatrix} \frac{\phi^2}{(1+\phi^2)^2} & \frac{\phi}{1+\phi^2} \\ \frac{\phi}{1+\phi^2} & 1 \end{pmatrix},$$

$$B = \begin{pmatrix} \frac{2\phi(1-\phi^2)}{(1+\phi^2)^3} & \frac{1-\phi^2}{(1+\phi^2)^2} \\ \frac{1-\phi^2}{(1+\phi^2)^2} & 0 \end{pmatrix},$$

and

$$g(\phi) = \frac{1+\phi^2+\phi^4}{1+\phi^2}.$$

Let λ_1^1, λ_1^2 be two eigenvalue-values of B, we have

$$\lambda_1^1 = \frac{\phi - \phi^5 + \sqrt{1+2\phi^2-\phi^4-6\phi^6-\phi^8+3\phi^{10}+\phi^{12}}}{1+4\phi^2+6\phi^4+4\phi^6+\phi^8},$$

and

$$\lambda_1^2 = \frac{\phi - \phi^5 - \sqrt{1+2\phi^2-\phi^4-6\phi^6-\phi^8+3\phi^{10}+\phi^{12}}}{1+4\phi^2+6\phi^4+4\phi^6+\phi^8}.$$

The estimating equation for ϕ is,

$$\begin{aligned} 0 \;=\; & \sum_{j=1}^{n}\left\{(\lambda_1^{1^3}+\lambda_1^{2^3})\frac{(1+\phi^2)^4}{(1-\phi^2)^2}y_j^2 + 2(\lambda_1^{1^2}+\lambda_1^{2^2})\frac{(1+\phi^2)^2}{(1-\phi^2)}y_jy_{j+1}\right. \\ & \left. +(\lambda_1^1+\lambda_1^2)y_{j+1}^2\right\}. \end{aligned} \tag{3.9}$$

The estimating equation for σ^2 is,

$$\sigma^2 = \frac{1}{n}\sum_{j=1}^{n}\frac{[\phi y_j + (1+\phi^2)y_{j+1}]^2}{(1+\phi^2)(1+\phi^2+\phi^4)}. \tag{3.10}$$

In the second example we choose $p = 1$ for the AR(1) model which is defined by the equation (2.10). In this example,

$$y_{j+1}|y_j \sim N\left(\rho y_j, \sigma^2\right).$$

Hence,

$$A = \begin{pmatrix} \rho^2 & -\rho \\ -\rho & 1 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 2\rho & -1 \\ -1 & 0 \end{pmatrix}.$$

The estimating equation of ρ is,

$$(4\rho^3 + 3\rho)y_1^2 + \rho y_{n+1}^2 + 4\rho(1+\rho^2)\sum_{i=2}^{n} y_i^2 = 2(2\rho^2+1)\sum_{i=1}^{n} y_i y_{i+1}. \tag{3.11}$$

It should be stressed that, different from the discrete ECF method, in the continuous ECF method we do not need to choose the transformation variables, $\boldsymbol{r}$'s, since they are simply integrated out.

2.4 Monte Carlo Studies

Up to now, we have reviewed the DECF method proposed by Knight and Satchell (1996) and proposed the OLS-CECF method, the WLS-CECF method, and the GLS-CECF method. Although Knight and Satchell (1996) estimate an MA(1) model by using the DECF method, the step function is arbitrarily chosen and hence the weight is not necessarily the best. In fact, the multi-step procedure they propose has not been used to estimate a stationary time series model in practice.

In this section Monte Carlo studies are designed to compare the ECF estimator with the MLE for the MA(1) model and the AR(1) model, and with the CMLE for the Gaussian ARMA(1, 1) model.

The first experiment involves 1,000 replications of generated data from an MA(1) model with $\phi = 0.6$, $\sigma^2 = 1.0$ and the number of observations set at $T = 100$. To use the DECF proposed by Knight and Satchell (1996), one has to determine the value of p, the value of q, and the criterion to obtain the optimal $\boldsymbol{r}_1, \cdots, \boldsymbol{r}_q$. In this Monte Carlo study we choose p to be 2 since the autocovariances are zero in an MA(1) for lags greater than one and q to be 5 to guarantee that the number of elements in the vector V_n and V is larger than the number of parameters. In this multi-parameter case, we choose the criterion function in two ways, i.e., minimizing the determinant of the asymptotic covariance matrix and minimizing the trace of the asymptotic covariance matrix. Without further restriction on $\boldsymbol{r}$'s, unfortunately, the minimization problem for obtaining optimal $\boldsymbol{r}$'s has 15 dimensions and hence is numerically very expensive. To simplify the matter and also for $\boldsymbol{r}$'s to be sufficiently fine and extended we have chosen $\boldsymbol{r}$'s in many ways. For example, we choose them in the following two ways,

$$\boldsymbol{r} = \begin{bmatrix} -4\tau & -2\tau & -\tau & 0 & 2\tau \\ -3\tau & -\tau & 0 & \tau & 3\tau \\ -2\tau & 0 & \tau & 2\tau & 4\tau \end{bmatrix}, \tag{4.1}$$

or

$$\boldsymbol{r} = \begin{bmatrix} -6\tau & -3\tau & -\tau & \tau & 4\tau \\ -5\tau & -2\tau & 0 & 2\tau & 5\tau \\ -4\tau & -\tau & \tau & 3\tau & 6\tau \end{bmatrix}, \tag{4.2}$$

where the behavior of $\boldsymbol{r}$'s is determined by only one parameter, i.e., τ.

By plotting the criterion function against τ, we find that two criterion functions with any one of two choices of $\boldsymbol{r}$'s are smooth and unimodal. This observation convinces us that the global minimization can be obtained numerically in all the cases.

In Table 2.1 we report the results from all four procedures of the DECF method. In c(1), $\boldsymbol{r}$'s are chosen to be (4.1) and the criterion function to be the determinant of the asymptotic covariance. In c(2) $\boldsymbol{r}$'s are chosen to be (4.1) and the criterion function to be the trace of the asymptotic covariance. In c(3) $\boldsymbol{r}$'s are chosen to be (4.2) and the criterion function to be the determinant of the asymptotic covariance. In c(4), $\boldsymbol{r}$'s are chosen to be (4.2) and the criterion function to be the trace of the asymptotic covariance. The exact MLE is presented as well.

By allowing two parameters in $\boldsymbol{r}$'s, we choose $\boldsymbol{r}$'s to be,

$$\boldsymbol{r} = \begin{bmatrix} -\tau_1 - 2\tau_2 & -\tau_1 - \tau_2 & -\tau_1 & -\tau_1 + \tau_2 & -\tau_1 + 2\tau_2 \\ -2\tau_2 & -\tau_2 & 0 & \tau_2 & 2\tau_2 \\ \tau_1 - 2\tau_2 & \tau_1 - \tau_2 & \tau_1 & \tau_1 + \tau_2 & \tau_1 + 2\tau_2 \end{bmatrix}, \qquad (4.3)$$

or

$$\boldsymbol{r} = \begin{bmatrix} -\tau_1 - 3\tau_2 & -\tau_1 - \tau_2 & -\tau_1 & -\tau_1 + \tau_2 & -\tau_1 + 3\tau_2 \\ -3\tau_2 & -\tau_2 & 0 & \tau_2 & 3\tau_2 \\ \tau_1 - 3\tau_2 & \tau_1 - \tau_2 & \tau_1 & \tau_1 + \tau_2 & \tau_1 + 3\tau_2 \end{bmatrix}. \qquad (4.4)$$

By plotting the criterion function against τ_1, τ_2, we find that two criterion functions with any one of two choices of $\boldsymbol{r}$'s are smooth and unimodal. This observation convinces us that the global minimization can be obtained numerically in all the cases.

In Table 2.2 we report the results from all four procedures of the DECF method. In c(1), $\boldsymbol{r}$'s are chosen to be (4.3) and the criterion function to be the determinant of the asymptotic covariance. In c(2) $\boldsymbol{r}$'s are chosen to be (4.3) and the criterion function to be the trace of the asymptotic covariance. In c(3) $\boldsymbol{r}$'s are chosen to be (4.4) and the criterion function to be the determinant of the asymptotic covariance. In c(4), $\boldsymbol{r}$'s are chosen to be (4.4) and the criterion function to be the trace of the asymptotic covariance. The exact MLE is presented as well.

A detailed examination of Table 2.1 and 2.2 reveals that the DECF estimates appear not to be much effected by the way of choosing $\boldsymbol{r}$'s, the criterion and how many parameters in $\boldsymbol{r}$'s. Furthermore, although the DECF estimates is a viable alternative to the MLE, the finite sample properties of the DECF estimates are dominated by those of the MLE. For instance, the variance of the DECF estimate of ϕ is at least three times larger than that of the MLE of ϕ. With one parameter allowed in $\boldsymbol{r}$'s, the averages of the optimal τ are 0.8, 0.59, 0.89, and 0.66 respectively. With two parameters allowed in $\boldsymbol{r}$'s, the averages of the optimal (τ_1, τ_2) are $(0.63, 1.16)$, $(0.63, 0.45)$, $(0.84, 0.88)$, and $(0.55, 0.21)$ respectively. All these numbers are small. The finding is consistent with the theory that the CF contains the most information around the origin.

To use the CECF method proposed in Chapter 1, one needs to determine the value of p and the weight function. For all the cases we first choose p to be 2. Table 2.3 reports the results from the CECF method. In c(1), the OLS-CECF is used. In c(2), the WLS-CECF with the weight $\exp(-18\boldsymbol{r}'\boldsymbol{r})$ is used. In c(3), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ is used. In c(4), the GLS-CECF with the optimal weight is used. The MLE is also reported for comparison.

From Table 2.3 it appears that both the OLS-CECF and the WLS-CECF are dominated by the GLS-CECF and MLE. The MLE performs better than the GLS. This finding suggests that $p = 2$ is not large enough for the moving blocks to retain the most information in the original sequence. Comparing Table 2.3 with Table 2.1 and 2.2 we note that there is a clear improvement over the DECF by using the GLS-CECF in terms of the mean square error, although the approach of choosing $\boldsymbol{r}$'s discretely is proposed by Knight and Satchell (1996), Quandt and Ramsey (1978), Schmidt (1982), and Tran (1994).

In Table 2.4 we report the estimates of ϕ by using the GLS-CECF method, however, the value of p has been increased to be $3, 4, \cdots, 10$. As we argued before, as p gets larger and larger, the asymptotic variance of the GLS-CECF estimators is getting closer and closer to that of the MLE which achieves the Cramér-Rao lower bound. Therefore, we should expect a larger p works better. This is confirmed by Table 2.4. From Table 2.4, however, we note that the GLS-CECF method with a small p can work quit well. For example, when comparing the GLS-CECF estimates with $p = 6$ and the MLE's, we note that the GLS-CECF method performs as well as the MLE. The variance and the skewness of them are almost identical, but the GLS-CECF provides a mean and median which are slightly closer to the true parameter value. Also see Figure 2.1.

The second experiment involves 1,000 replications of generated data from an MA(1) model with $\phi = 0.6$, $\sigma^2 = 1.0$ and the number of observations set at $T = 1,000$. This experiment is designed to demonstrate that with a small p both the OLS-CECF and the WLS-CECF method work well when we have a large number of observations. Table 2.5 reports the estimates of ϕ. In c(1), the OLS-CECF with $p = 1$ is used. In c(2), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 1$ is used. In c(3), the OLS-CECF with $p = 2$ is used. In c(4), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 2$ is used. In c(5), the OLS-CECF with $p = 3$ is used. In c(6), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 3$ is used. As we expect, with everything else being equal, the larger the value of p, the better the performance of the resulting estimates. Also the WLS-CECF performs slightly better than the OLS-CECF. Furthermore, $p = 2$ is large enough to confirm the viability of the OLS-CECF and the WLS-CECF method.

The third experiment involves 1,000 replications of generated data from an AR(1) model with $\rho = 0.6$, $\sigma^2 = 1.0$ and the number of observation set at $T = 100$. With $p = 1$ the GLS-CECF method is used to estimate the AR(1) model and compared with the maximum likelihood method presented in Section 2.2. See Table 2.6, Figure 2.2 and Figure 2.3. Interestingly, two approaches provide almost identical results. The results can be explained by the Markov property of the AR(1) model. Since the joint density of the AR(1) model for a length T has the following property

$$f(x_1, x_2, \cdots, x_T) = f(x_1)\Pi_{j=2}^{T} f(x_j|x_{j-1}),$$

it is not surprising that one-period-overlapping blocks can preserve all the information in the original sequence.

The fourth experiment involves 1,000 replications of generated data from an AR(1) model with $\rho = 0.6$, $\sigma^2 = 1.0$ and the number of observations set at $T = 1,000$. This experiment is designed to demonstrate that with a small p both the OLS-CECF and the WLS-CECF method work well when we have a large number of observations. Table 2.7 reports the ECF estimates of ρ. In c(1), the OLS-CECF with $p = 1$ is used. In c(2), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 1$ is used. In c(3), the OLS-CECF with $p = 2$ is used. In c(4), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 2$ is used. Note that there is no clear improvement by increasing p from 1 to 2. The finding can still be explained by the Markov property of the AR(1) model. The advantage of using the WLS-CECF method is clear and the viability of the OLS-CECF and the WLS-CECF method is confirmed.

The fifth experiment involves 1,000 replications of generated data from a Gaussian ARMA(1,1) model with $-\phi = \rho = 0.6$, $\sigma^2 = 1.0$ and the number of observations set at $T = 100$. It is easy to show that the inverse of Φ for the ARMA(1,1) model has no closed form even for $\phi = \rho$. Consequently, this experiment is designed to compare the performance of the GLS-CECF estimate with that of the CMLE presented in Section 2.2. Table 2.8 and Figure 2.4 report the results where in the GLS-CECF methods we choose $p = 2, 3$. Theoretically, the asymptotic variance of both the GLS-CECF estimate and the CMLE would converge to that of the MLE. In terms of the finite sample properties, from the Table 2.8, we note that the GLS-CECF estimate is clearly a viable alternative to the CMLE. For example, there appears a trade-off between the GLS-CECF estimates with $p = 2$ and the CMLE's. The mean and median of the GLS-CECF estimates are closer to the true parameter value and the GLS-CECF estimates are less skewed than the CMLE's while the variance of the GLS-CECF estimates is larger. However, comparing the GLS-CECF estimates with $p = 3$ to the CMLE's, we note that the variance and mean square error are almost identical, whereas the GLS-CECF estimates clearly dominate the CMLE's in terms of the other statistics. For instance, the GLS-CECF estimates with $p = 3$ have a mean and median which are closer to the true parameter value and show less skewness.

The findings can be explained as follows. If p is large enough, all the information in the original series. Theoretically, therefore, the GLS-CECF method results in an estimator which is asymptotically equivalent to the MLE. Furthermore, when $T \to \infty$ both estimators can achieve the Cramér-Rao lower bound. However, p does not necessarily to be too large in practice. A much smaller p sometimes provides good finite sample properties. On the other hand, some initial conditions must be assumed in order to obtain the CMLE. If the initial condition has an error, it is carried over into all the following stages by the recursive formula such as (2.18). Of course, the effect of such an error will diminish for the stationary processes and thus the CMLE is asymptotically equivalent to the MLE. However, the effect may not be negligible for a small number of observations.

The sixth experiment involves 1,000 replications of generated data from a Gaussian ARMA(1, 1) model with $-\phi = \rho = 0.6$, $\sigma^2 = 1.0$ and the number of observations set at $T = 1000$. This experiment is designed to demonstrate that with a small p both the OLS-CECF and the WLS-CECF method work well when we have a large number of observations. Table 2.9 reports the estimates of ρ. In c(1), the OLS-CECF with $p = 1$ is used. In c(2), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 1$ is used. In c(3) the OLS-CECF with $p = 2$ is used. In c(4), the WLS-CECF with the weight

$\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 2$ is used. In c(5), the OLS-CECF with $p = 3$ is used. In $c(6)$, the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 3$ is used. All cases confirm the viability of the OLS-CECF and the WLS-CECF method.

The seventh experiment involves 1,000 replications of generated data from a Gaussian ARMA$(1, 1)$ model with $-\phi = \rho = 0.9$, $\sigma^2 = 1.0$ and the number of observations set at $T = 100$. This experiment is designed to compare the performance of the GLS-CECF method with that of the CMLE presented in Section 2.2. Table 2.10 and Figure 2.5 report the results where we choose $p = 2, 3$ for the GLS-CECF method. Compared with the sequences generated in the fifth experiment, the sequences in this experiment have much longer memory. Consequently, the finite sample properties of both the CMLE and GLS-CECF will be affected. From Table 2.8, however, we note that the effect on the ECF method is smaller than that on the CMLE. For example, the ECF estimates for both $p = 2$ and $p = 3$ perform better than the CMLE's. The bias, the variance and the mean square error of the ECF estimates are smaller.

The advantages of using the ECF method are the following. Firstly, all the ECF methods presented in Section 2.3 do not require the calculation of the inverse of a $T \times T$ matrix. For most stationary processes, however, in order to use the full maximum likelihood estimation method, such inversion is needed but is infeasible in some situations. Secondly, implementation of the full maximum likelihood may require the closed form and boundness for the likelihood. No such requirement is needed for the DECF method, the OLS-CECF method, and the WLS-CECF method. Instead, the CF must have a tractable form. Such examples are listed in Chapter 1. Finally, even for the linear stationary processes such as the MA(1) model whose covariance matrix can be inverted analytically, the full maximum likelihood method is more computationally intensive than the GLS-CECF method. This is because the former method has to deal with a $T \times T$ matrix, while only a $(p + 1) \times (p + 1)$ matrix is required in the latter case.

Appendix A

Proof of Theorem 2.3.1

For any p, define $\boldsymbol{x}_j = (y_j, \cdots, y_{j+p})'$.

Let $\sigma^2\Phi_{(p+1)\times(p+1)}$ be the covariance matrix of $\boldsymbol{x}_j$, then we have

$$\begin{aligned}
&\int\cdots\int |c_n(\boldsymbol{r}) - c(\boldsymbol{r})|^2 \exp(-a\boldsymbol{r}'\boldsymbol{r})d\boldsymbol{r} \\
&= \int\cdots\int (c_n(\boldsymbol{r}) - c(\boldsymbol{r}))(\bar{c}_n(\boldsymbol{r}) - \bar{c}(\boldsymbol{r}))\exp(-a\boldsymbol{r}'\boldsymbol{r})\,d\boldsymbol{r} \\
&= \int\cdots\int \left[\frac{1}{n}\sum_{j=1}^{n}\exp(i\boldsymbol{r}'\boldsymbol{x}_j) - \exp(-\frac{\sigma^2}{2}\boldsymbol{r}'\Phi\boldsymbol{r})\right] \\
&\quad \left[\frac{1}{n}\sum_{j=1}^{n}\exp(-i\boldsymbol{r}'\boldsymbol{x}_j) - \exp(-\frac{\sigma^2}{2}\boldsymbol{r}'\Phi\boldsymbol{r})\right]\exp(-a\boldsymbol{r}'\boldsymbol{r})\,d\boldsymbol{r} \\
&= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\int\cdots\int \exp(i\boldsymbol{r}'\boldsymbol{x}_j)\exp(i\boldsymbol{r}'\boldsymbol{x}_k)\,d\boldsymbol{r} \qquad (A.1)\\
&\quad -\frac{2}{n}\sum_{j=1}^{n}\int\cdots\int \exp(i\boldsymbol{r}'\boldsymbol{x}_j)\exp(-\boldsymbol{r}'A\boldsymbol{r})\,d\boldsymbol{r} \\
&\quad +\int\cdots\int \exp(-\boldsymbol{r}'B\boldsymbol{r})\,d\boldsymbol{r},
\end{aligned}$$

where $A = \frac{\sigma^2}{2}\Phi + a \times I$ and $B = \sigma^2\Phi + a \times I$ with I as an identity matrix . Note that the first part in the above equation is a constant with respect to $\boldsymbol{\theta}$. Let $A = C^{-1}$. Considering the characteristic function of a random variable $N(0, \frac{1}{2}C)$, we have

$$\exp(-\frac{1}{4}\boldsymbol{x_j}'C\boldsymbol{x_j}) = \int\cdots\int \exp(i\boldsymbol{x_j}'\boldsymbol{r})\frac{\exp(-\frac{1}{2}\boldsymbol{r}'(\frac{C}{2})^{-1}\boldsymbol{r})}{(2\pi)^{(p+1)/2}|C|^{1/2}(1/2)^{(p+1)/2}}d\boldsymbol{r}. \qquad (A.2)$$

Thus the second part in $(A.1)$ is,

$$\begin{aligned}
&\int\cdots\int \exp(i\boldsymbol{r}'\boldsymbol{x}_j)\exp(-\boldsymbol{r}'A\boldsymbol{r})\,d\boldsymbol{r} \\
&= (2\pi)^{(p+1)/2}|A|^{-\frac{1}{2}}2^{-(p+1)/2}\exp(-\frac{1}{4}\boldsymbol{x_j}'A^{-1}\boldsymbol{x_j}) \\
&= \pi^{\frac{p+1}{2}}|A|^{-\frac{1}{2}}\exp(-\frac{1}{4}\boldsymbol{x_j}'A^{-1}\boldsymbol{x_j}). \qquad (A.3)
\end{aligned}$$

Similarly, for the third part in $(A.1)$ we have

$$\int\cdots\int \exp(-\boldsymbol{r}'B\boldsymbol{r})\,d\boldsymbol{r} = \pi^{\frac{p+1}{2}}|B|^{-\frac{1}{2}}. \qquad (A.4)$$

Substituting $(A.2)$, $(A.3)$ and $(A.4)$ into $(A.1)$, we have

$$I_n(\boldsymbol{\theta}) = constant - \frac{2}{n}\pi^{\frac{p+1}{2}}|A|^{-\frac{1}{2}}\sum_{j=1}^{n}\exp(-\frac{1}{4}\boldsymbol{x_j}'A^{-1}\boldsymbol{x_j}) + \pi^{\frac{p+1}{2}}|B|^{-\frac{1}{2}}. \quad \blacksquare$$

Appendix B

Proof of Theorem 2.3.2

$$(y_{j+p}|y_j,\cdots,y_{j+p-1}) \sim N\left(f_1(\boldsymbol{\rho})y_j+\cdots+f_q(\boldsymbol{\rho})y_{j+p-1},\sigma^2 g(\boldsymbol{\rho})\right)$$

implies

$$\log f(y_{j+p}|y_j,\cdots,y_{j+p-1}) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\log g(\boldsymbol{\rho}) - \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\boldsymbol{x}_j' A\boldsymbol{x}_j,$$

where A is defined in Theorem (2.3.2). If we take the derivative of the log conditional density function with respect to the parameters, we have

$$\frac{\partial \log f(y_{j+1}|y_j,\cdots,y_{j+p-1})}{\partial\sigma^2} = -\frac{1}{2\sigma^2}+\frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\boldsymbol{x}_j' A\boldsymbol{x}_j,$$

and

$$\frac{\partial \log f(y_{j+1}|y_j,\cdots,y_{j+p-1})}{\partial\rho_k} = -\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})}+\frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}\boldsymbol{x}_j' A\boldsymbol{x}_j - \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\boldsymbol{x}_j' B_k\boldsymbol{x}_j,$$

where g_k and B_k with $k=1,\cdots,l+m$ are defined in Theorem (2.3.2). Then the optimal weight function could be

$$\begin{aligned} w_{\sigma^2}(r) &= (\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\Big(-\frac{1}{2\sigma^2}+\frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\boldsymbol{x}_j' A\boldsymbol{x}_j\Big)\,dy_j\cdots dy_{j+p} \\ &= -\frac{1}{2\sigma^2}\delta(r^1)\cdots\delta(r^{p+1})+\frac{1}{2\sigma^4 g(\boldsymbol{\rho})}(\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\boldsymbol{x}_j' A\boldsymbol{x}_j \\ &\quad dy_j\cdots dy_{j+p}. \end{aligned} \tag{B.1}$$

Consider

$$\begin{aligned} &(\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\boldsymbol{x}_j' A\boldsymbol{x}_j\,dy_j\cdots dy_{j+p} \\ =\ &(\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\boldsymbol{x}_j' M\Lambda M'\boldsymbol{x}_j\,dy_j\cdots dy_{j+p} \\ =\ &(\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-i\boldsymbol{s}'\boldsymbol{z})\boldsymbol{z}'\Lambda\boldsymbol{z}dz_1\cdots dz_2 \\ =\ &(\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-\sum_{k=1}^{p+1} is^k z_k)\sum_{k=1}^{p+1}\lambda^k z_k^2\,dz_1\cdots dz_{p+1} \end{aligned}$$

$$
\begin{aligned}
&= (\frac{1}{2\pi})^{p+1}\int \exp(-is^1 z_1)\lambda^1 z_1^2\,dz_1 \prod_{k\neq 1}\Big\{\int \exp(-is^k z_k)\,dz_k\Big\}+\cdots \\
&\quad +(\frac{1}{2\pi})^{p+1}\int \exp(-is^{p+1}z_{p+1})\lambda^{p+1}z_{p+1}^2\,dz_{p+1}\prod_{k\neq p+1}\Big\{\int \exp(-is^k z_k)\,dz_k\Big\} \\
&= -\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1})-\cdots-\lambda^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1}),
\end{aligned}
$$

where M and Λ are defined in Theorem (2.3.2), $\boldsymbol{z} = M'\boldsymbol{x}_j$ and $\boldsymbol{s} = M'\boldsymbol{r}$. Hence, $(B.1)$ becomes

$$
\begin{aligned}
w_{\sigma^2}(\boldsymbol{r}) &= -\frac{1}{2\sigma^2}\delta(r^1)\cdots\delta(r^{p+1}) \qquad \text{(B.2)} \\
&\quad -\frac{1}{2\sigma^4 g(\boldsymbol{\rho})}[\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1})+\cdots+\lambda^{p+1}\delta(s_1)\cdots\delta(s_p)\delta''(s_{p+1})],
\end{aligned}
$$

and

$$
\begin{aligned}
w_{\boldsymbol{\rho}_k}(\boldsymbol{r}) &= (\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\Big(-\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})}+\frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}\boldsymbol{x}_j'A\boldsymbol{x}_j \\
&\quad -\frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\boldsymbol{x}_j'B_k\boldsymbol{x}_j\Big)\,dy_j\cdots dy_{j+p} \\
&= -\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})}\delta(r^1)\cdots\delta(r^{p+1}) \qquad \text{(B.3)} \\
&\quad -\frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}[\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1})+\cdots+\lambda^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1})] \\
&\quad +\frac{1}{2\sigma^2 g(\boldsymbol{\rho})}[\lambda_k^1\delta''(t_k^1)\delta(t_k^2)\cdots\delta(t_k^{p+1})+\cdots+\lambda_k^{p+1}\delta(t_k^1)\cdots\delta(t_k^p)\delta''(t_k^{p+1})],
\end{aligned}
$$

where $\boldsymbol{t_k} = H_k'\boldsymbol{r}$ for $k = 1,\cdots,l+m$.

