1994

Mixtures, Moments And Information: Three Essays In Econometrics

Kien C. Tran

MIXTURES, MOMENTS AND INFORMATION - THREE ESSAYS IN ECONOMETRICS

by

Kien C. Tran

Department of 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, 1994

ISBN 0-315-93191-4

Abstract

This thesis is a collection of three independent essays in econometrics.

The first essay uses the empirical characteristic function (ECF) procedure to estimate the parameters of mixtures of normal distributions and switching regression models. The ECF procedure was formally proposed by Feuerverger and Mureika (1977), Heathcote (1977). Since the characteristic function is uniformly bounded, the procedure gives estimates that are numerically stable. Furthermore, it is also shown that the finite sample properties of the ECF estimator are very good, even in the case where the popular maximum likelihood fails to exist.

The second essays applies White's (1982) information matrix (IM) test to a stationary and invertible autoregressive moving average (ARMA) process. Our result indicates that, for ARMA specification, the derived covariance matrix of the indicator vector is not block diagonal implying the algebraic structure of the IM test is more complicated than other cases previously analyzed in the literature (see for example Hall (1987), Bera and Lee (1993)). Our derived IM test turns out to be a joint specification test of parameter heterogeneity (i.e. test for random coefficient or conditional heteroskedasticity) of the specified model and normality.

The final essay compares, using Monte Carlo simulation, the generalized method of moments (GMM) and quasi-maximum likelihood (QML) estimators of the parameter of a simple linear regression model with autoregressive conditional heteroskedastic (ARCH) disturbances. The results reveal that GMM estimates are often biased (apparently due to poor instruments), statistically insignificant, and dynamically unstable (especially the parameters of the ARCH process). On the other hand, QML estimates are generally unbiased, statistically significant and dynamically stable. Asymptotic standard errors for QML are 2 to 6 times smaller than for GMM, depending on the choice of the instruments.

To My Dearest Grandmother

Acknowledgements

I would, first, like to thank the members of my thesis committee, Professor John Knight, Professor R.A.L. Carter, Professor Kim Balls and Professor Bruce Hansen for their helpful comments, suggestions and constructive criticism.

My gratitude towards my supervisor, Professor John Knight is infinite. Professor Knight provided unquantifiable amounts of assistance to me during the past two years for which, I doubt, I will ever be able to repay him. For all his help and advice, I am sincerely grateful.

I also would like to thank Professor Anil Bera for initiating the idea in Chapter Two. Financial assistance from the Department of Economics is greatly appreciated.

Finally, but certainly not least, I would like to thank my parents for their patience and emotional support over the years and to my good friends who have provided much support and encouragement: Allan Florizone, Steve and Pam Kritzer, Dan and Lisa Hupka, Tom Porter, Toni Gravelle, Phillip Gunby and Scott Hendry. To Eva Prytula, a very special thank for keeping me sane during tough times. To Paula and Melissa and Jane, thanks for teaching me great cards games and to Sue Brown for her administrative assistance.

None of the aforementioned individuals should be held responsible for any errors in this thesis; this I reserve for myself.

TABLE OF CONTENTS

CHAPTER 2

CHAPTER 3

LIST OF TABLES

CHAPTER 1:

CHAPTER 3:

LIST OF FIGURES

CHAPTER 1:

CHAPTER 3:

OVERVIEW

This thesis is a collection of three independent essays in econometrics. The first chapter deals with the estimation methodology of the mixtures of normal distributions and switching regression models. The second chapter derives the information matrix (IM) test for a stationary and invertible autoregressive moving average (ARMA) process. The final chapter compares various parametric estimation procedures for a linear regression model with autoregressive conditional heteroskedastic (ARCH) disturbances.

In Chapter 1, the empirical characteristic function (ECF) procedure is used to estimate the parameters of mixtures of normal distributions and switching regression models. The ECF procedure was formally proposed by Feuerverger and Mureika (1977), Heathcote (1977), Bryan and Paulson (1983). Since the characteristic function is uniformly bounded, the procedure gives estimates that are numerically stable. Furthermore, it is also shown that the finite sample properties of the ECF estimator are very good, even in the case where the popular maximum likelihood fails to exist.

In Chapter 2, we apply White's (1982) information matrix test to a stationary and invertible ARMA process. Our results indicate that, for the ARMA specification, the derived covariance matrix of the indicator vector is not block diagonal implying the algebraic structure of the IM test is more complicated than in Hall (1987) and Bera and Lee (1993). Our derived IM test turns out to be a joint specification test of parameter heterogeneity (i.e. test for random coefficient or conditional heteroskedasticity) of the specified model and normality.

Chapter 3 compares, using Monte Carlo simulation, the generalized method of moments (GMM) and quasi-maximum likelihood (QML) estimators of the parameters of a simple linear regression model with ARCH disturbances. The results reveal that GMM estimates are often biased (apparently due to poor instruments), statistically insignificant, and dynamically unstable (especially the parameters of the ARCH process). On the other hand, QML estimates are generally unbiased, statistically significant and dynamically stable. Asymptotic standard errors for QML are 2 to 6 times smaller than for GMM, depending on the choice of the instruments.

CHAPTER 1

EMPIRICAL CHARACTERISTIC FUNCTION PROCEDURE IN MIXTURES OF NORMAL DISTRIBUTIONS AND SWITCHING REGRESSION MODEL

I. INTRODUCTION

One of the best known problems that occurs in applied research is the mixtures of two or more normal distributions, see for example Bhattacharya (1966), Cohen (1967), Day (1969), Odell and Basu (1976), Hosmer (1973, 1978), Tan and Chang (1972), Quandt (1975), Quandt and Ramsey (1978), Schmidt (1982), and Titterington et al. (1985) for reference therein. Thus, in the general case, we have a random variable y such that

$$(1) \qquad y \sim N(\mu_i, \sigma_i^2) \quad \text{with Probability } \lambda_i, \qquad i = 1,2,\ldots,k$$

where $\sum_{i=1}^{k} \lambda_i = 1$, $(\lambda_i,\ \mu_i,\ \sigma_i^2)$ are $(3k - 1)$ unknown parameters.

Alternatively, one can extend the problem in (1) to the regression case where we allow for the means to depend on some explanatory variables in which case it is referred as the "switching regressions" problem. In the economic context, the normal mixture model such as (1) (or switching regressions) can be viewed as a "contaminated data" (or structural change) problem. Some of the contaminated data examples include Granger and Orr (1972) (see also Hamermesh (1970), Quandt (1975) for the economic application of the switching regressions case): (i) In a firm's monthly production series, the contamination may be due to a sudden strike, a sales promotion, or an annual vacation shutdown; (ii) In daily interest rate changes, the result of a governmental policy action, such as large open-market purchase, can be viewed as contamination.

For simplicity, we shall restrict our attention to the case where the number of mixtures is two. This problem is an old one and the history of attempts to solve the problem is a long one beginning with Newcomb (1886) and most recently Lindsay and Basak (1993), see Titterington et al. (1985) for extensive references up to 1983. Also, this problem is irregular in the sense that without any further restrictions the likelihood function is unbounded. Computational difficulties, therefore, may be encountered in practice. There are a couple of well known facts about the problem worth noting. First, unlike other mixtures, the parameters are identified (Teicher (1961, 1963)) and secondly, the parameters can be estimated consistently by method of moments (Cohen (1967), Day (1969)) and the method of moment generating function (Quandt and Ramsey (1978), Schmidt (1982)).

In this paper, we introduce an alternative method of estimating the parameters of normal mixtures and switching regressions. The procedure is similar to that of the Quandt and Ramsey (1978) method of moment generating function (hereafter MGF) except we replace the sample moment generating function by the characteristic function. This method was formally proposed by Feuerverger and Mureika (1977), Heathcote (1977). There are several advantages of using the characteristic function. One is the characteristic function is uniformly bounded and thus, it should lead to greater numerical stability. Furthermore, the characteristic function is applicable to cases where the moment generating function fails to exist, as with certain fat tailed distributions.

Section 2 reviews various methods of estimating the parameters and their drawbacks for the normal mixtures case. These include the method of moments, the method of maximum likelihood and the MGF method.

Section 3 proceeds to outline the alternative estimation procedures using the characteristic function approach and offers detailed discussion on its asymptotic properties. Section 4 provides some Monte Carlo evidence of the finite sample properties of the estimator discussed in section 3. Section 5 generalizes the framework in section 3 to the case of switching regression models. Finally, section 6 concludes the paper and offers some suggestions for future research.

II. VARIOUS METHODS OF ESTIMATING PARAMETERS OF NORMAL MIXTURES MODEL

For simplicity, we shall examine the case of a mixture of two univariate normal distributions. Thus consider we have a random sample $y_1, y_2, \ldots, y_n$ from the distribution specified in (1) for k = 2, with probability density function given by:

(2.1)

$$f(y;\lambda,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2) = \left[\lambda/(2\pi)^{1/2}\sigma_1\right] exp\left[-(1/2\sigma_1^2)(y-\mu_1)^2\right] + \left[(1-\lambda)/(2\pi)^{1/2}\sigma_2\right] exp\left[-(1/2\sigma_2^2)(y-\mu_2)^2\right]$$

and suppose we wish to estimate the five parameters λ, μ_1, μ_2, σ_1^2, σ_2^2.

2.1. Method of Moments (MOM)

The method of moments technique can be found in Cohen (1967), Day (1969), and others (see Titterington et al. (1985) for a list of references). Estimation of the five parameters in (2.1) using MOM is obtained by equating the first five sample moments to their population counterparts and, after much elimination, we are left with a problem of finding the negative root for the nonic equation[1].

$$\sum_{j=1}^{9} \gamma_j w^j = 0 \qquad (2.1.1)$$

where γ_j are known coefficients. Let m_s and k_s ($s = 1,2,...,5$) be the *sth* sample central moment and sample cumulant, respectively, and suppose that the negative root $\hat{w}$ of (2.1.1) exists, we then calculate the following quantities

$$u = (-6m_3\hat{w}^3+2k_5\hat{w}^2+9m_3k_4+6m_3^3)/(2\hat{w}^3+3k_4\hat{w}+4m_3^2), \qquad (2.1.2)$$

$$v = u-m_3, \qquad (2.1.3)$$

$$\rho = v/\hat{w}, \qquad (2.1.4)$$

$$z = (2v-m_3)/3\hat{w}, \qquad (2.1.5)$$

---

[1] Refer to Cohen (1967) for details.

and solve the quadratic equation

(2.1.6) $$d^2 - \rho d + \hat{w} = 0,$$

giving roots d_1 and d_2, with $d_1 > 0 > d_2$, say. We may express our estimates in the form

$$\begin{aligned} \hat{\mu}_j &= d_j + \bar{x}, \\ \hat{\sigma}_j &= d_j z + m_2 - d_j^2, \qquad\qquad j = 1,2, \\ \hat{\lambda} &= d_2/(d_1 - d_2). \end{aligned}$$

Note that the above estimates can be greatly simplified if knowledge about various restrictions of equal means or equal variances is given a priori. However, in general, there are a number of potential problems which occur frequently with the moment estimators.

(i) The parameter estimates may not be unique and may be non-feasible (i.e. lie outside the parameter space).

(ii) There may be more than one negative root or none at all in solving the nonic equation (2.1.1).

(iii) The quadratic form in (2.1.6) may not have real roots.

(iv) Although the parameter estimates are consistent, they are inefficient. Indeed, simulation work by Tan and Chang (1972) establish the lack of efficiency against alternative maximum likelihood estimators, especially when the two component densities are close together. Furthermore, computation of the exact covariance matrix of the estimated parameters is not usually possible.

2.2. Method of Maximum Likelihood (ML)

Alternatively, Quandt (1972), Hosmer (1973, 1974) proposed to estimate the parameters by using the popular maximum likelihood approach. It is popular because of its attractive asymptotic theory, i.e. strongly consistent, asymptoticaly normal and efficient, and because the estimates are often easy to compute. However, as we will see, this is not always the case. Given a sample of n independent observations from (2.1), the log-likelihood function is given by

$$(2.2.1) \qquad L^{*} = \ln L$$

$$= \sum_{j=1}^{n} \ln \Big[(\lambda/(2\pi)^{1/2}\sigma_1) \exp\{-(1/2\sigma_1^2)(y_j-\mu_1)^2\} + (1-\lambda)/(2\pi)^{1/2}\sigma_2^2) \exp\{-(1/2\sigma_2^2)(y_j-\mu_2)^2\} \Big]$$

Maximization of L^{*} with respect to the five parameters, λ, μ_1, μ_2, σ_1^2, σ_2^2 yields the maximum likelihood estimates. It is obvious that from (2.2.1) the set of likelihood equations cannot be solved explicitly and generally have multiple

(1978) demonstrated that the sequence of roots corresponding to the largest of the local maxima for each n is consistent, asymptotically normal and efficient with the last two results conditional on $\lambda \neq 0$ and $(\mu_i,\sigma_i^2) \neq (\mu_j,\sigma_j^2)$ for $(i \neq j = 1,2)$. In addition, the likelihood surface is littered with singularities. For example, Kiefer and Wolfowitz (1956) noted that if we set $\mu_1 = \mu_2 = \mu$, $\sigma_1^2 = 1$, $\sigma_2^2 = \sigma^2$, and $\lambda = 0.5$, then the supremum of L^* is almost always infinite and no ML estimator exists. Therefore, ML approach may encounter some difficulties in practice.

2.3. Method of Moment Generating Fuction (MGF)

Due to the difficulties of both MOM and ML approaches, alternative estimation procedures have been considered in the literature. One particular procedure of interest is the MGF proposed by Quandt and Ramsey (1978). The MGF method seems to work reasonably well against both MOM and ML (Hosmer (1978)) for the case of mixtures of two normal distributions. However, since the MGF is unbounded, the estimates may be numerically unstable (Quandt and Ramsey (1978)).

The moment generating function associated with (2.1) is given by:

$$(2.3.1) \qquad G(t,y) = E(e^{ty}) = \lambda exp(\mu_1 t + \tfrac{1}{2}\sigma_1^2 t^2) + (1-\lambda)exp(\mu_2 t + \tfrac{1}{2}\sigma_2^2 t^2)$$

with its empirical counterpart defined as

$$(2.3.2) \qquad \bar{g}_n(t) = n^{-1}\sum_{i=1}^{n} exp(ty_i)$$

Quandt and Ramsey (1978) propose to estimate the 5 parameters by minimizing the distance between these two functions (i.e. (2.3.1) and (2.3.2) for a given set of fixed grid points t_1, $t_2,\ldots,t_5$. That is, one chooses a set of 5 distinct values of $t_1,\ldots,t_5$ and finds $\hat{\theta}$ to minimize

$$(2.3.3) \qquad \varepsilon(\theta)'\varepsilon(\theta)$$

where $[\varepsilon(\theta)]_i = G(t_i,\theta) - \bar{g}_n(t_i)$, $i = 1,2,\ldots,5$, and $\theta = (\lambda,$ μ_1, μ_2, σ_1^2, $\sigma_2^2)$. Minimization of (2.3.3) is equivalent to a non-linear least squares problem and hence asymptotically,

$$n^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} N(0,\Psi)$$

where $\Psi = (A_n'A_n)^{-1}A_n'\Omega A_n(A_n'A_n)^{-1}$

with $[A_n]_{ij} = \partial G(t_i,\theta)/\partial\theta_j$

and $\Omega = \Omega(t,\theta) = G(t_i+t_j,\theta) - G(t_i-t_j,\theta)$

for all i,j = 1,2,...,5.

Schmidt (1982) pointed out that a more efficient estimator of θ, θ^{*}, can be found by minimizing the generalized least-squares criterion rather than (2.3.3), i.e.

$$\varepsilon(\theta)'\hat{\Omega}^{-1}\varepsilon(\theta) \tag{2.3.4}$$

where $\hat{\Omega}$ is a consistent estimator of Ω. Asymptotically,

$$n^{1/2}\,(\theta^{*} - \theta) \xrightarrow{d} N(0,[A_n'\hat{\Omega}^{-1}A_n]^{-1}).$$

It is interesting to note that Quandt and Ramsey's MGF method and the method of moments are related. Johnson (1978) points out that the use of the MGF method with a normal mixture is equivalent to the use of the method of moments with fractional moments on the mixture of lognormals and thus would suffer some of the same drawbacks as the method of moments.

III. EMPIRICAL CHARACTERISTIC FUNCTION PROCEDURE

3.1. Theoretical Aspects

The empirical characteristic function (hereafter ECF) procedure has been previously investigated by Paulson, Holcomb and Leitch (1975), Heathcote (1977), Feuerverger and Mureika (1977), Bryan and Paulson (1983), Feuerverger and McDunnough (1981a, 1981b), and most recently Feuerverger (1990)[2]. These papers are mostly confined to the theoretical properties of the procedure, and very few have examined the application of the technique with the exception of the stable law family which has been considered extensively. In what follows we will examine the application of the ECF procedure to the case of mixtures of normal distributions.

Suppose that a random sample $y_1, y_2, \ldots, y_n$ has been drawn from a population with the probability density function specified in (2.1). Define the characteristic function (CF) of y_j as follows

(3.1.1)

$$c(t,\theta) = E[exp(ity_j)]$$

---

2 Feuerverger and McDunnough (1981a, 1981b), Feuerverger (1990) extend the procedure to estimate the stationary time series and stationary stochastic process models.

$$= \lambda exp(i\mu_1 t - \frac{1}{2}\sigma_1^2 t^2) + (1-\lambda) exp(i\mu_2 t - \frac{1}{2}\sigma_2^2 t^2)$$

$$= \lambda cos(\mu_1 t) exp(-\frac{1}{2}\sigma_1^2 t^2) + (1-\lambda) cos(\mu_2 t) exp(-\frac{1}{2}\sigma_2^2 t^2) +$$

$$i\left[\lambda sin(\mu_1 t) exp(-\frac{1}{2}\sigma_1^2 t^2) + (1-\lambda) sin(\mu_2 t) exp(-\frac{1}{2}\sigma_2^2 t^2)\right]$$

and the empirical characteristic function (ECF) as

$$(3.1.2) \qquad C_n(t) = n^{-1}\sum_{j=1}^{n} exp(ity_j)$$

$$= n^{-1}\sum_{j=1}^{n} cos(ty_j) + i\left[n^{-1}\sum_{j=1}^{n} sin(ty_j)\right]$$

where $i = \sqrt{-1}$, the imaginary number, $\theta = (\lambda, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$, and t are the fixed grid points which can be discrete or continuous. Now separating $c(t,\theta)$ and $C_n(t)$ into their real (*Re*) and imaginary (*Im*) parts and evaluating at *m* grid points, $t_1, t_2, \ldots, t_m$, we have

$$Z_n = \begin{bmatrix} Re\ C_n(t_1) \\ \vdots \\ Re\ C_n(t_m) \\ Im\ C_n(t_1) \\ \vdots \\ Im\ C_n(t_m) \end{bmatrix} \quad \text{and} \quad F(\theta) = \begin{bmatrix} Re\ c(t_1,\theta) \\ \vdots \\ Re\ c(t_m,\theta) \\ Im\ c(t_1,\theta) \\ \vdots \\ Im\ c(t_m,\theta) \end{bmatrix}$$

where

$$Re\ c(t_k,\theta) = \lambda cos(\mu_1 t_k) exp(-\tfrac{1}{2}\ \sigma_1^2 t_k^2) + (1-\lambda) cos(\mu_2 t_k) exp(-\tfrac{1}{2}\ \sigma_2^2 t_k^2)$$

$$Im\ c(t_k,\theta) = \lambda sin(\mu_1 t_k) exp(-\tfrac{1}{2}\ \sigma_1^2 t_k^2) + (1-\lambda) sin(\mu_2 t_k) exp(-\tfrac{1}{2}\ \sigma_2^2 t_k^2)$$

$$Re\ C_n(t_k) = n^{-1} \sum_{j=1}^{n} cos(t_k y_j)$$

$$Im\ C_n(t_k) = n^{-1} \sum_{j=1}^{n} sin(t_k y_j)\ , \qquad \text{for } k = 1,2,\ldots,m$$

From Feuerverger and McDunnough (1981a) we know that $\sqrt{n}\ (Z_n - F(\theta))$ is asymptotically normal with mean zero and ($2m$ x $2m$) covariance matrix:

$$\Omega = \begin{bmatrix} \Omega_{RR} & \Omega_{RI} \\ \Omega_{IR} & \Omega_{II} \end{bmatrix} \tag{3.1.3}$$

where the elements in the partitions associated with t_l and t_k are given by (see the appendix A for a formal derivation)[3]

$$(\Omega_{RR})_{lk} = \frac{1}{2}\left\{Re\ c(t_l+t_k) + Re\ c(t_l-t_k)\right\} - Re\ c(t_l)Re\ c(t_k)$$
$$(\Omega_{RI})_{lk} = \frac{1}{2}\left\{Im\ c(t_l+t_k) - Im\ c(t_l-t_k)\right\} - Re\ c(t_l)Im\ c(t_k)$$
$$(\Omega_{II})_{lk} = \frac{1}{2}\left\{Re\ c(t_l-t_k) - Re\ c(t_l+t_k)\right\} - Im\ c(t_l)Im\ c(t_k)$$

Furthermore, if we define

$$\varepsilon(t) = Z_n(t) - F(t,\theta) \tag{3.1.4}$$

then (3.1.4) can be thought of as the non-linear regression with a non-scalar covariance matrix where $Z_n(t)$ serves as the dependent variable and $F(t,\theta)$ serves as the right hand side explanatory function. Hence, for a given set of grid points $t_1,t_2,...,t_m$, an efficient estimator of θ, which we will denote as ECF estimator $\tilde{\theta}$, can be found by minimizing $\varepsilon(t)'\hat{\Omega}^{-1}\varepsilon(t)$ where $\hat{\Omega}$ is a consistent estimator of Ω. The asymptotic properties of $\tilde{\theta}$ have been examined by Feuerverger

[3] For notational simplicity we suppress θ in $c(t,\theta)$.

and McDunnough (1981a, 1981b) with the basic result stated in the following theorem.