Therefore, based on the optimal weight functions, the estimating equations are

$$
\begin{aligned}
0 &= \int\cdots\int w_{\sigma^2}(\boldsymbol{r})c_n(\boldsymbol{r})\,d\boldsymbol{r} \\
&= \int\cdots\int\Big\{-\frac{1}{2\sigma^2}\delta(r^1)\cdots\delta(r^{p+1})-\frac{1}{2\sigma^4 g(\boldsymbol{\rho})}[\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1}) \\
&\quad +\cdots+\lambda^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1})]\Big\}c_n(\boldsymbol{r})\,d\boldsymbol{r} \\
&= -\frac{1}{2\sigma^2}c_n(\boldsymbol{0})-\frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\int\cdots\int\Big[\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1}) \\
&\quad +\cdots+\lambda^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1})\Big]c_n(\boldsymbol{r})\,d\boldsymbol{r} \\
&= -\frac{1}{2\sigma^2}-\frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\Big[\lambda^1\frac{\partial^2 c_n(M\boldsymbol{s})}{\partial s^{1\,2}}\Big|_{\boldsymbol{s}=\boldsymbol{0}}+\cdots+\lambda^{p+1}\frac{\partial^2 c_n(M\boldsymbol{s})}{\partial s^{p+1\,2}}\Big|_{\boldsymbol{s}=\boldsymbol{0}}\Big]
\end{aligned}
$$

$$= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\frac{1}{n}\sum_{j=1}^{n}\left[\lambda^1((M'\boldsymbol{x}_j)_{(1)})^2 + \cdots + \lambda^{p+1}((M'\boldsymbol{x}_j)_{(p+1)})^2\right]$$

$$= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\bar{P}(\boldsymbol{\rho}), \tag{B.4}$$

where

$$\bar{P}(\boldsymbol{\rho}) = \frac{1}{n}\sum_{j=1}^{n}\left[\lambda^1((M'\boldsymbol{x}_j)_{(1)})^2 + \cdots + \lambda^{p+1}((M'\boldsymbol{x}_j)_{(p+1)})^2\right],$$

and

$$\begin{aligned}
0 &= \int\cdots\int w_{\boldsymbol{\rho}_k}(\boldsymbol{r})c_n(\boldsymbol{r})\,d\boldsymbol{r} \\
&= \int\cdots\int\Big\{-\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})}\delta(r^1)\cdots\delta(r^{p+1}) \\
&\quad -\frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}[\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1}) + \cdots + \lambda^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1})] \\
&\quad +\frac{1}{2\sigma^2 g(\boldsymbol{\rho})}[\lambda_k^1\delta''(t_k^1)\delta(t_k^2)\cdots\delta(t_k^{p+1}) + \cdots + \lambda_k^{p+1}\delta(t_k^1)\cdots\delta(t_k^p)\delta''(t_k^{p+1})]\Big\}c_n(\boldsymbol{r})\,d\boldsymbol{r} \\
&= -\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})} + \frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}\bar{P}(\boldsymbol{\rho}) - \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\bar{Q}(\boldsymbol{\rho}),
\end{aligned} \tag{B.5}$$

where

$$\bar{Q}_k(\boldsymbol{\rho}) = \frac{1}{n}\sum_{j=1}^{n}(\lambda_k^1((H_k'\boldsymbol{x}_j)_{(1)})^2 + \cdots + \lambda_k^{p+1}((H_k\boldsymbol{x}_j')_{(p+1)})^2),$$

with $k = 1, \cdots, l+m$. If concentrating out σ^2 in $(B.4)$ and $(B.5)$, we have (3.8). ■

TRUE VALUES OF PARAMETERS $\phi = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

INITIAL VALUES OF VECTORS $\boldsymbol{r}$

−0.10	−0.05	−0.01	0.03	0.08
−0.09	−0.04	0.00	0.04	0.09
−0.08	−0.03	0.01	0.05	0.10

INITIAL VALUE OF τ 0.5

	$\phi = 0.6$					$\sigma^2 = 1.0$				
	c(1)	c(2)	c(3)	c(4)	MLE	c(1)	c(2)	c(3)	c(4)	MLE
MEAN	.5963	.5913	.5838	.5814	.6032	.9806	.9753	.9945	.9829	.9922
MED	.5730	.5388	.5456	.5243	.6026	.9728	.9541	1.006	.9605	.9838
MIN	.2739	.1876	.2163	.0371	.3472	.4526	.6086	.4358	.6125	.6411
MAX	1.00	1.00	1.00	1.00	.8758	1.455	1.528	1.53	1.569	1.466
SKEW	.6886	.6115	.5344	.2771	.0371	-.142	.0045	.4733	.4108	.2940
KURT	2.62	2.54	2.41	2.29	3.13	2.81	2.73	2.79	2.67	3.01
VAR	.043	.0487	.0486	.0667	.007	.0406	.0364	.0507	.0454	.0196
BIAS	.0037	.0087	.0162	.0186	.0032	.0394	.0247	.0055	.0171	.0078
MSE	.043	.0487	.0489	.0671	.007	.0421	.037	.0507	.0457	.0197

Table 2.1: Monte Carlo Comparison of DECF and MLE for an MA(1) Model When $\boldsymbol{r}$'s Depends on Only One Variable

TRUE VALUES OF PARAMETERS $\phi = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

INITIAL VALUES OF VECTORS $\boldsymbol{r}$

−0.10	−0.05	−0.01	0.03	0.08
−0.09	−0.04	0.00	0.04	0.09
−0.08	−0.03	0.01	0.05	0.10

INITIAL VALUES OF τ_1 AND τ_2: 0.5 0.5

	$\phi = 0.6$					$\sigma^2 = 1.0$				
	c(1)	c(2)	c(3)	c(4)	MLE	c(1)	c(2)	c(3)	c(4)	MLE
MEAN	.5961	.5858	.5900	.5920	.6032	.9776	.9869	.9963	.9850	.9922
MED	.5466	.5672	.5824	.5736	.6026	.9409	1.000	.9919	.9714	.9838
MIN	.2401	.2233	.2283	.3206	.3472	.5919	.5775	.4313	.5918	.6411
MAX	1.00	1.00	1.00	1.00	.8758	1.502	1.389	1.438	1.375	1.466
SKEW	.5852	.6262	.6047	.7664	.0371	.4116	-.192	.0342	.2586	.2940
KURT	2.55	2.37	2.94	3.21	3.13	2.77	2.62	2.75	2.63	3.01
VAR	.04096	.0308	.0500	.0271	.007	.0409	.0269	.0440	.0252	.0196
BIAS	.0039	.0142	.0100	.0080	.0032	.0224	.0131	.0037	.0150	.0078
MSE	.04100	.0310	.0501	.0272	.007	.0415	.0271	.0440	.0254	.0197

Table 2.2: Monte Carlo Comparison of DECF and MLE for an MA(1) Model When $\boldsymbol{r}$'s Depends on Two Variables

TRUE VALUES OF PARAMETERS $\phi = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

INITIAL VALUES OF ϕ AND σ^2: 0.3 0.7

	$\phi = 0.6$				
	c(1)	c(2)	c(3)	c(4)	MLE
MEAN	.6479	.5981	.6475	.6269	.6032
MED	.6102	.5325	.6076	.5960	.6026
MIN	.0281	.1588	.1508	.2382	.3472
MAX	1.000	1.000	1.000	1.000	.8758
SKEWNESS	.3798	.9556	.4070	.7960	.0371
KURTOSIS	2.917	2.749	2.134	3.244	3.126
VAR	.0566	.04258	.0456	.026	.007
BIAS	.0479	.0019	.0475	.0269	.0032
MSE	.0589	.04259	.0479	.0268	.0070

Table 2.3: Monte Carlo Comparison of CECF and MLE for an MA(1) model

TRUE VALUES OF PARAMETERS $\phi = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

INITIAL VALUES OF ϕ AND σ^2: 0.3 0.7

	$\phi = 0.6$								
	p=3	p=4	p=5	p=6	p=7	p=8	p=9	p=10	MLE
MEAN	.6133	.6061	.6037	.6024	.6028	.6025	.6026	.6023	.6032
MED	.6011	.5992	.6004	.6009	.6018	.6015	.6006	.6016	.6026
MIN	.3138	.3446	.3401	.3493	.338	.3441	.3274	.323	.3472
MAX	1.000	1.000	1.000	.8861	.943	.9055	.9002	.9415	.8757
SKEW	.8491	.5473	.3802	.0398	.0253	-.012	.0054	.0693	.0371
KURT	4.211	4.183	3.960	3.148	3.240	3.054	3.073	3.284	3.126
VAR	.015	.010	.0085	.007	.0075	.007	.007	.0076	.007
BIAS	.0133	.0061	.0037	.0024	.0028	.0025	.0026	.0023	.0032
MSE	.015	.010	.0085	.007	.0075	.007	.007	.0076	.0070

Table 2.4: Monte Carlo Comparison of GLS-CECF and MLE for an MA(1) Model

TRUE VALUES OF PARAMETERS $\phi = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=1,000

	$\phi = 0.6$					
	c(1)	c(2)	c(3)	c(4)	c(5)	c(6)
MEAN	.6350	.6200	.6128	.6095	.6068	.6072
MED	.6061	.6045	.6028	.6030	.5994	.6014
MIN	.3404	.3907	.3746	.4090	.4104	.4204
MAX	1.000	1.000	1.000	1.000	.9900	1.000
SKEWNESS	.9929	1.243	.9365	.7593	.6927	.6765
KURTOSIS	3.637	5.551	5.329	4.846	4.864	4.971
VAR	.0211	.0101	.0078	.0049	.0049	.0039
BIAS	.035	.020	.0128	.0095	.0068	.0072
MSE	.0223	.0105	.0080	.0050	.0049	.0040

Table 2.5: Monte Carlo Comparison of OLS-CECF and WLS-CECF with Different Values of p for an MA(1) Model

TRUE VALUES OF PARAMETERS $\rho = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

INITIAL VALUES OF ρ AND σ^2: 0.3 0.7

	$\rho = 0.6$		$\sigma^2 = 1.0$	
	GLS-CECF	MLE	GLS-CECF	MLE
MEAN	.5907	.5906	.9924	.9923
MED	.5957	.5982	.9850	.9838
MIN	.2208	.2200	.6425	.6410
MAX	.7878	.7797	1.443	1.516
SKEWNESS	-.4960	-.5190	.2953	.3111
KURTOSIS	3.517	3.544	3.021	3.076
VAR	.0063	.0063	.0199	.0198
BIAS	.0093	.0094	.0076	.0077
MSE	.0064	.0064	.01999	.01991

Table 2.6: Monte Carlo Comparison of GLS-CECF and MLE for an AR(1) Model

TRUE VALUES OF PARAMETERS $\rho = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=1,000

	$\rho = 0.6$			
	c(1)	c(2)	c(3)	c(4)
MEAN	.5996	.5997	.5998	.6002
MED	.6018	.6008	.6025	.6010
MIN	.4465	.4693	.4300	.4679
MAX	.7051	.6841	.7060	.6885
SKEWNESS	-.3101	-.2246	-.3566	-.1916
KURTOSIS	3.083	3.196	3.021	3.089
VAR	.00149	.00092	.00162	.0010
BIAS	.0004	.0003	.0002	.0002
MSE	.00149	.00092	.00162	.0010

Table 2.7: Monte Carlo Comparison of OLS-CECF and WLS-CECF with Different Values of p for an AR(1) Model

TRUE VALUES OF PARAMETERS $-\phi = \rho = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

	$\rho = 0.6$		
	$p = 2$	$p = 3$	CMLE
MEAN	.5972	.6000	.5895
MED	.5973	.6000	.5909
MIN	.3785	.3991	.3491
MAX	.7878	.7681	.7351
SKEWNESS	-.0906	-.0672	-.2132
KURTOSIS	2.991	3.056	3.115
VAR	.003767	.003060	.003035
BIAS	.0028	.0000	.0105
MSE	.003775	.003060	.003145

Table 2.8: Monte Carlo Comparison of GLS-CECF and MLE for a Gaussian ARMA(1,1) Model

TRUE VALUES OF PARAMETERS $-\phi = \rho = 0.6 \; \sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=1000

	$\rho = 0.6$					
	c(1)	c(2)	c(3)	c(4)	c(5)	c(6)
MEAN	.6007	.6010	.6012	.6014	.6017	.6019
MED	.6013	.6007	.6018	.6008	.6017	.6019
MIN	.4488	.4841	.4671	.4797	.4777	.4734
MAX	.7112	.6902	.7162	.7046	.6993	.7093
SKEWNESS	-.1993	-.0829	-.1067	-.0844	-.0802	-.1047
KURTOSIS	3.182	3.003	3.235	2.992	3.163	3.032
VAR	.0014	.0010	.0011	.0012	.0010	.0013
BIAS	.0007	.0010	.0012	.0014	.0017	.0019
MSE	.0014	.0010	.0011	.0012	.0010	.0013

Table 2.9: Monte Carlo Comparison of OLS-CECF and WLS-CECF with Different Values of p for a Gaussian ARMA$(1, 1)$ Model

TRUE VALUES OF PARAMETERS $-\phi = \rho = 0.9$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

	$\rho = 0.9$		
	$p = 2$	$p = 3$	CMLE
MEAN	.8846	.8881	.8512
MED	.8904	.8949	.8601
MIN	.6374	.6569	.6015
MAX	.9739	.9820	.9690
SKEWNESS	-.9348	-.9060	-.7173
KURTOSIS	4.307	4.186	3.577
VAR	.002333	.002315	.003202
BIAS	.0154	.0119	.0488
MSE	.002571	.002440	.005583

Table 2.10: Monte Carlo Comparison of GLS-CECF and MLE for a Gaussian ARMA(1, 1) Model