Theorem 3.1. Let $t_1, t_2, \ldots, t_m$ *be distinct fixed grid points. The* ECF *estimator of* θ, $\tilde{\theta}$ *is strongly consistent and asymptotically normal with covariance matrix* $n^{-1}\left[A_n'\Omega^{-1}A_n\right]^{-1}$ *where* $A_n = \dfrac{\partial F(\theta)}{\partial \theta}$.

Proof: See Feuerverger and McDunnough (1981a, 1981b).

From theorem 3.1 it is clear that the asymptotic efficiency of the ECF procedure depends essentially on the choice of $\{t_j\}$. Feuerverger and McDunnough (1981b) argue that for some cases, one can obtain full asymptotic efficiency of the procedure (in terms of achieving Cramer-Rao lower bound) by selecting the grid points $\{t_j\}$ to be sufficiently fine and extended. A more detailed discussion on the choice of $\{t_j\}$ is given on the next subsection.

3.2. Choice of $\{t_j\}$

Like the MGF procedure, ECF procedure also requires specification of the grid points $t_1, t_2, \ldots, t_m$. Schmidt (1982) gives a reasonable detailed discussion regarding the choice of t's including how to obtain the optimal t values for a given *m* and the value of *m*. He suggests one possible criterion that might be useful in practice is to choose t that minimizes the size of the asymptotic covariance matrix (or its determinant) evaluated at some preliminary value of the estimated parameters. However, to further simplify the optimal choice of t in the minimization problem, Feuerverger and McDunnough (1981b) suggested that t should be chosen to be equally spaced, that is, taking $t_j = \tau j$, for $j = 1,2,\ldots,m$ and τ is some *real* constant. This reduces the minimization to one dimension.

We remark that in determining the optimal value for τ, one should keep in mind that both *negative* and *positive* value of τ would produce the same estimates of the parameters when using ECF procedure. This would not be the case for the MGF estimator. This type of 'symmetry' essentially comes from the properties of the trigonometric functions. Also, some initial experiences (not reported here) show that the initial choice for τ equal to 0.25 to 0.65 seem to work reasonably well in terms of the number of iterations it takes to convergence.

For the value of *m*, Schmidt (1982) showed that for the MGF estimator, as the number of *t's* increases the asymptotic variance of the estimator declines. In fact, as *m* approaches infinity, the asymptotic variance of the estimator approaches the Cramer-Rao lower bound. Thus, the task of determining the optimal value of *m* will not be trivial and as Schmidt has pointed out it may depend on the finite sample properties of the estimator which at the present are unknown. Nevertheless, for practical purposes, m = 10 is sufficient. We found similar results for the case of the ECF estimator as one might expect since MGF and ECF methods are both based on the data transformation, that is, Laplace transformation for the MGF and Fourier Transformation for the ECF. Some of the Monte Carlo results will be presented in the next section.

IV. MONTE CARLO EXPERIMENTS

4.1. Experimental Design

To evaluate the performance of the ECF technique in finite samples, several sampling experiments were carried out. Samples, of size n = 50 and 100, of the random variable were generated according to (2.1). Table 1 summarizes the seven experiments undertaken specifying their parameter values and sample size[4]. Figures 1-6 plot the density of the mixture for each case. Note that if the mixing weights are equal and the variances of the two components are equal, then the mixture density is bimodal if $|\mu_1 - \mu_2|/\sigma > 2$ (Case 3 and 4). Other bimodal densities include Case 1, 5 and 6.

---

[4] All the cases examine here are previouly studied either by Quandt and Ramsey (1978) or Schmidt (1982).

Each experiment was replicated as many times as required to successfully produce 1000 replications, and the minimization was performed with the DFP (Davidon-Fletcher-Powell) algorithm and the computation was terminated if the length of the gradient fell below 10^{-5}. All the computations were done using GAUSS386 Version 2.1. on a 486DXII-50 PC. A small experiment on the effect of the starting values of θ, showed that the ECF procedure was insensitive to reasonable initial guesses. Consequently, the true parameter values were used as the starting values. Also, in all experiments, the value of τ was set to 0.4.

4.2. Simulation Results

4.2.1. *Finite Sample Results*

Table 2 (A)-(G) displays the summary statistics of the sampling distributions of the ECF estimates of the five parameters for the seven cases. Examination of Table 2 will reveal that the ECF estimator performs very well. The means are close to the medians which are close to the true values and the quantiles are more or less symmetrically placed around the true parameter values. The interquartile ranges (IQR) are close to their expected values of 1.35 times the true asymptotic standard errors. Table 2 also contains the means and standard deviations of the estimates over 1000 replications. Furthermore, there are other interesting features in Table 2 worth noting. First, increasing the variance of one component, the quality of the estimates deteriorates both in terms of mean squared errors (MSE) and median absolute deviation (MAD) (Case 1 and 2). Second, increasing the sample size improves the accuracy of the estimates (Case 3 and 4). Third, asymmetrical mixtures with common variances (Case 5 and 6) generally increase (albeit small) the inaccuracy of the estimate of both the mean and the variance of the component with the lower mixture proportion. Finally, the MAD was always smaller than root MSE for all cases.

Table 3 presents a comparison of the root mean squared errors (RMSE) of the ECF and MLE estimates for the seven cases . First note that when the two components of the mixtures are not well separated, as in Case 7, MLE procedure seems to have serious problems in estimating the parameters. Previous simulation studies by Hosmer (1973, 1974) have shown that one would need a sample size of at least n = 250 in order to get reliable estimates. As for the remaining six cases, out of 30 possible comparisons of the RMSE, the ECF method out performs the MLEs in 19 of 30 possible comparisons. These results illustrate the superiority of the ECF procedure over the MLE in small samples. It is also worth noting that we did not compare the ECF procedure with the MGF method since, comparison of the methods using the same grid points is rarely possible. The spacing between the grid points for the MGF is required to be small whereas the spacing between grid points for the ECF procedure must be much larger. Closely spaced grid points for the ECF procedure yields singularity in the covariance matrix . Also, since the characteristic function of mixtures of normals contains both *real* and *imaginary* parts, the grid points used in estimation are twice as numerous as with the MGF method. Consequently, one would expect the ECF procedure to perform better than the MGF method in both finite samples and asymptotically.

Finally, some experiments were run in which the value of *m* (number of grid points) was increased to 6, 7 or 10 to see if the sampling distributions of the estimates could offer any guidances as to the choice of *m*. Intuitively, the ECF procedure can be thought of as the generalized method of moments (GMM) applied to the exponentiated data with the instruments being the grid points. However, while the instruments used in the GMM technique come from the relevant data sets, the grid points in the ECF procedure are arbitrary. Tauchen (1986) studied the finite sample behavior of the GMM estimator and showed that there is strong evidence of a bias/variance trade-off as the number of instruments increases. Thus, in practice, the instrument list should be kept small[5]. Unfortunately, this result does not carry over to the ECF procedure. Our results (not reported here but available upon request) indicate that there is evidence of a bias/variance trade-off for *some*, but not all, of the parameters. Consequently, the choice of grid size *m* needs further study.

4.2.2. *Asymptotic Results*

One standard for evaluating an estimation method is the reliability of its interval estimates of the parameters. Table 4 provides the coverage rate for 95 percent confidence intervals of the ECF estimators. The 95 percent nominal confidence intervals were constructed using the estimates of the asymptotic standard errors from Theorem 1. We then compute the number of times in 1000 replications that individual 95 percent confidence intervals for each of the five parameters cover their respective underlying true values. Our results show that, in almost all cases, the intervals are extremely reliable as the coverage rates are very close to the expected value of 0.95.

Another standard on which any estimation method should be evaluated is the validity of the estimator's asymptotic

[5] There was a bias/variance trade-off in the sense that, as the number of instruments increased, the measure of dispersion decreased and the bias increased.

distribution. Table 5 reports the Kolmogorov-Smirnov test for normality of the ECF estimators. At a 5 percent significance level, normality is rejected in 23 out of 35 possible cases. However, a comparison of Case 3 with Case 4 shows that doubling the sample size greatly improves the fit of the normal distribution. Normality is strongly rejected in Case 6 in which the mixtures are highly asymmetrical with equal variances.

Table 6 displays how the choice of m (number of t's) affects the asymptotic variance of the estimated parameters for Case 1 with a sample size of n = 100. The column labelled $\hat{\tau}$ gives the "optimal" value for the spacing parameter τ from minimizing the determinant of the asymptotic covariance matrix. It is clear from Table 6 that the asymptotic efficiency of the ECF estimators increases as we increase the number of t's. This result seems reasonable since, as Schmidt (1982) pointed out for the case of the MGF estimator, increasing m is equivalent to adding extra observations to the generalized least squares regression and hence the asymptotic efficiency could never decrease. In fact, as $m \longrightarrow \infty$, Schmidt conjectured (and also provided some evidence) that the asymptotic variance of the MGF estimators approach the Cramer-Rao lower bound[6]. Our result also

[6] Following Schmidt (1982) the information matrix is evaluated by simulation. Specifically, 50,000 drawings were

supports this conjecture for the case of the ECF estimators. Also, it is interesting to note that as m increases the optimal spacing between the t values uniformly declines. This striking result is somewhat counterintuitive since for smaller τ, the asymptotic covariance matrix, $(A'\Omega^{-1}A)^{-1}$, is almost singular. However, the determinant of $(A'\Omega^{-1}A)^{-1}$ is small when τ is small so as m increases smaller values of τ can indeed be optimal.

V. THE SWITCHING REGRESSION MODEL

Now consider a more general case where we let the means of the two normal distributions depend on the values of some of the explanatory variables. That is,

$$
\text{(4.1)} \qquad
\begin{aligned}
y_j &\sim N(\mu_{j1},\sigma_1^2) \quad \text{with probability } \lambda \\
y_j &\sim N(\mu_{j2},\sigma_2^2) \quad \text{with probability } (1-\lambda)
\end{aligned}
$$

with

$$
\begin{aligned}
\mu_{j1} &= x_j'\beta_1 \\
\mu_{j2} &= x_j'\beta_2, \qquad j = 1,2,\ldots,n
\end{aligned}
$$

where x_j is a Q-vector of non-stochastic, observable

---

made from the particular mixture distribution, and for each drawing the second derivative matrix was calculated; these 50,000 second derivative matrices were then average to get the information matrix.

explanatory variables. Let $\theta = (\beta_1, \beta_2, \sigma_1^2, \sigma_2^2, \lambda)$ which is of dimension $Q^* = (2Q+3)$. Also let us redefine the notation for the characteristic function, with the subscript k = 1,2,...,*m* and j = 1,2,...,n.

$$Re\ c(t_k,\theta,x_j) = \lambda\cos(t_k x_j'\beta_1)\exp(-\tfrac{1}{2}\sigma_1^2 t_k^2) + (1-\lambda)\cos(t_k x_j'\beta_2)\exp(-\tfrac{1}{2}\sigma_2^2 t_k^2)$$

$$Im\ c(t_k,\theta,x_j) = \lambda\sin(t_k x_j'\beta_1)\exp(-\tfrac{1}{2}\sigma_1^2 t_k^2) + (1-\lambda)\sin(t_k x_j'\beta_2)\exp(-\tfrac{1}{2}\sigma_2^2 t_k^2)$$

$$Re\ C_{nj}(t_k) = \cos(t_k y_j)$$

$$Im\ C_{nj}(t_k) = \sin(t_k y_j)$$

$$\varepsilon_j(t) = Z_{nj} - F_j(t,\theta)$$

where Z_{nj}, and $F_j\ (t,\theta)$ are defined as

$$Z_{nj} = \begin{bmatrix} Re\ C_{nj}(t_1) \\ \vdots \\ Re\ C_{nj}(t_m) \\ Im\ C_{nj}(t_1) \\ \vdots \\ Im\ C_{nj}(t_m) \end{bmatrix}, \quad F_j(t,\theta) = \begin{bmatrix} Re\ c(t_1,\theta,x_j) \\ \vdots \\ Re\ c(t_m,\theta,x_j) \\ Im\ c(t_1,\theta,x_j) \\ \vdots \\ Im\ c(t_m,\theta,x_j) \end{bmatrix}$$

Also, since μ_{jl} $(l = 1,2)$ depends on x_j, we need to modify the covariance structure of $\varepsilon_j = (\varepsilon_{j1}, \varepsilon_{j2},\ldots,\varepsilon_{jm})$, so that

$$(4.2) \qquad E[\varepsilon_j \varepsilon_j'] = \Omega_j = \begin{bmatrix} \Omega^j_{RR} & \Omega^j_{RI} \\ \Omega^j_{RI} & \Omega^j_{II} \end{bmatrix} \qquad j = 1,2,\ldots n.$$

where Ω_j is a (2m x 2m) matrix associated with observation j.

Following Schmidt (1982), to obtain the ECF estimator for θ, we minimize the following sum of squares criterion:

$$(4.3) \qquad S_n = n^{-1} \sum_{j=1}^{n} \varepsilon_j' \hat{\Omega}_j^{-1} \varepsilon_j$$

where $\hat{\Omega}_j$ is a consistent estimator for Ω_j based on any consistent estimate of θ. The asymptotic properties of ECF estimator for this case are a straight forward generalization of theorem 3.1 and hence it is easy to verify that,

$$(\tilde{\theta} - \theta) \xrightarrow{d} N(0,\Phi_n)$$

where $\Phi_n = \lim_{n\to\infty} n\left[W'(\hat{\Omega}^*)^{-1}W\right]^{-1}$

$\hat{\Omega}^*$ is an $(nm \times nm)$ block diagonal matrix whose jth block is $\hat{\Omega}_j$ defined in (4.2), j = 1,2,...,n.

$W' = (A_1,A_2,\ldots,A_n)$ is a $(2nm \times Q^*)$ matrix with

$$A_j = \left[\frac{\partial F_j(t,\theta)}{\partial\theta}\right].$$

VI. CONCLUSION

This paper has considered the empirical characteristic function prccedure for estimating the parameters of normal mixtures and switching regressions. The Monte Carlo study showed that the procedure produces estimates with good finite sample properties even in the case where the maximum likelihood estimator fails to exist. One important problem remaining is the choice of the number of values of the fix grid points t. This has shown to be a difficult problem even in the finite sample since asymptotically more t's are preferred to less. Thus, the technique used in this paper can be viewed as providing an alternative estimator (or it can be used as a starting values) to the popular maximum likelihood approach.

References

Amemiya, T. (1977), "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator of the General Non- linear Simultaneous Equation Model," *Econometrica*, 45, 955-968.

Bhattacharya, C. G. (1966), "A Simple Method of Resolution of a Distribution into Gaussian Components," *Biometrics*, 23, 115-135.

Bryant, J. L. and A. S. Paulson (1983), "Estimation of Mixing Proportions via Distance Between Characteristic Functions," *Communications in Statistics-Theory and Method*, 12, 1009- 1029.

Clarke, B. R. and C. R. Heathcote (1978), "Comment on 'Estimating Mixture of Normal Distributions and Switching Regressions', " *Journal of the American Statistical Association*, 73, 749- 750.

Cohen, A. C. (1967), "Estimation in Mixtures of Two Normal Distributions," *Biometrika*, 56, 15-28.

Day, N. E. (1969), "Estimating the Components of a Mixture of Normal Distributions," *Biometrika*, 56, 463-474.

Feuerverger, A., and R. Mureika (1977), "The Empirical Characteristic Function and Its Applications," *The Annals of Statistics*, 5, 88-97.

Feuerverger, A. and P. McDunnough (1981a), "On Some Fourier methods for Inference," *Journal of the American Statistical Association*, 76, 379-386.

Feuerverger, A. and P. McDunnough (1981b), "On the Efficiency of Empirical Characteristic Function Procedures," *Journal of The Royal Statistical Society*, B, 43, 147-156.

Feuerverger, A. (1990), "An Efficiency Result for the Empirical Characteristic Function in Stationary Time-Series Models," *Canadian Journal of Statistics*, 18, 155-161.

Granger, C. W. J. and D. Orr (1972), "'Infinite Variance' and research Strategy in Time Series Analysis," *Journal of the American Statistical Association*, 67, 275-285.

Hamermesh, D. S. (1970), "Wage Bargains, Threshold Effects, and the Phillip Curve," *Quarterly Journal of Economics*, 84, 501-517.

Heathcote, C. R. (1977), "Integrated Mean Square Error Estimation of Parameters," *Biometrika*, 64, 255-264.

Hosmer, D. W. Jr. (1973), "On the MLE of the Parameters of a mixture of Two Normal Distributions When Sample Size is Small," *Communications in Statistics*, 1, 217-227.

Hosmer, D. W. Jr. (1974), "Maximum Likelihood Estimates of the Parameters of a Mixture of Two Regression Lines," *Communications in Statistics*, 3, 995-1006.

Hosmer, D. W. Jr. (1978), "Comment on 'Estimating Mixtures of Normal Distributions and Switching Regressions'," *Journal of the American Statistical Association*, 73, 741-744.

Johnson, N. L. (1978), "Comment on 'Estimating Mixture of Normal Distributions and Switching Regressions'," *Journal of the American Statistical Association*, 73, 750.

Kiefer, J., and J. Wolfowitz (1956), "Consistency on the Maximum Likelihood Estimator in the Presence of many Nuisance Parameters," *The Annals of Mathematical Statistics*, 27, 887-906.

Kiefer, N. M. (1978), "Discrete Parameter Variation: Efficiency Estimation of a Switching Regression Model," *Econometrica*, 46, 427-434.

Kiefer, N. M. (1978), "Comment on 'Estimating Mixture of Normal Distributions and Switching Regressions'," *Journal of the American Statistical Association*, 73, 744-745.

Lindsay, B. G. and P. Basak (1993), "Multivariate Normal Mixtures: A Fast Consistent Method of Moment," *Journal of the American Statistical Association*, 88, 468-476.

Newcomb, S. (1886), "A Generalize Theory of the Combination of Observations so as to Obtain the Best Result," *American Journal of Mathematics*, 8, 343-366.

Odell, P. L., and J. P. Basu (1976), "Concerning Several

Methods for Estimating Crop Acreages using Remote Sensing Data," *Communications in Satatistics, A*, 5, 1091-1114.

Quandt, R. E. (1972), "A New Approach to Estimating Switching Regressions," *Journal of the American Statistical Association*, 67, 306-310.

Quandt, R. E. and J. B. Ramsey (1978), "Estimating Mixtures of Normal Distributions and Switching Regressions," *Journal of the American Statistical Association*, 73, 730-738.

Schmidt, P. (1982), "An Improved Version of the Quandt-Ramsey MGF Estimator for Mixtures of Normal Distributions and Switching Regressions," *Econometrica*, 50, 501-516.

Tan, W. Y. and W. C. Chang (1972), "Some Comparisons of the method of Moments and the Method of Maximum Likelihood in Estimating Parameters of a Mixture of Two Normal Densities," *Journal of the American Statistical Association* 67, 702-708.

Tauchen, G. (1986), "Statistical Properties of Generalized Method -of-Moments Estimators of Structual Parameters Obtained From Financial Market Data," *Journal of Business and Economic Statistics* 4, 397-416.

Teicher, H. (1961), "Identifiability of Mixtures," *The Annals of Mathematical Statistics*, 32, 244-248.

Teicher, H. (1963), "Identifiability of Finite Mixtures," *The Annals of Mathematical Statistics*, 34, 1265-1269.

Titterington, D. M., A. F. M. Smith and U. E. Makov (1985), *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Son Ltd.