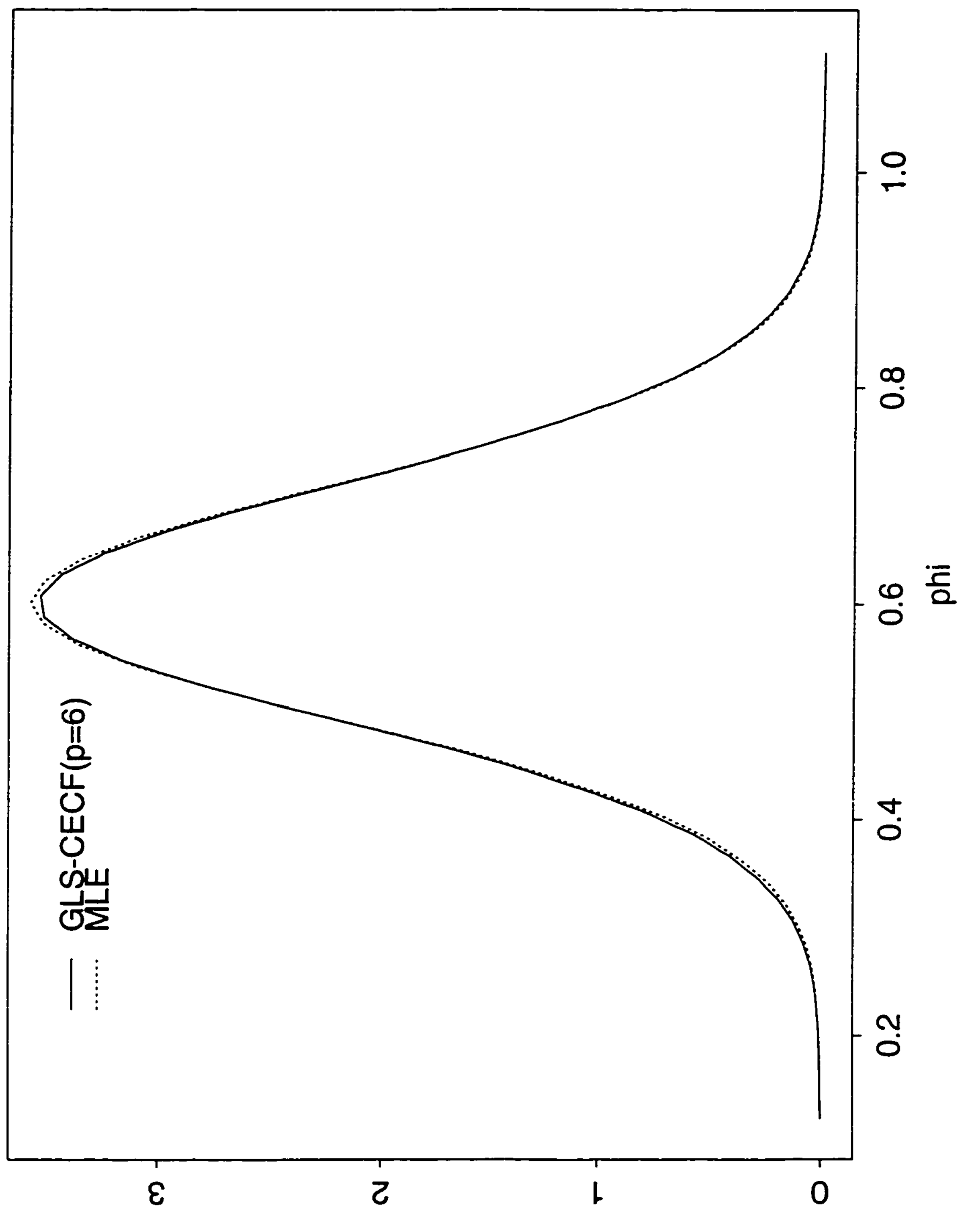

Figure 2.1: Density Function of Estimator of ϕ in an MA(1) Model

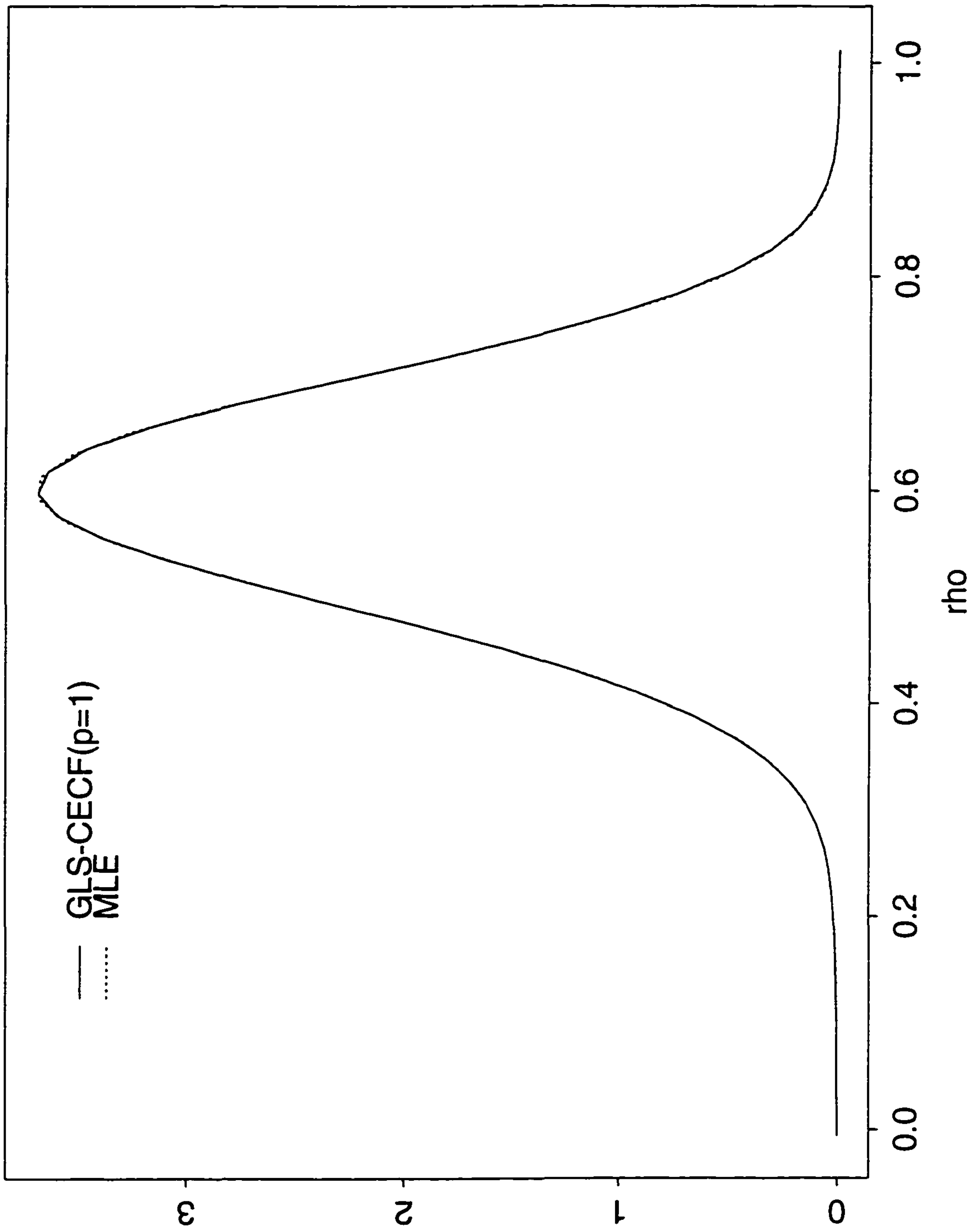

Figure 2.2: Density Function of Estimator of ρ in an AR(1) Model

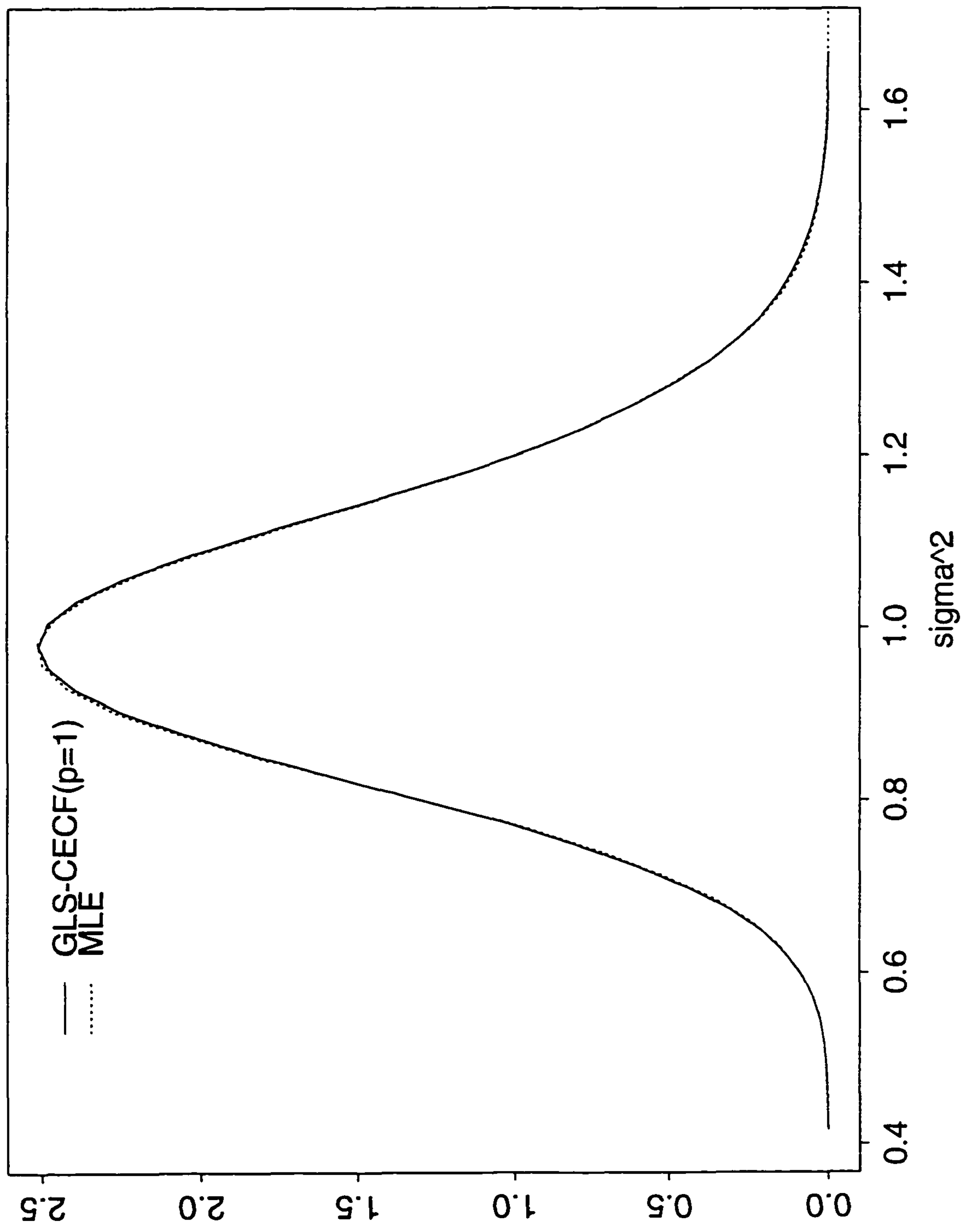

Figure 2.3: Density Function of Estimator of σ^2 in an AR(1) Model

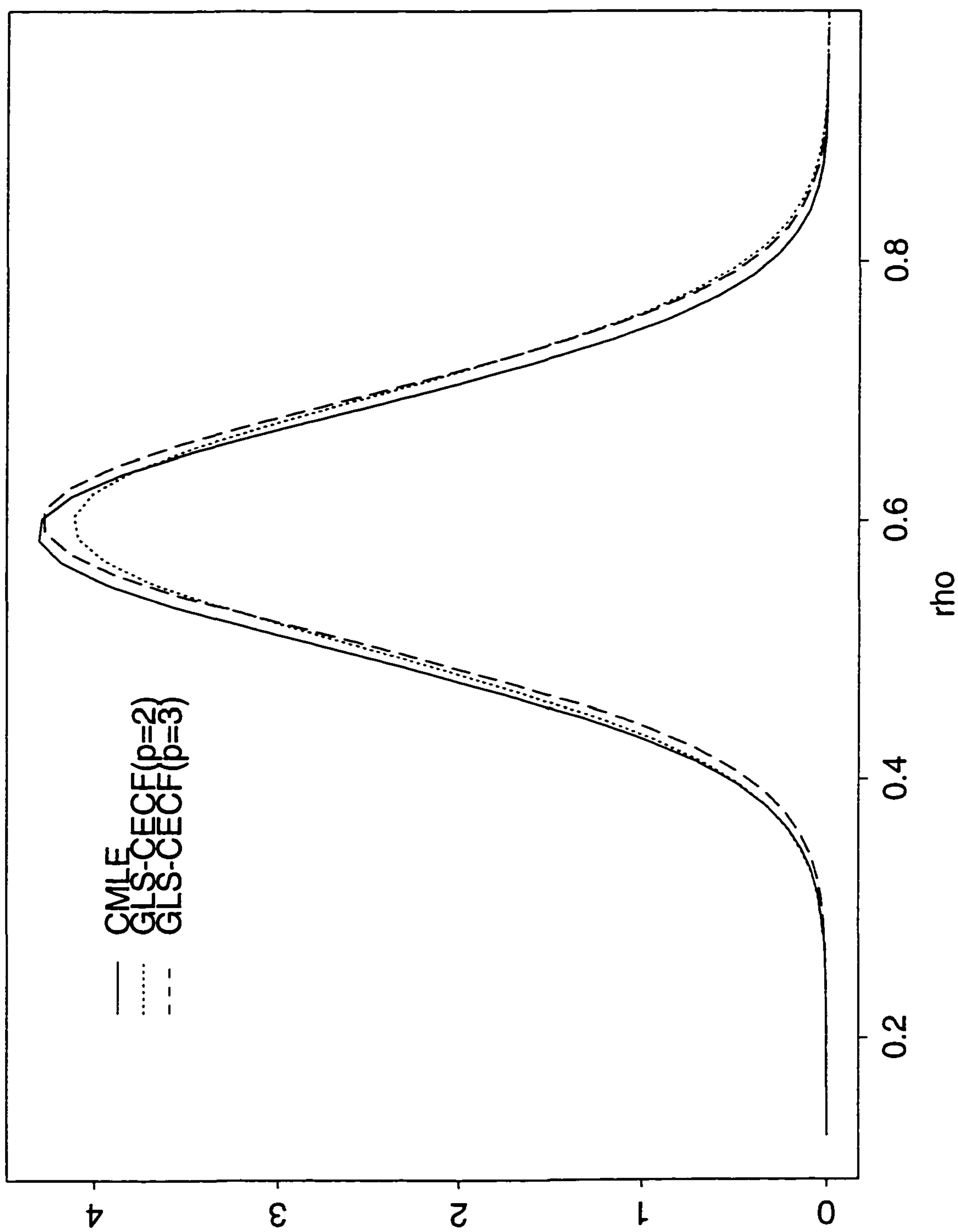

Figure 2.4: Density Function of Estimator of ρ in a Gaussian ARMA(1,1) Model

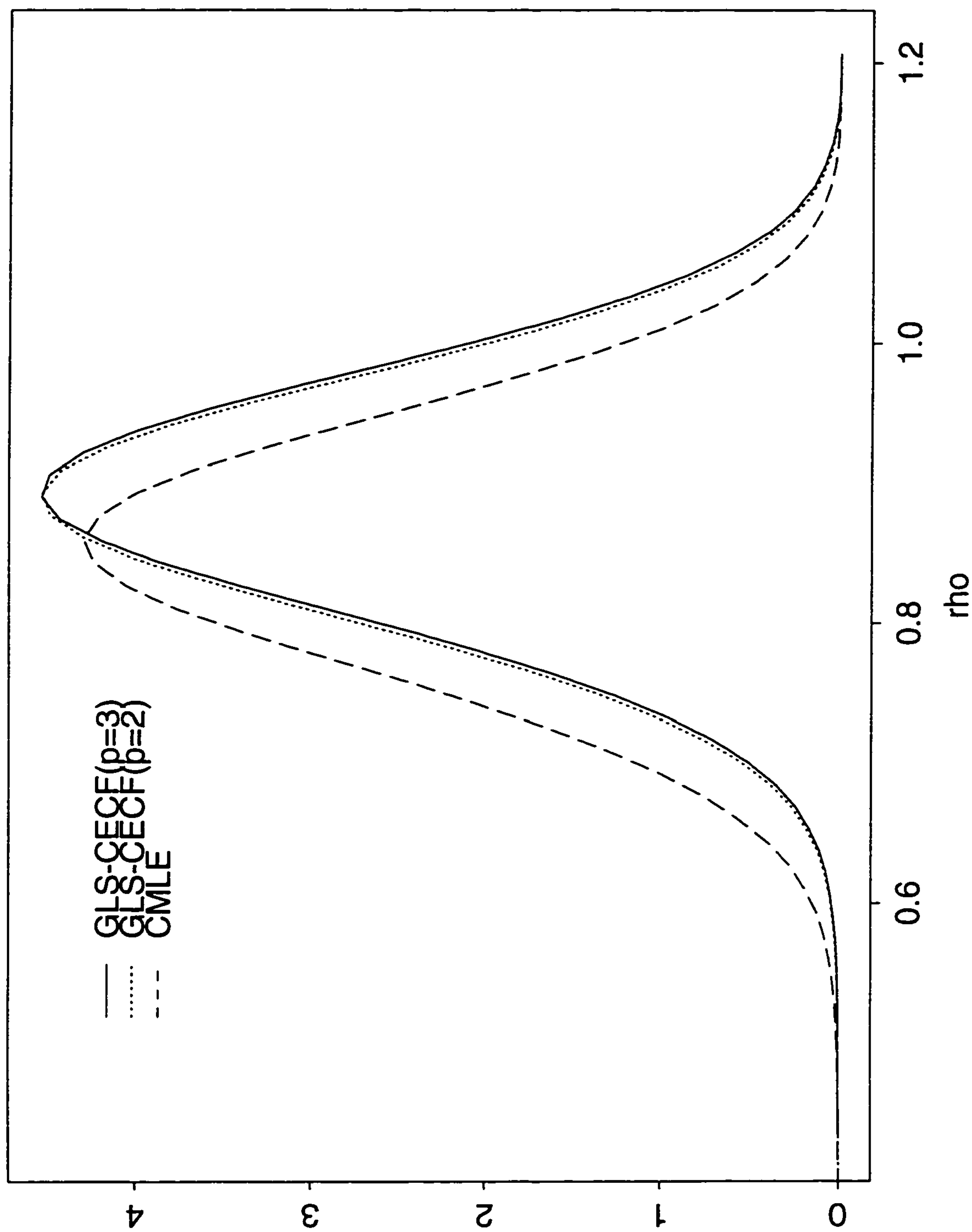

Figure 2.5: Density Function of Estimator of ρ in a Gaussian ARMA(1,1) Model

Chapter 3

ESTIMATION OF THE STOCHASTIC VOLATILITY MODEL VIA EMPIRICAL CHARACTERISTIC FUNCTION

3.1 Introduction

Modeling the volatility of financial and macroeconomic time series has attracted a lot of attention since the introduction of autoregressive conditional heteroskedasticity (ARCH) models (Engle (1982)). A feature of the ARCH type model is that the conditional variance is driven by the past observable variables. As an alternative setup to the ARCH-type model, the Stochastic Volatility (SV) model is supposed to describe the financial time series better than the ARCH-type model, since it essentially involves two noise processes. This added dimension makes the model more flexible, for example, the SV model can explain not only volatility clustering but also leverage effects. For further discussion, see Ghysels, Harvey, and Renault (1996) . Unfortunately, the density function for the SV model has no closed form and hence neither does the likelihood function, even for the simplest version of the SV model. It is a consequence of this that direct maximum-likelihood estimation is impossible. Therefore, alternative estimation methods to the maximum likelihood have been proposed to estimate the SV models, which we discuss next.

Melino and Turnbull (1990) use generalized method of moments (GMM) for the discrete SV model. A more efficient GMM is proposed by Andersen and Sorensen (1993) . For the continuous time SV model, a GMM approach is developed by Hansen and Scheinkman (1995). The idea is to match a finite number of sample moments and theoretical moments. Alternatively, the quasi maximum likelihood (QML) approach is suggested by Nelson (1988), Ruiz (1994) and Harvey, Ruiz and Shephard (1994). The main idea is to treat non Gaussian disturbances as if they are normal and then maximize the quasi likelihood function. Often estimation methods involve the whole family of simulation based methods, including simulated MM/GMM proposed by Duffie and Singleton (1993), indirect inference proposed by Gouriéroux, Monfort and Renault (1993), simulated maximum likelihood (SML) proposed by Danielsson (1994b), and Markov Chain Monte Carlo (MCMC) proposed by Jacquier, Polson and Rossi (1994). The SV model has become a central model to describe financial time series and to compare the relative merits of estimation procedures.

Although most of these methods are consistent under appropriate regularity conditions, in general they are not efficient. For example, by using only a finite number of moment conditions, MM/GMM may ignore important information contained in the realizations. The QML approach simply approximates the true information. Not surprisingly, such an approximation could lose substantial amounts of information. The simulation based methods decrease the efficiency by introducing an extra random error. This raises the question as to whether we can find a methodology with efficiency equivalent to maximum likelihood.

The present chapter uses such an alternative approach to estimate the stochastic volatility model - via the empirical characteristic function. The rationale for using the characteristic function is that there is a one to one correspondence between the characteristic function and the distribution function. Consequently, the empirical characteristic function (ECF) should contain the same amount of information as the empirical distribution function (EDF). Theoretically, therefore, inference based on the characteristic function should perform as well as inference based on the likelihood. Moreover, by using the characteristic function, we can overcome the difficulties arising from ignorance of the true density function or the true likelihood function. This chapter is organized as follows. The next section introduces a canonical SV model and explains why the model is difficult to estimate. Section 3.3 presents a discussion on ECF estimation for the SV model; the characteristic function of the SV model is obtained as well. Section 3.4 discusses the implementation of the ECF method as well as a Monte Carlo study and an empirical application. All proofs are collected in the Appendix.

3.2 The Model

The formulation of the discrete time stochastic volatility model is similar to that of the ARCH-type models. That is, the conditional variance is directly modeled. However, in contrast to the ARCH-type models, the stochastic volatility model allows a random component in the transition equation. By doing so, the model can explain why large changes can follow stable periods. The model is of the form,

$$x_t = \sigma_t e_t, \; t = 1, 2, \cdots, T, \tag{2.1}$$

where σ_t^2 is the conditional variance based on the information at the end of time t, and e_t is a series of i.i.d. random disturbances which are assumed to be a standard normal distribution. We define

$$\sigma_t = \exp(0.5h_t) \tag{2.2}$$

and assume h_t follows a Gaussian AR(1) process, i.e.,

$$h_t = \lambda + \alpha h_{t-1} + v_t, \; v_t \sim iidN(0, \sigma^2), \tag{2.3}$$

where $\boldsymbol{\theta} = (\alpha, \lambda, \sigma^2)$ are the unknown parameters. It is well-known that if $|\alpha| < 1$, this process is invertible and stationary. Heuristically, we can say that the conditional variance depends on past conditional variance and a random component. When the effect of the past conditional variance is strong, volatility clustering will appear in the series. However, if the random innovation is not dominated, it can bring a large change into a stable period and can smooth large booms and crashes as well. Without including the random component, the transition equation is deterministic and the model exhibits time-varying but deterministic volatility. Finally, we assume e_t and v_t are independent, we shall return to this assumption later.

Some statistical properties of x_t are determined by h_t since x_t is a simple function of h_t. For example, h_t is stationary for $|\alpha| < 1$, thus x_t is stationary as well. Furthermore, x_t is a martingale difference because h_t is a martingale difference; see Ghysels, Harvey and Renault (1996). We also note that x_t has finite moments of all orders and in particular the second and fourth moments are given by $E(x_t^2) = \exp(\frac{\sigma^2}{2(1-\alpha^2)})$, and $E(x_t^4) = 3\exp(\frac{2\sigma^2}{1-\alpha^2})$. The kurtosis of x_t is therefore $3\exp(\frac{\sigma^2}{1-\alpha^2})$, so x_t exhibits more kurtosis than a constant variance normal model. Furthermore, Harvey (1993) derives the moments of powers of the absolute value of x_t,

$$E|x_t|^c = 2^{c/2}\frac{\Gamma(c/2+1/2)}{\Gamma(1/2)}\exp(\frac{c^2\sigma^2}{8(1-\alpha^2)}),\ c > -1, c \neq 0, \tag{2.4}$$

and

$$Var|x_t|^c = 2^c\left\{\frac{\Gamma(c+\frac{1}{2})}{\Gamma(1/2)} - \left[\frac{\Gamma(\frac{c}{2}+\frac{1}{2})}{\Gamma(1/2)}\right]^2\right\}\exp(\frac{c^2\sigma^2}{2(1-\alpha^2)}),\ c > -\frac{1}{2}, c \neq 0. \tag{2.5}$$

Since x_t is a non-linear function of an AR(1) process, however, the process is difficult to work with. For example, there is no closed form expression for the characteristic function of x_t. Observing that the dependence of x_t is completely characterized by the dependence of h_t, we define y_t to be the logarithm of x_t^2. Then we have

$$y_t = \log\sigma_t^2 + \log e_t^2 = h_t + \epsilon_t,\ t = 1, 2, \cdots, T, \tag{2.6}$$

where $\epsilon_t = \log e_t^2$ is the logarithm of the chi square random variable with 1 degree of freedom. Hence, the new process y_t still depends on the AR(1) process h_t, but in a linear form. Since the process h_t contains all the parameters of interest, y_t loses no information from the estimation point of view, the only loss of information being the sign of e_t which for a symmetric distribution, uncorrelated in ϵ_t and v_t, contributes nothing to volatility estimation. This is why most of the estimation procedures in the literature are based on y_t, not x_t.

Unfortunately, neither y_t nor x_t has a closed form expression for the likelihood function. This property makes the estimation based on the likelihood extremely difficult. However, from (2.6) we know that y_t is the convolution of an AR(1) process and an iid logarithmic $\chi^2_{(1)}$ sequence, and hence there is a closed form expression for the characteristic function of y_t which we will derive in the next section. Since the CF contains the same amount of information as the distribution function, the model is fully and uniquely parameterized by the CF. Therefore, inference based on the ECF can achieve efficiency.

3.3 ECF Estimation

The model we are going to estimate via the ECF is the one defined by (2.6) since we can derive the closed form of the characteristic function. In order to use the ECF method, of course, we need to find the expression of the joint characteristic function. First, the characteristic function for the logarithm of the $\chi^2_{(1)}$ distribution is given in Theorem 3.3.1. And then the joint characteristic function for $y_t, \cdots, y_{t+k-1}$ is obtained in Theorem 3.3.2.

Theorem 3.3.1 *Suppose ϵ_t is the logarithm of the $\chi^2_{(1)}$ distribution. The character-*

istic function of ϵ_t, $c(r)$, *is*

$$c(r) = \frac{\Gamma(0.5 + ir)}{\Gamma(0.5)} 2^{ir}, \tag{3.1}$$

where i *is the imaginary number and defined by* $\sqrt{-1}$.

Proof: See Appendix A.

Theorem 3.3.2 *Suppose* $\{y_t\}_{t=1}^T$ *is defined by* (2.6). *The joint characteristic function of* $y_t, y_{t+1}, \cdots, y_{t+k-1}$ *is*

$$\begin{aligned} c(r_1, \cdots, r_k, \boldsymbol{\theta}) &= \exp[i\frac{\lambda}{1-\alpha}\sum_{j=1}^{k}\alpha^{j-1}r_j + i\lambda\sum_{j=2}^{k}\frac{1-\alpha^{j-1}}{1-\alpha}r_j - \frac{1}{2}(\sum_{j=1}^{k}\alpha^{j-1}r_j)^2\frac{\sigma^2}{1-\alpha^2} \\ &\quad -\frac{1}{2}\sum_{l=2}^{k}(\sum_{j=l}^{k}\alpha^{j-l}r_j)^2\sigma^2]\frac{\prod_{j=1}^{k}\Gamma(\frac{1}{2}+ir_j)}{\Gamma^k(\frac{1}{2})}2^{i\sum_{j=1}^{k}r_j}. \end{aligned} \tag{3.2}$$

Proof: See Appendix B.

Using the joint characteristic function we can easily obtain the joint cumulant generating function and consequently the autocorreltion function. The autocorrelation function of $\{y_t\}_{t=1}^T$ is given in the following theorem.

Theorem 3.3.3 *Suppose* $\{y_t\}_{t=1}^T$ *is defined by* (2.6). *The autocorrelation function of* $\{y_t\}_{t=1}^T$ *is,*

$$\rho_k = \frac{\frac{\alpha^{k-1}\sigma^2}{1-\alpha^2}}{\frac{\sigma^2}{1-\alpha^2} + \frac{\Gamma''(0.5)}{\Gamma(0.5)} - (\frac{\Gamma'(0.5)}{\Gamma(0.5)})^2}, \quad k = 1, 2, \cdots. \tag{3.3}$$

Proof: See Appendix C.

The y_t process defined by (2.6) is the sum of an AR(1) and white noise, it is well-known that this process has the similar dynamic behavior as the Gaussian ARMA(1,1) (hence we still call it ARMA(1,1)). This is confirmed by the formula in (3.3). Furthermore, the ρ mixing condition of the SV model is ensured by (3.3).

In order to use the ECF method to estimate the SV model (2.6), we have to choose a value for p. Although our process is not Markovian, being an ARMA(1,1), we shall choose $p = 1$ at first. Our reason relates to the results of a Monte Carlo study in Chapter 2 where the ECF method is used to estimate a linear ARMA(1,1) process and we found that $p = 1$ works quite well (see Table 2.9 in Chapter 2). For the SV model (2.6) we note that with $p = 1$,

$$\begin{aligned} c(r_1, r_2, \boldsymbol{\theta}) &= \exp[i\lambda\frac{r_1 + r_2}{1-\alpha} - \frac{1}{2}(r_1 + \alpha r_2)^2 \frac{\sigma_v^2}{1-\alpha^2} - \frac{1}{2} r_2^2 \sigma_v^2] \\ &\quad \frac{\Gamma(\frac{1}{2} + ir_1)\Gamma(\frac{1}{2} + ir_2)}{\Gamma^2(\frac{1}{2})} 2^{ir_1 + ir_2}, \end{aligned} \tag{3.4}$$

and

$$c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^{n} \exp(ir_1 y_j + ir_2 y_{j+1}). \tag{3.5}$$

Defining $Re\, c(r_1, r_2, \boldsymbol{\theta})$, $Re\, c_n(r_1, r_2)$, $Im\, c(r_1, r_2, \boldsymbol{\theta})$ and $Im\, c_n(r_1, r_2)$ to be the real and imaginary parts of $c(r_1, r_2)$ and $c_n(r_1, r_2)$ respectively, we have,

$$Re\, c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^{n} cos(r_1 y_j + r_2 y_{j+1}), \tag{3.6}$$

and

$$Im\, c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^{n} sin(r_1 y_j + r_2 y_{j+1}). \tag{3.7}$$

As we mentioned before, a clear advantage of choosing the transformation variable continuously is that we do not need to choose q. Furthermore, since the same Monte Carlo study conducted in Chapter 2 shows that the continuous ECF method works better than the discrete ECF method, we use the continuous ECF method to estimate the SV model. However, the optimal weight function in the continuous ECF method is not readily obtained because the conditional score function has no closed form expression for the SV model. Instead the exponential weight is considered. The exponential function is chosen because it puts more weight on the points around the origin, consistent with the recognition that the CF contains the most information around the origin. Therefore, the procedure is to choose $(\hat{\alpha}, \hat{\sigma}^2, \hat{\lambda})$ to minimize,

$$\int\int \Big[(Re\, c(r_1, r_2, \boldsymbol{\theta}) - \tfrac{1}{n}\textstyle\sum_{j=1}^{n} cos(ir_1 y_j + ir_2 y_{j+1}))^2 \tag{3.8}$$

$$+(Im\, c(r_1, r_2, \boldsymbol{\theta}) - \tfrac{1}{n}\textstyle\sum_{j=1}^{n} sin(ir_1 y_j + ir_2 y_{j+1}))^2\Big] \exp(-ar_1^2 - ar_2^2)\, dr_1 dr_2,$$

where $c(r_1, c_2)$ is given by (3.4) and a is an arbitrary positive constant which is chosen to be 32.5 in the Monte Carlo studies and the application.

We could check that the appropriate regularity conditions given in Chapter 1 hold for the weight function. Therefore, the resulting estimators are consistent and asymptotically normal. The asymptotic covariance matrix of the estimators is given below,

$$\frac{1}{n}\left\{\int\int\left[\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}+\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right]g(\boldsymbol{r})dr_1dr_2\right\}^{-1}\times$$
$$\Sigma^2\times\left\{\int\int\left[\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}+\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right]g(\boldsymbol{r})dr_1dr_2\right\}^{-1}.$$

In appendix D, the expression of Σ^2 is given as well as the process of calculating it.

We should note that the joint characteristic function of y_t is of different functional form if v_t and e_t are correlated. However, the ECF method can be still used in the same way as the uncorrelated case. Of course, the the joint characteristic function of y_t in the correlated case is more difficult.

3.4 Implementation, Simulation and Application

3.4.1 Implementation

The implementation of the ECF method essentially requires minimizing (3.8), and thus involves double integrals. Unfortunately, no analytical solutions for either the double integrals or the optimization are available. Consequently, we will numerically evaluate the multiple integral (3.8), followed by numerical minimization of (3.8) with respect to $\boldsymbol{\theta}$. The numerical solutions are the desired estimators.

A 40-point Gauss-Kronod algorithm is used to approximate the two dimensional integrations in (3.8). Since there is no analytical expression for the derivative of the objective functions, a quasi-Newton method is used to find the minimum. The starting point in the optimization is chosen to be the quasi-maximum likelihood estimates proposed by Ruiz (1994).[1] All computations were done in double precision.

By using the implementation procedure, we examine the performance of the ECF method in the estimation of a SV model in a Monte Carlo study. We also apply the procedure to a real dataset.

3.4.2 Monte Carlo Simulation

The Monte Carlo study is designed to check the viability of the ECF method. We choose the same parameter setting as Jacquier, Polson and Rossi (1994) did in one of their Monte Carlo studies, that is, $\alpha = 0.9$, $\sigma = 0.3629$, $\lambda = -0.736$. The number of observations set at $T = 2,000$ and the number of replications set at 500.

Table 3.1 reports the simulation results. The table shows the mean, the minimum, the maximum, the mean square error (MSE) and the root mean square error (RMSE) for all three estimators, and serves to illustrate that the ECF method works well.

In Table 3.2 we duplicate the results in Table 9 of Jacquier, Polson and Rossi (1994), where the same experimental design is used but the three alternative methods are employed, i.e., the GMM, QML and MCMC. We also report the simulation results provided by Danielsson (1994b) based on the SML for the same experiment. Of course our random numbers may not be the same as those generated by Jacquier, Polson and Rossi (1994) and by Danielsson (1994). However, we believe that the experiments should be comparable. The finite sample performance of the ECF method is better than that of the QML and GMM method, while the MCMC and SML method outperform the ECF method. This can be accounted for by the use of the non-optimal weight function. How to choose a better weight function is of future

[1] Actually in the empirical application many other initial point have been used.

interest. Nevertheless, the ECF method provides a viable alternative.

3.4.3 Application

Data

The data we used was supplied by George Tauchen[2] and is the same as that used by Danielsson (1994b). It consists of eight years (2,022 observations) of daily geometric returns (defined as $100(\log P_{t+1} - \log P_t)$) for the S&P 500 index covering period 1980-1987. The data are adjusted as detailed in Danielsson (1994b).

Empirical Results

The empirical results are reported in Table 3.3, along with the MCMC estimates obtained by Jacquier, Polson and Rossi (1994) using the same data set. To obtain the ECF estimates the initial values are chosen to be the QML estimates, as well as the MCMC estimates and other starting values. This serves to show that the global minimum is achieved. From Table 3.3, we note that the ECF estimates are very different from the MCMC estimates. For example, $\hat{\alpha}_{ECF}$ is close to 0 while the $\hat{\alpha}_{MCMC}$ is close to 1; $\hat{\sigma}^2_{ECF}$ is 20 times larger than $\hat{\sigma}^2_{MCMC}$. Since the empirical results are so different, the comparison of the goodness of fit is of particular interest.

To compare the goodness of fit, we simulate two sequences by using the ECF estimates and the MCMC estimates. In Figure 3.1, we plot the empirical density of the real data and densities of two simulated data sets. Figure 3.1 demonstrates that the ECF estimates give a better fit than the MCMC estimates. The Kolmogorov-Smirnov test is performed to test for the goodness of fit. The results are reported in Table 3.4. For the ECF the Kolmogorov-Smirnov test statistic (0.0257) is much smaller and the p-value (0.498) is very large while the p-value for the MCMC is 0. The

[2]We would like to thank Tauchen for supplying the dataset.

MCMC has been rejected at any significant level and the ECF can not be rejected. Therefore, the ECF method is significantly better than the MCMC in the sense of fitting the steady state distribution. This result is very intuitive because the ECF method basically matches all the moments and hence the density.

We next discuss the implication of our results. Firstly, a much smaller α implies smaller volatility clustering, that is less persistence. Consequently, there is not much dependence for the variances between two consecutive trading days. This contrasts with the implication of large α. Secondly, a much larger σ^2 suggests that a large change can possibly follow a stable period and a stable period can follow an unstable period. This happens because with the large variance the random innovation v_t may dominate the deterministic term and hence bring in a significant change. Thus whilst the estimated models have similar means their persistence characteristics seem dramatically different. Evaluating our objective function, given by (3.8), for the two sets of converged estimates in Table 3.3 we find that the ECF estimates result in a value of 1.6354×10^{-7} while the MCMC estimates give 9.73823×10^{-5}. The latter is nearly 600 times larger than the former!

Finally we should stress that we have also chosen larger values of p for the ECF method. Theoretically, we know that with a larger p the moving blocks preserve more information and hence the ECF method can be more efficient. On the other hand, however, a larger p is computationally more time-consuming since higher dimensional numerical integrations are involved. In Table 3.5, we report the ECF estimates for $p = 2, 3, 4, 5$ where we fit the SV model to the same data set.[3] From this table we note that the empirical results remain almost unchanged for different values of p and

[3]Of course higher dimensional integrations are required to be approximated for larger values of p. A DCUHRE algorithm proposed by Bernsten, Espelid and Genz (1993) is used to approximate such integrations.

are very close to those for $p = 1$.

Appendix A

Proof of Theorem 3.3.1

According to the definition of the characteristic function, and with $\epsilon_t = \log(\chi^2_{(1)})$, we have

$$
\begin{aligned}
c(r) &= E[\exp(ir\epsilon_t)] \\
&= E[(\exp(\epsilon_t))^{ir}] \\
&= E[(\exp(\log(\chi^2_{(1)})))^{ir}] \\
&= E[(\chi^2_{(1)})^{ir}] \\
&= \int x^{ir} f(x)\, dx \\
&= \int x^{ir} \frac{1}{\Gamma(1/2)\sqrt{2}} x^{-1/2} e^{-x/2}\, dx \\
&= \int x^{ir-\frac{1}{2}} \frac{1}{\Gamma(1/2)\sqrt{2}} e^{-x/2}\, dx \\
&= \frac{\Gamma(\frac{1+2ir}{2}) 2^{\frac{1+2ir}{2}}}{\Gamma(\frac{1}{2}) 2^{\frac{1}{2}}} \\
&= \frac{\Gamma(\frac{1}{2}+ir) 2^{ir}}{\Gamma(\frac{1}{2})},
\end{aligned} \tag{A.1}
$$

where $f(x)$ is the density function of $\chi^2_{(1)}$ and is given by

$$
f(x) = \frac{1}{\Gamma(1/2)\sqrt{2}} x^{-1/2} e^{-x/2}. \quad \blacksquare
$$

Appendix B

Proof of Theorem 3.3.2

Since y_t is a convolution of a Gaussian AR(1) process and an iid sequence with $\chi^2_{(1)}$ distribution, we have

$$\begin{aligned}
c(r_1, \cdots, r_k, \boldsymbol{\theta}) &= E[\exp(ir_1 y_t + ir_2 y_{t+1} + \cdots + ir_k y_{t+k-1})] \\
&= E[\exp(ir_1 h_t + ir_1 \epsilon_t + ir_2 h_{t+1} + ir_2 \epsilon_{t+1} + \cdots + ir_k h_{t+k-1} + ir_k \epsilon_{t+k-1}] \\
&= E[\exp(ir_1 h_t + ir_2 h_{t+1} + \cdots + ir_k h_{t+k-1}] \prod_{j=1}^{k} E[\exp(ir_j \epsilon_{t+j-1})] \\
&= E[\exp(ih_t \sum_{j=1}^{k} \alpha^{j-1} r_j + i\lambda \sum_{j=2}^{k} \frac{1-\alpha^{j-1}}{1-\alpha} r_j + \sum_{l=2}^{k} v_{t+l-1} \sum_{j=l}^{k} r_j \alpha^{j-l})] \\
&\quad \prod_{j=1}^{k} E[\exp(ir_j \epsilon_{t+j-1})] \\
&= \exp[i\frac{\lambda}{1-\alpha} \sum_{j=1}^{k} \alpha^{j-1} r_j + i\lambda \sum_{j=2}^{k} \frac{1-\alpha^{j-1}}{1-\alpha} r_j - \frac{1}{2}(\sum_{j=1}^{k} \alpha^{j-1} r_j)^2 \frac{\sigma^2}{1-\alpha^2} \\
&\quad -\frac{1}{2} \sum_{l=2}^{k} (\sum_{j=l}^{k} \alpha^{j-l} r_j)^2 \sigma^2] \frac{\prod_{j=1}^{k} \Gamma(\frac{1}{2} + ir_j)}{\Gamma^k(\frac{1}{2})} 2^{i\sum_{j=1}^{k} r_j}. \quad \blacksquare
\end{aligned}$$

Appendix C

Proof of Theorem 3.3.3

Defined as the logarithm of the CF, the cumulant generating function is of the form,

$$\begin{aligned}
\phi(r_1,\cdots,r_k) &= \log(c(r_1,\cdots,r_k,\boldsymbol{\theta})) \\
&= i\frac{\lambda}{1-\alpha}\sum_{j=1}^{k}\alpha^{j-1}r_j + i\lambda\sum_{j=2}^{k}\frac{1-\alpha^{j-1}}{1-\alpha}r_j - \frac{1}{2}(\sum_{j=1}^{k}\alpha^{j-1}r_j)^2\frac{\sigma^2}{1-\alpha^2} \\
&\quad -\frac{1}{2}\sum_{l=2}^{k}(\sum_{j=l}^{k}\alpha^{j-l}r_j)^2\sigma^2 + \sum_{j=1}^{k}\log(\Gamma(\frac{1}{2}+ir_j)) + i\sum_{j=1}^{k}r_j\log 2.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
var(y_t) &= \frac{\partial^2\phi(r_1,\cdots,r_k)}{\partial r_1^2}|_{r_1=0} \\
&= \frac{\sigma^2}{1-\alpha^2} + \frac{\Gamma''(0.5)}{\Gamma(0.5)} - (\frac{\Gamma'(0.5)}{\Gamma(0.5)})^2,
\end{aligned}$$

and

$$\begin{aligned}
cov(y_t, y_{t+k-1}) &= \frac{\partial^2\phi(r_1,\cdots,r_k)}{\partial r_1\partial r_k}|_{r_1=0,r_k=0} \\
&= \frac{\alpha^{k-1}\sigma^2}{1-\alpha^2},\ k=1,2,\cdots
\end{aligned}$$

Hence the autocorrelation functions are,

$$\rho_k = \frac{\frac{\alpha^{k-1}\sigma^2}{1-\alpha^2}}{\frac{\sigma^2}{1-\alpha^2} + \frac{\Gamma''(0.5)}{\Gamma(0.5)} - (\frac{\Gamma'(0.5)}{\Gamma(0.5)})^2},\ k=1,2,\cdots \quad \blacksquare$$

Appendix D

Asymptotic Covariance Matrix of the ECF estimator

With $p = 1$, we can define $\boldsymbol{r} = (r_1, r_2), \boldsymbol{s} = (s_1, s_2)$. By the way we defined $\Psi_k(\boldsymbol{r}, \boldsymbol{s})$ in Chapter 1, we have

$$\begin{aligned}
\Psi_k(\boldsymbol{r}, \boldsymbol{s}) &= E[\exp(ir_1 y_1 + ir_2 y_2 + is_1 y_{k+1} + is_2 y_{k+2})] \\
&= E[\exp(ir_1 h_1 + ir_2 \alpha h_1 + is_1 \alpha^k h_1 + is_2 \alpha^{k+1} h_1)] \\
&\quad E[\exp(ir_1 \epsilon_1 + ir_2 \epsilon_2 + is_1 \epsilon_{k+1} + is_2 \epsilon_{k+2})] \\
&\quad E[\exp(ir_2 v_1 + is_1 \sum_{j=1}^{k} \alpha^{j-1} v_{k+1-j} + is_2 \sum_{j=1}^{k+1} \alpha^{j-1} v_{k+2-j})]. \qquad (D.1)
\end{aligned}$$

For the stochastic volatility model,

$$\begin{aligned}
\Psi_0(\boldsymbol{r}, \boldsymbol{s}) &= \exp\left[i\frac{\lambda(r_1 + \alpha r_2 + s_1 + \alpha s_2)}{1-\alpha} - \frac{\sigma^2(r_1 + \alpha r_2 + s_1 + \alpha s_2)^2}{2(1-\alpha^2)}\right] \\
&\quad \frac{\Gamma(0.5 + i(r_1 + s_1))\Gamma(0.5 + i(r_2 + s_2))}{\Gamma^2(0.5)} 2^{i(r_1+s_1+r_2+s_2)} \\
&\quad \exp(-0.5(r_2 + s_2)^2\sigma^2)\exp(ir_2\lambda + is_2\lambda), \qquad (D.2)
\end{aligned}$$

$$\begin{aligned}
\Psi_1(\boldsymbol{r}, \boldsymbol{s}) &= \exp\left[i\frac{\lambda(r_1 + \alpha r_2 + \alpha s_1 + \alpha s_2^2)}{1-\alpha} - \frac{\sigma^2(r_1 + \alpha r_2 + \alpha s_1 + \alpha s_2^2)^2}{2(1-\alpha^2)}\right] \\
&\quad \frac{\Gamma(0.5 + ir_1)\Gamma(0.5 + is_1 + ir_2)\Gamma(0.5 + is_2)}{\Gamma^3(0.5)} 2^{i(r_1+s_1+r_2+s_2)} \exp(-0.5 s_2^2\sigma^2) \\
&\quad \exp(-0.5(r_2 + s_1 + \alpha s_2)^2\sigma^2)\exp(ir_2\lambda + is_1\lambda + is_2(1+\alpha)\lambda), \qquad (D.3)
\end{aligned}$$

and

$$\begin{aligned}
\Psi_k(\boldsymbol{r}, \boldsymbol{s}) &= \exp\left[i\frac{\lambda(r_1 + \alpha r_2 + \alpha^k s_1 + \alpha^{k+1} s_2^2)}{1-\alpha} - \frac{\sigma^2(r_1 + \alpha r_2 + \alpha^k s_1 + \alpha^{k+1} s_2^2)^2}{2(1-\alpha^2)}\right] \\
&\quad \frac{\Gamma(0.5 + ir_1)\Gamma(0.5 + ir_2)\Gamma(0.5 + is_1)\Gamma(0.5 + is_2)}{\Gamma^4(0.5)} 2^{i(r_1+s_1+r_2+s_2)} \\
&\quad \exp(-0.5(r_2 + \alpha^{k-1}s_1 + \alpha^k s_2)^2\sigma^2)\exp(-0.5 s_2^2\sigma^2) \\
&\quad \exp(ir_2\lambda + is_1\frac{1-\alpha^k}{1-\alpha}\lambda + is_2\frac{1-\alpha^{k+1}}{1-\alpha}\lambda) \\
&\quad \prod_{j=2}^{k}\{\exp(-0.5(\alpha^{k-j}s_1 + \alpha^{k+1-j}s_2)^2\sigma^2)\}, \qquad (D.4)
\end{aligned}$$

where $k \geq 2$ in $(D.4)$. Define Σ^2 as

$$\Sigma^2 = \lim_{n\to\infty} \frac{1}{n^2} Var(K_1(\theta_0) + \cdots + K_n(\theta_0)). \tag{D.5}$$

Consequently, we can calculate Σ^2 as follows,

$$\begin{aligned}
\Sigma^2 &= \lim_{n\to\infty} \int\!\!\int\!\!\int\!\!\int \Big\{ \frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\boldsymbol{\theta}} \frac{\partial Re\, c(\boldsymbol{s};\boldsymbol{\theta})}{\boldsymbol{\theta}^T} \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} Cov\Big(cos(\boldsymbol{r}'\boldsymbol{x}_j), cos(\boldsymbol{s}'\boldsymbol{x}_k)\Big) \\
&+ \frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\boldsymbol{\theta}} \frac{\partial Im\, c(\boldsymbol{s};\boldsymbol{\theta})}{\boldsymbol{\theta}^T} \frac{2}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} Cov\Big(cos(\boldsymbol{r}'\boldsymbol{x}_j), sin(\boldsymbol{s}'\boldsymbol{x}_k)\Big) \\
&+ \frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\boldsymbol{\theta}} \frac{\partial Im\, c(\boldsymbol{s};\boldsymbol{\theta})}{\boldsymbol{\theta}^T} \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} Cov\Big(sin(\boldsymbol{r}'\boldsymbol{x}_j), sin(\boldsymbol{s}'\boldsymbol{x}_k)\Big) \Big\} \\
& g(\boldsymbol{r}) g(\boldsymbol{s})\, d\boldsymbol{r} d\boldsymbol{s},
\end{aligned} \tag{D.6}$$

where the covariances in the integrand can be obtained by

$$\begin{aligned}
& \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} Cov\Big(cos(\boldsymbol{r}'\boldsymbol{x}_j), cos(\boldsymbol{s}'\boldsymbol{x}_k)\Big) \\
= & \frac{1}{2n}(Re\, c(\boldsymbol{r}+\boldsymbol{s}) + Re\, c(\boldsymbol{r}-\boldsymbol{s})) - Re\, c(\boldsymbol{r}) Re\, c(\boldsymbol{s}) + \frac{1}{2n^2} \sum_{k=1}^{n-1} (n-k)(Re\, \Psi_k(\boldsymbol{r},\boldsymbol{s}) \\
& + Re\, \Psi_k(\boldsymbol{r},-\boldsymbol{s}) + Re\, \Psi_k(\boldsymbol{s},\boldsymbol{r}) + Re\, \Psi_k(\boldsymbol{s},-\boldsymbol{r})),
\end{aligned} \tag{D.7}$$

$$\begin{aligned}
& \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} Cov\Big(cos(\boldsymbol{r}'\boldsymbol{x}_j), sin(\boldsymbol{s}'\boldsymbol{x}_k)\Big) \\
= & \frac{1}{2n}(Im\, c(\boldsymbol{r}+\boldsymbol{s}) - Im\, c(\boldsymbol{r}-\boldsymbol{s})) - Re\, c(\boldsymbol{r}) Re\, c(\boldsymbol{s}) + \frac{1}{2n^2} \sum_{k=1}^{n-1} (n-k)(Im\, \Psi_k(\boldsymbol{r},\boldsymbol{s}) \\
& - Im\, \Psi_k(\boldsymbol{r},-\boldsymbol{s}) + Im\, \Psi_k(\boldsymbol{s},\boldsymbol{r}) + Im\, \Psi_k(\boldsymbol{s},-\boldsymbol{r})),
\end{aligned} \tag{D.8}$$

$$\begin{aligned}
& \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} Cov\Big(sin(\boldsymbol{r}'\boldsymbol{x}_j), sin(\boldsymbol{s}'\boldsymbol{x}_k)\Big) \\
= & \frac{1}{2n}(Re\, c(\boldsymbol{r}+\boldsymbol{s}) + Re\, c(\boldsymbol{r}-\boldsymbol{s})) - Im\, c(\boldsymbol{r}) Im\, c(\boldsymbol{s}) + \frac{1}{2n^2} \sum_{k=1}^{n-1} (n-k)(Re\, \Psi_k(\boldsymbol{r},-\boldsymbol{s}) \\
& - Re\, \Psi_k(\boldsymbol{r},\boldsymbol{s}) + Re\, \Psi_k(\boldsymbol{s},-\boldsymbol{r}) - Re\, \Psi_k(\boldsymbol{s},\boldsymbol{r})),
\end{aligned} \tag{D.9}$$

with $c(\boldsymbol{r}) = c(\boldsymbol{r};\boldsymbol{\theta})$ which is defined by (3.4). Therefore, Σ^2 can be calculated by using $(D.2)$, $(D.3)$, $(D.4)$, $(D.6)$, $(D.8)$, $(D.9)$, and $(D.10)$. Finally, based on Σ^2, the asymptotic covariance matrix is

$$\frac{1}{n}\left\{\int\int\left[\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}+\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right]g(\boldsymbol{r})d\,r_1 d\,r_2\right\}^{-1}\times$$
$$\Sigma^2\times\left\{\int\int\left[\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}+\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right]g(\boldsymbol{r})d\,r_1 d\,r_2\right\}^{-1}\quad . \blacksquare$$

True Values of Parameters $\alpha = 0.9$ $\sigma = 0.3629$, $\lambda = 0.736$

No. of Replications=500 No. of Observations=2,000

	$\alpha = 0.9$	$\sigma = 0.3629$	$\lambda = -0.736$
MEAN	.892	.3812	.-7962
MED	.895	.3763	-.774
MIN	.75	.1985	-1.843
MAX	.95	.6399	-.3401
RMSE	.03	.067	.231

Table 3.1: A Monte Carlo Study of the ECF Method of a SV Model

True Values of Parameters $\alpha = 0.9$ $\sigma = 0.3629$, $\lambda = -0.736$

No. of Replications=500 No. of Observations=2,000[4]

Method	$\alpha = 0.9$	$\sigma = 0.3629$	$\lambda = -0.736$
GMM	.88(.06)	.31(.10)	-.86(.42)
QML	.88(.06)	.383(.11)	-.853(.46)
MCMC	.896(.02)	.359(.034)	-.763(.15)
SML	.902(.02)	.359(.039)	-.721(.15)

Table 3.2: Monte Carlo Comparison of GMM, QML, MCMC, and SML Estimates

[4]The table shows the mean and RMSE(in parentheses).