APPENDIX : DERIVATION OF COVARIANCE MATRIX Ω_{ε}.

$$E\left[\varepsilon(t)\varepsilon(t)'\right] = \Omega = \begin{bmatrix} \Omega_{R,R} & \Omega_{R,I} \\ \Omega_{I,R} & \Omega_{I,I} \end{bmatrix}$$

where

(1) $$\Omega_{R,R} = E\left\{ [Re\ C_n(t_k) - Re\ c(t_k)][Re\ C_n(t_l) - Re\ c(t_l)] \right\}$$

$$= E\left\{ Re\ C_n(t_k) Re\ C_n(t_l) \right\} - Re\ c(t_k) Re\ c(t_l)$$

Now:

$$E\left\{ Re\ C_n(t_k) Re\ C_n(t_l)) \right\} = n^{-2} E\left\{ \sum_{j=1}^{n} cos(t_k y_j) \sum_{j=1}^{n} sin(t_l y_j) \right\}$$

$$= n^{-2} \sum_{j=1}^{n} E\left\{ cos(t_k y_j) cos(t_l y_j) \right\} +$$

$$2n^{-2} \sum_{j} \sum_{r} E\left\{ cos(t_k y_j) cos(t_l y_r) \right\} \quad \text{(A1)}$$

The E(.) of the first term of (A1) can be written as:

$$E\left\{ cos(t_k y_j) cos(t_l y_j) \right\} = E\left\{ \frac{1}{4} (e^{it_k x_j} + e^{-it_k x_j}) (e^{it_l x_j} + e^{-it_l x_j}) \right\}$$

by using the fact that $cos(\alpha) = \frac{1}{2}(e^{i\alpha} + e^{-i\alpha})$. Hence

$$= \frac{1}{4} E\left\{ e^{i(t_k + t_l) y_j} + e^{i(t_k - t_l) y_j} + \right.$$

$$e^{-i(t_k+t_l)y_j} + e^{-i(t_k-t_l)y_j} \Big\}$$

$$= \frac{1}{4}\Big\{ c(t_k+t_l) + c(t_k-t_l) + c(-(t_k+t_l)) + c(-(t_k-t_l)) \Big\}$$

$$= \frac{1}{2}\Big\{ Re\ c(t_k+t_l) + Re\ c(t_k-t_l) \Big\} \qquad (A2)$$

Similarily, the E(.) of the second term in (A1) can be written as:

$$E\Big\{ cos(t_k y_j) cos(t_l y_r) \Big\}$$

$$= E\Big\{ \frac{1}{4}\ e^{it_k y_j} + e^{-it_k y_j})(e^{it_l y_r} + e^{-it_l y_r}) \Big\}$$

$$= E\Big\{ e^{i(t_k y_j + t_l y_r)} + e^{i(t_k y_j - t_l y_r)} +$$

$$e^{-i(t_k x_j + t_l x_r)} + e^{-i(t_k x_j - t_l x_r)} \Big\}$$

$$= \frac{1}{4}\Big\{ c(t_k)c(t_l) + c(t_k)c(-t_l) +$$

$$c(-t_k)C(-t_l) + c(-t_k)c(t_l) \Big\}$$

$$= \frac{1}{4}\Big\{ [c(t_k) + c(-t_k)][c(t_l) + c(-t_l)] \Big\}$$

$$= Re\ c(t_k) Re\ c(t_l) \qquad (A3)$$

Now by subtituting (A3) and (A2) into (A1) ans since $\bar{z}(t)$ are i.i.d., we have:

$$E\Big\{ Re\ C_n(t_k) Re\ C_n(t_l) \Big\} = (2n)^{-1}\Big\{ Re\ c(t_k+t_l) +$$

$$Re\ c(t_k-t_l) \Big\} + (1 - 1/n) Re\ c(t_k) Re\ c(t_l)$$

Thus,

$$\Omega_{R,R} = n^{-1}\Big\{ \frac{1}{2}(Re\ c(t_k+t_l) + Re\ c(t_k-t_l)) -$$

$$Re\ c(t_k) Re\ c(t_l) \Big\}$$

$$(2)\ \Omega_{I,I} = E\Big\{ [Im\ C_n(t_k) - Im\ c(t_k)]$$

$$[Im\ C_n(t_k)) - Im\ c(t_l)] \Big\}$$

$$= E\left\{ Im\ C_n(t_k) Im\ C_n(t_l) \right\} - Im\ c(t_k) Im\ c(t_l)$$

Now,

$$E\left\{ Im\ C_n(t_k) Im\ C_n(t_l) \right\} = n^{-2} E\left\{ \sum_{j=1}^{n} sin(t_k y_j) \sum_{r=1}^{n} sin(t_l y_r) \right\}$$

$$= n^{-2}\ E\left\{ \sum_{j=1}^{n} sin(t_k y_j) sin(t_l y_j) + \sum\sum_{j \neq r} sin(t_k y_j) sin(t_l y_r) \right\}$$

$$= n^{-2} \sum_{j=1}^{n} E\left\{ sin(t_k y_j) sin(t_l y_j) \right\} + n^{-2} \sum\sum_{j \neq r} E\left\{ sin(t_k y_j) sin(t_k y_r) \right\} \qquad (A4)$$

Furthermore, by using $sin(\alpha) = \dfrac{e^{i\alpha} - e^{-i\alpha}}{2i}$, the E(.) of the first term in (A4) becomes:

$$E\left\{ sin(t_k y_j) sin(t_l y_j) \right] = E\left\{ -\frac{1}{4} (e^{it_k y_j} - e^{-it_k y_j}) (e^{it_l y_j} - e^{it_l y_j}) \right\}$$

$$= -\frac{1}{4}\ E\left\{ e^{i(t_k+t_l)y_j} + e^{-i(t_k+t_l)y_j} - e^{i(t_k-t_l)y_j} - e^{-i(t_k-t_l)y_j} \right\}$$

$$= -\frac{1}{4} \left\{ c(t_k+t_l) + c(-(t_k+t_l)) - c(t_k-t_l) - c(-(t_k-t_l)) \right\}$$

$$= \frac{1}{2} \left\{ -Re\ c(t_k+t_l) + Re\ c(t_k-t_l) \right\} \qquad (A5)$$

Simialrily, the E(.) in the second term of (A4) can be shown to have

$$E\left\{ sin(t_k y_j) sin(t_l y_r) \right\} = Im\ c(t_k) Im\ c(t_l)) \qquad (A6)$$

Hence, by subtituting (A6) and (A5) into (A4), we get

$$E\left\{ Im\ C_n(t_k) Im\ C_n(t_l) \right\} = (2n)^{-1}\left\{ -Re\ c(t_k+t_l) + Re\ c(t_k-t_l) \right\} - (1 - 1/n) Im\ c(t_k) Im\ c(t_l)$$

Therefore,

$$\Omega_{I,I} = n^{-1}\left\{ \frac{1}{2} [-Re\ c(t_k+t_l) + Re\ c(t_k-t_l)] - Im\ c(t_k) Im\ c(t_l) \right\}$$

(3) $$\Omega_{R,I} = E\left\{ [Re\ C_n(t_k) - Re\ c(t_k)][Im\ C_n(t_l) - Im\ c(t_l)] \right\}$$
$$= E\left\{ Re\ C_n(t_k) Im\ C_n(t_l) \right\} - Re\ c(t_k) Im\ c(t_l)$$

$$E\left\{ Re\ C_n(t_k) Im\ C_n(t_l) \right\} = n^{-2} E\left\{ \sum_{j=1}^{n} cos(t_k y_j) \sum_{r=1}^{n} sin(t_l y_r) \right\}$$
$$= n^{-2} \sum_{j=1}^{n} E\left\{ cos(t_k y_j) sin(t_l y_j) \right\} + n^{-2} \sum_{j \neq r}\sum E\left\{ cos(t_k y_j) sin(t_l y_r) \right\} \quad (A7)$$

Now the E(.) in the first term of (A7) can be evaluate as follow:

$$E\left\{ cos(t_k y_j) sin(t_l y_j) \right\}$$
$$= E\left\{ \frac{1}{4i}(e^{it_k y_j}+e^{-it_k y_j})(e^{it_l y_j}-e^{-it_l y_j}) \right\}$$
$$= \frac{1}{4i}\left\{ E(e^{i(t_k+t_l)y_j}-e^{-i(t_k+t_l)y_j}) - \right.$$

$$E(e^{i(t_k-t_l)y_j}-e^{-i(t_k-t_l)y_j})\Big\}$$

$$= \frac{1}{2}\Big\{ Im\ c(t_k+t_l) - Im\ c(t_k-t_l) \Big\} \qquad \text{(A8)}$$

$$E\Big\{ cos(t_k y_j) sin(t_l y_r) \Big\} = E\Big\{ \frac{1}{4i}(e^{it_k y_j}+e^{-it_k y_j})$$

$$(e^{it_l y_r}-e^{-it_l y_r}) \Big\}$$

$$= \frac{1}{4i}\Big\{ c(t_k)c(t_l) + c(-t_k)c(t_l) -$$

$$c(t_k)c(-t_l) - c(-t_k)c(-t_l) \Big\}$$

$$= \frac{1}{4i}\Big\{ [c(t_k) + c(-t_k)][c(t_l) - c(-t_l)] \Big\}$$

$$= Re\ c(t_k) Im\ c(t_l) \qquad \text{(A9)}$$

Now subtituting (A9) and (A8) into (A7) we have:

$$E\Big\{ Re\ C_n(t_k) Im\ C_n(t_l) \Big\} = (2n)^{-1}\Big\{ Im\ c(t_k+t_l) -$$

$$Imc(t_k-t_l) \Big\} + (1 - 1/n) Re\ c(t_k) Im\ c(t_l)$$

Therefore,

$$\Omega_{R,I} = n^{-1}\Big\{ \frac{1}{2}\ [Im\ c(t_k+t_l)) - Im\ c(t_k-t_l)] -$$

$$Re\ c(t_k) Im\ c(t_l) \Big\}$$

Table 1. Summary Characteristics of Experiments

	Case						
Parameters	1	2	3	4	5	6	7
λ	0.5	0.5	0.5	0.5	0.6	0.75	0.5
μ_1	-3.0	-3.0	0.0	0.0	0.0	0.0	0.0
μ_2	3.0	3.0	4.0	4.0	4.0	4.0	0.0
σ_1^2	1.0	1.0	1.0	1.0	1.0	1.0	1.0
σ_2^2	3.0	16.0	1.0	1.0	1.0	1.0	10.0
n	50	50	50	100	50	50	100

Table 2. Quantiles, Sampling Statistics and Measures of Performance

(A) <u>CASE 1</u>: $\lambda = 0.5$, $\mu_1 = -3.0$, $\mu_2 = 3.0$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 3.0$, $n = 50$.

	Parameters				
	λ	μ_1	μ_2	σ_1^2	σ_2^2
Quantiles					
$Q_{.01}$	.3044	-3.5923	1.1144	.2413	1.1752
$Q_{.05}$	.3652	-3.3988	2.1113	.4551	1.4985
Med.	.5025	-3.0162	2.9966	.9591	2.9913
$Q_{.95}$	.6343	-2.6378	3.6176	1.8213	6.3395
$Q_{.99}$	.6999	-2.4691	4.0142	2.5661	11.5580
Other Statistics					
IQR	.1145	.3227	.5474	.5481	1.6126
Mean	.4995	-3.0158	2.9477	1.0339	3.3751
Std.	.0851	.2373	.5179	.4694	1.8123
*Bias and MSE**					
Bias	-.0005	-.0158	-.0523	.0339	.3751
	(.0027)	(.0075)	(.0146)	(.0148)	(.0573)
MSE	.0072	.0565	.2707	.2213	3.4221
	(.0004)	(.0028)	(.0247)	(.0185)	(.4037)
MAD	.0670	.1817	.3632	.3421	1.1454

* Standard errors are in parenthesis. IQR = Interquartile Range; Std. = Sample standard deviation; MAD = Median Absolute Deviation

(B) CASE 2: $\lambda = 0.5$, $\mu_1 = -3.0$, $\mu_2 = 3.0$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 16.0$, $n = 50$.

	Parameters				
	λ	μ_1	μ_2	σ_1^2	σ_2^2
Quantiles					
$Q_{.01}$	.1999	-3.6321	-2.1959	.0849	1.3342
$Q_{.05}$	.2879	-3.4102	-1.0104	.2608	3.3399
Med.	.5226	-2.9865	3.1114	.8912	12.4810
$Q_{.95}$	.7348	-2.4560	5.7693	2.3441	25.4575
$Q_{.99}$	.8361	-2.1518	12.0455	4.2156	31.1895
Other Statistics					
IQR	.1786	.3567	2.4691	.7764	9.9360
Mean	.5156	-2.9681	2.8948	1.0738	12.9927
Std.	.1345	.2951	2.2948	.9301	6.9153
Bias and MSE					
Bias	.0156	.0319	-.1052	.0738	-3.0073
	(.0043)	(.0136)	(.0726)	(.0294)	(.2187)
MSE	.0183	.0880	5.2716	.8696	56.8169
	(.0008)	(.0052)	(.3756)	(.2598)	(2.1542)
MAD	.1068	.2250	1.6494	.5225	5.5912

(C) CASE 3: $\lambda = 0.5$, $\mu_1 = 0.0$, $\mu_2 = 4.0$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 1.0$, $n = 50$.

	Parameters				
	λ	μ_1	μ_2	σ_1^2	σ_2^2
Quantiles					
$Q_{.01}$	.2817	-.7062	3.0723	.2699	.2912
$Q_{.05}$	.3562	-.4373	3.5085	.4210	.4469
Med.	.5035	-.0312	4.0135	.9496	.9505
$Q_{.95}$	.6460	.4776	4.4012	1.9020	2.0066
$Q_{.99}$	.7077	.7969	4.6131	2.9931	2.6879
Other Statistics					
IQR	.1178	.3561	.3277	.5729	.5467
Mean	.5005	-.0107	3.9875	1.0383	1.0504
Std.	.0900	.2890	.2932	.5069	.5151
Bias and MSE					
Bias	.0005	-.0107	-.0125	.0383	.0504
	(.0028)	(.0091)	(.0093)	(.0160)	(.0163)
MSE	.0081	.0835	.0860	.2582	.2676
	(.0004)	(.0046)	(.0056)	(.0214)	(.0274)
MAD	.0713	.2205	.2183	.3647	.3604

(D) CASE 4: $\lambda = 0.5$, $\mu_1 = 0.0$, $\mu_2 = 4.0$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 1.0$, $n = 100$.

	Parameters				
	λ	μ_1	μ_2	σ_1^2	σ_2^2
Quantiles					
$Q_{.01}$	.3548	-.4219	3.4912	.4689	.4720
$Q_{.05}$	.4057	-.3272	3.6525	.5827	.5844
Med.	.5017	-.0166	4.0020	.9702	.9713
$Q_{.95}$	.5961	.3212	4.3049	1.6761	1.6544
$Q_{.99}$	.6472	.5453	4.4540	2.2162	2.1171
Other Statistics					
IQR	.0758	.2544	.2682	.3991	.4203
Mean	.5011	-.0085	3.9944	1.0375	1.0256
Std.	.0593	.2022	.2007	.3892	.3410
Bias and MSE					
Bias	.0011	-.0085	-.0056	.0375	.0256
	(.0019)	(.0064)	(.0063)	(.0123)	(.0108)
MSE	.0035	.0409	.0402	.1527	.1168
	(.0002)	(.0029)	(.0021)	(.0250)	(.0092)
MAD	.0464	.1541	.1577	.2630	.2527

(E) CASE 5: $\lambda = 0.6$, $\mu_1 = 0.0$, $\mu_2 = 4.0$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 1.0$, $n = 50$.

	Parameters				
	λ	μ_1	μ_2	σ_1^2	σ_2^2
Quantiles					
$Q_{.01}$	.3882	-.5477	2.9811	.3474	.2337
$Q_{.05}$	.4597	-.3581	3.4517	.4922	.3868
Med.	.6005	.0038	4.0230	.9688	.9063
$Q_{.95}$	.7353	.4204	4.4928	1.8349	2.0043
$Q_{.99}$	.7948	.7007	4.7649	2.5143	2.8312
Other Statistics					
IQR	.1116	.3079	.3752	.5208	.6211
Mean	.6007	.0098	4.0100	1.0363	1.0148
Std.	.0854	.2515	.3383	.4495	.5364
Bias and MSE					
Bias	.0007	.0098	.0100	.0363	.0148
	(.0027)	(.0079)	(.0107)	(.0142)	(.0170)
MSE	.0073	.0633	.1144	.2032	.2877
	(.0004)	(.0040)	(.0108)	(.0187)	(.0232)
MAD	.0669	.1920	.2428	.3250	.3909

(F) CASE 6: $\lambda = 0.75$, $\mu_1 = 0.0$, $\mu_2 = 4.0$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 1.0$, $n = 50$.

	Parameters				
	λ	μ_1	μ_2	σ_1^2	σ_2^2
Quantiles					
$Q_{.01}$	.4327	-.5690	2.2314	.3794	.0964
$Q_{.05}$	.5911	-.3948	3.0211	.5254	.2193
Med.	.7469	-.0003	4.0755	.9834	.8157
$Q_{.95}$	.8631	.3579	4.6916	1.7801	2.4320
$Q_{.99}$	.9485	.5823	5.0406	2.3973	3.6065
Other Statistics					
IQR	.1017	.2914	.5259	.4427	.7242
Mean	.7420	-.0058	4.0028	1.0478	.9897
Std.	.1012	.3100	.5477	.3962	.7296
Bias and MSE					
Bias	-.0080	-.0058	.0028	.0478	-.0103
	(.0032)	(.0098)	(.0173)	(.0125)	(.0231)
MSE	.0103	.0960	.2996	.1591	.5319
	(.0017)	(.0261)	(.0270)	(.0144)	(.0476)
MAD	.0675	.1856	.3708	.2853	.4964

(G) CASE 7: $\lambda = 0.5$, $\mu_1 = 0.0$, $\mu_2 = 0.0$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 10.0$, $n = 100$.

	Parameters				
	λ	μ_1	μ_2	σ_1^2	σ_2^2
Quantiles					
$Q_{.01}$	.2129	-.5866	-4.5552	.1609	1.6623
$Q_{.05}$	.2923	-.4054	-1.3612	.3266	5.9550
Med.	.5133	.0083	.0178	.8633	9.8364
$Q_{.95}$	.7543	.3533	1.4804	1.9707	15.3315
$Q_{.99}$	.9490	.5290	5.9672	3.0842	18.2695
Other Statistics					
IQR	.1911	.2936	.8883	.5771	3.6404
Mean	.5190	-.0028	.0578	.9746	10.1123
Std.	.1463	.2326	1.3253	.5608	3.0523
Bias and MSE					
Bias	.0190	-.0028	.0578	-.0254	.1123
	(.0046)	(.0074)	(.0419)	(.0177)	(.0965)
MSE	.0217	.0540	1.7577	.3248	9.3202
	(.0017)	(.0261)	(.0270)	(.0144)	(.0476)
MAD	.1147	.1812	.7333	.3875	2.3137

Table 3. Comparison of the RMSE of ECF and MLE Estimates.

Par	Method	Case 1	2	3	4	5	6	7
λ	ECF	.0851	.1354	.0900	.0593	.0853	.1015	.1474
	MLE	.0912	.1050	.0960	.0593	.0991	.2104	——[a]
μ_1	ECF	.2377	.2966	.2890	.2023	.2516	.3099	.2325
	MLE	.2400	.2597	.2473	.2142	.2519	.3107	——
μ_2	ECF	.5203	2.2960	.2933	.2006	.3383	.5474	1.3258
	MLE	.6017	1.3203	.2969	.2013	.3501	.5503	——
σ_1^2	ECF	.4704	.9325	.5081	.3908	.4508	.3989	.5611
	MLE	.4013	.9622	.5090	.3963	.4623	.4003	——
σ_2^2	ECF	1.8499	7.5377	.5173	.3418	.3250	.2853	3.0529
	MLE	1.8110	6.1285	.4802	.3553	.1581	.1818	——

[a] The MLEs fail to converge due to singularity in the likelihood surface (failure rate for this case is more than 90%).

Table 4. Coverage Rate for 95% Confidence Interval

	Case						
Parameter	1	2	3	4	5	6	7
λ	.974	.946	.974	.979	.970	.947	.967
μ_1	.950	.956	.945	.956	.954	.954	.947
μ_2	.941	.889*	.943	.952	.947	.929	.934
σ_1^2	.934	.922	.923	.945	.935	.945	.932
σ_2^2	.947	.876*	.935	.948	.920	.890*	.955

* Differ from expected value of 0.950.

Table 5. Kolmogorov-Smirnov Test of Normality

	Case						
Param	1	2	3	4	5	6	7
λ	.647*	.911*	.532*	.640*	.880*	2.661	.819*
μ_1	.577*	1.320*	1.556**	1.265*	1.406**	3.344	1.163*
μ_2	2.625	2.247	1.931	.764*	1.784	3.100	5.117
σ_1^2	2.408	4.675	2.761	1.150*	2.646	2.613	3.477
σ_2^2	4.398	1.728	3.425	2.531	2.971	4.135	1.450**

* Normality is *not* rejected at 5% significance level (critical value = 1.36).

** Normality is *not* rejected at 1% significance level (critical value = 1.67).

Table 6. Asymptotic Variances of ECF Estimator for Various value of m.

Case 1: $\lambda = 0.5$, $\mu_1 = -3.0$, $\mu_2 = 3.0$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 3.0$, $n = 100$.

		Asymptotic Variance				
m	$\hat{\tau}$	λ	μ_1	μ_2	σ_1^2	σ_2^2
6	.3361	.00285	.02663	.08210	.05817	.71227
8	.3343	.00281	.02658	.07610	.05796	.60971
10	.3298	.00269	.02657	.07534	.05721	.59605
15	.3110	.00265	.02651	.07532	.05678	.59306
20	.2951	.00265	.02650	.07532	.05619	.59275
$\mathcal{I}^{-1}$		.00263	.02409	.07524	.05548	.59016

Fig. 1. PDF of Mixtures of Normal Distributions—Case 1

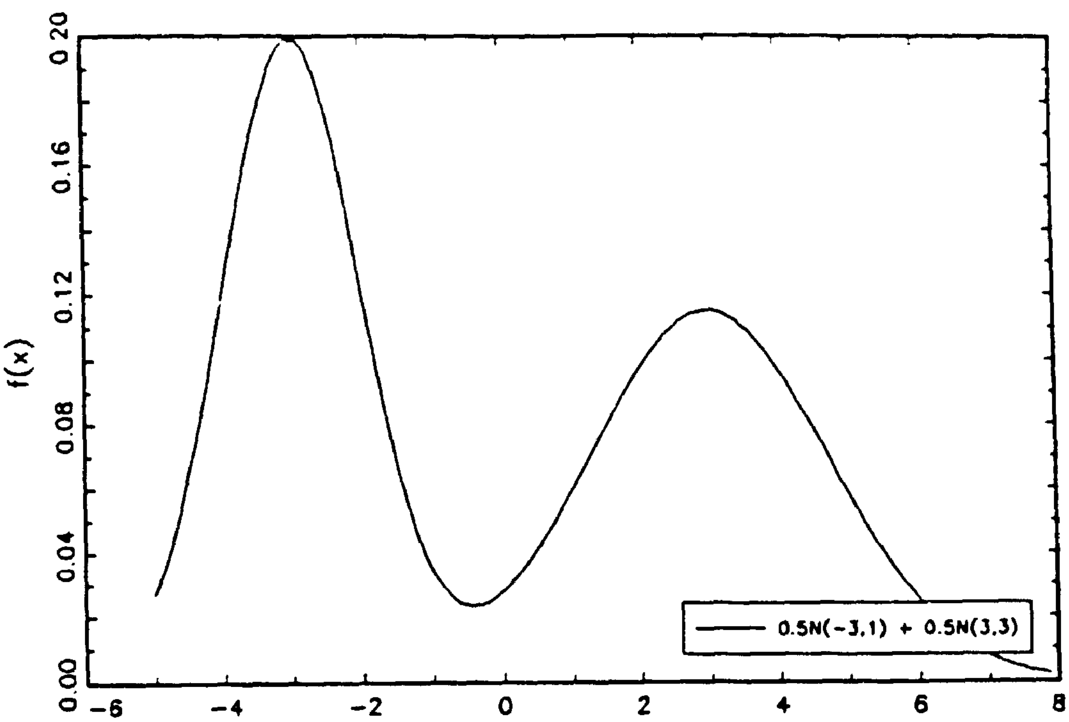


Fig. 2. PDF of Mixtures of Normal Distributions—Case 2

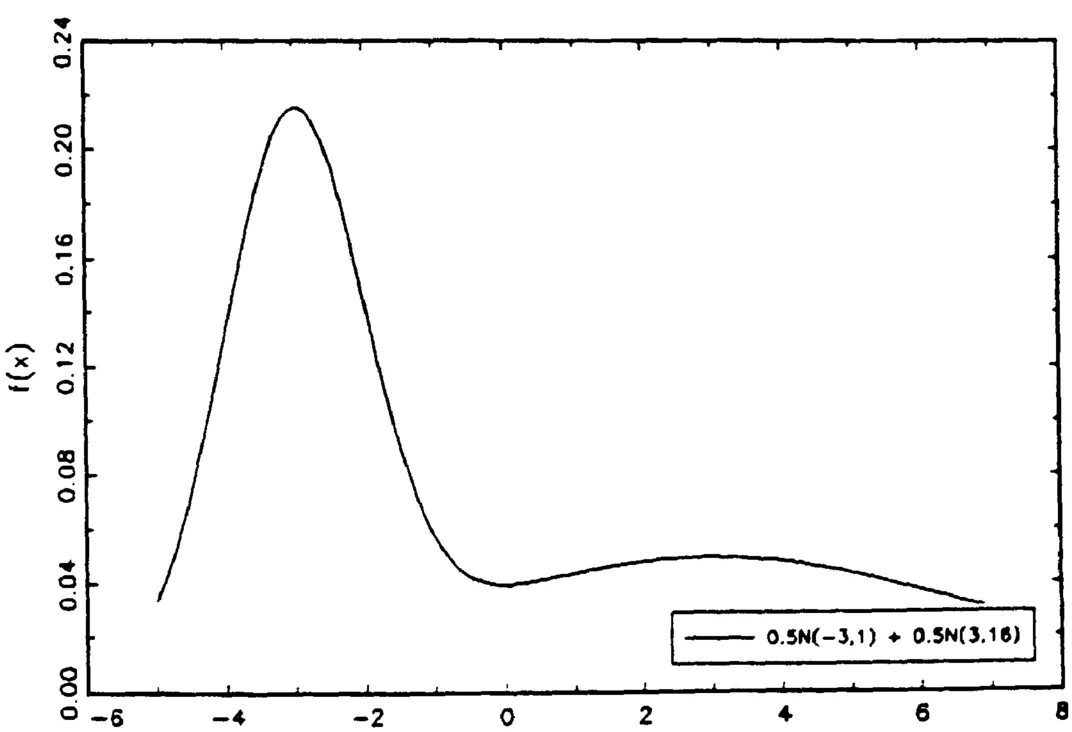


Fig. 3. PDF of Mixtures of Normal Distributions—Case 3 and 4

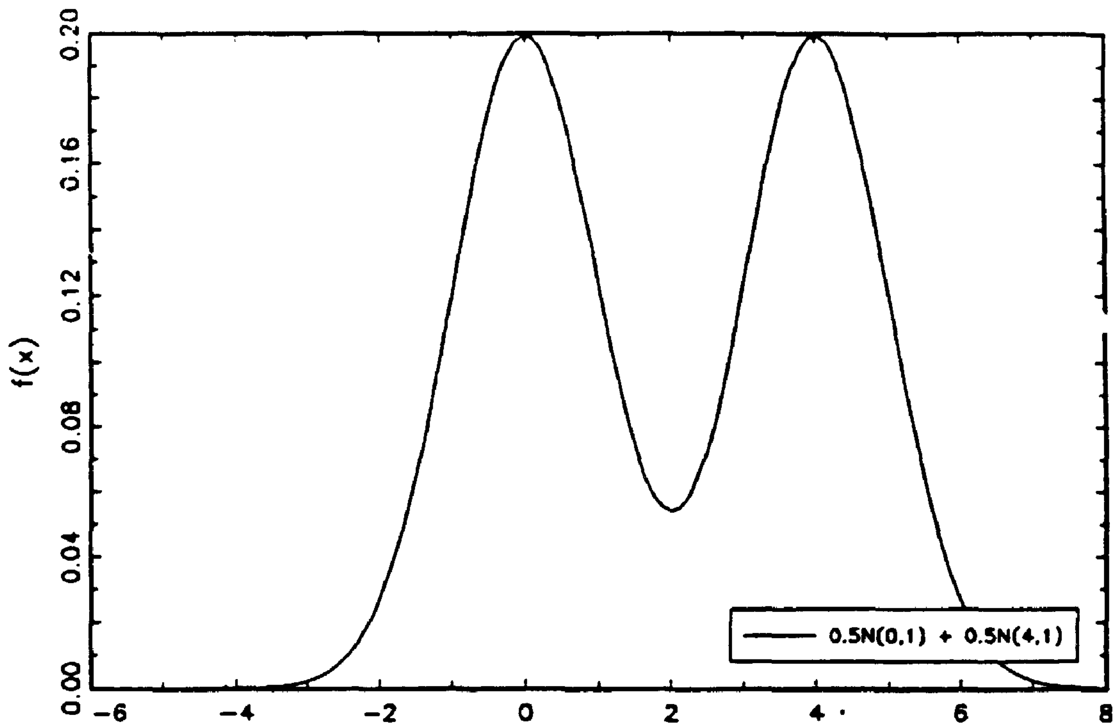


Fig. 4. PDF of Mixtures of Normal Distributions—Case 5

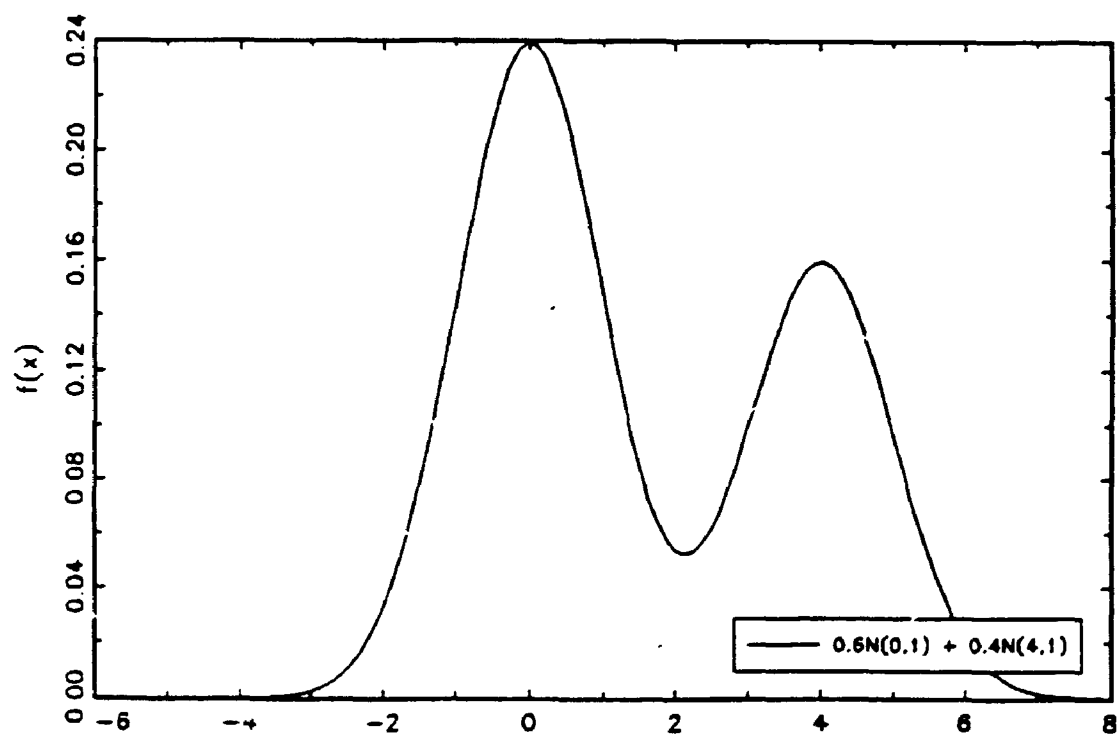


Fig. 5. PDF of Mixtures of Normal Distributions−Case 6

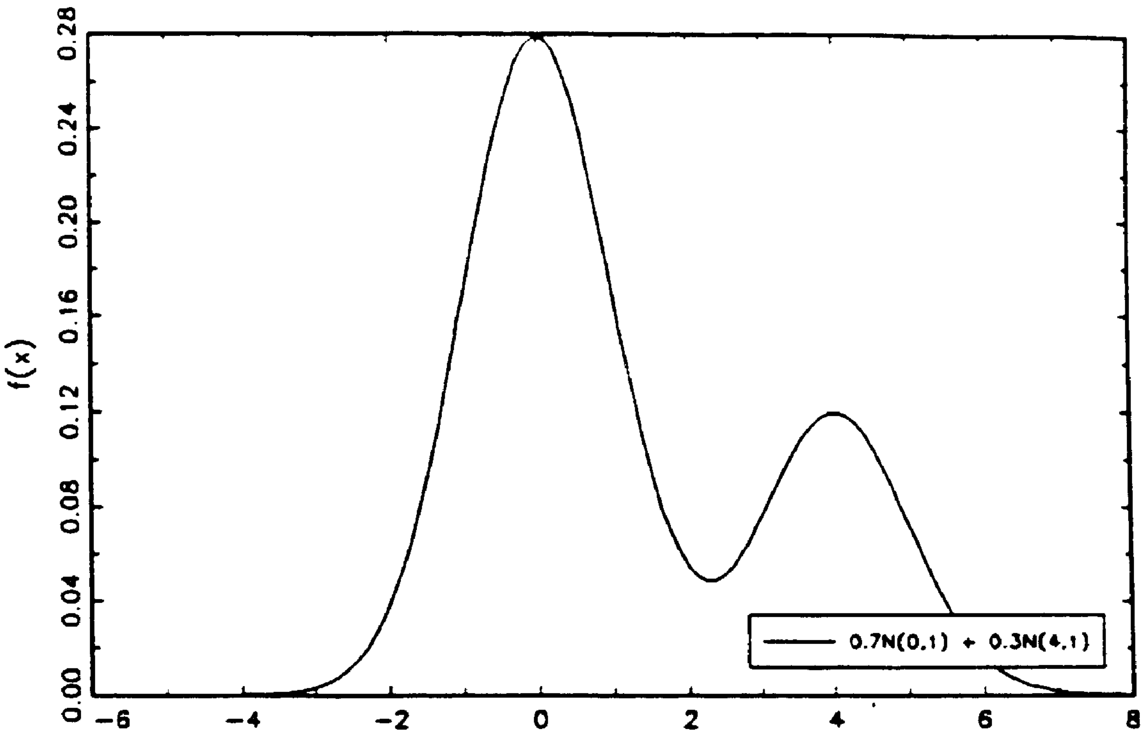
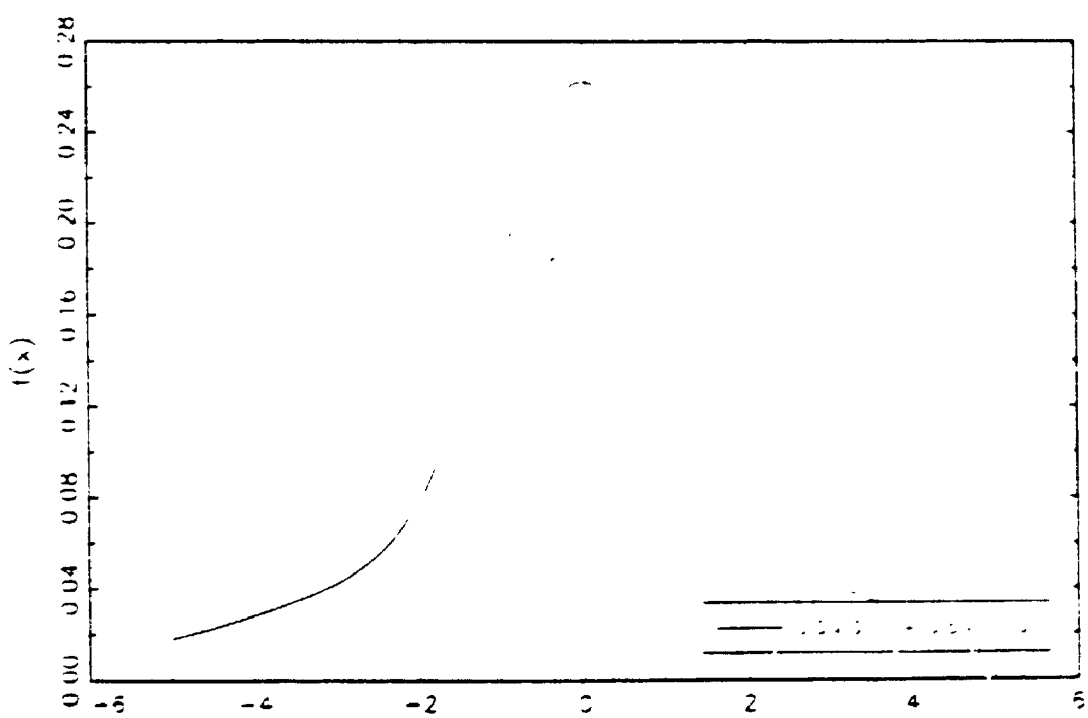


Fig. 6. PDF of Mixtures of Normal Distributions−Case 7



CHAPTER 2

INFORMATION MATRIX TEST FOR ARMA MODEL

I. INTRODUCTION

Recently, the information matrix (IM) test proposed by White (1982) as a test for general model specification has received a lot of attention (see for example Chesher (1983, 1984), Lancaster (1984), Chesher and Irish (1987), Hall (1987), Davidson and MacKinnon (1992) Bera and Lee (1993), Bera and Zuo (1993), Santos Silva (1993). In this paper, we use the IM framework to derive the test for a pure stationary ARMA process.

The paper is organised as follows. In section 2 we specify the model and derive the algebraic structure of the IM test. Unlike Hall (1987), and Bera and Lee (1993), who derived the IM test for the linear regression model without and with autoregressive disturbances respectively, our derived covariance matrix of the indicator vector no longer has a block diagonal structure implying the algebraic structure of the IM test is much more complicated. Our derived IM test can be separated into two general tests: a general test for parameter heterogeneity (i.e. test for random coefficient or conditional heteroskedaticity) of the specified model and a general test for normality. In section 3, the results derived in section 2 are interpreted while section 4 offers a concluding summary. All of the detailed algebraic derivations are given in the appendices.

II. MODEL AND INFORMATION MATRIX TEST

For the sake of simplicity, we consider the stationary and invertible ARMA(p,q) model of the form:

$$(1) \quad y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \gamma_1 \varepsilon_{t-1} + \ldots + \gamma_q \varepsilon_{t-q} + \varepsilon_t$$

where $p,q \geq 1$, $\varepsilon_t \sim N(0,\sigma^2)$, $t = 2,3,\ldots,T$, y_1 is taken to be fixed, and $\varepsilon_1 = 0$. Alternatively, we could start the recursion in (1) at $t = 1$ with y_0 and ε_0 set equal to zero. However, this would not be recommended since arbitrarily setting $y_0 = 0$ introduces a distortion into the calculation. The ARMA(p,q) stated in (1) can be written more elegantly by defining associated polynomials in the lag operator. If

(2) $$\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$$

and

(3) $$\gamma(L) = 1 + \gamma_1 L + \dots + \gamma_q L^q$$

then (1) becomes:

(4) $$\phi(L)y_t = \gamma(L)\varepsilon_t$$

We assume that the AR(p) and MA(q) polynomials in (4) do not have a *common factor*, i.e. they do not have a root which is the same. Otherwise, the model would be over-parameterized and hence would not be identifiable and possibly suffer from computational problems. Rearranging (1) we have

$$\varepsilon_t = y_t - \sum_{i=1}^{p}\phi_i y_{t-i} - \sum_{j=1}^{q}\gamma_j \varepsilon_{t-j}$$

(5) $$= y_t - \underline{y}_{t-1}\phi' - \underline{\varepsilon}_{t-1}\gamma'$$

where $\underline{y}_{t-1} = (y_{t-1}, y_{t-2}, \dots, y_{t-p})$,

$\underline{\varepsilon}_{t-1} = (\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q})$,

$\phi = (\phi_1, \phi_2, \dots, \phi_p)$, $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_q)$.

Assuming that $\underline{\varepsilon}_{t-1}$ is given, the conditional likelihood function for this model is given by:

$$(6) \quad L(\theta) = \sum_{t=2}^{T} \ell_t(\theta) = -\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\varepsilon_t^2$$

where $\theta = (\phi', \gamma', \sigma^2)$ is a $k\times 1$ vector of parameters with $k = (p+q+1)$. Let $\hat{\theta}$ denote the maximum likelihood estimate (MLE) of θ. Then White's IM test can be constructed based on

$$(7) \qquad d(\hat{\theta}) = vech\, \mathcal{C}(\hat{\theta}) = vech(\mathcal{A}(\hat{\theta}) + \mathcal{B}(\hat{\theta}))$$

where *vech*(.) is the *vector-half* operator, which stacks the lower triangular elements of a matrix, and

$$\mathcal{C}(\hat{\theta}) = \frac{1}{T-1}\sum_{t=2}^{T}\left[\left(\frac{\partial^2 \ell_t(\hat{\theta})}{\partial\theta\partial\theta'}\right) + \left(\frac{\partial \ell_t(\hat{\theta})}{\partial\theta}\right)\left(\frac{\partial \ell_t(\hat{\theta})}{\partial\theta}\right)'\right] = \mathcal{A}(\hat{\theta}) + \mathcal{B}(\hat{\theta})$$

A consistent estimator of the variance matrix of $\sqrt{T}d(\hat{\theta})$ is given in White (1982, p.11) as

$$(8) \qquad \mathcal{V}(\hat{\theta}) = \frac{1}{T-1}\sum_{t=2}^{T} a_t(\hat{\theta})a_t(\hat{\theta})'$$

where

$$a_t(\hat{\theta}) = d_t(\hat{\theta}) - \nabla d(\hat{\theta})\mathcal{A}(\hat{\theta})^{-1}\nabla\ell_t(\hat{\theta})$$

$$\nabla d(\hat{\theta}) = \frac{1}{T-1}\sum_{t=2}^{T}\frac{\partial d_t(\hat{\theta})}{\partial \theta}$$

$$\nabla \ell_t(\hat{\theta}) = \frac{\partial \ell_t(\hat{\theta})}{\partial \theta}$$

Then White's IM test takes the form

(9) $$\mathcal{T}_w = Td'(\hat{\theta})V(\hat{\theta})^{-1}d(\hat{\theta})$$

Under the null hypothesis that model (1) is correctly specified, $\mathcal{T}_w$ is asymptotically distributed as χ^2 with $\frac{k(k+1)}{2}$ degrees of freedom. It is important to note that White's IM test holds under fairly general conditions. For our case, the mixing conditions stated in White (1987) are satisfied and hence our derivation of the IM test here remains valid.