Method	α	σ	λ
ECF	-.0676	.747	-0.29
MCMC	0.97	.15	-.002

Table 3.3: Empirical Comparison of ECF and MCMC Estimates

	KS statistic	p-value
ECF	0.0257	0.498
MCMC	0.0875	0

Table 3.4: Kolmogorov-Smirnov Goodness-of-Fit Test of ECF and MCMC Estimates

Method	α	σ	λ
$p = 2$	-0.0719	0.743	-0.45
$p = 3$	-0.0822	0.740	-0.31
$p = 4$	-0.0927	0.738	-0.38
$p = 5$	-0.0742	0.731	-0.40

Table 3.5: Empirical Results of ECF Estimates with Different Values of p

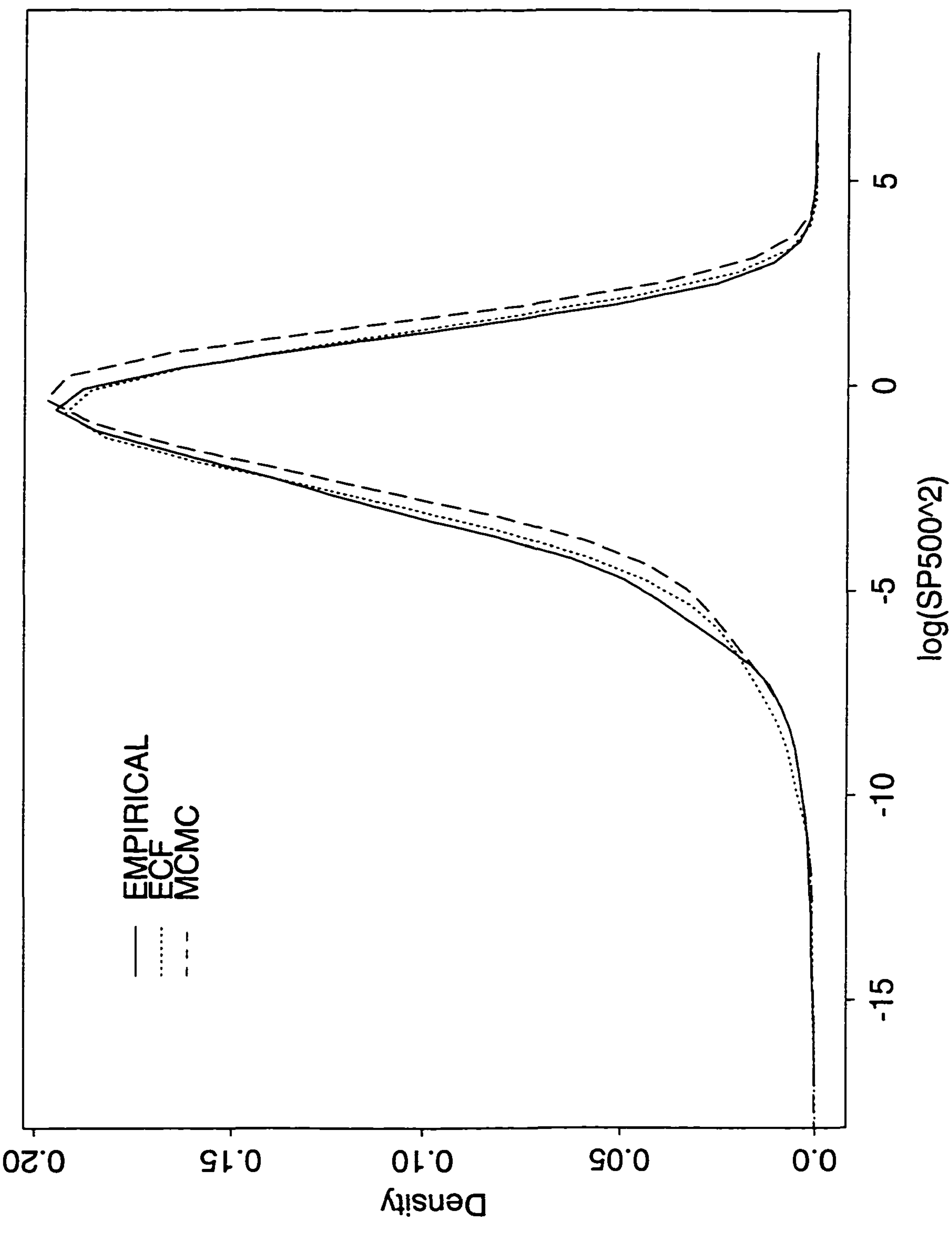

Figure 3.1: Empirical Density and Densities from ECF and MCMC Estimates

Chapter 4

ESTIMATION OF THE DIFFUSION JUMP MODEL VIA EMPIRICAL CHARACTERISTIC FUNCTION

4.1 Introduction

Motivated by the non-linearity in financial time series, such as volatility clustering, researchers have resorted to the processes which can generate such a property. Among them are ARCH models proposed by Engle (1982), GARCH models proposed by Bollerslerv (1986), Stochastic Volatility (SV) models proposed by Clark (1973), Tauchen and Pitts (1983) and Taylor (1986), and the diffusion jump model with a self-exciting intensity proposed by Knight and Satchell (1993). A common feature for these models is that they allow for time dependence. Also, all these models can be representable as a martingale difference and hence consistent with the market efficiency in the weak sense. Among these three types of models, the ARCH-type models have thus far attracted the most attention for at least one reason. The reason is that the estimation of the ARCH-type models is relatively easy to implement. Since the ARCH-type models have Markovian properties, the conditional maximum likelihood (CML) estimation method is usually employed. Unfortunately there are still some difficulties involved in the ARCH-type models. One of them is the model's inability to explain that large changes are not unusual after stable periods, since the volatility evolves according to a deterministic mechanism. Another difficulty is that the model can not well explain the dis-continuity existing in the sample path of most financial time series; see Jorion (1988).

By introducing another error term and hence treating the volatility as a stochastic process, the SV model can overcome the first difficulty involved in the ARCH-type models. However, a new problem is arising. Even for the simplest SV model, neither the exact likelihood function nor the conditional likelihood function has closed form. Therefore, the likelihood based estimation method is extremely difficult to implement. Consequently, some alternative estimation methods were discussed in the recent literature. The methods include the generalized method of moments (GMM), quasi-maximum likelihood (QML) method, Markov Chain Monte Carlo (MCMC), simulated maximum likelihood (SML) method and efficient method moment (EMM).

An alternative way to overcome the difficulty involved in the ARCH-type model is to compound a Brownian motion (BM) and a Poisson jump process. The jump component is first introduced by Merton (1976) and extended by Jorion (1988) to explain the dis-continuity. Furthermore, to allow for non-linearity, the intensity parameter in the Poisson process is assumed to be self exciting and hence time dependent, making the model different from the model proposed by Jorion (1988). The model was first proposed by Knight and Satchell (1993) and then applied to fit the UK stock data by Knight, Satchell and Yoon (1993). Unfortunately, the maximum likehood method is not applicable to this model since the likelihood function has no closed form. Instead Knight, Satchell and Yoon (1993) employ the GMM method to estimate the model. When fitting the UK data to the model, however, they find that the GMM estimates sometimes do not make sense. For example, the estimate of the variance parameter for some stocks is negative. The finding is due to the poor finite sample properties of the GMM estimator. In this thesis, we use an alternative approach to estimate this model - via the empirical characteristic function. This chapter is organized as follows. The next section introduces the diffusion jump model where the intensity in the Poisson jump process is self-exciting and explains why the model is difficult to estimate. Section 4.3 presents a discussion on ECF estimation for this model; the characteristic function of the model is obtained as well. Section 4.4 discusses the implementation of the ECF method as well as a Monte Carlo study and an empirical application. The proof of the theorem is in the Appendix.

4.2 The Model

It is common in financial literature to assume that the price of an asset at time t, $P(t)$, follows a geometric Brownian Motion (BM),

$$dP(t) = \alpha P(t)dt + \sigma P(t)dB(t), \tag{2.1}$$

where $B(t)$ is standard Brownian motion, α is the instaneous return and σ^2 is the instaneous volatility. By including the jump component, Knight and Satchell (1993) assume that the price follows the mixed Brownian-Poisson process,

$$dP(t) = \alpha P(t)dt + \sigma P(t)dB(t) + P(t)(\exp(Q(t)) - 1)dN(t), \tag{2.2}$$

where $Q(t)$ is a normal variate with mean μ_Q and variance σ_Q^2 in the interval $(t, t+\Delta t]$, $N(t)$ is a Poisson process with intensity parameter $\lambda(t)$, and $Q(t)$ is independent with $N(t)$. By using Ito's lemma, we can solve the stochastic equation (2.2) for the stock return $X(t)(= \log(P(t)/P(t-1))$,

$$\begin{aligned} X(t) &= (\alpha - \frac{\sigma^2}{2})t + \sigma B(1) + \sum_{n=1}^{\Delta N(t)} Q(n) \\ &= \mu t + \sigma B(1) + \sum_{n=1}^{\Delta N(t)} Q(n), \end{aligned} \tag{2.3}$$

where $\mu = \alpha - \frac{\sigma^2}{2}$. Hence the behavior of $X(t)$ depends not only on the continuous diffusion part $(\alpha - \frac{\sigma^2}{2})t + \sigma B(1)$, but also a dis-continuous jump part $\sum_{n=1}^{\Delta N(t)} Q(n)$.

The continuous part is responsible for the usual day-to-day price movement, such as temporary imbalance between supply and demand or firm-specific information that only marginally affects the value of the stock. The dis-continuous jump part corresponds to the arrival of new important information to the market. When the jump part has a significant effect, a large change can certainly follow stable periods. Furthermore, the Poisson process $N(t)$ is assumed to have an intensity function $\lambda(t)$ which is self-exciting as follows,

$$\lambda(t) = \sum_{i=1}^{m} \alpha_i \nu^2(t-i) + \sum_{j=1}^{l} \beta_j Var(X(t-j)|I(t-j-1)), \tag{2.4}$$

where $\nu(t)$ is $N(0,1)$ conditional on $N(t)$, and $I(t)$ is information up to the close of the market on day t. The motivation for the model is the idea that the flow of information at day t's trading is conditioned by the news known at the close of trading at day $t-1$ or prior to opening at day t. Also, the motivation comes from the idea that the expected number of jumps, which corresponds to the arrival of new important information to the market, depends upon past volatility and derivation from fundamentals $B^2(1)$ $(= \nu^2(t))$. Due to the dependency of $\lambda(t)$, $X(t)$ could be time dependent. Considering the conditional variance of $X(t)$, we have

$$\begin{aligned} &Var(X(t)|I(t-1)) = \\ &\quad \sigma^2 + (\mu_Q^2 + \sigma_Q^2)\left(\sum_{i=1}^{m} \alpha_i \nu^2(t-i) + \sum_{j=1}^{l} \beta_i Var(X(t-j)|I(t-j-1))\right). \end{aligned} \tag{2.5}$$

Obviously the conditional variance is time dependent and volatility clustering can be accounted for by the model. Unfortunately, $X(t)$ has no closed form expression for the likelihood function because it compounds three processes, the BM, a Poisson process and an i.i.d. $\chi^2_{(1)}$ sequence. This lack of tractable form for the likelihood makes the ML method difficult to implement.

For simplicity, Knight, Satchell and Yoon (1993) consider the model with $m =$

$l = 1$ in Equation (2.4), i.e.,

$$\lambda(t) = \alpha\nu^2(t-1) + \beta Var(X(t-1)|I(t-2)). \qquad (2.6)$$

In Appendix A, we will show that Equation (2.6) is equivalent to

$$\lambda(t) = \beta\sigma^2 + \beta(\mu_Q^2 + \sigma_Q^2)\lambda(t-1) + \alpha\nu^2(t-1), \qquad (2.7)$$

or

$$Var(X(t)|I(t-1)) = \sigma^2 + \beta(\mu_Q^2+\sigma_Q^2)Var(X(t-1)|I(t-2)) + \alpha(\mu_Q^2+\sigma_Q^2)\nu^2(t-1). \qquad (2.8)$$

Knight, Satchell and Yoon (1993) claim that $m = l = 1$ is found to suffice in most applications. From Equation (2.8) we can tell that the model is similar to GARCH(1, 1) model proposed by Bollerslev (1986). For the model based on Equation (2.3) and Equation (2.6), some conditions on the parameters are needed for the stationarity of the process. See Knight, Satchell and Yoon (1993) for details. Furthermore, they give the formula to calculate up to the fifth cumulant and the covariance between $X(t)$ and $X(t-s)$. Based on the results on moments, the GMM estimator can be obtained. However, since the GMM only uses a few moment conditions, it is not surprising that the finite sample performance is not good. In the next section we will derive the characteristic function of $X(t)$ and then perform the estimation based on the empirical characteristic function. Since the characteristic function contains the same amount of information as the distribution function, the model is fully and uniquely parameterized by the CF. Therefore, inference based on the ECF can outperform that based on the GMM.

4.3 ECF Estimation

The model we are going to estimate via the ECF is defined by (2.3) and (2.6). In order to use the ECF method, of course, we need to find the expression of the joint characteristic function. The joint characteristic function for $X(t), \cdots, X(t-k)$ is obtained in the theorem below. The joint characteristic function for the model defined by (2.3) and (2.4) can be obtained in the similar way.

Theorem 4.3.1 *If a random process $\{X(t)\}_{t=1}^{n}$ is a compound Poisson process with self-exciting intensity which is defined by Equations (2.3) and Equation (2.6), then the joint* CF *of $X(t), \cdots, X(t-k)$ is,*

$$c(r_1, \cdots, r_{k+1}; \boldsymbol{\theta}) = \exp(i\mu \sum_{j=1}^{k+1} r_j - \frac{1}{2}\sigma^2 \sum_{j=1}^{k+1} r_j^2) \exp\left\{\frac{\beta\sigma^2}{1-\phi} \sum_{j=1}^{k+1} G(r_j)\right\} \prod_{l=0}^{\infty}\left\{1 - 2\alpha\phi^l \sum_{j=1}^{k+1} \phi^{k+1-j} G(r_j)\right\}^{-1/2} \prod_{j=2}^{k+1}\left\{1 - 2\alpha \sum_{l=1}^{j} \phi^{j-l} G(r_l)\right\}^{-1/2}. \quad (3.1)$$

where $\phi = \beta(\mu_Q^2 + \sigma_Q^2)$, $G(r) = \exp(ir_j\mu_Q - \frac{r_j^2\sigma_Q^2}{2}) - 1$, and $\boldsymbol{\theta} = (\mu, \sigma^2, \alpha, \beta, \mu_Q, \sigma_Q^2)$ is the parameter.

Proof: See Appendix A.

Using this theorem, we can easily obtain the covariance for the returns and square of the returns. For example, $Cov(X(t), X(t-s)) = 2\mu_Q^2\alpha^2\theta^s/(1-\theta^2)$. Hence $X(t)$ and $X(t-s)$ are uncorrelated when $\mu_Q = 0$. However, $X(t)$ and $X(t-1)$ are not independent in any case since $c(r_1, r_2) \neq c(r_1)c(r_2)$.

In order to use the ECF method to estimate the model defined by (2.3) and (2.6), we have to choose a value for p. As we argued before, a larger value of p works better than a smaller value of p, however, we shall choose $p = 1$ at first. For the model defined by (2.3) and (2.6) we note that with $p = 1$,

$$\begin{aligned} c(r_1, r_2, \boldsymbol{\theta}) &= \exp\left\{\frac{\beta\sigma^2}{1-\phi}(G(r_1) + G(r_2)) + i\mu(r_1 + r_2) - \frac{1}{2}\sigma^2(r_1^2 + r_2^2)\right\} \\ &\quad \prod_{l=0}^{\infty}\left\{1 - 2\alpha\phi^l(\phi G(r_1) + G(r_2))\right\}^{-1/2} (1 - 2\alpha G(r_1))^{-\frac{1}{2}}, \end{aligned} \quad (3.2)$$

and

$$c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^{n} \exp(ir_1 y_j + ir_2 y_{j+1}). \tag{3.3}$$

Defining $Re\, c(r_1, r_2, \boldsymbol{\theta})$, $Re\, c_n(r_1, r_2)$, $Im\, c(r_1, r_2, \boldsymbol{\theta})$ and $Im\, c_n(r_1, r_2)$ to be the real and imaginary parts of $c(r_1, r_2)$ and $c_n(r_1, r_2)$ respectively, we have,

$$Re\, c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^{n} cos(r_1 y_j + r_2 y_{j+1}), \tag{3.4}$$

and

$$Im\, c_n(r_1, r_2) = \frac{1}{n}\sum_{j=1}^{n} sin(r_1 y_j + r_2 y_{j+1}). \tag{3.5}$$

As we mentioned before, a clear advantage of choosing the transformation variable continuously is that we do not need to choose q. Furthermore, since the Monte Carlo study conducted in Chapter 2 shows that the continuous ECF method works better than the discrete ECF method, we use the continuous ECF method to estimate the model. However, the optimal weight function in the continuous version is not readily obtained because the conditional score function has no closed form expression for the model. Instead we use the OLS of the continuous ECF method with a constant weight function chosen over an interval $[0, 1]$. Therefore, the procedure is to choose $(\hat{\mu}, \hat{\sigma}^2, \hat{\alpha}, \hat{\beta}, \hat{\mu}_Q, \hat{\sigma}_Q^2)$ to minimize,

$$\begin{aligned} \textstyle\int_0^1 \int_0^1 \Big[&(Re\, c(r_1, r_2, \boldsymbol{\theta}) - \tfrac{1}{n}\textstyle\sum_{j=1}^n cos(ir_1 y_j + ir_2 y_{j+1}))^2 \\ &+(Im\, c(r_1, r_2, \boldsymbol{\theta}) - \tfrac{1}{n}\textstyle\sum_{j=1}^n sin(ir_1 y_j + ir_2 y_{j+1}))^2 \Big] dr_1 dr_2, \end{aligned} \tag{3.6}$$

where $c(r_1, c_2)$ is given by (3.2).

We can check that the appropriate regularity conditions given in Chapter 1 hold. Therefore, the resulting estimators are consistent and asymptotically normal. The asymptotic covariance matrix of the estimators is given below,

$$\frac{1}{n}\left\{\int_0^1\int_0^1 \left[\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} + \frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right] dr_1 dr_2\right\}^{-1} \times$$
$$\Sigma^2 \times \left\{\int_0^1\int_0^1 \left[\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} + \frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right] dr_1 dr_2\right\}^{-1}.$$

4.4 Implementation, Simulation and Application

4.4.1 Implementation

The implementation of the ECF method essentially requires minimizing (3.6), and thus involves double integrals. Unfortunately, no analytical solutions for either the double integrals or the optimization are available. Consequently, we will numerically evaluate the multiple integral (3.6), followed by numerical minimization of (3.6) with respect to $\boldsymbol{\theta}$. The numerical solutions are the desired estimators.

A DCUHRE algorithm proposed by Bernsten, Espelid and Genz (1993) is used to approximate the two dimensional integrations in (3.6). Since there is no analytical expression for the derivative of the objective functions, the Powell's conjugate direction algorithm (see Powell (1964)) is used to find the global minimum. All computations were done in double precision.

By using the implementation procedure, we examine the performance of the ECF method in the estimation of the diffusion jump model in a Monte Carlo study. We also apply the procedure to a real dataset.

4.4.2 Monte Carlo Simulation

The Monte Carlo study is designed to check the viability of the ECF method. For simplicity, in the model defined by (2.3) and (2.6), we let $\mu = \mu_Q = \sigma = 0, \alpha = 1$, and thus $\phi = \beta\sigma_Q^2$. Therefore, the model can be rewritten as,

$$X(t) = \sum_{n=1}^{\Delta N(t)} Q(n), \tag{4.1}$$

where $Q(n)|\Delta N(t) \sim i.i.d.\, N(0, \sigma_Q^2), \Delta N(t) \sim P(\lambda(t))$, and

$$\lambda(t) = \phi\lambda(t-1) + \nu^2(t-1). \tag{4.2}$$

Obviously, for the process $\{\lambda(t)\}$ to be stationary, $|\phi|$ must be smaller than 1. We choose parameters as $\sigma_Q^2 = 1$ and $\beta = 0.5$, which implies $\phi = 0.5$. The number of observations is set at $T = 2,000$ and the number of replications is set at 500. For the numerical optimization, the starting point is chosen to be the GMM estimates. In the GMM, we use the second and fourth cumulants and the weight function is chosen to be the identity matrix.

Table 4.1 reports the simulation results. The table shows the mean, median, minimum, maximum, mean square error (MSE) and root mean square error (RMSE) for both estimates, and serves to illustrate that the ECF method outperforms the GMM. Also see Figure 4.1 and Figure 4.2. From the table and two figures, we know that, with the same weight function for both methods, the ECF works much better than the GMM. For example, the MSE of $\hat{\beta}$ for GMM is at least 15 times larger than that for the ECF. The results for the estimates of σ_Q^2 also favor the ECF method.

4.4.3 Application

Data

The data we used is the same as that we used in Chapter 3. We plot it in Figure 4.3. The October 1987 stock-market crash can be clearly identified in the graph.

Empirical Results

In this empirical study, we fit the model defined by (2.2) and (2.6) to the data. The parameters of interest are $\boldsymbol{\theta} = (\mu, \sigma^2, \alpha, \beta, \mu_Q, \sigma_Q^2)$. Again we let $\phi = \beta(\mu_Q^2 + \sigma_Q^2)$. To simply the model, we shall assume that $X(t)$ is stationary. Consequently, $\lambda(t)$ has to be stationary. From (2.7) we know that $\lambda(t)$ is stationary only if $|\phi| < 1$. We further require $\alpha, \beta \geq 0$ to guarantee the intensity function to be non-negative. Two constraints are $\sigma^2 \geq 0$ and $\sigma_Q^2 \geq 0$. The empirical results are reported in Table 4.2.

Firstly, from Table 4.2, we can calculate $\phi = \beta(\mu_Q^2 + \sigma_Q^2) = 0.29$, which implies a positive correlation of the conditional variance between two consecutive trading days. Furthermore, the conditional variance at day t has less dependency on the deviation from market fundamentals $\nu(t-1)$. This dependency is captured by the $\alpha(\mu_Q^2 + \sigma_Q^2)$, which equals to 0.001. Also note that the corresponding parameters in the GARCH(1, 1) model are usually much larger. Therefore, our estimates suggest much lower volatility clustering and hence less inefficiency in the market.

Secondly, we note that the instaneous mean and variance in the continuous part are 0.055 and 0.79 respectively. As we mentioned before, the continuous part is due to general economic information which only has marginal effect. Therefore, on average the log return is about 0.055% because of the normal economic activities.[1]

Finally, since μ_Q and σ_Q^2 corresponds to the mean and variance of the normal variate in the dis-continuous part, $\mu_Q = -0.16$ implies a negative return due to the impact of important information and $\sigma_Q^2 = 6.18$ suggests a much bigger volatility an average important information can incur than that the usual information can incur.

[1]This is because $X(t) = 100(\log(P(t+1)) - \log(P(t)))$.