Following the notation of Bera and Lee (1993), after some tedious algebra, collecting and rearranging terms in $d(\hat{\theta})$, we can write (see Appendices A and B for derivations)

(10) $$\hat{d} = (\hat{d}_1', \hat{d}_2', \hat{d}_3', \hat{d}_4', \hat{d}_5', \hat{d}_6')'$$

where we suppress θ for notational simplicity and write $\hat{d}$ for $d(\hat{\theta})$. The element $\hat{d}_1$ is a $\frac{p(p+1)}{2} \times 1$ vector of the difference of two estimates of the variance of $\hat{\phi}$, $\hat{d}_2$ is a

$\frac{q(q+1)}{2} \times 1$ vector of the difference of two estimates of the variance of $\hat{\gamma}$, $\hat{d}_3$ is a scalar difference of two estimates of the variance of $\hat{\sigma}^2$, $\hat{d}_4$ is a $pq \times 1$ vector of the difference of two estimates of the covariance between $\hat{\phi}$ and $\hat{\gamma}$, $\hat{d}_5$ is a $p\times 1$ vector of the difference of two estimates of the covariance between $\hat{\phi}$ and $\hat{\sigma}^2$ and $\hat{d}_6$ is a $q\times 1$ vector of the difference of two estimates of the covariance between $\hat{\gamma}$ and $\hat{\sigma}^2$. The typical elements of $\hat{d}_i$, i = 1,2,...,6 are given below

$$\hat{d}_1: \quad \left[\frac{1}{(T-1)\hat{\sigma}^4} \sum_{t=2}^{T} (\hat{\varepsilon}_t^2 - \hat{\sigma}^2) \left(\frac{\partial \hat{\varepsilon}_t}{\partial \phi_i}\right) \left(\frac{\partial \hat{\varepsilon}_t}{\partial \phi_j}\right)' \right]_{\substack{i,j=1,\dots,p \\ i\le j}}$$

$$\hat{d}_2: \quad \left[\frac{1}{(T-1)\hat{\sigma}^4} \sum_{t=2}^{T} (\hat{\varepsilon}_t^2 - \hat{\sigma}^2) \left(\frac{\partial \hat{\varepsilon}_t}{\partial \gamma_i}\right) \left(\frac{\partial \hat{\varepsilon}_t}{\partial \gamma_j}\right)' - \frac{1}{(T-1)\hat{\sigma}^2} \sum_{t=2}^{T} \hat{\varepsilon}_t \frac{\partial^2 \hat{\varepsilon}}{\partial \gamma_i \partial \gamma_j} \right]_{\substack{i,j=1,\dots,p \\ i\le j}}$$

$$\hat{d}_3: \quad \left[\frac{1}{4(T-1)\hat{\sigma}^8} \sum_{t=2}^{T} (\hat{\varepsilon}_t^4 - 6\hat{\sigma}^2\hat{\varepsilon}_t^2 + 3\hat{\sigma}^4) \right] = \left[\frac{1}{4(T-1)\hat{\sigma}^8} \sum_{t=2}^{T} (\hat{\varepsilon}_t^4 - 3\hat{\sigma}^4) \right]$$

$$\hat{d}_4: \quad \left[\frac{1}{(T-1)\hat{\sigma}^4} \sum_{t=2}^{T} (\hat{\varepsilon}_t^2 - \hat{\sigma}^2) \left(\frac{\partial \hat{\varepsilon}_t}{\partial \gamma_i}\right) \left(\frac{\partial \hat{\varepsilon}_t}{\partial \phi_j}\right)' - \frac{1}{(T-1)\hat{\sigma}^2} \sum_{t=2}^{T} \hat{\varepsilon}_t \frac{\partial^2 \hat{\varepsilon}_t}{\partial \gamma_i \partial \phi_j} \right]_{\substack{i=1,\ldots,q \\ j=1,\ldots,p}}$$

$$\hat{d}_5: \quad \left[\frac{1}{2(T-1)\hat{\sigma}^6} \sum_{t=2}^{T} \frac{\partial \hat{\varepsilon}_t}{\partial \phi_j} (3\hat{\sigma}^2 \hat{\varepsilon}_t - \hat{\varepsilon}_t^3) \right]_{j = 1,\ldots,p}$$

$$\hat{d}_6: \quad \left[\frac{1}{2(T-1)\hat{\sigma}^6} \sum_{t=2}^{T} \frac{\partial \hat{\varepsilon}_t}{\partial \gamma_i} (3\hat{\sigma}^2 \hat{\varepsilon}_t - \hat{\varepsilon}_t^3) \right]_{i = 1,\ldots,q}$$

where $\dfrac{\partial \hat{\varepsilon}_t}{\partial \phi_j} = -y_{t-j} - \dfrac{\partial \hat{\varepsilon}_{t-j}}{\partial \phi_j}$, and $\dfrac{\partial \hat{\varepsilon}_t}{\partial \gamma_i} = -\hat{\varepsilon}_{t-i} - \dfrac{\partial \hat{\varepsilon}_{t-i}}{\partial \gamma_i}$.

Note that our expressions for $\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_6$ are identical to those of Bera and Lee (1993) if we put $\hat{\gamma} = 0$ in our model and $\beta = 0$ in their model. As White (1982) and Hall (1987) pointed out, if we are only interested in testing a certain direction, we can premultiply $\hat{d}$ by a selection matrix whose elements are either zero or unity.

To obtain the IM test we need the asymptotic variance matrix of $\hat{d}$. Unfortunately, unlike in Hall (1987) and Bera and Lee (1993) the variance matrix of $\hat{d}$ is *not* block diagonal and hence the exact expression for the variance of $\hat{d}$ is extremely complicated (see Appendix B for derivation).

Therefore, the derived IM test for the specification of (1) can not be written as the sum of quadratic terms.

Alternatively, as in Bera and Zuo (1993), we can use a result from Pierce (1982) to obtain the asymptotic variance of the components that are of interest to us, namely, $\hat{d}_1$, $\hat{d}_2$ and $\hat{d}_4$. The indicator vector $\hat{d}_1$ is related solely to the AR(p) parameter vector ϕ. Likewise, the indicator vector $\hat{d}_2$ is related solely to the MA(q) parameter vector γ. We will discuss these two components here and leave the remaining to the next section. But first, we briefly describe Pierce's result.

Suppose there is a sequence of random variables, $y_1, y_2, \ldots, y_n$ whose joint distribution depends on a parameter θ. Let $\mathcal{T}_n = \mathcal{T}_n(y_1, y_2, \ldots, y_n, \theta)$ be a sequence of statistics which has a known limiting normal distribution, and suppose that we wish to determine the limiting distribution of $\hat{\mathcal{T}}_n = \mathcal{T}_n(y_1, y_2, \ldots, y_n, \hat{\theta}_n)$ where $\hat{\theta}_n = \hat{\theta}_n(y_1, y_2, \ldots, y_n)$ is an asymptotically normal and efficient sequence of estimators. Assume that for every θ, we have the following joint convergence in distribution

$$\begin{bmatrix} \sqrt{n}\mathcal{T}_n \\ \sqrt{n}(\hat{\theta} - \theta) \end{bmatrix} \sim \mathcal{N} \left[0, \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \right]$$

and that there exists a matrix $W = \lim E\left[\frac{\partial \mathcal{T}_n}{\partial \theta}\right]$, such that

$$\sqrt{n}\hat{\mathcal{T}}_n = \sqrt{n}\mathcal{T}_n + W\sqrt{n}(\hat{\theta} - \theta) + o_p(1)$$

Under the above assumptions, Pierce (1982) showed that

$$\mathrm{Cov}\left[\sqrt{n}\mathcal{T}_n + W\sqrt{n}(\hat{\theta} - \theta),\ \sqrt{n}(\hat{\theta} - \theta)\right] = 0$$

and thus,

$$\sqrt{n}\hat{\mathcal{T}}_n \sim \mathcal{N}(0,\ V_{11} - WV_{22}W')$$

Using the above results, an estimate of the variance of $\hat{d}_1$ is (the derivation is given in Appendix C2, Part I)

$$V(\hat{d}_1) = \hat{V}_1 = \frac{2}{(T-1)\hat{\sigma}^4}\sum_{t=2}^{T}\left(\frac{\partial \hat{\varepsilon}_t}{\partial \phi_i}\right)^2\left(\frac{\partial \hat{\varepsilon}_t}{\partial \phi_j}\right)^2 + \nabla\hat{d}_{13}\hat{\mathcal{A}}^{33}\nabla\hat{d}_{13}'$$

$$i,j = 1,\ldots,p;\ i \le j$$

where $\nabla\hat{d}_{13} = -\frac{1}{\hat{\sigma}^4}\lim_{T\to\infty}\frac{1}{T-1}\sum_{t=2}^{T}\left(\frac{\partial \hat{\varepsilon}_t}{\partial \phi_i}\right)\left(\frac{\partial \hat{\varepsilon}_t}{\partial \phi_j}\right)$, $i,j = 1,\ldots,p;\ i \le j$, and $\hat{\mathcal{A}}^{33}$ is the lower right-hand corner block of $\hat{\mathcal{A}}^{-1}$. Thus, the test statistic can be written as

$$\mathcal{T}_1 = \hat{d}_1'\hat{V}_1^{-1}\hat{d}_1$$

which has an asymptotic χ^2 distribution with $\frac{p(p+1)}{2}$ degrees of freedom. It is important to recognize that the test, $\mathcal{T}_1$, is asymptotically equivalent to $(T-1)R^2$ of an ordinary least squares (OLS) regression of $\hat{u}$ on $\hat{Z}_1$, i.e.

$$\mathcal{T}_1 = \frac{1}{2}\,\hat{u}'\hat{Z}_1(\hat{Z}_1'\hat{Z}_1)^{-1}\hat{Z}_1'\hat{u}$$

where $\hat{u} = (\hat{u}_2,\ldots,\hat{u}_T)'$ is a $(T-1)\times 1$ vector with $\hat{u}_t = \left[\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2} - 1\right]$ and

$$(11) \qquad \hat{Z}_1 = \begin{bmatrix} \frac{\partial\hat{\varepsilon}_2}{\partial\phi_1}\frac{\partial\hat{\varepsilon}_2}{\partial\phi_1} & \frac{\partial\hat{\varepsilon}_2}{\partial\phi_1}\frac{\partial\hat{\varepsilon}_2}{\partial\phi_2} & \cdots & \frac{\partial\hat{\varepsilon}_2}{\partial\phi_1}\frac{\partial\hat{\varepsilon}_2}{\partial\phi_p} & \cdots & \frac{\partial\hat{\varepsilon}_2}{\partial\phi_p}\frac{\partial\hat{\varepsilon}_2}{\partial\phi_p} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \frac{\partial\hat{\varepsilon}_T}{\partial\phi_1}\frac{\partial\hat{\varepsilon}_T}{\partial\phi_1} & \frac{\partial\hat{\varepsilon}_T}{\partial\phi_1}\frac{\partial\hat{\varepsilon}_T}{\partial\phi_2} & \cdots & \frac{\partial\hat{\varepsilon}_T}{\partial\partial\phi_1}\frac{\partial\hat{\varepsilon}_T}{\partial\phi_p} & \cdots & \frac{\partial\hat{\varepsilon}_T}{\partial\phi_p}\frac{\partial\hat{\varepsilon}_T}{\partial\phi_p} \end{bmatrix}$$

is a $(T-1) \times \frac{p(p+1)}{2}$ matrix of derivatives of $\hat{\varepsilon}$ with respect to the AR parameter ϕ (for detail derivation see Appendix C2, Part II).

Similarly, we can derive the test statistic $\mathcal{T}_2$ as

$$\mathcal{T}_2 = \hat{d}_2' \hat{V}_2^{-1} \hat{d}_2$$

which is asymptotically distributed as χ^2 with $\frac{q(q+1)}{2}$ degrees of freedom. Furthermore, $\mathcal{T}_2$ can also be obtained by using $(T-1)R^2$ from the auxiliary regression of $\hat{u}$ on $\hat{Z}_2$ where $\hat{u}$ defined as before and

$$(12) \qquad \hat{Z}_2 = \begin{bmatrix} \frac{\partial\hat{\varepsilon}_2}{\partial\gamma_1}\frac{\partial\hat{\varepsilon}_2}{\partial\gamma_1} & \frac{\partial\hat{\varepsilon}_2}{\partial\gamma_1}\frac{\partial\hat{\varepsilon}_2}{\partial\gamma_2} & \cdots & \frac{\partial\hat{\varepsilon}_2}{\partial\gamma_1}\frac{\partial\hat{\varepsilon}_2}{\partial\gamma_2} & \cdots & \frac{\partial\hat{\varepsilon}_2}{\partial\gamma_q}\frac{\partial\hat{\varepsilon}_2}{\partial\gamma_q} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \frac{\partial\hat{\varepsilon}_T}{\partial\gamma_1}\frac{\partial\hat{\varepsilon}_T}{\partial\gamma_1} & \frac{\partial\hat{\varepsilon}_T}{\partial\gamma_1}\frac{\partial\hat{\varepsilon}_T}{\partial\gamma_2} & \cdots & \frac{\partial\hat{\varepsilon}_T}{\partial\gamma_1}\frac{\partial\hat{\varepsilon}_T}{\partial\gamma_q} & \cdots & \frac{\partial\hat{\varepsilon}_T}{\partial\gamma_q}\frac{\partial\hat{\varepsilon}_T}{\partial\gamma_q} \end{bmatrix}$$

is a $(T-1) \times \frac{q(q+1)}{2}$ matrix of derivatives of $\hat{\varepsilon}$ with respect to the MA parameter vector γ. It is important to recognize that we have omitted the second term in $\hat{d}_2$ since under the normality assumption of ε_t, this term is asymptotically zero.

III. INTERPRETATION OF THE INFORMATION MATRIX TEST

Given the analysis of Chesher (1984), Hall (1987), Bera and Lee (1993), and Bera and Zuo (1993), we are ready to give the interpretation of each of the components of the indicator vector $\hat{d}$. As mentioned in the previous section, the component $\hat{d}_1$ is related solely to the AR parameter vector, so the test statistic $\mathcal{T}_1$ is a test for randomness of the AR parameter, $\phi = (\phi_1, \phi_2, \ldots, \phi_p)'$. Suppose that the AR parameters, ϕ are fluctuating around a mean with finite variance. This can be formulated as $\phi_t \sim (\phi, \Omega)$ where $\phi_t = (\phi_{1t}, \phi_{2t}, \ldots, \phi_{pt})'$. Then $\mathcal{T}_1$ is the LM test for testing the hypothesis H_0: $\Omega = 0$. Note that if we put $\gamma = 0$ in our model, the test statistic $\mathcal{T}_1$ is exactly identical to the test statistic T_2 in the Bera and Lee (1993) model (with $\beta = 0$).

Like $\hat{d}_1$, the component $\hat{d}_2$ is related solely to the MA parameter vector and hence the statistic $\mathcal{T}_2$ is a test for randomness of the MA parameters, $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_q)'$. That is, if $\gamma_t \sim (\gamma_t, \Sigma)$ where $\gamma_t = (\gamma_{1t}, \gamma_{2t}, \ldots, \gamma_{qt})'$ and Σ is finite, then $\mathcal{T}_2$ is the LM test for testing H_0: $\Sigma = 0$.
Next, in close connection with $\hat{d}_1$ and $\hat{d}_2$, the component $\hat{d}_4$ is based on the correlation between the AR and the MA parameter vectors. Like $\hat{d}_2$, under the normality assumption, $E(\hat{\varepsilon}_t^3) = E(\hat{\varepsilon}_t) = 0$ and this allows us to omit the second term

in $\hat{d}_4$. Then, $\hat{d}_4$ simplifies to $\tilde{d}_4 = \frac{1}{(T-1)\hat{\sigma}^4} \sum_{t=2}^{T} (\hat{\varepsilon}_t^2 - \hat{\sigma}^2)\left(\frac{\partial\hat{\varepsilon}_t}{\partial\gamma_i}\right)\left(\frac{\partial\hat{\varepsilon}_t}{\partial\phi_j}\right)$, i = 1,...,q; j = 1,...,p. This can be easily seen to be a description of the relationship between $\hat{u}$ and $\hat{Z}_{12}$ where the elements of the matrix $\hat{Z}_{12}$ are the cross products between the first derivatives of $\hat{\varepsilon}_t$ with respect to AR parameter vector, ϕ, and the first derivatives of $\hat{\varepsilon}_t$ with respect to the MA parameter vector, γ. Thus, this allows us to test for randomness of the ARMA parameters.

Finally, it is worth pointing out that each of the three components, $\hat{d}_1$, $\hat{d}_2$ and $\hat{d}_4$, contain the term $\hat{u}_t = \left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2} - 1\right)$. For obtaining $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_4$ we run the regressions of $\hat{u}_t$ on $\hat{Z}_{1t}$, $\hat{u}_t$ on $\hat{Z}_{2t}$ and $\hat{u}_t$ on $\hat{Z}_{12t}$, where $\hat{Z}_{1t}$, $\hat{Z}_{2t}$, and $\hat{Z}_{12t}$ are respectively the tth row of $\hat{Z}_1$, $\hat{Z}_2$, and $\hat{Z}_{12}$. Moreover, as byproduct of our analysis we have a simple test for heteroskedaticity in the presence of ARMA. To see this, simply regress $\hat{u}_t$ on $\hat{Z}_{1t}$, $\hat{Z}_{2t}$ and $\hat{Z}_{12t}$ we would get some sort of White's (1980) test for (static) heteroskedasticity. As for the remaining components, $\hat{d}_3$ is exactly the same as in Bera and Lee (1993) and hence can be interpreted as a test for variation in σ^2, i.e. it is a pure test for kurtosis. This can be easily calculated by taking $(T-1)R^2$ from the auxiliary regression of $\hat{v}_t = (\hat{\varepsilon}_t^4 - 3\hat{\sigma}^4)$ on a constant. The components $\hat{d}_5$ and $\hat{d}_6$ each involve the third sample moment of $\hat{\varepsilon}_t$ and thus can be given an interpretation as a test for skewness. In Bera and Lee (1993) jargon, $\hat{d}_5$ and $\hat{d}_6$ are tests for *heteroclíticity*.

Since the covariance matrix $\hat{V}$ is not block diagonal, all these tests are definitely correlated with each other and hence to get the overall test of the model it is necessary to have a joint specification test. This can be obtained using the results derived in Appendix B2.

IV. CONCLUSION

In this paper, we have presented an application of White's IM test to a pure stationary ARMA model. We provide the computation and interpretation of the resulting test. Due to the complex structure of the ARMA process, the framework of the information matrix test is much more complicated. Because of the non-block diagonal nature of the information matrix, the estimated variance of the indicator vector is also not a block diagonal matrix. This implies that the components of the indicator vector are no longer asymptotically independent. Consequently, to get the overall test for correct specification of the model, the joint test is needed.

Finally, it is worth pointing out that although there are many applications of the IM test, very few have examined its size and power, see for example Taylor (1987), Orme (1990), Chesher and Spady (1991), Davidson and McKinnon (1992) and recently Horowitz (1994). Thus it would be useful to examine these issues, specifically, the power of the IM test, since, as Horowitz (1994) points out "...obtaining high power and getting the finite-sample size right are different objectives and that achieving one does not insure that the other is also achieved."

REFERENCES

Bera, A. K., and Lee, S. (1993), "Information Matrix Test, Parameter Heterogeneity and ARCH: A Synthesis," *Review of Economic Studies*,

Bera, A. K., and Zuo, X-L. (1993), "Specification Test for a Linear Regression Model with ARCH process," *BEBR Working Paper* #93-0112, College of Commerce and Business Administration, University of Illinois at Urbana-Champaign.

Chesher, A. (1983), "The Information Matrix Test: Simplified Calculation via a Score Test Interpretation," *Economics Letters*, 13, 45-48.

Chesher, A. (1984), "Testing Neglected Heterogeneity," *Econometrica*, 52, 865-872.

Chesher, A., and Irish, M. (1987), "Residual Analysis in the Grouped and Censor Normal Linear Model," *Journal of Econometrics*, 34, 33-62.

Chesher, A. and R. Spady (1991), "Asymptotic Expansions of the Information Matrix Test Statistic," *Econometrica*, 59, 787-815.

Davidson, R., and MacKinnon (1992), "A New Form of the Information Matrix Test," *Econometrica*, 60, 145-157.

Hall, A. (1987), "Information Matrix Test for the Linear Model," *Review of Economic Studies*, 54, 257-263.

Harvey, A. C. (1993), *Time Series Models*, Second Edition, The MIT Press, Cambridge, Massachusetts, Chapter 3.

Horowitz, J. L. (1994), "Bootstrap-Based Critical Values for the Information Matrix Test," *Journal of Econometrics*, 61, 395-411.

Lancaster, T. (1984), "The Covariance Matrix of the Information Matrix Test," *Econometrica*, 52, 1051-1053.

Orme, C. (1990), "The Small-Sample Performance of the Information Matrix Test," *Journal of Econometrics*, 46, 309-331.

Pierce, D. A. (1982), "The Asymptotic Effect of Substituting Estimators for Pararerters in Certain Types of Statistics," *The Annals of Statistics*, 10, 475-478.

Santos Silva, J. M. C. (1993), "A Note on the Score Test for Neglected Heterogeneity in the Truncated Normal Regression Model," *Economics Letters*, 43, 11-14.

Taylor, L. W. (1987), "The Size Bias of White's Information Matrix Test," *Economics Letters*, 24, 63-67.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.

White, H. (1987), "Specification Testing in Dynamics Models," in *Advances in Econometrics:* Fifth World Congress, Volume I, edited by T. F. Bewley, Cambridge: Cambridge University Press, 1-58.

APPENDIX A2

The Derivatives of the Log-likelihood Function

For parameter vector $\theta = (\phi', \gamma', \sigma^2)'$ and an information set $\mathcal{F}_{t-1}$, the conditional log-likelihood function for the t-th observation is given by

$$\ell_t(\theta) = \frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \varepsilon_t^2$$

The first derivatives are

$$\frac{\partial \ell_t(\theta)}{\partial \phi} = - \frac{1}{\sigma^2} \varepsilon_t \frac{\partial \varepsilon_t}{\partial \phi} = \frac{1}{\sigma^2} \varepsilon_t (\underline{y}'_{t-1} + \frac{\partial \underline{\varepsilon}_{t-1}}{\partial \phi})$$

$$\frac{\partial \ell_t(\theta)}{\partial \gamma} = - \frac{1}{\sigma^2} \varepsilon_t \frac{\partial \varepsilon_t}{\partial \gamma} = \frac{1}{\sigma^2} \varepsilon_t (\underline{\varepsilon}'_{t-1} + \gamma \frac{\partial \underline{\varepsilon}_{t-1}}{\partial \gamma})$$

$$\frac{\partial \ell_t(\theta)}{\partial \sigma^2} = \cdot \frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} \varepsilon_t^2$$

and the second order derivatives are

$$\frac{\partial^2 \ell_t(\theta)}{\partial\phi\partial\phi'} = -\frac{1}{\sigma^2}\left[\frac{\partial\varepsilon_t}{\partial\phi}\right]\left[\frac{\partial\varepsilon_t}{\partial\phi}\right]'$$

$$\frac{\partial^2 \ell_t(\theta)}{\partial\gamma\partial\gamma'} = -\frac{1}{\sigma^2}\left[\frac{\partial\varepsilon_t}{\partial\gamma}\right]\left[\frac{\partial\varepsilon_t}{\partial\gamma}\right]' - \frac{1}{\sigma^2}\varepsilon_t\frac{\partial^2\varepsilon_t}{\partial\gamma\partial\gamma'}$$

$$\frac{\partial^2 \ell_t(\theta)}{\partial(\sigma^2)^2} = \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}\varepsilon_t^2$$

$$\frac{\partial^2 \ell_t(\theta)}{\partial\phi\partial\gamma'} = -\frac{1}{\sigma^2}\left[\left[\frac{\partial\varepsilon_t}{\partial\phi}\right]\left[\frac{\partial\varepsilon_t}{\partial\gamma}\right]' + \varepsilon_t\frac{\partial^2\varepsilon_t}{\partial\phi\partial\gamma'}\right]$$

$$\frac{\partial^2 \ell_t(\theta)}{\partial\phi\partial\sigma^2} = \frac{1}{\sigma^4}\varepsilon_t\frac{\partial\varepsilon_t}{\partial\phi}$$

$$\frac{\partial^2 \ell_t(\theta)}{\partial\gamma\partial\sigma^2} = \frac{1}{\sigma^4}\varepsilon_t\frac{\partial\varepsilon_t}{\partial\gamma}$$

The outer product of the gradients (OPG) are

$$\frac{\partial\ell_t}{\partial\phi}\frac{\partial\ell_t'}{\partial\phi} = \frac{1}{\sigma^4}\varepsilon_t^2\left[\frac{\partial\varepsilon_t}{\partial\phi}\right]\left[\frac{\partial\varepsilon_t}{\partial\phi}\right]'$$

$$\frac{\partial \ell_t}{\partial \gamma} \frac{\partial \ell_t}{\partial \gamma}' = \frac{1}{\sigma^4} \varepsilon_t^2 \left[\frac{\partial \varepsilon_t}{\partial \gamma}\right] \left[\frac{\partial \varepsilon_t}{\partial \gamma}\right]'$$

$$\frac{\partial \ell_t}{\partial \sigma^2} \frac{\partial \ell_t}{\partial \sigma^2} = \frac{1}{4\sigma^8} (\varepsilon_t^4 - 2\sigma^2\varepsilon_t^2 + \sigma^4)$$

$$\frac{\partial \ell_t}{\partial \phi} \frac{\partial \ell_t}{\partial \gamma}' = \frac{1}{\sigma^4} \varepsilon_t^2 \left[\frac{\partial \varepsilon_t}{\partial \phi}\right] \left[\frac{\partial \varepsilon_t}{\partial \gamma}\right]'$$

$$\frac{\partial \ell_t}{\partial \phi} \frac{\partial \ell_t}{\partial \sigma^2}' = \frac{1}{2\sigma^4} \varepsilon_t \frac{\partial \varepsilon_t}{\partial \phi} (1 - \frac{1}{\sigma^2} \varepsilon_t^2)$$

$$\frac{\partial \ell_t}{\partial \gamma} \frac{\partial \ell_t}{\partial \sigma^2}' = \frac{1}{2\sigma^4} \varepsilon_t \frac{\partial \varepsilon_t}{\partial \gamma} (1 - \frac{1}{\sigma^2} \varepsilon_t^2)$$

APPENDIX B2

Covariance Matrix for the Information Matrix Test

A consistent estimator of the covariance matrix for the IM test is given in White (1987) and can be written as

$$V(\hat{\theta}) = \frac{1}{(T-1)} \sum_{t=2}^{T} a_t(\hat{\theta}) a_t(\hat{\theta})' \tag{B.1}$$

where $a_t(\hat{\theta}) = d_t(\hat{\theta}) - \nabla d(\hat{\theta}) \mathcal{A}(\hat{\theta})^{-1} \nabla \ell_t(\hat{\theta})$. Let us start with the indicator vector $d(\hat{\theta})$ which is

$$d(\hat{\theta}) = vech\ \mathcal{C}(\hat{\theta}) = vech(\mathcal{A}(\hat{\theta}) + \mathcal{B}(\hat{\theta})) \tag{B.2}$$

where

$$\mathcal{A}(\hat{\theta}) = \frac{1}{(T-1)} \sum_{t=2}^{T} \left[\frac{\partial^2 \ell_t(\hat{\theta})}{\partial\theta\partial\theta'} \right] = \frac{1}{(T-1)} \sum_{t=2}^{T}$$

$$\begin{bmatrix} -\frac{1}{\sigma^2} z_{1t}z'_{1t} & -\frac{1}{\sigma^2}(z_{1t}z'_{2t} + \varepsilon_t\zeta_t) & \frac{1}{\sigma^4}\varepsilon_t z_{1t} \\ -\frac{1}{\sigma^2}(z_{2t}z'_{1t} + \varepsilon_t\zeta_t) & -\frac{1}{\sigma^2}(z_{2t}z'_{2t} + \varepsilon_t\nu_t) & \frac{1}{\sigma^4}\varepsilon_t z_{2t} \\ \frac{1}{\sigma^4}\varepsilon_t z_{1t} & \frac{1}{\sigma^4}\varepsilon_t z_{2t} & \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}\varepsilon_t^2 \end{bmatrix}_{\theta=\hat{\theta}}$$

and $\mathcal{B}(\hat{\theta}) = \frac{1}{(T-1)} \sum_{t=2}^{T} \left(\frac{\partial \ell_t(\hat{\theta})}{\partial \theta}\right)\left(\frac{\partial \ell_t(\hat{\theta})}{\partial \theta}\right)' = \frac{1}{(T-1)} \sum_{t=2}^{T}$

$$\begin{bmatrix} \frac{1}{\sigma^4}\varepsilon_t^2 z_{1t}z'_{1t} & \frac{1}{\sigma^4}\varepsilon_t^2 z_{1t}z'_{2t} & \frac{1}{2\sigma^6}\varepsilon_t z_{1t}(\sigma^2-\varepsilon_t^2) \\ \frac{1}{\sigma^4}\varepsilon_t^2 z_{2t}z'_{1t} & \frac{1}{\sigma^4}\varepsilon_t^2 z_{2t}z'_{2t} & \frac{1}{2\sigma^6}\varepsilon_t z_{2t}(\sigma^2-\varepsilon_t^2) \\ \frac{1}{2\sigma^6}\varepsilon_t z_{1t}(\sigma^2-\varepsilon_t^2) & \frac{1}{2\sigma^6}\varepsilon_t z_{2t}(\sigma^2-\varepsilon_t^2) & \frac{1}{4\sigma^8}(\varepsilon_t^4 - 2\varepsilon_t^2\sigma^2 + \sigma^4) \end{bmatrix}_{\theta=\hat{\theta}}$$

where

$$z_{1t} = \frac{\partial \varepsilon_t}{\partial \phi}, \quad z_{2t} = \frac{\partial \varepsilon_t}{\partial \gamma}, \quad \zeta_t = \frac{\partial^2 \varepsilon_t}{\partial \phi \partial \gamma}, \quad \nu_t = \frac{\partial^2 \varepsilon_t}{\partial \gamma \partial \gamma'}.$$

From $\mathcal{A}(\hat{\theta})$ and $\mathcal{B}(\hat{\theta})$ we can obtain $\mathcal{C}(\hat{\theta})$ as

$$\mathcal{C}(\hat{\theta}) = \mathcal{A}(\hat{\theta}) + \mathcal{B}(\hat{\theta}) = \frac{1}{(T-1)} \sum_{t=2}^{T}$$

$$\begin{bmatrix} -\frac{1}{\sigma^4}(\varepsilon_t^2-\sigma^2) z_{1t} z_{1t}' & -\frac{1}{\sigma^4}(\varepsilon_t^2-\sigma^2) z_{1t} z_{2t}' - \frac{1}{\sigma^2}\varepsilon_t \zeta_t & \frac{1}{2\sigma^6} \varepsilon_t z_{1t}(3\sigma^2-\varepsilon_t^2) \\ -\frac{1}{\sigma^4}(\varepsilon_t^2-\sigma^2) z_{2t} z_{1t}' - \frac{1}{\sigma^2}\varepsilon_t \zeta_t & -\frac{1}{\sigma^4}(\varepsilon_t^2-\sigma^2) z_{2t} z_{2t}' - \frac{1}{\sigma^2}\varepsilon_t \nu_t & \frac{1}{2\sigma^6}\varepsilon_t z_{2t}(3\sigma^2-\varepsilon_t^2) \\ \frac{1}{2\sigma^6}\varepsilon_t z_{1t}(3\sigma^2-\varepsilon_t^2) & \frac{1}{2\sigma^6}\varepsilon_t z_{2t}(3\sigma^2-\varepsilon_t^2) & \frac{1}{4\sigma^8}(\varepsilon_t^4-6\varepsilon_t^2\sigma^2+3\sigma^4) \end{bmatrix}_{\theta=\hat{\theta}}$$

Therefore, $d(\hat{\theta})$ is given by

$$d(\hat{\theta}) = \frac{1}{(T-1)} \sum_{t=2}^{T} d_t(\hat{\theta})$$

where

$$d_t(\hat{\theta}) = (\hat{d}_{t2}', \hat{d}_{t3}', \ldots, \hat{d}_{t6}')'$$

and

$$\hat{d}_{t1} = \Big[\hat{\sigma}^{-4}(\hat{\varepsilon}_t^2 - \hat{\sigma}^2)\frac{\partial\hat{\varepsilon}_t}{\partial\phi_1}\frac{\partial\hat{\varepsilon}_t}{\partial\phi_1}, \ldots, \hat{\sigma}^{-4}(\hat{\varepsilon}_t^2 - \hat{\sigma}^2)\frac{\partial\hat{\varepsilon}_t}{\partial\phi_p}\frac{\partial\hat{\varepsilon}_t}{\partial\phi_p},$$

$$\hat{\sigma}^{-4}(\hat{\varepsilon}_t^2 - \hat{\sigma}^2)\frac{\partial\hat{\varepsilon}_t}{\partial\phi_1}\frac{\partial\hat{\varepsilon}_t}{\partial\phi_2}, \ldots, \hat{\sigma}^{-4}(\hat{\varepsilon}_t^2 - \hat{\sigma}^2)\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{p-1}}\frac{\partial\hat{\varepsilon}_t}{\partial\phi_p} \Big]$$

is a $\frac{p(p+1)}{2} \times 1$ vector,

$$\hat{d}_{t2} = \Big[\ \hat{\sigma}^{-4}(\hat{\varepsilon}_t^2 - \hat{\sigma}^2)\frac{\partial\varepsilon_t}{\partial\gamma_1}\ \frac{\partial\hat{\varepsilon}_t}{\partial\gamma_1} - \hat{\sigma}^{-2}\hat{\varepsilon}_t\ \frac{\partial^2\hat{\varepsilon}_t}{\partial\gamma_1\partial\gamma_1}\ \ldots,$$

$$\hat{\sigma}^{-4}(\hat{\varepsilon}_t^2 - \hat{\sigma}^2)\frac{\partial\hat{\varepsilon}_t}{\partial\gamma_p}\ \frac{\partial\hat{\varepsilon}_t}{\partial\gamma_p} - \hat{\sigma}^{-2}\hat{\varepsilon}_t\ \frac{\partial^2\hat{\varepsilon}_t}{\partial\gamma_1\partial\gamma_q}\ \ldots,$$

$$\hat{\sigma}^{-4}(\hat{\varepsilon}_t^2 - \hat{\sigma}^2)\frac{\partial\hat{\varepsilon}_t}{\partial\gamma_{p-1}}\ \frac{\partial\hat{\varepsilon}_t}{\partial\gamma_p} - \hat{\sigma}^{-2}\hat{\varepsilon}_t\ \frac{\partial^2\hat{\varepsilon}_t}{\partial\gamma_{q-1}\partial\gamma_q}\ \Big]$$

is a $\frac{q(q+1)}{2} \times 1$ vector,

$$\hat{d}_{t3} = (4\hat{\sigma}^8)^{-1}(\hat{\varepsilon}_t^4 - 6\hat{\sigma}^2\varepsilon_t^2 + 3\hat{\sigma}^4)$$

is a scalar,

$$\hat{d}_{t4} = \Big[\ \hat{\sigma}^{-4}(\hat{\varepsilon}_t^2 - \hat{\sigma}^2)\frac{\partial\hat{\varepsilon}_t}{\partial\phi_1}\ \frac{\partial\hat{\varepsilon}_t}{\partial\gamma_1} - \hat{\sigma}^{-2}\hat{\varepsilon}_t\ \frac{\partial^2\hat{\varepsilon}_t}{\partial\phi_1\partial\gamma_1}\ ,\ldots,$$

$$\hat{\sigma}^{-4}(\hat{\varepsilon}_t^2 - \hat{\sigma}^2)\frac{\partial\hat{\varepsilon}_t}{\partial\phi_1}\ \frac{\partial\hat{\varepsilon}_t}{\partial\gamma_q} - \hat{\sigma}^{-2}\hat{\varepsilon}_t\ \frac{\partial^2\hat{\varepsilon}_t}{\partial\phi_1\partial\gamma_q}\ ,\ldots,$$

$$\hat{\sigma}^{-4}(\hat{\varepsilon}_t^2 - \hat{\sigma}^2)\frac{\partial\hat{\varepsilon}_t}{\partial\phi_p}\ \frac{\partial\hat{\varepsilon}_t}{\partial\gamma_q} - \hat{\sigma}^{-2}\hat{\varepsilon}_t\ \frac{\partial^2\hat{\varepsilon}_t}{\partial\phi_p\partial\gamma_q}\ \Big]$$

is a $pq \times 1$ vector,

$$\hat{d}_{t5} = \left[(2\hat{\sigma}^6)^{-1}(3\hat{\sigma}^2\hat{\varepsilon}_t - \hat{\varepsilon}_t^3) \frac{\partial \hat{\varepsilon}_t}{\partial \phi_1}, \ldots, (2\hat{\sigma}^6)^{-1}(3\hat{\sigma}^2\hat{\varepsilon}_t - \hat{\varepsilon}_t^3) \frac{\partial \hat{\varepsilon}_t}{\partial \phi_p} \right]$$

is a $p \times 1$ vector,

$$\hat{d}_{t6} = \left[(2\hat{\sigma}^6)^{-1}(3\hat{\sigma}^2\hat{\varepsilon}_t - \hat{\varepsilon}_t^3) \frac{\partial \hat{\varepsilon}_t}{\partial \gamma_1}, \ldots, (2\sigma^6)^{-1}(3\hat{\sigma}^2\hat{\varepsilon}_t - \hat{\varepsilon}_t^3) \frac{\partial \hat{\varepsilon}_t}{\partial \gamma_q} \right]$$

is a $q \times 1$ vector. Next we consider

$$\nabla d(\theta_0) = \lim_{T\to\infty} \frac{1}{T-1} \sum_{t=2}^{T} E\left[\frac{\partial d_t(\theta_0)}{\partial \theta}\right].$$

Using the normality assumption of ε_t and by taking the expectation conditional on the information set $\mathcal{F}_{t-1}$ iteratively, after some algebraic simplification we get the following simple form of $\nabla d(\theta_0)$

$$\nabla d(\theta_0) = \begin{bmatrix} 0 & 0 & \nabla d_{13} \\ 0 & 0 & \nabla d_{23} \\ 0 & 0 & 0 \\ \nabla d_{41} & \nabla d_{42} & \nabla d_{43} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where $\nabla d_{13} = (ar_{11},\ldots,ar_{pp},ar_{12},\ldots ar_{(p-1)p})'$ is a $\frac{p(p+1)}{2} \times 1$ vector with

$$ar_{ij} = -\frac{1}{\sigma^4}\lim_{T\to\infty}\frac{1}{T-1}\sum_{t=2}^{T}\left[\frac{\partial\varepsilon_t(\theta_0)}{\partial\phi_i}\right]\left[\frac{\partial\varepsilon_t(\theta_0)}{\partial\phi_j}\right],$$

$$i,j = 1,\ldots,p; i \le j.$$

$\nabla d_{23} = (ma_{11},\ldots,ma_{qq},\ldots,ma_{12},\ldots,ma_{(q-1)q})'$ is a $\frac{q(q+1)}{2} \times 1$ vector with

$$ma_{ij} = -\frac{1}{\sigma^4}\lim_{T\to\infty}\frac{1}{T-1}\sum_{t=2}^{T}\left[\frac{\partial\varepsilon_t(\theta_0)}{\partial\gamma_i}\right]\left[\frac{\partial\varepsilon_t(\theta_0)}{\partial\gamma_j}\right],$$

$$i,j = 1,\ldots,q; i \le j.$$

$\nabla d_{41} = (x_{11},x_{12},\ldots,x_{pq})'$ is a $pq \times p$ matrix with

$$x_{ij} = -\frac{1}{\sigma^2}\lim_{T\to\infty}\frac{1}{T-1}\sum_{t=2}^{T}\left[\frac{\partial\varepsilon_t(\theta_0)}{\partial\phi_j}\,\frac{\partial^2\varepsilon_t(\theta_0)}{\partial\gamma_i\partial\phi_j}\right], \qquad \begin{matrix} i = 1,\ldots,q \\ j = 1,\ldots,p \end{matrix}$$

$\nabla d_{42} = (w_{11},w_{12},\ldots,w_{qp})'$ is a $qp \times q$ matrix with

$$w_{ij} = -\frac{1}{\sigma^2}\lim_{T\to\infty}\frac{1}{T-1}\sum_{t=2}^{T}\left[\frac{\partial\varepsilon_t(\theta_0)}{\partial\gamma_i}\,\frac{\partial^2\varepsilon_t(\theta_0)}{\partial\gamma_i\partial\phi_j}\right], \qquad \begin{matrix} i = 1,\ldots,q \\ j = 1,\ldots,p \end{matrix}$$

$\nabla d_{43} = (z_{11}, z_{21}, \ldots, x_{pq})'$ is a $pq \times 1$ matrix with

$$z_{ij} = -\frac{1}{\sigma^2} \lim_{T\to\infty} \frac{1}{T-1} \sum_{t=2}^{T} \left[\frac{\partial \varepsilon_t(\theta_0)}{\partial \gamma_i} \frac{\partial \varepsilon_t(\theta_0)}{\partial \phi_j}\right], \qquad \begin{array}{l} i = 1,\ldots,q \\ j = 1,\ldots,p \end{array}$$

Thus, we can consistently estimate $\nabla d(\theta_0)$ by replacing θ_0 with its consistent estimate, $\hat{\theta}$ (MLE). Therefore

$$\nabla d(\theta_0) = \begin{bmatrix} 0 & 0 & \nabla\hat{d}_{13} \\ 0 & 0 & \nabla\hat{d}_{23} \\ 0 & 0 & 0 \\ \nabla\hat{d}_{41} & \nabla\hat{d}_{42} & \nabla\hat{d}_{43} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and

$$\mathcal{I}(\hat{\theta}) = \begin{bmatrix} -\frac{1}{(T-1)\hat{\sigma}^2}\sum_{t=2}^{T} \hat{z}_{1t}\hat{z}'_{1t} & -\frac{1}{(T-1)\hat{\sigma}^2}\sum_{t=2}^{T} (\hat{z}_{1t}\hat{z}'_{2t}+\hat{\varepsilon}_t\hat{\zeta}_t) & 0 \\ -\frac{1}{(T-1)\hat{\sigma}^2}\sum_{t=2}^{T} (\hat{z}_{1t}\hat{z}'_{2t}+\hat{\varepsilon}_t\hat{\zeta}_t) & -\frac{1}{(T-1)\hat{\sigma}^2}\sum_{t=2}^{T} (\hat{z}_{2t}\hat{z}'_{2t}+\hat{\varepsilon}_t\hat{\nu}_t) & 0 \\ 0 & 0 & \frac{1}{2\hat{\sigma}^4} \end{bmatrix}$$

where

$$\hat{z}_{1t} = \frac{\partial\hat{\varepsilon}_t}{\partial\phi}, \quad \hat{z}_{2t} = \frac{\partial\hat{\varepsilon}_t}{\partial\gamma}, \quad \hat{\zeta}_t = \frac{\partial^2\hat{\varepsilon}_t}{\partial\phi\partial\gamma}, \quad \hat{\nu}_t = \frac{\partial^2\hat{\varepsilon}_t}{\partial\gamma\partial\gamma'}, \quad \hat{\varepsilon}_t = \varepsilon_t(\hat{\theta})$$

Finally,
$$\nabla\ell_t(\hat{\theta}) = \frac{\partial\ell_t(\hat{\theta})}{\partial\theta}$$

$$\nabla\ell_t(\hat{\theta}) = \begin{bmatrix} -\frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t\hat{z}_{1t} \\ -\frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t\hat{z}_{2t} \\ -\frac{1}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\hat{\varepsilon}_t^2 \end{bmatrix}$$

Now, having obtained the expressions for $d(\hat{\theta})$, $\nabla d(\hat{\theta})$, $\nabla\ell_t(\hat{\theta})$, and $\mathcal{A}(\hat{\theta})$, using formula (B.1) we can obtain an estimate of the covariance matrix for the IM test. As mentioned in the main text, the information matrix, $\mathcal{A}(\hat{\theta})$, is *not* block diagonal, thus implying that $V(\hat{\theta})$ will also be non-diagonal. Consequently, the final expression for $V(\hat{\theta})$ is extremely complicated and will not be pursued here.