Appendix A

Proof of Theorem 4.3.1

We start the proof by showing that Equation (2.6), Equation (2.7) and Equation (2.8) are equivalent. From the definition of $I(t-1)$, we have

$$Var(X(t)|I(t-1)) = Var(\mu+\sigma B(1) + \sum_{n=1}^{\Delta N(t)} Q(n)|\lambda(t)) = \sigma^2 + \lambda(t)(\mu_Q^2+\sigma_Q^2). \quad (A.1)$$

Substituting out $\lambda(t)$ in Equation (2.6), we then have,

$$Var(X(t)|I(t-1)) = \sigma^2 + \alpha(\mu_Q^2+\sigma_Q^2)\nu(t-1)^2 + \beta(\mu_Q^2+\sigma_Q^2)Var(X(t-1)|I(t-2)). \quad (A.2)$$

Therefore, the condition variance of this model follows a GARCH(1, 1). However, by substituting out $Var(X(t-1)|I(t-2))$, we have,

$$\lambda(t) = \beta\sigma^2 + \beta(\mu_Q^2+\sigma_Q^2)\lambda(t-1) + \alpha\nu^2(t-1), \quad (A.3)$$

which is Equation (2.7). Since the intensity represents how fast new information arrives, Equation (2.7) means that the speed of the arrival of new information today depends on how frequent new information has arrived yesterday, as well as a random component. By applying backward induction to Equation (2.7), we have

$$\lambda(t) = \frac{\beta\sigma^2}{1-\phi}(1-\phi^{t-1}) + \phi^{t-1}\lambda(1) + \alpha\sum_{j=0}^{t-2}\phi^j\nu^2(t-j-1). \quad (A.4)$$

If $|\phi| < 1$, as $t \to \infty$,

$$\lambda(t) = \frac{\beta\sigma^2}{1-\phi} + \sum_{j=0}^{\infty}\phi^j\nu^2(t-j-1). \quad (A.5)$$

Consequently, the characteristic function of $\lambda(t)$ is,

$$\begin{aligned}\Psi(r) &= E(\exp(ir\lambda(t))) \\ &= E\left\{\exp\left(is(\frac{\beta\sigma^2}{1-\phi} + \sum_{j=1}^{\infty}\phi^j\nu^2(t-j-1))\right)\right\} \\ &= \exp(ir\frac{\beta\sigma^2}{1-\phi})\prod_{j=0}^{\infty}(1-2i\alpha\phi^j)^{-\frac{1}{2}}. \quad (A.6)\end{aligned}$$

Considering

$$\sum_{n=1}^{\Delta N(t-l)} Q(n)|\Delta N(t-l) \sim N(\mu_Q \Delta N(t-l), \sigma_Q^2 \Delta N(t-l)),$$

we have,

$$\begin{aligned}
& E\left\{\prod_{l=1}^{k+1} \exp\left(ir_l \sum_{n=1}^{\Delta N(t-l)} Q(n)\right)\right\} \\
= \; & E\left\{E\left\{E\left[\prod_{l=1}^{k+1} \exp(ir_l \sum_{n=1}^{\Delta N(t-l)} Q(n))|\Delta N(t-l), I(t-k)\right]\right\}\right\} \\
= \; & E\left\{E\left[\prod_{l=1}^{k+1} \exp(i\mu_Q r_l \Delta N(t-l) - \frac{r_l^2}{2}\sigma_Q^2 \Delta N(t-l))|I(t-k)\right]\right\} \\
= \; & E\left\{\prod_{l=1}^{k+1} \exp[\lambda(t-l)(\exp(i\mu_Q r_l - \frac{r_l^2}{2}\sigma^2) - 1)]\right\} \\
= \; & E\{\prod_{l=1}^{k+1} \exp[(\frac{\beta\sigma^2}{1-\phi}(1-\phi^{k-l}) + \phi^{k-l}\lambda(t-k) \\
& +\alpha(\nu^2(t-j-1) + \cdots + \phi^{k-l-1}\nu^2(t-k)))(\exp(i\mu_Q r_l - \frac{r_l^2}{2}\sigma^2) - 1)]\} \\
= \; & \exp\left[\sum_{l=1}^{k+1} \frac{\beta\sigma^2}{1-\phi}(1-\phi^{k-l})(\exp(i\mu_Q r_l - \frac{r_l^2}{2}\sigma^2) - 1)\right] \\
& E\left\{\exp\left[\lambda(t-k)\sum_{l=1}^{k+1}\left(\phi^{k+1-l}(\exp(i\mu_Q r_l - \frac{r_l^2}{2}\sigma^2) - 1)\right)\right]\right\} \\
& E\left\{\prod_{l=1}^{k+1} \exp\left[\alpha(\nu^2(t-j-1) + \cdots + \phi^{k-l-1}\nu^2(t-k))(\exp(i\mu_Q r_l - \frac{r_l^2}{2}\sigma^2) - 1)\right]\right\} \\
= \; & \exp\left[\sum_{l=1}^{k+1} \frac{\beta\sigma^2}{1-\phi}(1-\phi^{k-l})(\exp(i\mu_Q r_l - \frac{r_l^2}{2}\sigma^2) - 1)\right] \\
& \exp\left\{\frac{\beta\sigma^2}{1-\phi}\sum_{l=1}^{k+1}\left[\phi^{k+1-l}(\exp(i\mu_Q r_l - \frac{r_l^2}{2}\sigma^2) - 1)\right]\right\} \\
& \prod_{j=0}^{\infty}\left\{1 - 2\alpha\phi^j \sum_{l=1}^{k+1}\left[\phi^{k+1-l}(\exp(i\mu_Q r_l - \frac{r_l^2}{2}\sigma^2) - 1)\right]\right\} \\
& E\left\{\exp\left[\sum_{j=2}^{k+1}\left(\alpha\sum_{l=1}^{j}\phi^{j-l}(\exp(i\mu_Q r_l - \frac{r_l^2}{2}\sigma^2) - 1)\right)\right]\right\} \\
= \; & \exp\left[\sum_{l=1}^{k+1} \frac{\beta\sigma^2}{1-\phi}(1-\phi^{k-l})G(r_l)\right] \exp\left\{\frac{\beta\sigma^2}{1-\phi}\sum_{l=1}^{k+1}\left[\phi^{k+1-l}G(r_l)\right]\right\} \\
& \prod_{l=0}^{\infty}\left\{1 - 2\alpha_1\phi^l \sum_{j=1}^{k+1}\phi^{k+1-j}G(r_j)\right\}^{-1/2} \prod_{j=2}^{k+1}\left\{1 - 2\alpha\sum_{l=1}^{j}\phi^{j-l}G(r_l)\right\}^{-1/2}
\end{aligned}$$

$$= \exp\left\{\frac{\beta\sigma^2}{1-\phi}\sum_{j=1}^{k+1} G(r_j)\right\}\prod_{l=0}^{\infty}\left\{1-2\alpha\phi^l\sum_{j=1}^{k+1}\phi^{k+1-j}G(r_j)\right\}^{-1/2}$$

$$\prod_{j=2}^{k+1}\left\{1-2\alpha\sum_{l=1}^{j}\phi^{j-l}G(r_l)\right\}^{-1/2}. \tag{A.7}$$

Thus, the characteristic function of $X(t), \cdots, X(t-k)$ is,

$$c(r_1, \cdots, r_{k+1}; \boldsymbol{\theta})$$

$$= E\left\{\exp(ir_1X(t)+\cdots+ir_{k+1}X(t-k))\right\}$$

$$= E\left\{\exp(ir_1(\mu+\sigma B(1)+\sum_{n=1}^{\Delta N(t-l)}Q(n))+\cdots+ir_{k+1}(\mu+\sigma B(1)+\sum_{n=1}^{\Delta N(t-l)}Q(n)))\right\}$$

$$= E\left\{\exp(i\mu\sum_{l=1}^{k+1}r_l)\exp(i\sigma(r_1W(1)+\cdots+r_{k+1}W(1)))\exp(\sum_{l=1}^{k+1}ir_l\sum_{n=1}^{\Delta N(t-l)}Q(n))\right\}$$

$$= \exp(i\mu\sum_{j=1}^{k+1}r_j-\frac{1}{2}\sigma^2\sum_{j=1}^{k+1}r_j^2)E\left\{\prod_{l=1}^{k+1}\exp\left(ir_l\sum_{n=1}^{\Delta N(t-l)}Q(n)\right)\right\}$$

$$= \exp(i\mu\sum_{j=1}^{k+1}r_j-\frac{1}{2}\sigma^2\sum_{j=1}^{k+1}r_j^2)\exp\left\{\frac{\beta\sigma^2}{1-\phi}\sum_{j=1}^{k+1}G(r_j)\right\}$$

$$\prod_{l=0}^{\infty}\left\{1-2\alpha\phi^l\sum_{j=1}^{k+1}\phi^{k+1-j}G(r_j)\right\}^{-1/2}\prod_{j=2}^{k+1}\left\{1-2\alpha\sum_{l=1}^{j}\phi^{j-l}G(r_l)\right\}^{-1/2}. \quad \blacksquare$$

True Values of Parameters $\beta = 0.5$ $\sigma_Q^2 = 1.0$

No. of Replications=500 No. of Observations=2,000

	$\beta = 0.5$		$\sigma_Q^2 = 1.0$	
	ECF	GMM	ECF	GMM
MEAN	.5098	.5602	.9792	.9792
MED	.5078	.5408	.9746	.9512
MIN	.3408	.0958	.6342	.4213
MAX	.6481	1.804	1.340	1.706
MSE	.00276	.0469	.0184	.0407
RMSE	.0525	.2165	.1358	.2016

Table 4.1: Monte Carlo Comparison of ECF and GMM for a Diffusion Jump Model

Method	μ	σ^2	α	β	μ_Q	σ_Q^2
ECF	0.055	.79	0.00016	0.047	-0.16	6.18

Table 4.2: Empirical Results when Fitting the Diffusion Jump Model to SP500 Daily Returns

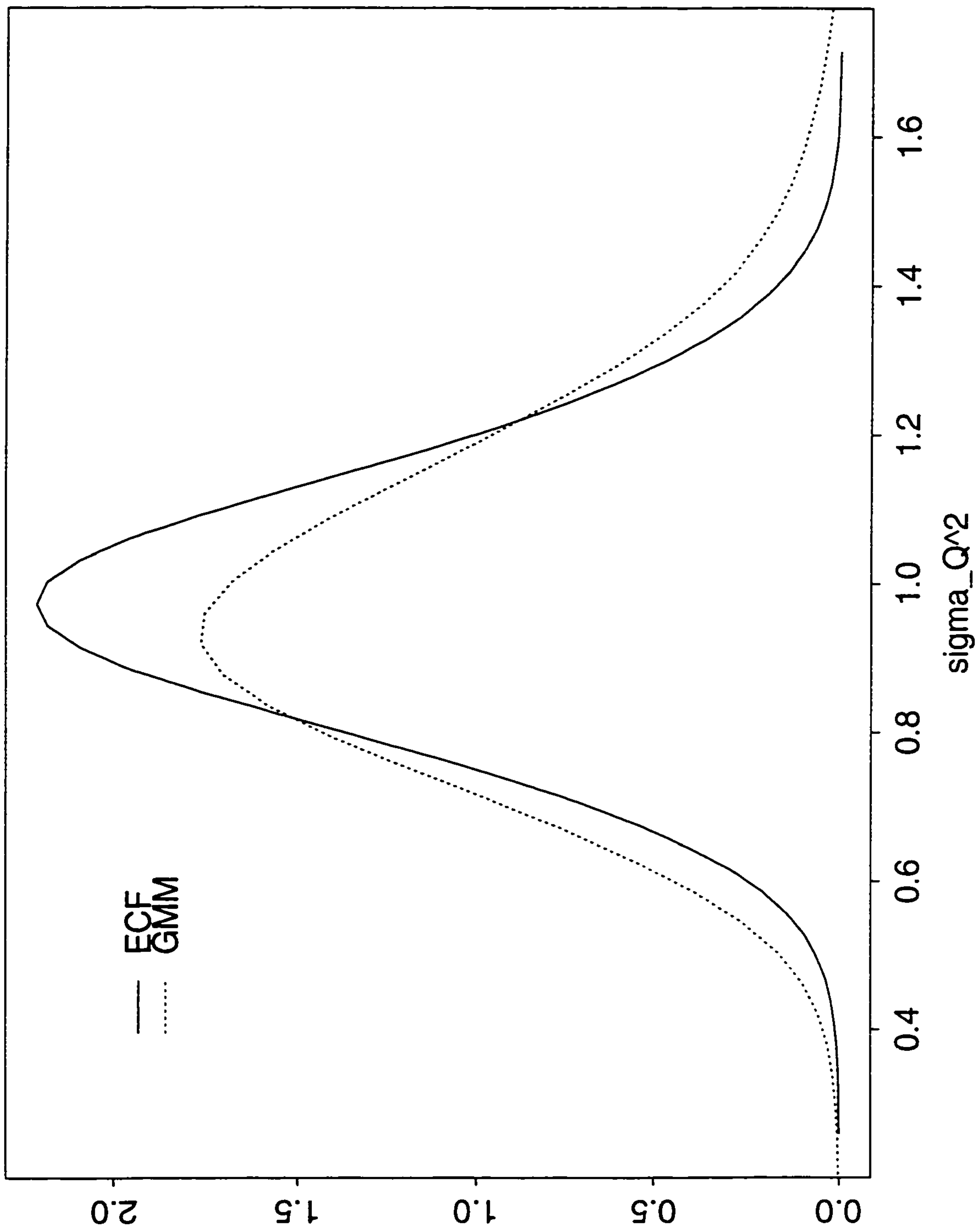

Figure 4.1: Density Function of Estimator of σ_Q^2 in the Diffusion Jump Model

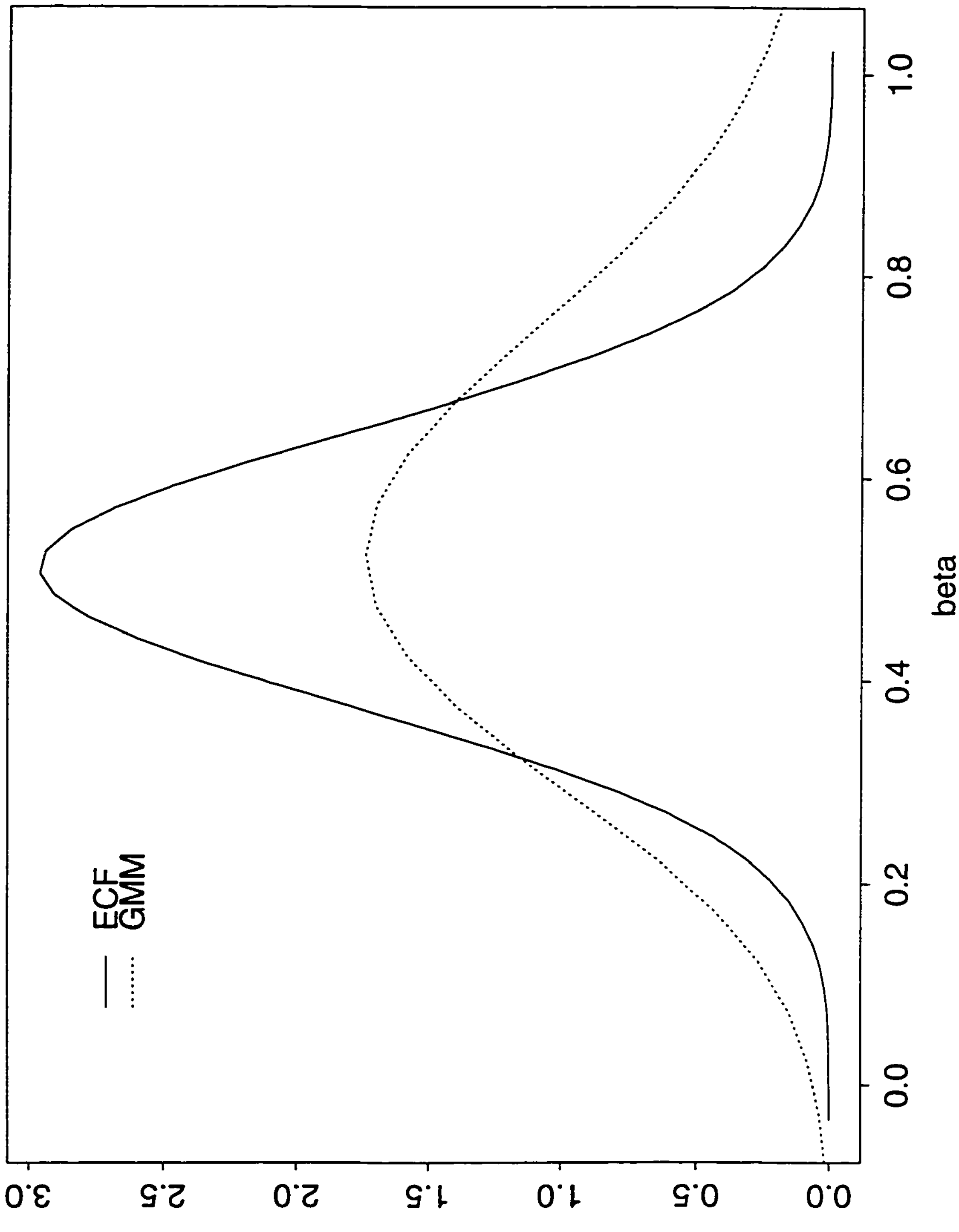

Figure 4.2: Density Function of Estimator of β in the Diffusion Jump Model

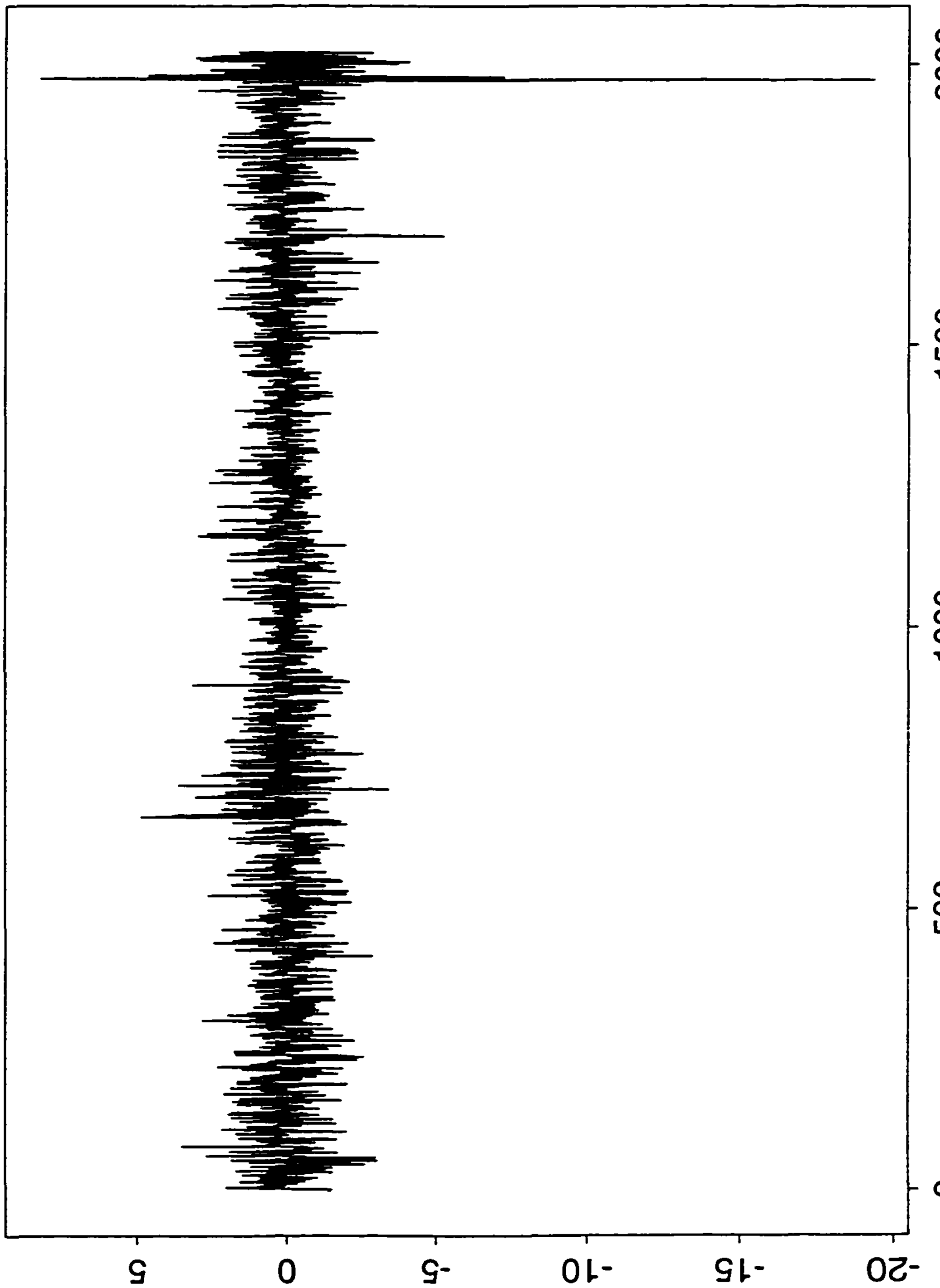

Figure 4.3: SP500 Daily Returns

Chapter 5

A TEST STATISTIC AND ITS APPLICATION IN FINANCIAL MODELLING

5.1 Introduction

Modelling stock returns has been a very interesting topic for a long time. One reason is that some important models in financial theory critically rely on the distribution form for the returns of underlying stocks, such as the option pricing model. In the search for satisfactory descriptive models of stock returns, many distributions have been tried and some new distributions have been created over past several decades. All these alternative models can be categorized by two families. One is finite-variance distributions. Examples include the normal distribution by Osborne (1959), the Student t distribution by Blattberg and Gonedes (1974), the mixture of normals (MN) by Kon (1984), the compound log-normal and normal (LN) distribution by Clark (1973), the mixed diffusion-jump (MDJ) model by Press (1967) and more recent one, the Weibull distribution by Mittnik and Rachev (1993). The other family has infinite-variance, such as the Stable distribution by Mandlebrot (1963).

The Stable distribution has been appreciated as a possible alternative to describe the stock returns for both statistical and economic reasons. Statistically speaking, the Stable distribution has domain of attraction and belongs to their domain of attraction. Economically speaking, the stable distribution has unbounded variation and allows arbitrage, and hence is consistent with continuous-time equilibrium in competitive markets (see McCulloch (1978)). Furthermore, a sample path generated by the Stable distribution almost surely contains an infinite number of dis-continuities, consistent with the efficient market hypothesis (see McCulloch (1978)).

Despite these appealing properties, the Stable distribution is less commonly used today. It has fallen out of favor, partly because of the difficulties involved in theoretical modelling; standard financial theory, such as option theory, almost always requires finite variance of returns. Furthermore, evidence has been found against the Stable distribution. Firstly, by using the likelihood ratio test, Blattberg and Gonedes (1974) found that the Student t distribution has greater descriptive validity than the symmetric Stable distribution, and Tucker (1992) found that finite-variance models outperform the asymmetric Stable distribution. By using the Komogorov-Smirnov test, Mittnik and Rachev (1993) found that the Weibull distribution is a suitable candidate. Secondly, when the tail behavior was investigated, Akgiray and Booth (1987) found that the tails of Stable distribution are too thick to fit the empirical data. Thirdly, Lau, Lau and Wingender (1990) found that as the sample size gets big the sample variance seems to converge while the Stable distribution implies that sample variance should blow up rapidly. Finally, the evidence provided by Blattberg and Gonedes (1974) indicates that the distribution of monthly returns conforms well to the normal distribution, while the Stable distribution implies that long horizon (for example, monthly) returns will be just as non-normal as short-horizon (for example, daily) returns.

The purpose of this chapter is to re-examine the descriptive power of the finite-variance distribution family and the infinite-variance distribution family as models of daily stock returns. However, instead of using overall goodness of fit testing methodology, we concentrate on studying the variance behavior for chosen distribution families.

To be more specific, we propose a test statistic to distinguish finite-variance families against infinite-variance families for stock returns. Particular attention is paid to the variance due to two reasons. Firstly, as far as the variance is concerned, an infinite-variance model is fundamentally riskier than a finite-variance one. Secondly and more importantly, many financial models critically depend on the assumption on the second moment. Examples include the capital asset pricing model (CAPM) and the Black-Scholes model. As a result, finite variance and infinite variance could have very different implications for theoretical and empirical analysis. Unfortunately, testing for finite variance or infinite variance based on a sample without choosing specific distribution families will probably never be possible to have a test with uniformly good power. Instead of directing the test on variance itself, we test a specific finite-variance model against a specific infinite-variance model.

This chapter is organized as follows. The next section introduces the test statistic, motivates the intuition behind it, and obtains the statistical properties of it. Section 5.3 briefly summarizes the candidate models of the stock returns, including finite-variance family and infinite-variance family. The proposed statistic is used to distinguish these two families. Section 5.4 discusses the implementation of the test as well as Monte Carlo studies and an empirical application. Section 5.5 concludes. All the proofs are collected in Appendix A and B.

5.2 Proposed Statistic and Its Properties

DuMoucher (1973) states that if a sample has a standard deviation many times as large as the interquartile range, the Data Generating Process (DGP) could have an infinite variance. However, he does not give a statistical analysis to indicate when the DGP has an infinite variance. Despite this we find that his statement is quite intuitive and a study along this line serves our purpose to distinguish finite variance models and infinite variance models. In other words, the statistical properties of the relative magnitude of the sample standard deviation with the sample interquartile range should be investigated.