APPENDIX C2

Part I.

Recall,

$$d_1 = \frac{1}{(T-1)\sigma^4} \sum_{t=2}^{T} (\varepsilon_t^2 - \sigma^2) \left[\frac{\partial \varepsilon_t}{\partial \phi_i}\right] \left[\frac{\partial \varepsilon_t}{\partial \phi_j}\right] \qquad i,j = 1,\ldots,p;\ i \le j$$

By using

$$V(x) = E[V(x|y)] + V[E(x|y)]$$

we obtain

$$V(d_1) = V_1 = V_{11} + WV_{22}W'$$

$$= E\left\{ \frac{2}{(T-1)^2\sigma^4} \sum_{t=2}^{T} \left[\frac{\partial \varepsilon_t}{\partial \phi_i}\right]^2 \left[\frac{\partial \varepsilon_t}{\partial \phi_j}\right]^2 \right\} + \nabla d_{13} \mathcal{A}^{33} \nabla d_{13}'$$

$$i,j = 1,\ldots,p$$

where $\mathcal{A}^{33}$ denotes the lower right-hand corner block of the matrix $\mathcal{A}^{-1}$, and ∇d_{13} defined as before.

Hence

$$V(\hat{d}_1) = \hat{V}_1 = \frac{2}{(T-1)^2\hat{\sigma}^4} \sum_{t=2}^{T} \left[\frac{\partial \hat{\varepsilon}_t}{\partial \phi_i}\right]^2 \left[\frac{\partial \hat{\varepsilon}_t}{\partial \phi_j}\right]^2 + \nabla \hat{d}_{13} \hat{\mathcal{A}}^{33} \nabla \hat{d}_{13}'$$

$$i,j = 1,\ldots,p$$

finally, we have

$$\underset{T\to\infty}{plim}\; V(\hat{d}_1) = V(d_1)$$

$$\mathcal{T}_1 = \hat{d}_1'\hat{V}_1^{-1}\hat{d}_1 \sim \chi^2_\nu \qquad \text{where } \nu = \frac{p(p+1)}{2}$$

We can obtain $V(\hat{d}_2)$ and $V(\hat{d}_4)$ in the same way but this is omitted in the interest of saving space.

Also note that

$$\frac{\partial d_1}{\partial \phi} = \frac{1}{\sigma^4}\left[2\varepsilon_t \frac{\partial \varepsilon_t}{\partial \phi'} \frac{\partial \varepsilon_t}{\partial \phi} \frac{\partial \varepsilon_t}{\partial \phi'}\right]$$

$$\frac{\partial d_1}{\partial \gamma} = \frac{1}{\sigma^4}\left[2\varepsilon_t \frac{\partial \varepsilon_t}{\partial \gamma} \frac{\partial \varepsilon_t}{\partial \phi} \frac{\partial \varepsilon_t}{\partial \phi'} + 2(\varepsilon_t^2-\sigma^2) \frac{\partial^2 \varepsilon_t}{\partial \phi \partial \gamma} \frac{\partial \varepsilon_t}{\partial \phi'}\right]$$

$$\frac{\partial d_1}{\partial \sigma^2} = -\frac{1}{\sigma^6}(2\varepsilon_t^2-\sigma^2)\frac{\partial \varepsilon_t}{\partial \phi}\frac{\partial \varepsilon_t}{\partial \phi'}$$

and $E\left[\frac{\partial d_1}{\partial \phi}\right] = 0, \quad E\left[\frac{\partial d_1}{\partial \gamma}\right] = 0, \quad E\left[\frac{\partial d_1}{\partial \sigma^2}\right] = -\frac{1}{\sigma^4}\frac{\partial \varepsilon_t}{\partial \phi}\frac{\partial \varepsilon_t}{\partial \phi'}$

Part II.

The test statistics $\mathcal{T}_1$ and $\mathcal{T}_2$ which we derived in section 2 can be written as $(T-1)R_i^2$ by running the OLS regression $\hat{u}$ on $\hat{Z}_i$, i = 1,2.

$$\hat{d}_i = \frac{1}{(T-1)\hat{\sigma}^2} \hat{Z}_i' \hat{u}$$

where $\hat{u} = (\hat{u}_2, \ldots, \hat{u}_T)$ is $(T-1)\times 1$ vector with $\hat{u}_t = \left[\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2} - 1\right]$ and $\hat{Z}_1$, $\hat{Z}_2$ are respectively defined in (11) and (12) in the main text. From Part I we know that

$$\begin{aligned} V(\hat{d}_i) &= E[V(\hat{d}_i | \mathcal{F}_{t-1})] \\ &= \frac{1}{(T-1)\hat{\sigma}^2} \hat{Z}_i' E(\hat{u}\hat{u}')\hat{Z}_i \\ &= \frac{2}{(T-1)^2\hat{\sigma}^4} \hat{Z}_i'\hat{Z}_i \end{aligned}$$

$$\begin{aligned} \mathcal{T}_i &= \hat{d}_i' V(\hat{d}_i)^{-1}\hat{d}_i \\ &= \frac{1}{2} \hat{u}'\hat{Z}_i(\hat{Z}_i'\hat{Z}_i)^{-1}\hat{Z}_i'\hat{u} \end{aligned}$$

$$(T-1)R_i^2 = (T-1)\frac{\hat{u}'\hat{Z}_i(\hat{Z}_i'\hat{Z}_i)^{-1}\hat{Z}_i'\hat{u}}{\hat{u}'\hat{u}}$$

$$= \frac{1}{2}\,\hat{u}'\hat{Z}_i(\hat{Z}_i'\hat{Z}_i)^{-1}\hat{Z}_i'\hat{u}$$

since

$$\underset{T\longrightarrow\infty}{plim}\ \frac{\hat{u}'\hat{u}}{T} = 2.$$

Thus,

$$\mathcal{T}_i = (T-1)R_i^2, \qquad\qquad i = 1,2.$$

where R_i^2 is the uncentered coefficient of determination from regression $\hat{u}$ on $\hat{Z}_i$.

CHAPTER 3

ALTERNATIVE ESTIMATORS FOR ARCH MODELS: MONTE CARLO COMPARISON

I. INTRODUCTION

Recently, the concept of autoregressive conditional heteroskedasticity (ARCH) introduced by Engle (1982), has been popularised by the empirical work in financial economics associated with the modeling of time-varying variances. See for example, Bollerslev, Chou, and Kroner (1992) for a recent survey. The ARCH model specifies the conditional error variance as a linear function of the past squared realizations, and suggests estimation by maximum likelihood (ML). Bollerslev and Wooldridge (1992) showed that, albeit the incorrect specification of the conditional distribution of the disturbances, the ML procedure still can produce strongly consistent estimators if the first two conditional moments of a random vector of the disturbances are correctly specified. However, convergent results for nonlinear estimation are not easily obtained by standard numerical maximization routines. Consequently, researchers tend to turn to alternative estimation procedures such as Hansen's (1982) generalized method of moment (GMM) discussed in Rich, Raymond and Butler (1991, 1992) (other procedures include nonparametric and semiparametric methods considered by Pagan and Ullah (1988), Pagan and Hong (1988), Engle and Gonzalez (1991)).

There are several advantages in applying the GMM estimation procedure rather than ML. First, the GMM procedure is robust to the distributional assumption for the disturbances of the ARCH model and yields consistent estimates of the parameters and their variance-covariance matrix. Secondly, the GMM procedure is more tractable and less computationally expensive, in this case, than ML. However, in finite samples the properties of the GMM estimator are virtually unknown and hence deserve study.

In this paper, the finite sample performance of (Quasi) ML and GMM estimators is examined through Monte-Carlo simulation. Particularly, we study: the biases of estimators, the ratios of the estimated standard errors to the sample standard deviations, the coverage of the confidence intervals, Wilcoxon matched-pairs signed-ranks tests, Kolmogorov-Smirnov tests of normality and the relative efficiency.

Section 2 briefly discusses the (Quasi) ML and GMM estimation procedures and their asymptotic properties. The design of the Monte-Carlo experiments and the results are presented in section 3. Generally speaking, the results are quite discouraging for the GMM estimator in the ARCH regression setting. Section 4 offers some concluding remarks.

II. THE ARCH PROCESS AND ESTIMATION STRATEGIES

The pth order linear ARCH model introduced by Engle (1982) is given by:

(1) $$y_t = E(y_t|\Psi_{t-1}) + \varepsilon_t = x_t\beta + \varepsilon_t, \quad t = 1,2,\ldots,T$$

(2) $$E(\varepsilon_t|\Psi_{t-1}) = 0$$

(3) $$E(\varepsilon_t^2|\Psi_{t-1}) = h_t = \alpha_0 + \sum_{j=1}^{p}\alpha_j\varepsilon_{t-j}^2,$$

$$\alpha_0 > 0, \ \sum_{j=1}^{p}\alpha_j < 1, \ 0 < \alpha_j < 1.$$

where y_t is the dependent variable, x_t is a vector of explanatory variables in the information set $\Psi_{t-1} = \{ y_{t-j}, x_{t-j+1}\}_{j=1}^{\infty}$, and β is a vector of parameters of interest. Notice that the mean function in $E(\varepsilon_t|\Psi_{t-1})$ in (1) can also be generalized to any type of regression model including an autoregressive integrated moving average (ARIMA)[7]. The ARCH model focuses upon the distinction between unconditional and conditional second moments. Specifically, it generalizes the second moment to allow for the variance to depend on the past forecast errors but leaves the specification of the mean unchanged. Estimation of the parameters of the ARCH(p) model can be done by applying ordinary least square (OLS) or by maximum likelihood (ML). Engle (1982) shows that there is a large gain in efficiency from using ML rather than OLS and thus we focus here on the ML procedure.

2.1. Maximum likelihood and Quasi-maximum likelihood estimators

Let $\theta = (\beta,\alpha)$ and assume that the conditional probability distribution of ε's in (1) is $g(\varepsilon_t|\Psi_{t-1})$, then the log-likelihood function for a sample $\varepsilon_1, \varepsilon_2,\ldots,\varepsilon_T$ is, apart from some initial conditions, given by:

---

[7] Bollerslev (1986) proposed a generalized autoregressive conditional heteroskedasticity, GARCH(p,q), in which h_t in (3) is modified to allow for the additional effect from the past conditional variances, i.e. $h_t = \omega + \sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{q}\beta_j h_{t-j}$. For more detail on GARCH(p,q) process see Bollerslev (1986).

(4) $$L_T = \sum_{t=1}^{T} \ell_t = \sum_{t=1}^{T} \log\ g(\varepsilon_t|\Psi_{t-1})$$

In most applications, $g(\varepsilon_t|\Psi_{t-1})$ is taken to be conditionally normal because of its simplicity and well known properties under certain ideal conditions. Now under the assumption that $g(\varepsilon_t|\Psi_{t-1})$ is $N(0,h_t)$, it follows that the density of y_t conditional upon Ψ_{t-1} is $N(x_t\beta,h_t)$ and so (4) would be

(5) $$L_T = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log(h_t) - \frac{1}{2}\sum_{t=1}^{T}(y_t - x_t\beta)^2 h_t^{-1}$$

By substituting h_t in (3) into (5), the joint estimation of the parameters β and α can be obtained by maximizing (5) with respected to β and α subject to the restrictions listed in (3). While the estimation algorithms are detailed in Engle (1982), we concentrate on the asymptotic property of MLE which can be stated as follows.

Proposition 1: (Engle (1982)). If the regularity conditions set out in Crowder (1976) hold, then, with $\hat{\theta}$ the MLE from (5),

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0,A_T^{-1})$$

where $A_T = \mathcal{J}_{\theta\theta} = - T^{-1} \sum_{t=1}^{T} E[\partial S_t(\theta_0)/\partial\theta]$, with $S_t = \partial L_T(\theta_0)/\partial\theta$.

Notice that the above proposition holds true if the conditionally distributed errors are normal and $\hat{\theta}$ is the MLE of θ. However, when the conditionally normal distributed errors is violated, but the normal log-likelihood is falsely maximized, then $\hat{\theta}$ is referred to as a quasi-maximum likelihood estimator (QMLE) of θ and denoted by $\tilde{\theta}$. Moreover, the asymptotic properties of $\tilde{\theta}$ are still valid under a slightly modified set of regularity conditions (Bollerslev and Wooldridge (1992)).

Weiss (1986) provided the first study of the asymptotic properties of Quasi-maximum likelihood estimation of the univariate ARCH models. Bollerslev and Wooldridge (1992) extend the results to the multivariate GARCH models. Bollerslev and Wooldridge also investigated the finite sample properties of the QMLE and found that QMLE performs reasonably well in small samples. The consistency and asymptotic normality of QMLE can be summarized in the following proposition.

Proposition 2 : (Bollerslev and Wooldridge (1992)). If the regularity conditions in appendix A1 hold and for some $\theta \in$ Int Θ, then $\sqrt{T}(\tilde{\theta} - \theta_0) \stackrel{a}{\sim} N(0,\Omega_T)$ where $\Omega_T = A_T^{-1}B_TA_T^{-1}$, and

$$B_T = T^{-1} \sum_{t=1}^{T} E\left[S_t(\theta_0)S_t(\theta_0)'\right], \qquad S_t(\theta_0) = \partial\ell_t(\theta_0)/\partial\theta$$

$$A_T = T^{-1} \sum_{t=1}^{T} E\left[a_t(\theta_0)\right], \qquad a_t(\theta_0) = -E\left[\partial S_t(\theta_0)/\partial\theta | x_t\right]$$

Moreover, $$\tilde{A}_T = T^{-1} \sum_{t=1}^{T} E\left[a_t(\tilde{\theta}_0)\right] \xrightarrow{p} A_T \text{ and}$$

$$\tilde{B}_T = T^{-1} \sum_{t=1}^{T} E\left[S_t(\tilde{\theta}_0)'S_t(\tilde{\theta}_0)\right] \xrightarrow{p} B_T$$

The proof of the above proposition can be found in Weiss(1986) and in Bollerslev and Wooldridge (1992) with the first author considering the univariate ARCH model and the second authors considering the multivariate GARCH case. Note that while the consistency and asymptotic normality still hold for $\tilde{\theta}_T$, the asymptotic covariance matrix of $\tilde{\theta}_T$ takes the form of $A_T^{-1}B_TA_T^{-1}$ due to the incorrect specification of the likelihood function.

2.2. Generalized Method of Moments (GMM)

An alternative method of estimation to MLE and QMLE is Hansen's (1982) GMM and the idea has been implemented by Glosten, Jagannathan, and Runkle (1989), Harvey (1989), Mark (1990), Simon (1989), Rich, Raymond and Butler (1991, 1992). To obtain GMM estimates, we rewrite (1) and (2) as follows:

$$\varepsilon_t = y_t - x_t\beta \tag{6}$$

$$\eta_t = (y_t - x_t\beta)^2 - \alpha_0 - \sum_{i=1}^{p}\alpha_i(y_{t-i} - x_{t-i}\beta)^2 \tag{7}$$

where $\eta_t = \varepsilon_t^2 - h_t$, and $E(\eta_t|\Psi_{t-1}) = 0$.

The system of equations (6) and (7) contains a cross equation restriction on β since β appears in both the specification of the conditional mean and the conditional variance equations.

Let $u_t(\theta_0) = \{\varepsilon_t(\theta_0), \eta_t(\theta_0)\}$ be (1x2) vector of innovations to the ARCH model where θ_0 is the true value for θ. Define the following $[2x(m_1+m_2)]$ block diagonal matrix

$$Z_{t-1} = \begin{bmatrix} z_{1,t-1} & 0 \\ 0 & z_{2,t-1} \end{bmatrix}$$

where $z_{1,t-1}$ and $z_{2,t-1}$ are respectively $(1xm_1)$ and $(1xm_2)$ vectors of observable variables that are in the information set at time *t-1*. Under the assumption that u_t are uncorrelated with the set of information available at time *t-1*, it follows that

$$(8) \qquad E\left[u_t(\theta_0)Z_{t-1}\right] = E\left[g_t(\theta_0)\right] = 0$$

Equation (8) is the basis of the GMM procedure. It represents a set of (m_1+m_2) orthogonality conditions which are used to estimate θ_0 with Z_{t-1} serving as instruments for the regressors in the ARCH model. Note that if we let $\bar{g}_T(\theta_0) = T^{-1}\sum_{t=1}^{T} g_t(\theta_0)$, then under regularity conditions $\bar{g}_T(\theta_0) \longrightarrow E\left[g_t(\theta_0)\right]$ almost surely, and since $E\left[g_t(\theta_0)\right] = 0$, the GMM estimator of θ can be found by minimizing the following criterion function:

$$(9) \qquad J_T(\theta) = \bar{g}_T(\theta)'\, W_T\, \bar{g}_T(\theta)$$

where W_T is a $[(m_1+m_2)x(m_1+m_2)]$ symmetric nonsingular weighting matrix that satisfies $W_T \rightarrow W$ almost surely, where W is also symmetric and nonsingular. The GMM estimation of θ involves a two-step procedure and Hansen (1982) provides the details of the estimation algorithm including how to obtain the optimal weighting matrix, W_T^* .

Proposition 3 : (Hansen (1982)) If the conditions set out in appendix A2 hold, then $\sqrt{T}(\theta_T^{\bullet} - \theta_0)$ is distributed asymptotically as $N(0,\Omega)$ where $\Omega = (D'W_T^{\bullet} D)$ with[8]

$$D = E\left[\partial g_t(\theta_0)/\partial\theta\right] \quad \text{and} \quad W_T^{\bullet} = \left\{ E\left[g_t(\theta_0)'g_t(\theta_0)\right] \right\}^{-1}$$

The detailed proof of proposition 3 is given in Hansen(1982).

Now we have presented thus far the three estimators and their asymptotic properties. The next problem is to examine whether these properties would carry over to finite samples since how these estimators behave in finite samples will have special interest to an applied researcher.

The finite sample properties of MLE and QMLE for the univariate ARCH model have been studied in detail by Engle, Hendry and Trumble (1985), Bollerslev and Wooldridge (1992), Engle and González-Rivera (1991). All these papers show that when the underlying density is nonnormal, the biases and MSE are very small. Tauchen (1986) provide a small sample property for the GMM estimator in a Markov chain model applied to the generated data on asset returns from

[8] Here we use the optimal GMM estimator, i.e. formula (10) of Hansen's Theorem 3.2 (p. 1048) is used, rather than more general formula of Theorem 3.1 (p.1042).

stochastic exchange economies. He shows that there is a bias/variance trade off regarding the lag length used to form the instruments.

In the following section we present detailed finite sample behavior of the three estimators though mean of Monte Carlo simulations.