Suppose $\{X_i\}_{i=1}^{n}$ be a sequence of observations with common distribution function $F(x)$, common density function $f(x)$, mean μ and variance σ^2. Let

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

be the sample variance, where $\hat{\mu}$ is the sample mean. Denote the quantile process by $Q_n(t)$ (see Chapter 6, Csörgő and Horváth (1993)). The proposed test statistic is then defined as,

$$T_n(\theta_0) = \frac{s_n}{Q_n(1-\theta_0) - Q_n(\theta_0)}, \tag{2.1}$$

where $0 < \theta_0 < 0.5$. Hence the denominator is the θ_0-quartile range and indeed the interquartile range when $\theta_0 = 0.25$. Therefore, $T_n(0.25)$ is basically the ratio of the sample standard deviation and sample interquartile range. If the true DGP has an infinite variance, more observations must be from the tails and $s_n \to \infty$ as $n \to \infty$. On the other hand, however, both $Q_n(0.75)$ and $Q_n(0.25)$ are finite for any n. This implies the unboundness of $T_n(0.25)$. If the true DGP has a finite variance, less observations come from the tails. Hence $s_n \to \sigma$ as $n \to \infty$ and T_n converges to a finite number as $n \to \infty$. Consequently, it is reasonable to believe that a large T_n comes from a DGP with infinite variance rather than a DGP with finite variance. Thus we set up the hypothesis as the following,

$$\begin{cases} H_0: & \text{DGP is a \textit{certain} finite variance distribution,} \\ H_1: & \text{DGP is a \textit{certain} infinite variance distribution.} \end{cases} \tag{2.2}$$

If H_0 is rejected, the model in H_0 should not be used as a candidate model.

In this subsection we assume $X_1, X_2, ..., X_n$ to be iid random variables. The properties of T_n are established in this section. Their proofs are found in Appendix A.

Theorem 5.2.1 *T_n is invariant for a scale-location family.*

This is an indeed appealing property. For a scale-location family, no matter how big the scale is, the expectation of the statistic always takes the same value. In other words, if we think of T_n as a measure of risk, the risk associated with a scale-location family is a constant. Because of this property, any scale-location family can be treated as one model.

Theorem 5.2.2 *If $\sigma^2 < +\infty$, and $Q(t)$ is continuous at θ_0 and $1-\theta_0$, then*

$$T_n \to T = \frac{\sigma}{q_1 - q_0} < \infty \quad a.s., \tag{2.3}$$

where $q_1 = Q(1-\theta_0)$, $q_0 = Q(\theta_0)$ with $Q(t) = inf\{x : F(x) \geq t\}$.

This is the result of the strong law of large numbers (S.L.L.N.).

Theorem 5.2.3 *Assume that*

(i) $f(q_1) > 0$, $f(q_0) > 0$.

(ii) $f(x)$ is continuous in a neighborhood of q_1 and q_0.

If $E|X_1|^4 < \infty$, then,

$$\sqrt{n}(T_n - T) \xrightarrow{d} N(0, \Sigma^2), \tag{2.4}$$

that is,

$$T_n \overset{a}{\sim} N(T, \frac{\Sigma^2}{n}), \tag{2.5}$$

where

$$\Sigma^2 = (q_1 - q_0)^{-4} E\left\{ \frac{(q_1 - q_0)}{2\sigma}((X_1-\mu)^2 - \sigma^2) - \frac{\sigma(I\{X_1 > q_1\} - \theta_0)}{f(q_1)} - \frac{\sigma(I\{X_1 \leq q_0\} - \theta_0)}{f(q_0)} \right\}^2$$

This central limit theorem (C.L.T.) is the main result of the chapter since the hypothesis test formulated is based on it. If the model in H_0 has good descriptive power, followed by the C.L.T., it must yield a T value which is close to the empirical T_n. Although T_n is invariant for a scale-location family, it is important to note that in general both T and Σ^2 depend on f and hence H_0. Therefore in general our statistic cannot be used to test the following hypothesis,

$$\begin{cases} H_0: & \text{DGP is } \textit{any} \text{ finite variance distribution,} \\ H_1: & \text{DGP is } \textit{any} \text{ infinite variance distribution.} \end{cases} \tag{2.6}$$

Instead T_n is a non-nested test of a specific finite variance distribution against a specific infinite variance distribution.

Theorem 5.2.4 *Under assumptions of the theorem* (5.2.3), *if f is symmetric, then*

$$\Sigma^2 = \frac{K-1}{4a^2} + \frac{2\theta_0(1-2\theta_0)}{a^4b^2} + \frac{\theta_0 c_1 - (1-\theta_0 c_2)}{2a^3b}, \tag{2.7}$$

where K is the kurtosis of X, $a = \frac{1}{T}$, $b = \sigma f(q), c_1 = \int_{-\infty}^{q_1} (\frac{x-\mu}{\sigma})^2 f(x)\, dx$, and $c_2 = \int_{q_1}^{\infty} (\frac{x-\mu}{\sigma})^2 f(x)\, dx$.

Theorem 5.2.5 *T_n is consistent.*

This property guarantees a good power of the proposed test statistic when the number of observations is big enough.

5.3 Candidate Models for Daily Stock Returns

In this section we introduce the most well-known time-independent models for daily stock returns, briefly review the properties of the candidate models, and discuss the relevant estimation method and numerical algorithm if necessary. In the finite-variance family, the normal distribution, the Student t distribution, the mixture of normals, mixed diffusion-jump model, the compound log-normal and normal model, and the Weibull distribution are presented, while the Stable distribution represents infinite-variance family.

5.3.1 Normal Distribution

The first model used in the literature to describe daily stock returns is the normal distribution proposed by Bachelier (1900) and extended by Osborne (1959). Black and Scholes (1973) provide a formula to price a option assuming the normality of underlying asset. Although the assumption of normality greatly simplified the theoretical modelling, many empirical studies have shown evidence against it (see Blattberg and Gonedes (1974), Clark (1973), Kon (1984) and Niederhoffer and Osborne(1966)). For example, empirical daily stock returns exhibit fatter tails and greater kurtosis than the normal distribution. Despite this evidence, in this paper we still choose it as a competing model because we want to check the validity of this assumption by using our test statistic. We note that all moments for the normal distribution exist and the kurtosis for the normal family is three. Furthermore, since the normal distribution belongs to a scale-location family, T_n is invariant with respect to both μ and σ^2 and hence parameter estimation is not necessary.

5.3.2 Student Distribution

The Student distribution is first proposed to model the stock returns by Blattberg and Gonedes (1974). Its density is,

$$g(x) = \frac{\Gamma[(1+\nu)/2]\nu^{\nu/2}\sqrt{H}}{\Gamma(1/2)\Gamma(\nu/2)}[\nu + H(x-m)^2]^{-(\nu+1)/2}, \qquad (3.1)$$

where $\nu \geq 2$, and H, m, ν are the scale parameter, location parameter, and degrees-of-freedom parameter. Therefore, T_n is invariant to both H and m, but depends on ν. Furthermore, when $\nu > 4$ the Student distribution has a finite fourth moment and hence the C.L.T. in Section 2 can be applied. The model is estimated by the maximum likelihood method using a Quasi-Newton algorithm.[1]

5.3.3 Mixture of Normals

Kon (1984) proposes to use the mixture of normals to model stock returns, i.e., the stock return X_i come from $N(\mu_j, \sigma_j^2)$ with probability α_j and $\alpha_1 + \cdots + \alpha_k = 1$. A characteristic of this model is that it can capture the structural change. The density function is,

$$g(x) = \sum_{j=1}^{k} \alpha_j \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{ -\frac{(x-\mu_j)^2}{2\sigma_j^2} \right\}. \tag{3.2}$$

All moments exist for the mixture of normals. However, in this paper we only consider the mixture of two normals due to two reasons. Firstly, Tucker (1992) found the mixture of two normals has the greatest descriptive power among the family of the mixture of normals. Secondly, we want to avoid a model with too many parameters. The parameters of interest for the mixture of two normals are $\alpha, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ and T_n depends on all of them. The maximum likelihood method is employed using a program proposed by Venables and Ripley (1994) based on Newton-Raphson algorithm.

5.3.4 Mixed Diffusion-Jump Process

Press (1967) and Merton (1976) propose a process which mixes Brownian motion and a compound Poisson process to model the movement of stock prices,

$$dP(t) = \alpha P(t)dt + \sigma_D P(t)dB(t) + P(t)(\exp(Q) - 1)dN(t). \tag{3.3}$$

where $B(t)$ is a standard Brownian motion (BM). $N(t)$ is a homogeneous Poisson process with parameter λ. Q is a normal variate with mean μ_Q and variance σ_Q^2 and independent with $N(t)$.

[1]With little effort, we can show that $\hat{m}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Therefore, only parameters H and ν are considered in the numerical algorithm.

Using Ito's Lemma, we can solve the stochastic differential equation (3.3) for the stock return $X(t)(=\log(P(t)/P(t-1)))$,

$$X(t) = \mu_D + \sigma_D B(1) + \sum_{n=1}^{\Delta N(t)} Q_n, \tag{3.4}$$

where $\mu_D = \alpha - \frac{\sigma_D^2}{2}$. The density function for the process is,

$$g(x) = \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} \exp\Big(- \frac{(x - \mu_D - n\mu_Q)^2}{2(\sigma_D^2 + n\sigma_Q^2)}\Big) \frac{1}{\sqrt{2\pi(\sigma_D^2 + n\sigma_Q^2)}}. \tag{3.5}$$

All moments are finite for this density and T_n depends on all five parameters μ_D, σ_D^2, μ_Q, σ_Q^2 and λ. The maximum likelihood estimates are found by using a Quasi-Newtom algorithm. However, to numerically maximize the likelihood, we have to truncate the infinite sum in the equation (3.5) after some value of N. In practice, we choose $N = 11$ which provides satisfactory accuracy.

5.3.5 Compound Log-normal and Normal

This model was first proposed by Clark (1973). Instead of modelling returns as drawn from a single distribution or a mixture of two distributions, Clark (1973) assumes the returns to be conditional normal, conditional on a variance parameter which is itself stochastic. To be more specific, he assumes $X_i|Z \sim N(0, Z\sigma_1^2)$ and $\log(Z) \sim N(\alpha, \sigma_2^2)$. The density is then,

$$g(x) = \int_0^{\infty} \{\frac{1}{\sqrt{2\pi z\sigma_1^2}} \exp(-\frac{x^2}{2z\sigma_1^2})\}\{\frac{1}{z\sqrt{2\pi\sigma_2^2}} \exp(-\frac{(\log z - \alpha)^2}{2\sigma_2^2})\}\, dz. \tag{3.6}$$

It is easy to show that α and σ_1^2 can be only identified jointly. See Appendix B for the proof. Consequently, we assume $X_i|Z \sim N(0, Z\sigma_1^2)$ and $\log(Z) \sim N(0, \sigma_2^2)$. The density is then,

$$g(x) = \int_0^{\infty} \{\frac{1}{\sqrt{2\pi z\sigma_1^2}} \exp(-\frac{x^2}{2z\sigma_1^2})\}\{\frac{1}{z\sqrt{2\pi\sigma_2^2}} \exp(-\frac{(\log z)^2}{2\sigma_2^2})\}\, dz. \tag{3.7}$$

All moments exist for this density and T_n is invariant to σ_1^2. The estimates are obtained by the maximum likelihood method using a Quasi-Newton algorithm.

5.3.6 Weibull Distribution

Mittnik and Rachev (1993) were the first to propose the use of the Weibull distribution to model stock daily returns. The Weibull distribution is attractive since it is one type of min-stable distribution (Mittnik and Rachev (1993)). More specifically, suppose $m_n = \min\{X_1, \cdots, X_n\}$, where $X_1, \cdots, X_n$ are iid. If, for some constants $c_n > 0$ and $d_n \in R$, $c_n m_n + d_n \xrightarrow{d} Z$, where Z is a random variable with non-degenerate distribution function m, then m could be a Weibull distribution.

The density function for the Weibull distribution is,

$$f(x) = \begin{cases} 0 & \text{if } x < b \\ \frac{\alpha}{a}(\frac{x-b}{a})^{\alpha-1} \exp\{-(\frac{x-b}{a})^{\alpha}\} & \text{if } x \geq b \end{cases},$$

where α is the index parameter, b is the location parameter and a is the scale parameter and thus T_n is invariant to both a and b. Furthermore, the density has finite all order of moments, for example, $E(X) = a\Gamma(\frac{1}{\alpha} + 1) + b, Var(X) = a^2\left\{\Gamma(\frac{2}{\alpha} + 1) - \left(\Gamma(\frac{1}{\alpha} + 1)\right)^2\right\}$. The estimates are obtained by the maximum likelihood method using a Quasi-Newton algorithm.

5.3.7 Stable Distribution

Mandlebrot (1963) is the first person who proposes the Stable distribution to model stock returns. The Stable distribution is usually characterized by the characteristic function. The characteristic function of the general Stable distribution is given by,

$$c(t) = \exp\{iat - c|t|^{\alpha}[1 + i\beta\frac{t}{|t|}tan(\frac{\pi\alpha}{2})]\}, \tag{3.8}$$

where index(α), skewness(β), scale(c), and location(a) are parameters. Therefore, T_n is invariant to both c and a. If $1 < \alpha < 2$, which is the case for almost every financial series, the tails of the stable are fatter than those of the normal and the variance is infinite. Unfortunately, the density function has no closed form for $1 < \alpha < 2$. The maximum likelihood method is difficult to implement. Instead in this paper we obtain the estimates of the model by using the method proposed by McCulloch(1986).

5.4 Implementation, Simulation and Application

The dataset we use is daily returns for the Standard and Poor 500 (S&P500) stock market composite raw index. We consider three different periods. The first one is pre-crash sample covering the period from January 1976 to March 1985 with 2,400 observations. The second one also has 2,400 observations but covers the period after the crash from May 1988 to July 1997. The entire sample from January 1976 to July 1997 with 5,614 observations is also examined. Table 5.1 reports T_n with $\theta_0 = 0.25$ for these three samples. We note that the post-crash sample shows a larger value of T_n than the pre-crash sample. Furthermore, since the entire sample includes October, 1987 — stock-market crash days, it is not surprising that the associated T_n is largest.

As we argued before, the hypothesis we are going to test is the one given by (2.2). Since all the competing models except the Stable distribution have finite variance, we set H_0 to be one of these models and H_1 to be a Stable distribution. When T_n is parameter free under H_0, we can choose H_0 to be one distribution family, such as the normal family. Unfortunately, in most cases T_n is not completely parameter free. Consequently, H_0 has to be a certain model with parameters specified.[2] Furthermore, since H_1 is Stable distribution whose parameters α and β affect the value of T_n, α and β have to be specified in H_1. Therefore, to implement the test, we have to first fit the models in both H_0 and H_1 to the data sets. The relevant estimation method for each candidate model was presented in Section 3. After setting up the hypothesis, we can obtain the asymptotic mean and asymptotic variance for T_n based on Theorem 5.2.3 in Section 5.2. The p-value is then calculated.

[2] Actually only those parameters on which T_n depends are needed to be specified.

In Table 5.2 we report the estimates of all competing models for each data set. Since μ, σ^2 in the normal distribution, H, m in the Student distribution, σ_1^2 in the compound log-normal and normal model, a, b in the Weibull distribution, and a, c in the Stable distribution can not change the value of T_n, the estimates of these parameters are not reported. Moreover, the estimates of ν in the Student model are less than 4 for both the post-crash sample and the entire sample, the C.L.T. of T_n is not applicable in either situation. Table 5.3 reports the asymptotic distributions and Table 5.4 shows the associated p-values.

A Monte Carlo study is presented to obtain the finite sample distribution of T_n and the power of the test. 3,000 replications are generated under H_0 and H_1 respectively according to the estimates reported in Table 5.2. T_n is calculated for each replication and thus the finite sample distribution of T_n under H_0 and H_1 is obtained. Following the finite sample distributions, we calculate the critical value and power of the test. In Table 5.4 we present the finite sample distribution of T_n under H_0 for all three samples. We report the 95% critical value in Table 5.6 and the power of the test in Table 5.7. We also perform a Monte Carlo study to obtain the sizes of the test in finite samples and compare them with the nominal sizes. 3,000 replications are generated under H_0 according to the estimates reported in the first column of Table 5.2 and each replication has 2,400 observations. The nominal sizes are chosen to be $0.1\%, 0.5\%, 1\%, 5\%, 10\%, 20\%$ and 50%. The sizes are reported on Table 5.8 and plotted in Figure 5.1.

A detailed examination of Table 5.3 and Table 5.5 reveals that the asymptotic distribution of T_n is very close to the finite sample distribution of T_n. Not surprisingly, therefore, we should expect the same conclusion from both Table 5.4 and Table 5.6. Table 5.4 indicates that, for all three samples, the normal distribution can be easily rejected by the proposed test statistic, consistent with most empirical results when some other methods, such as the sample kurtosis, are used. Furthermore, for both the pre-crash sample and the post-crash sample, most finite variance distributions can not be rejected. For example, for the pre-crash sample the Student distribution, the mixture of normals, the mixed diffusion-jump process and the compound log-normal and normal model can not be rejected at 5% significant level. For the post-crash sample the mixture of normals, the mixed diffusion-jump process and the compound log-normal and normal model can not be rejected at 5% significant level. This finding is consistent with what is normally found in most recent literature; see Tucker(1992), Kon(1984), Blattberg and Gonedes (1974). However, for the entire sample all the finite-variance models are rejected at 5% and even smaller significant levels. This finding is significant and suggests that when the value of T_n gets bigger and bigger, it is harder and harder for the data to be modeled by the existing finite-variance models. The result is not surprising since a finite-variance model is prone to generate a value of T which is not large enough to match the empirical T_n. If we interpret T_n as a measure of risk, the above finding means that the existing finite-variance models have difficulties to explore the high risk that the actual stock markets have. Finally, Table 5.7 provides the evidence that our test has good power. From Table 5.6 and Figure 5.1, we note that in terms of the size of the test, it works quite well for the normal distribution, the Student t distribution, the mixture of normal distribution, the compound log-normal and normal model, and the Weibull distribution. Although the test under-rejects the mixed diffusion jump model, the differences between the sizes and the nominal sizes are not large.

5.5 Conclusion

This paper has considered a test to compare the competing models for daily stock returns with particular concern about the variance behavior. In the recent literature, the likelihood ratio test and the Komogorov-Smirnov test are used to compare the descriptive power of the competing models. Both tests suggest that distributions with finite variance outperform the distribution with infinite variance. A common feature for these two tests is that all the observations receive the same weight. Our test statistic, however, assigns different observations different weights. Obviously in our test statistic more extreme observations receive larger weight than less extreme ones. Consequently, our test statistic prefers a distribution whose tail behavior is closer to the empirical distribution to a distribution whose near-origin behavior is closer to the empirical distribution. Therefore, the empirical evidence suggests that most existing finite-variance models can not generate enough extreme observations although they may fit the empirical distribution well around the origin. The conclusion is that these finite-variance models are not always good candidates to describe stock returns. Therefore, one may expect a good candidate model could be either a distribution with infinite variance or a non-parametric model. Although most finite variance models have good descriptive power for both pre-crash sample and after-crash sample, they do not perform well for the entire sample. This provides the evidence that a structural change occurred due to the crash. The finding suggests that a good candidate among the finite variance family should be the one which can incorporate the structural change. Theoretically speaking, the mixture of normals is able to explain the structural change, however, it has been rejected for the entire sample. Therefore, a suitable structural change model which can serve as a good candidate for the entire sample has to be more complicated than the mixture of normals.

We have to point out that, observing that stock returns are not independent, many researchers have investigated the time dependent models such as ARCH-type models (see Bollerslev, Chou and Kroner (1992)), stochastic volatility models (see Ghysels, Harvey, and Renault (1996)). Although we will generalize our test statistic into the dependent case in the future research, the comparison within the iid framework is still quite interesting due to at least two reasons. Firstly, Mittnik and Rachev (1993) argue that the iid assumption is not as crucial as it appears. Secondly, although the time-dependent model has had great practical success, the time-independent model is still attractive because of the associated mathematical convenience, for example, for pricing an option.

Appendix A

Proof of Theorem 5.2.1

Since the random variable X belongs a scale-location family, we assume that

$$F(x) = G(\frac{x-\mu}{\sigma}),$$

where F is the distribution function of X, μ is the location of X, and σ is the scale of X. Define $Y = \frac{X-\mu}{\sigma}$, then $G(y)$ is the distribution function of Y. With the new notations, we have,

$$\begin{aligned} s_n^2(X) &= \frac{1}{n-1}\sum_{i=1}^{n}(Y_i\sigma - \bar{Y}\sigma)^2 \\ &= \frac{\sigma^2}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \\ &= \sigma^2 s_n^2(Y), \end{aligned}$$

and

$$\begin{aligned} F_n^{-1}(\theta_0) &= \inf\{x : F_n(x) \geq \theta_0\} \\ &= \inf\{x : G_n(\frac{x-\mu}{\sigma}) \geq \theta_0\} \\ &= \inf\{\sigma y : G_n(y) \geq \theta_0\} \\ &= \sigma G_n^{-1}(\theta_0). \end{aligned}$$

Therefore,

$$\begin{aligned} T_n(X, \theta_0) &= \frac{s_n(X)}{F_n^{-1}(1-\theta_0) - F_n^{-1}(\theta_0)} \\ &= \frac{\sigma s_n(Y)}{\sigma G_n^{-1}(1-\theta_0) - \sigma G_n^{-1}(\theta_0)} \\ &= T_n(Y, \theta_0). \quad \blacksquare \end{aligned}$$

Proof of Theorem 5.2.2

The theorem follows immediately from the strong law of large number, since $S_n \to$

σ *a.s.* and $Q_n(1-\theta_0) - Q_n(\theta_0) \to q_1 - q_0$ *a.s.* ∎.

Proof of Theorem 5.2.3

Letting $q_{1,0} = q_1 - q_0$, note that

$$\frac{s_n}{Q_n(1-\theta_0) - Q_n(\theta_0)} - \frac{\sigma}{q_1 - q_0} = \frac{q_{1,0}s_n - \sigma(Q_n(1-\theta_0) - Q_n(\theta_0)}{(Q_n(1-\theta_0) - Q_n(\theta_0))q_{1,0}} \tag{A.1}$$

and

$$\begin{aligned} s_n &= \left(\frac{1}{n-1}\{\sum_{i=1}^n (X_i-\mu)^2 - n(\bar{X}-\mu)^2\}\right)^{1/2} \\ &= \sigma\left\{1 + \frac{1}{(n-1)\sigma^2}\sum_{i=1}^n\{(X_i-\mu)^2 - \sigma^2\} - \frac{n(\bar{X}-\mu)^2 - \sigma^2}{(n-1)\sigma^2}\right\}^{1/2} \\ &= \sigma\left\{1 + \frac{1}{2(n-1)\sigma^2}\sum_{i=1}^n\{(X_i-\mu)^2 - \sigma^2\} - \frac{n(\bar{X}-\mu)^2 - \sigma^2}{2(n-1)\sigma^2}\right\} \\ &\quad + O_P\left(\left(\frac{1}{n}\sum_{i=1}^n\{(X_i-\mu)^2-\sigma^2\}\right)^2 + \left(\frac{n(\bar{X}-\mu)^2-\sigma^2}{n}\right)^2\right) \\ &= \sigma + \frac{1}{2n\sigma}\sum_{i=1}^n\{(X_i-\mu)^2 - \sigma^2\} + O_P(1/n). \end{aligned} \tag{A.2}$$

Therefore,

$$\begin{aligned} &q_{1,0}s_n - \sigma(Q_n(1-\theta_0) - Q_n(\theta_0)) \\ &= \frac{q_{1,0}}{2n\sigma}\sum_{i=1}^n\{(X_i-\mu)^2 - \sigma^2\} - \sigma\left\{Q_n(1-\theta_0) - Q_n(\theta_0) - q_{1,0}\right\} + O_P(1/n) \end{aligned} \tag{A.3}$$

According to the Bahadur representation, we have

$$Q_n(1-\theta_0) - q_1 = -\frac{1}{nf(q_1)}\sum_{i=1}^n\{I\{X_i \le q_1\} - (1-\theta_0)\} + o_P(n^{-1/2})$$

and

$$Q_n(\theta_0) - q_0 = -\frac{1}{nf(q_0)}\sum_{i=1}^n\{I\{X_i \le q_0\} - \theta_0)\} + o_P(n^{-1/2})$$

Putting the above statements together yields

$$\begin{aligned}
&\sqrt{n}\Big(q_{1,0}s_n - \sigma(Q_n(1-\theta_0) - Q_n(\theta_0))\Big) \qquad (A.4)\\
= \;& n^{-1/2}\Bigg\{\frac{q_{1,0}}{2\sigma}\sum_{i=1}^{n}\{(X_i-\mu)^2-\sigma^2\}\\
&+\frac{\sigma}{f(q_1)}\sum_{i=1}^{n}\{I\{X_i\le q_1\}-(1-\theta_0)\} - \frac{\sigma}{f(q_0)}\sum_{i=1}^{n}\{I\{X_i\le q_0\}-\theta_0)\}\Bigg\} + o_P(1)\\
= \;& n^{-1/2}\sum_{i=1}^{n}\left\{\frac{q_{1,0}}{2\sigma}((X_i-\mu)^2-\sigma^2) - \frac{\sigma(I\{X_i>q_1\}-\theta_0)}{f(q_1)} - \frac{\sigma(I\{X_i\le q_0\}-\theta_0)}{f(q_0)}\right\}\\
+ \;& o_P(1)
\end{aligned}$$

This proves (2.4). (2.5) simply follows (2.4). ■

Proof of Theorem 5.2.4

Expanding the expression for Σ^2, we have,

$$\begin{aligned}
\Sigma^2 = \;& \left(\frac{\sigma}{2(q_1-q_0)}\right)^2 E\left(\left(\frac{X_1-\mu}{\sigma}\right)^4 - 1\right) + \frac{\sigma^2\theta_0(1-\theta_0)}{(q_1-q_0)^4}\left(\frac{1}{f^2(q_0)}+\frac{1}{f^2(q_1)}\right)\\
&-\frac{\sigma^2}{(q_1-q_0)^3 f(q_1)}E\left\{\left(\frac{X_1-\mu}{\sigma}\right)^2(I\{X_1>q_1\}-\theta_0)\right\}\\
&-\frac{\sigma^2}{(q_1-q_0)^3 f(q_0)}E\left\{\left(\frac{X_1-\mu}{\sigma}\right)^2(I\{X_1\le q_0\}-\theta_0)\right\}\\
&-\frac{2\sigma^2\theta_0^2}{(q_1-q_0)^4 f(q_0)f(q_1)}.
\end{aligned}$$

Since f is symmetric about μ, we have $f(q_0) = f(q_1)$. A simplification of above expression gives us (2.7). ■

Proof of Theorem 5.2.5

Under H_0, the critical value (CV)is always a finite number. Then,

$$\lim_{n\to+\infty} Pr(\text{Reject\_}H_0|H_1\text{\_is true}) = \lim_{n\to+\infty} Pr(T_n > \text{CV}|\sigma = +\infty) = 1. \quad ■$$

Appendix B

In this appendix we prove that α and σ_1^2 can be only identified jointly in (3.6). According to the assumption, we have

$$Z \sim \exp(N(\alpha, \sigma_2^2)) = \exp(\alpha + N(0, \sigma_2^2)) = e^{\alpha} \exp(N(0, \sigma_2^2)).$$

Therefore,

$$X_i | Z \sim N(0, Z\sigma_1^2) = N(0, \sigma_1^2 e^{\alpha} \exp^{N(0, \sigma_2^2)}).$$

Obviously, α and σ_1^2 can not be identified if $\sigma_1^2 \times e^{\alpha}$ equals to a constant. ∎

	Sample 1: 76-85	Sample 2: 88-97	Sample 3: 76-97
$T_n(\theta_0 = 0.25)$	0.8406	0.9694	1.0174

Table 5.1: T_n in the Empirical Samples

	Sample 1: 76-85	Sample 2: 88-97	Sample 3: 76-97
Student	$\nu = 6.387879$	$\nu = 3.994152$	$\nu = 3.938247$
MN	$\mu_1 = -2.58732 \times 10^{-4}$ $\mu_2 = 1.39901 \times 10^{-3}$ $\sigma_1 = 6.16532 \times 10^{-3}$ $\sigma_2 = 1.16697 \times 10^{-2}$ $\alpha = 0.673595$	$\mu_1 = 6.2079 \times 10^{-4}$ $\mu_2 = 3.758 \times 10^{-4}$ $\sigma_1 = 4.21049 \times 10^{-3}$ $\sigma_2 = 1.03509 \times 10^{-2}$ $\alpha = 0.567403$	$\mu_1 = 5.0964 \times 10^{-4}$ $\mu_2 = -1.39886 \times 10^{-3}$ $\sigma_1 = 7.20942 \times 10^{-3}$ $\sigma_2 = 2.64926 \times 10^{-2}$ $\alpha = 0.952783$
LN	$\sigma_2^2 = 0.4576217$	$\sigma_2^2 = 0.8810937$	$\sigma_2^2 = 0.9063072$
MDJ	$\mu = -3.732 \times 10^{-4}$ $\mu_Q = 7.06 \times 10^{-4}$ $\sigma^2 = 2.59 \times 10^{-5}$ $\sigma_Q^2 = 4.73 \times 10^{-5}$ $\lambda = 0.92847$	$\mu = 4.7576 \times 10^{-4}$ $\mu_Q = 3.05 \times 10^{-5}$ $\sigma^2 = 7.42 \times 10^{-6}$ $\sigma_Q^2 = 3.72 \times 10^{-5}$ $\lambda = 1.2796$	$\mu = 5.168 \times 10^{-4}$ $\mu_Q = -2.047 \times 10^{-4}$ $\sigma^2 = 2.527 \times 10^{-5}$ $\sigma_Q^2 = 9.25 \times 10^{-5}$ $\lambda = 0.515686$
Weibull	$\alpha = 5.06930572$	$\alpha = 9.00619902$	$\alpha = 20.3287225$
Stable	$\alpha = 1.697953$ $\beta = 0.2132737$	$\alpha = 1.501834$ $\beta = -0.01664654$	$\alpha = 1.547262$ $\beta = 0.05912446$

Table 5.2: Estimates of the Competing Models

	Sample 1: 76-85	Sample 2: 88-97	Sample 3: 76-97
Normal	$\mathcal{N}(0.7413, 1.97 \times 10^{-4})$	$\mathcal{N}(0.7413, 1.97 \times 10^{-4})$	$\mathcal{N}(0.7413, 8.42 \times 10^{-5})$
Student	$\mathcal{N}(0.8440, 4.50 \times 10^{-4})$	Not Applicable	Not Applicable
MN	$\mathcal{N}(0.8402, 3.25 \times 10^{-4})$	$\mathcal{N}(0.9802, 5.57 \times 10^{-4})$	$\mathcal{N}(0.8953, 4.18 \times 10^{-4})$
LN	$\mathcal{N}(0.8543, 4.11 \times 10^{-4})$	$\mathcal{N}(0.9688, 8.01 \times 10^{-4})$	$\mathcal{N}(0.9754, 3.34 \times 10^{-4})$
MDJ	$\mathcal{N}(0.8511, 2.75 \times 10^{-4})$	$\mathcal{N}(0.9648, 4.32 \times 10^{-4})$	$\mathcal{N}(0.9425, 1.57 \times 10^{-4})$
Weibull	$\mathcal{N}(0.7307, 1.86 \times 10^{-4})$	$\mathcal{N}(0.7569, 2.35 \times 10^{-4})$	$\mathcal{N}(0.7854, 1.31 \times 10^{-4})$

Table 5.3: Asymptotic Distribution of T_n under H_0

	Sample 1: 76-85	Sample 2: 88-97	Sample 3: 76-97
Normal	0	0	0
Student	0.5628	Not Applicable	Not Applicable
MN	0.4907	0.6757	0
LN	0.7511508	0.4915044	0.01162844
MDJ	0.7368	0.4120	0
Weibull	0	0	0

Table 5.4: p-values of the Test

	Sample 1: 76-85	Sample 2: 88-97	Sample 3: 76-97
Normal	$(0.7420, 1.96 \times 10^{-4})$	$(0.7420, 1.96 \times 10^{-4})$	$(0.7418, 8.98 \times 10^{-5})$
Student	$(0.8442, 4.44 \times 10^{-4})$	Not Applicable	Not Applicable
MN	$(0.8316, 3.46 \times 10^{-4})$	$(0.9760, 5.76 \times 10^{-4})$	$(0.8990, 4.30 \times 10^{-4})$
LN	$(0.8551, 4.10 \times 10^{-4})$	$(0.9691, 8.07 \times 10^{-4})$	$(0.9759, 3.5 \times 10^{-4})$
MDJ	$(0.8520, 3.92 \times 10^{-4})$	$(0.9663, 6.18 \times 10^{-4})$	$(0.9434, 2.38 \times 10^{-4})$
Weibull	$(0.7313, 1.92 \times 10^{-4})$	$(0.7575, 2.40 \times 10^{-4})$	$(0.7855, 1.26 \times 10^{-4})$

Table 5.5: Finite Sample Distribution of T_n under H_0

	Sample 1: 76-85	Sample 2: 88-97	Sample 3: 76-97
Normal	0.7653077	0.7653077	0.7576618
Student	0.8814489	Not Applicable	Not Applicable
MN	0.8622002	1.015521	0.9331136
LN	0.8904581	1.01819	1.006542
MDJ	0.8836779	1.007142	0.9692391
Weibull	0.754372	0.7837439	0.8034553

Table 5.6: Critical Value of the Finite Sample Distribution

	Sample 1: 76-85	Sample 2: 88-97	Sample 3: 76-97
Normal	1	1	1
Student	1	Not Applicable	Not Applicable
MN	1	1	1
LN	1	1	1
MDJ	1	1	1
Weibull	1	1	1

Table 5.7: Power of the Test

Nominal size	0.001	0.005	0.01	0.05	0.1	0.2	0.5
Normal	0.003	0.009	0.012	0.056	0.114	0.223	0.507
Student	0.002	0.0067	0.0110	0.065	0.107	0.192	0.484
MN	0.001	0.0077	0.0117	0.0517	0.096	0.196	0.490
LN	0.0	0.0087	0.0197	0.0637	0.112	0.193	0.501
MDJ	0.007	0.0217	0.0250	0.0937	0.1450	0.257	0.513
Weibull	0.001	0.006	0.0137	0.062	0.114	0.218	0.519

Table 5.8: Size of the Test

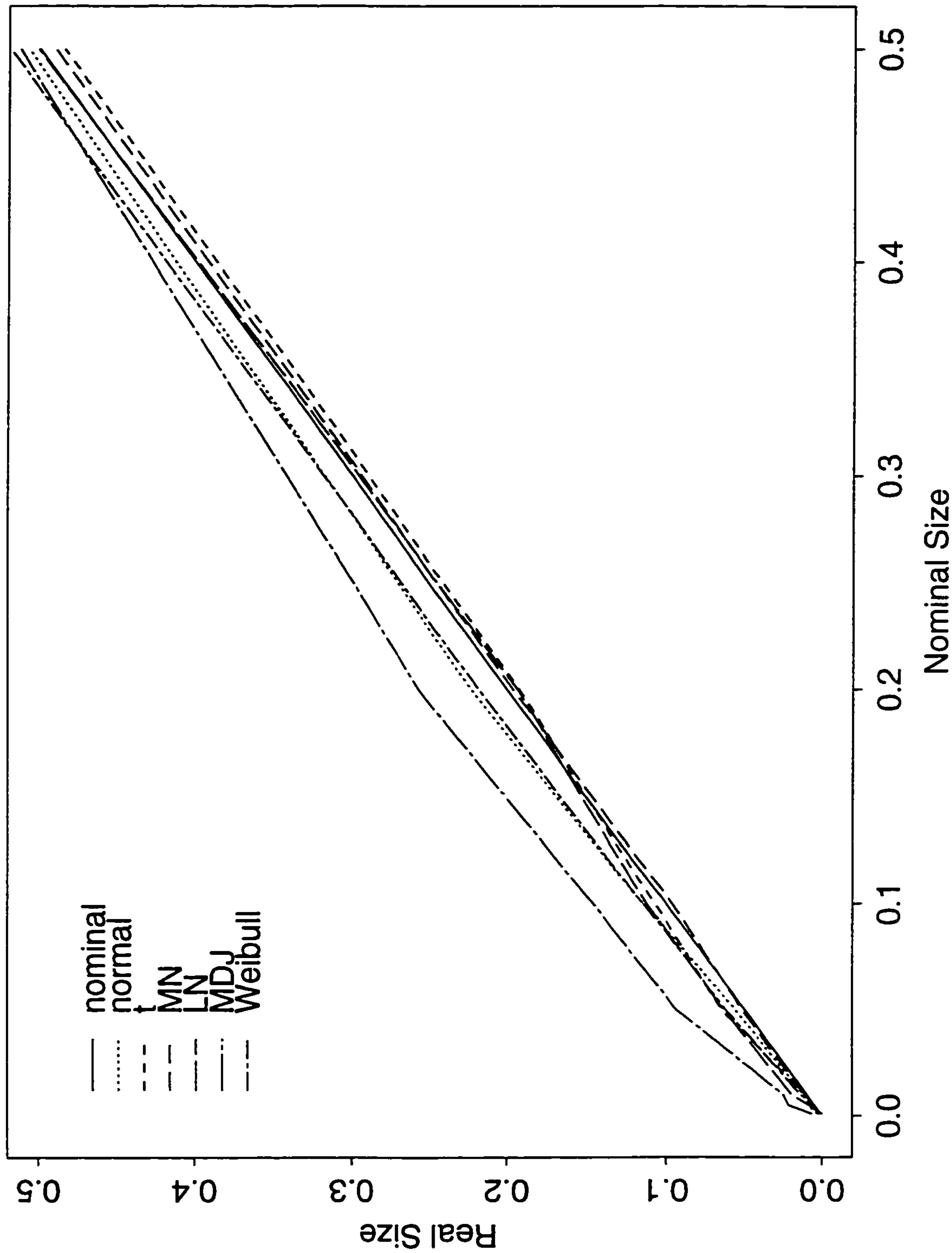

Figure 5.1: Size of the Test

REFERENCES

[1] Akgiray, V. and G. Booth. (1987). Compound distribution of models of stock returns: An empirical comparison. *Journal of Financial Research.* **10** 269–280.

[2] Andersen, T. and B. Sorensen. (1994). GMM estiomation of a stochastic volatility models: A Monte Carlo study. *Journal of Business and Economic Statistics.* **14** 329–352.

[3] Bachelier, L. (1900). *Theorie de la Speculation.* Gauthier-Villars, Paris.

[4] Bersten, J., T.O. Espelid, and Genz, A. (1991). Algorithm 698: DCUHRE-an adaptive multidimensional integration routine for a vector of integrals. *ACM Trans. Math. Software.* **17** 452–456.

[5] Billingsley, P. (1968). *Convergence of Probability Measure.* Wiley, New York.

[6] Black, F. and M. Scholes. (1973). The pricing of options and corporate liabilities. *Journal of Political Economics.* **81** 637–653.

[7] Blattberg, R. and N. Gonedes. (1974). A comparison of stable and student distribution. *Journal of Business.* **47** 244–280.

[8] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics.* **31** 307–327.

[9] Bollerslev, T., R.Y. Chou, and K.F. Kroner. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics.* **52** 5–59.

[10] Clark, P.K. (1973). A subordinated stochastic process model with finite variance for speculative prices (com: P157-159). *Econometrica.* **41** 135–155.

[11] Csörgő, M. and L. Horváth. (1993). *Weighted approximations in probability and statistics.* Wiley, New York.

[12] Csörgő, S. (1981). Limit behaviour of the empirical characteristic function. *The Annals of Probability.* **9** 130–144.

[13] Danielsson, J. (1994). Comment on "Bayesian analysis of stochastic volatility models". *Journal of Business and Economic Statistics.* **12** 393–395.

[14] Danielsson, J. (1994). Stochastic volatility in asset prices. Estimation with simulated maximum likelihood. *Journal of Econometrics.* **64** 375–400.

[15] Duffie, D. and K.J. Singleton. (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica.* **61** 929–952.

[16] DuMouchel, W.H. (1973). On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *The Annals of Statistics.* **1** 948–957.

[17] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation (stma V24 501). *Econometrica.* **50** 987–1007.

[18] Fama, E. (1965). The behavior of stock prices. *Journal of Business.* **47** 244–280.

[19] Fan, Y. (1996). Goodness-of-fit tests for a multivariate distribution by the empirical characteristic function. Journal of Multivariate Analysis, Forthcoming.

[20] Feuerverger, A. (1990). An efficiency result for the empirical characteristic function in stationary time-series models. *The Canadian Journal of Statistics.* **18** 155–161.

[21] Feuerverger, A. and P. McDunnough. (1981). On some Fourier methods for inference. *Journal of the American Statistical Association.* **76** 379–387.

[22] Feuerverger, A. and P. McDunnough. (1981). On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society, Series B, Methodological.* **43** 20–27.

[23] Feuerverger, A. and R.A. Mureika. (1977). The empirical characteristic function and its applications. *The Annals of Statistics.* **5** 88–97.

[24] Gel'Fan, A. (1964). *Generalized Functions*, volume 1. Academic Press.

[25] Ghysels, E., A.C. Harvey, and E. Renault. (1996). *Stochastic Volatility.* In *Handbook of Stististics-14.*

[26] Gourieroux, C., A. Monfort, and E. Renault. (1993). Indirect inference. *Journal Applied Econometrics.* **85** 85–118.

[27] Granger, C.W.J. and D. Orr. (1972). "Infinite variance" and research strategy time in series analysis. *Journal of the American Statistical Association.* **67** 275–285.

[28] Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica.* **50** 1029–1054.

[29] Hansen, L.P. and J.A. Scheinkman. (1995). Back to future: Generating moment implications for continuous-time markov processes. *Econometrica.* **63** 767–804.

[30] Harvey, A.C. (1993). Long memory in stochastic volatility. London School of Economics.

[31] Harvey, A.C., E. Ruiz, and N. Shephard. (1994). Multivariate stochastic variance models. *Review of Economic Studies.* **61** 247–264.

[32] Heathcote, C.R. (1977). The integrated squared error estimation of parameters. *Biometrika.* **64** 255–264.

[33] Huber, P. (1981). *Robust Statistics.* Wiley, New York.

[34] Ibragimov, I. and Y. Linnik. (1971). *Independent and Stationary Sequences of Random Variables.* Groningen, Wolters-Noordhoff.

[35] Jacquier, E, N.G. Polson, and P.E. Rossi. (1994). Bayesian analysis of stochastic volatility models (disc: P389-417). *Journal of Business and Economic Statistics.* **12** 371–389.

[36] Jorion, P. (1988). On jump processes in the foreign exchange and stock markets. *Review of Financial Studies.* **1** 427–445.

[37] Knight, J.L. and S.E. Satchell. (1993). Garch processes - some exact results, some difficulties and a suggested remedy. The University of Western Ontario.

[38] Knight, J.L. and S.E. Satchell. (1996). Estimation of stationary stochastic processes via the empirical characteristic function. The University of Western Ontario.

[39] Knight, J.L. and S.E. Satchell. (1997). The cumulant generating function estimation method. *Econometric Theory.* **13** 170–184.

[40] Knight, J.L., S.E. Satchell, and Y. Yoon. (1993). A model of a continuous time garch process applied to uk stocks. The University of Western Ontario.

[41] Kon, S. (1984). Model of stock returns–a comparison. *Journal of Finance.* **39** 147–165.

[42] Lau, A. and H. Lau andJ. Wingender. (1990). The distribution of stock returns: New evidence against the stable model. *Journal of Business and Economic Statistics.* **8** 217–223.

[43] Lukacs, E. (1970). *Characteristic Functions.* Griffin, New York.

[44] Mandelbrot, B. (1963). The vatiations of certain speculative prices. *Journal of Business.* **36** 394–419.

[45] Martin, R.D. and V.J. Yohai. (1986). Influence functionals for time series (c/r: P819-855). *The Annals of Statistics.* **14** 781–818.

[46] McCulloch, J. (1978). Continuous time processes with stable increments. *Journal of Business.* **51** 601–619.

[47] McCulloch, J. (1986). Simple consistent estimators of stable distribution parameters. *Comm. Statist. B-Simulation Comut.* **15** 1109–1136.

[48] Melino, A. and S.M. Turnbull. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics.* **45** 239–265.

[49] Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics.* **3** 125–144.

[50] Mittnik, S. and S.T. Rachev. (1993). Modeling asset returns with alternative stable distributions (disc: P331-389). *Econometric Reviews.* **12** 261–330.

[51] Nelson, D.B. (1988). *Time Series Behavior of Stock Market Volatility and Returns.* PhD thesis, Massachusetts Institution of Technology.

[52] Niederhoffer, V. and M.F.M. Osborne. (1966). Market making and reversal on the stock exchange. *Journal of the American Statistical Association.* **61** 897–916.

[53] Osborne, M. (1959). Brownian motions in the stock market. *Operation Research.* **7** 145–173.

[54] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics.* **33** 1065–1076.

[55] Paulson, A.S., E.W. Holcomb, and R.A. Leitch. (1975). The estimation of the parameters of the stable laws. *Biometrika.* **62** 163–170.

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

[57] Press, S.J. (1967). A compound events models of security prices. *Journal of Business.* **40** 317–335.

[58] Press, S.J. (1972). Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association.* **67** 842–846.

[59] Quandt, R.E. and J.B. Ramsey. (1978). Estimating mixtures of normal distributions and switching regressions (c/r: P738-752; V74 p52-56). *Journal of the American Statistical Association.* **73** 730–738.

[60] Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics.* **63** 289–306.

[61] Schmidt, P. (1982). An improved version of the Quandt-Ramsey MGF estimator for mixtures of normal distributions and switching regressions. *Econometrica.* **50** 501–524.

[62] Tauchen, G. and M. Pitts. (1983). The price variability-volume relationship on speculative markets. *Econometrica.* **51** 485–505.

[63] Taylor, S. (1986). *Modelling Financial Time Series.* John Wiley, New York.

[64] Titterington, D.M., A.F.M. Smith, and U.E. Makov. (1985). *Statistical Analysis of Finite Mixture Distributions.* Chichester, Toronto.

[65] Tran, K. (1994). *Mixture, Moment and Information-Three Essays in Econometrics.* PhD thesis, The University of Western Ontario.

[66] Tucker, A. (1992). A reexamination of finite- and infinite-variance distributions as models of daily stock returns. *Journal of Business and Economic Statistics.* **10** 73–81.

[67] Venables, W. and R. Ripley. (1994). *Modern Applied Statistics with S-Plus.* Springer-Verlag.

[68] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica.* **50** 1–26.

[69] Whittle, P. (1983). *Prediction and Regulation.* University of Minnesota Press.

[70] Yao, Q. and B. Morgan. (1996). Empirical transformation estimation. University of Kent.

[71] Zinde-Walsh, V. (1988). Some exact formulae for autoregressive moving average processes. *Econometric Theory.* **4** 384–402.

VITA

NAME:	Jun Yu
PLACE OF BIRTH:	Ezhou, P.R. CHINA
YEAR OF BIRTH:	1969
POST-SECONDARY EDUCATION AND DEGREES:	University of Western Ontario London, Ontario, CANADA Ph.D. 1998
	University of Western Ontario London, Ontario, CANADA M.A. 1994
	Wuhan University Wuhan, Hubei, P.R.CHINA B.Sc. and B.A. 1990
HONOURS AND AWARDS:	Natural Sciences and Engineering Research Council of CANADA Doctoral Fellowship 1997-1999
	Social Sciences and Humanities Research Council of CANADA Doctoral Scholarship 1997-1998

Ontario Graduate Scholarship
1997-1998

University of Western Ontario
Sir Arthur Currie Memorial Scholarship
1995-1996

University of Western Ontario
Special University Scholarship
1993-1998

University of Western Ontario
Graduate Tuition Scholarship
1994-1998

RELATED WORK EXPERIENCE:

Instructor, Summer 1994, 1995, 1996
Department of Economics
University of Western Ontario
London, Ontario, CANADA

Teaching Assistant, 1993-1998
Department of Economics
University of Western Ontario
London, Ontario, CANADA

IMAGE EVALUATION TEST TARGET (QA-3)

150mm

6"

APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989