III. MONTE CARLO EXPERIMENTS

3.1. Data Generating Process

To evaluate the finite sample performance of MLE, QMLE and GMM estimators described in the previous section, and their relative efficiencies, we carried out Monte-Carlo experiments. All the simulated models were nested within the following model:

(10) $$y_t = \delta y_{t-1} + \beta x_t + \varepsilon_t$$

(11) $$x_t = \lambda x_{t-1} + \omega_t, \qquad |\lambda| < 1 \text{ and } \omega_t \sim N(0,1)$$

(12) $$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \qquad t = 1,2, \quad ,T$$

(13) $$\varepsilon_t = \zeta_t h_t^{1/2}, \qquad \zeta_t \sim \text{i.i.d. } t_\nu$$

with ω_t and ζ_t independent and t_ν is the standard t-distribution with ν degrees of freedom. We assume that $\alpha_0 + \alpha_1 = 1$, and that the intercept term in (10) is equal to zero. However, the restrictions on α's are not imposed in estimation. The series $\{x_t\}$ were kept fixed in repeated samples within experiments but generated separately between experiments. The t_ν distributed random variables were formed as the squared root of $(\nu-2)$ times an $N(0,1)$ random variable divided by the square root of χ^2_ν variate generated by forming $-2\ln(\prod_{i=1}^{n} u_i)$ where n = df/2 and u_i are independent random variables from a uniform (0,1) distribution.[9]

The experimental design is governed by the parameters of the data generating process (10)-(13); i.e. $\theta = (\delta, \beta, \lambda, \alpha_0, \alpha_1)$. When $\beta = 0$, the simulated model is an AR(1)-ARCH(1) and we choose $\theta = (\delta, \alpha_0, \alpha_1)$ as (0.8, 0.6, 0.4) and (0.6, 0.8, 0.2). We denote this model as model A. When $\delta = 0$, the simulated model is that of Engle, Hendry and Trumble (1985) and we set $\theta = (\beta, \lambda, \alpha_0, \alpha_1)$ as (1.0, 0.8, 0.6, 0.4) and (1.5. 0.8, 0.8, 0.2), and this is model B. Note that both models chosen here have been empirically studied by Lee and Tse (1992). The parameter values used for these experiments are similar to those values used by Engle, Hendry and Trumble (1985). The sample size T was set equal to 500 or 1000 and replicated 1000 times.[10] Also, to

---

9 When the df ν is an odd number, the χ^2 variate is generated by forming $-2\ln(\prod_{i=1}^{n} u_i) + z^2$ where $n = (\nu-1)/2$ and z is a standard normal.

10 The sample sizes chosen for our study are not uncommon since most financial series contain large data points, e.g. daily stock returns series from 1980-1990.

eliminate any influence of the initial condition, the first 100 observations were discarded in each replication. In all Monte-Carlo simulations, TSP 4.2A was employed, using the default seed, running on a 486DXII-50 PC. There were no convergence problems, however there is a substantial time difference in obtaining the results. For example, with MLE and QMLE, it took approximately 120 minutes to complete one experiment while it took 65 minutes for GMM with a sample size of 500.

Note that for GMM estimation, four different instrument sets were used for each model: the first three are "naive" instruments and the last one is "optimal" instrument. These are listed as follow.

For model A:

NAV1 uses instruments $Z_1 = Z_2 = \left[1, y_{t-1}, y_{t-1}^2, (\Delta y_{t-1})^2\right].$

NAV2 ———————— $Z_1 = Z_2 = \left[1, y_{t-1}, y_{t-2}, (\Delta y_{t-1})^2, (\Delta y_{t-2})^2\right].$

NAV3 ———————— $Z_1 = \left[1, y_{t-1}, y_{t-2}, (\Delta y_{t-1})^2, (\Delta y_{t-2})^2\right].$

$$Z_2 = \left[1, \Delta y_{t-1}, \Delta y_{t-2}, (\Delta y_{t-1})^2, (\Delta y_{t-2})^2\right].$$

OPT uses "optimal" instruments $Z_t^0 = H_t' \Phi_t$ where $H_t' = E\left[\partial u_t(\theta_0)/\partial \theta'\right]$ and $\Phi_t = E\left[u_t(\theta_0) u_t(\theta_0)'\right]$.

For model B:

NAV1 uses instruments $Z_1=Z_2=\left[1, x_{t-1}, y_{t-1}, (\Delta x_{t-1})^2, (\Delta y_{t-1})^2\right].$

NAV2 ———————— $Z_1=Z_2=\left[1, x_{t-1}, y_{t-1}, (\Delta x_{t-1})^2, (\Delta y_{t-1})^2\right],$

$i = 1,2.$

NAV3 ———————— $Z_1=\left[1, x_t, x_{t-1}, y_{t-1}\right]$

$Z_2=\left[1, \Delta x_t, (\Delta x_t)^2, (\Delta y_{t-1}), (\Delta y_{t-1})^2\right].$

OPT uses "optimal" instruments described above.

It should be noted that the so-call "optimal" instruments can not be used in practice due to the fact that the instruments Z_t^0 depend on the underlying parameter vector θ_0, which is unknown. However, for the purpose of comparison we include these instruments in the simulations.

3.2. Simulation Results

The estimated biases and mean square errors (MSE) of MLE, QMLE and GMM estimators using four different sets of instruments including the "optimal" instruments are presented in table 1 and 2. For both model A and B, the results indicate that with MLE and QMLE, the biases of the estimators of all parameters are quite small. The estimators of δ (for model A), β (for model B), and α_1 show negative biases. Engle, Hendry and Trumble (1985) showed a similar pattern but found much larger negative biases for α_1 for sample sizes smaller than 100. For GMM estimation, with the exception of OPT, we found severe biases of the estimators of α_1. The biases show the same direction (negative) but are very large ranging between 32% to 50% relative to the true value depending on the choice of the instruments. The larger the instrument list the larger the biases. With the exception of Tauchen (1985), there has not been much study of the finite sample behavior of the GMM estimator. Consequently, our results suggest that the problem of bias is due to the quality of the instruments used in the estimation. Finally, different parameter values do not seem to affect the results. Furthermore, as the sample size increases from 500 to 1000, the biases and MSE of estimators for both model A and B in most cases, decrease.

To further examine the finite sample properties of these estimators, we investigate whether the estimated standard errors, averaged across replications for each experiment, reasonably approximate the sample standard deviations. The estimated standard errors were calculated as the squared root of the asymptotic variances evaluated at the estimated values of the parameter. These results are tabulated in table 3 and 4. There are several interesting features of the results worth noting. First, for MLE and

QMLE, the ratios range from 0.900 to 1.089 for both model A and B, indicating that the estimated standard errors are a reasonable approximation of the sample standard deviations. Second, except for OPT, the estimated standard error of GMM estimators with different instrument sets seriously underestimate the sample standard deviations of all parameters in model A. The ratios range from 0.570 to 0.912. For model B, the problem of underestimation is less serious in some cases. The ratios are seen to improve somewhat and one possible explanation for this is that model B contains extra information coming from the exogenous regressor x_t, i.e. the instrument lists used in the GMM estimation now include both exogenous and predetermined variables. For example, the ratios of estimated standard errors to standard deviations for β are close to 1 but far less than 1 for the ARCH parameters α_0 and α_1, although these ratios are higher than for those in model A. Consequently, as observed earlier, these serious underestimation problems are due to a poor choice of instruments. For OPT, the ratios show good approximation ranging from 0.940 to 1.030.

Table 5 and 6 present the coverage confidence interval of the estimators. We compute the 95 percent nominal confidence intervals of estimators and then calculate the number of times in 1000 replications that the confidence interval includes the true values of parameters.[11] Table 5 and 6 show that for MLE and QMLE, the coverage of confidence intervals is very close to 95 percent of the 1000 replications. Table 3 and 4 show that the GMM of the ARCH parameters α_0 and α_1 are often underestimated. As a result, the t-statistics in GMM become upward biased. Table 5 and 6 reveal this phenomenon: the 95 percent confidence intervals include the true value much less than 95 percent of the time. In some cases, the confidence interval is as low as 34.1 percent of the 1000 replications. It is also interesting to note that the confidence interval for GMM of β is close to 0.95. Also, for OPT, the 95 percent confidence intervals of all parameters are very close to 0.95.

To get an idea of how one estimator performs in finite samples relative to another in terms of absolute bias, we conducted the Wilcoxon matched-pairs signed-ranks tests and table 7 and 8 report these results. First, the column denoted by "Rank(-)" in each table is the mean rank for the estimator with the smaller bias relative to another estimator (and hence the (-) sign). Second, the "Z" column denotes the Wicoxon statistic for testing the null of no difference between the two estimators. Third, in what follows, unless otherwise noted, we choose NAV1 and OPT for

[11] The asymptotic standard errors were use to calculate the 95% nominal confidence interval.

model A and NAV3 and OPT for model B as the candidates for the GMM estimator since the results are unaltered if we use any other GMM than the one mentioned. Overall, as expected, the MLE estimator is significantly better than the other two (QMLE and GMM) in almost all cases. Between QMLE and GMM, for both model A and B, GMM seem to be either significantly worse than QMLE (except for "optimal" GMM, OPT), especially for ARCH's parameters α_0 and α_1, or there is no significant difference at 5% significance level. Note that we used Wilcoxon test rather than the standard paired-t test because of its robustness to the departure of normality assumption. Table 9 and 10 show the Kolmogorov-Smirnov (K-S) test of normality of the parameters for the two models. First, observe that increasing the sample sizes or changing the parameter values do not seem to change the results much. Secondly, for the MLE estimator, K-S test failed to reject the null hypothesis of normality at (i) 5% level of significance for all cases, (ii) 1% for some cases. For QMLE and GMM, with the exception of OPT, there is a tendency of departure from normality in finite samples, especially for the ARCH parameters. For almost all cases, the K-S test reject the normality at the standard 5% level of significance for both estimators. With the regression parameters (δ for model A and β for model B), the normality seems to be valid for GMM but not for QMLE since the K-S rejects the null of normality for QMLE but cannot reject the null for GMM. Figures 1-4 provide some evidences for these conclusions. For conservation of space we plot out one case in model B as being representative (all other plots have similar pattern).[12] In each figure, the empirical distribution of the standardized estimates along with the cumulative distribution of the standard normal is plotted.[13] These plots confirm the results reported in table 9 and 10, namely (i) for MLE the asymptotic normality of the estimated parameters are valid in finite sample; (ii) the departure from normality is found with the QMLE estimate of both β and the ARCH parameters, and (iii) the evidence for normality of the GMM estimate's sampling distribution of β is weak and a complete departure from normality for the ARCH parameters (except for the "optimal" GMM).

Finally, we examine the relative efficiency of estimators. For an applied researcher, relative efficiency of estimators is an important issue since it represents the trade off between efficiency and computational costs. The ratios of sample MSE of pairs of estimators provide a good measure of relative efficiency. These ratios are presented in table 11 and 12 which also provide the ratio of

---

[12] The case chosen here is for T = 500, $\beta = 1.0$, $\alpha_0 = 0.6$, $\alpha_1 = 0.4$ and $\lambda = 0.8$.

[13] The estimate is standardized by the corresponding sample mean and standard deviation.

asymptotic variance of pairs of estimators. To get an idea of efficiency gain from using QMLE versus the GMM procedure, we compare these with the true MLE. The first three rows of table 11 and 12 depict these results. There seems to be a better gain in efficiency from using QMLE rather than GMM when the instruments used are other than the "optimal" instruments[14]. For model A, the relative efficiency of GMM1 to MLE (in terms of ratios of MSE) is particularly low when the ARCH coefficient, α_1 is large. Changing the parameters values of δ and α_1 do tend to increase this ratio (particularly, for smaller δ and α_1). The second last row of table 11 and 12 show the relative efficiency between QMLE and the GMM estimators. The results reveal that QMLE dominates GMM estimator in almost all cases. In some cases, the ratios of MSE and AV show an improvement in efficiency between 30 percent and 70 percent over GMM[15]. However when the "optimal" GMM is used, the improvement in efficiency is reversed.

In another study, Engle and González-Rivera (1991) show that, in their Monte-Carlo simulation, there is no gain in semi-parametric procedures over QMLE when the true

---

[14] However, as noted earlier, GMM using "optimal" instruments requires that the instrument list be a function of the unknown parameters and hence can not be used in practice.

[15] With "optimal" instruments, OPT dominates the QMLE in all cases.

conditional density is Student-t due to the poor nonparametric estimation of the tails of the density. However, when the true conditional density is Gamma, they find a large gain in efficiency. We have run some experiments (the results are not reported here) with conditional Gamma distributed errors and find that the results are unaltered, i.e. there is no gain in doing GMM over QMLE. This is not surprising since the GMM procedure is robust to any distributional assumption on the innovation of the model.

IV. CONCLUSION

In this paper, we have presented detailed finite sample properties of estimators for the univariate ARCH models. These estimators include MLE, QMLE, and GMM with 4 different instrument sets. In general, the results are quite discouraging for the GMM estimator. The results from the Monte-Carlo study suggest that the GMM estimators (except for the optimal GMM) are generally biased and the magnitude of the biases depend on the choice of the instrument list. The estimated standard errors of GMM estimator are essentially downward biased, resulting in an upward bias in the t statistics. Therefore, the question that arises immediately from our analysis is: can we derive a pratical

GMM estimator which does almost as well as the optimal one? This would be an interesting topic for future research.

On the other hand, the results for QMLE are quite encouraging and prove it to be more efficient than GMM estimation.

Overall, even though GMM is computationally less expensive than QMLE, care must be taken when applied researchers wish to use the GMM procedure in estimating ARCH type models. Also note that the results reported here should be taken as indicative rather than definitive.

REFERENCES

Bollerslev, T. P. (1986), "Generalized autoregressive conditional Heteroskedasticity", *Journal of Econometrics*, 31, 309-328.

Bollerslev, T. P., 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.

Bollerslev, T. P., and J. M. Wooldridge (1992), "Quasi-maximum likelihood estimation and inference in dynamics models with time-varying covariances", *Econometric Reviews*, 11, 143-172.

Crowder, M. J. (1976), "Maximum likelihood estimation for dependent observations", *Journal of the Royal Statistical Society, Series B*, 38, 45-53.

Engle, R. F. (1982), "Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation", *Econometrica*, 50, 987-1008.

Engle, R. F., and G. González-Rivera (1991), "Semiparametric ARCH models", *Journal of Business and Economic Statistics*, 9, 345-359.

Engle, R. F., D. M. Hendry, and D. Trumble (1985), "Small sample properties of ARCH estimators and tests", *Canadian Journal of Economics*, 18, 66-93.

Glosten, L. R., R. Jagannathan, and D. Runkle (1989), "Relationship between the expected value and the volatility of the excess nominal return on stocks", Unpublished manuscript, J. L. Kellogg Graduate School, Northwestern University.

Hansen, L. P. (1982), "Large sample properties of generalized method of moments estimators", *Econometrica*, 50, 1029-1054.

Hansen, L. P., and R. J. Hodrick (1980), " Forward exchange rates as optimal predictors of future spot rates: an econometric analysis", *Journal of Political Economy*, 88, 829-853.

Harvey C. (1989), "Is the expected compensation for market volatility constant through time?" Unpublished manuscript, Fuqua School of Business, Duke University.

Lee, T. K., and Y. K. Tse (1991), "Term structure of interest rates in the Singapore Asian dollar market", *Journal of Applied Econometrics*, 6, 143-152.

Mark, N (1988), "Time varying betas and risk premia in the pricing of forward foreign exchange contracts", *Journal of Financial Economics*, 22, 335-354.

Pagan, A.R., and Y. S. Hong (1989),"Nonparametric estimation and risk premium", in W. Barnett, J. Powell and G. Tauchen, eds., *Semiparametric and nonparametric method in economics and Statistics* (Cambridge University Press, Cambridge).

Pagan, A. R., and A. Ullah (1988),"The econometric analysis of models with risk terms", *Journal of Applied Econometrics*, 3, 87-105.

Rich, R. W., J. Raymond, and J. S. Butler (1991), "Generalized instrumental variables estimation of autoregressive conditionally heteroskedasticity models", *Economics Letters*, 35, 179-185.

Rich, R. W., J. Raymond, and J. S. Butler (1992),"The relationship between forecast dispersion and forecast uncertainty:evidence from survey data-ARCH model", *Journal of Applied Econometrics*, 7, 131-148.

Simon, D. P. (1989), "Expectations and risk in the treasury bill market: an instrumental variable approach", *Journal of Financial and Quantitative Analysis*, 24, 357-366.

Tauchen, G. (1986), "Statistical properties of generalized method of moments estimators of structural parameters using financial market data", *Journal of Business and Economics Statistics*, 4, 397-416.

Weiss, A. A. (1986), "Asymptotic theory for ARCH models: estimation and testing", *Econometric Theory*, 2, 107-131.

APPENDIX A1.

Regularity conditions for QMLE:

1. Θ is a compact subspace of Euclidean Space and has non empty interior.

2. The conditional mean and conditional variance are measurable for all $\theta \in \Theta$ and twice continuously differentiable on int.Θ.

3. (a) $\{\ell_t(\theta),\ t=1,2,\ldots\}$ is the log-likelihood function of observation t and it assumed to satisfy the uniform weak law of large numbers.

(b) θ_0 is the identifiable unique minimizer of $E(\sum_{t=1}^{T}\ell_t(\theta))$.

4. (a) Both $\partial^2\ell_t/\partial\theta\partial\theta'$, and $E(\partial^2\ell_t/\partial\theta\partial\theta')$ are assumed to satisfy the uniform of weak law of large numbers.

(b) $A_T = -T^{-1}\sum_{t=1}^{T}E(\partial^2\ell_t/\partial\theta\partial\theta')$ is uniformly positive definite.

5. $(\partial\ell_t/\partial\theta)'(\partial\ell_t/\partial\theta)$ satisfies the uniform weak law of large numbers.

6. (a) $B_T = T^{-1}\sum_{t=1}^{T}E(\partial\ell_t/\partial\theta)'(\partial\ell_t/\partial\theta)$ is uniformly positive definite.

(b) $T^{-1/2}\, B_T \sum_{t=1}^{T}(\partial\ell_t/\partial\theta)' \xrightarrow{A} \mathcal{N}(0,\mathcal{I}_p)$.

APPENDIX A2.

Conditions for GMM:

1. $\{(y_t, x_t): -\infty < t < +\infty\}$ is stationary and ergodic.

2. S is an open subset of $\mathbb{R}^p$ that contains θ_0.

3. $g(.,\theta)$ and $\partial g(.,\theta)/\partial\theta$ are Borel measurable for each $\theta \in S$ and $\partial g(.,\theta)/\partial\theta$ is continuous on S.

4. $\partial g_1/\partial\theta$ is the first moment continuous at θ_0, and $E[\partial g(x_1,\theta_0)/\partial\theta]$ exists, is finite, and has full rank.

Table 1. Biases and MSE of estimators*

Model A: $y_t = \delta y_{t-1} + \varepsilon_t$

$h_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1} \quad t = 1,2,\ldots,T$

$\varepsilon_t = \xi_t h_t^{1/2}\ , \ \xi_t$ i.i.d. $t(5)$

	δ		α_0		α_1	
	Bias	MSE	Bias	MSE	Bias	MSE
$T = 500,\ \delta = 0.8,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4.$						
MLE	-0.002	0.0007	0.005	0.0044	-0.005	0.0130
	(.0008)	(.0000)	(.0021)	(.0002)	(.0030)	(.0008)
QMLE	-0.002	0.0009	-0.000	0.0084	-0.012	0.0256
	(.0013)	(.0003)	(.0036)	(.0013)	(.0036)	(.0010)
NAV1	-0.004	0.0020	0.062	0.0220	-0.180	0.0480
	(.0014)	(.0004)	(.0042)	(.0013)	(.0039)	(.0012)
NAV2	-0.005	0.0017	0.049	0.0203	-0.185	0.0506
	(.0013)	(.0003)	(.0042)	(.0014)	(.0041)	(.0013)
NAV3	-0.006	0.0017	0.061	0.0215	-0.183	0.0502
	(.0013)	(.0003)	(.0042)	(.0015)	(.0041)	(.0013)
OPT	-0.002	0.0008	-0.020	0.0078	-0.042	0.0184
	(.0009)	(.0001)	(.0026)	(.0004)	(.0036)	(.0009)
$T = 500,\ \delta = 0.6,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2.$						
MLE	-0.001	0.0012	0.010	0.0063	-0.004	0.0085
	(.0011)	(.0000)	(.0025)	(.0004)	(.0015)	(.0004)
QMLE	-0.002	0.0018	-0.002	0.0123	-0.011	0.0150
	(.0012)	(.0002)	(.0035)	(.0006)	(.0025)	(.0005)
NAV1	-0.001	0.0022	0.009	0.0137	-0.090	0.0159
	(.0015)	(.0002)	(.0037)	(.0008)	(.0028)	(.0006)
NAV2	-0.002	0.0020	-0.017	0.0137	-0.091	0.0166
	(.0014)	(.0001)	(.0036)	(.0008)	(.0029)	(.0005)
NAV3	-0.002	0.0020	-0.010	0.0142	-0.087	0.0158
	(.0014)	(.0001)	(.0038)	(.0009)	(.0029)	(.0005)
OPT	-0.001	0.0016	-0.021	0.0113	-0.038	0.0100
	(.0012)	(.0001)	(.0032)	(.0006)	(.0025)	(.0004)

Table 1. Continue...

	$T = 1000,\ \delta = 0.8,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4.$					
MLE	-0.001	0.0003	0.002	0.0022	-0.002	0.0065
	(.0005)	(.0000)	(.0015)	(.0001)	(.0026)	(.0003)
QMLE	-0.001	0.0004	-0.000	0.0044	-0.005	0.0121
	(.0007)	(.0001)	(.0034)	(.0024)	(.0032)	(.0010)
NAV1	-0.000	0.0010	0.077	0.0211	-0.161	0.0410
	(.0010)	(.0003)	(.0039)	(.0024)	(.0038)	(.0014)
NAV2	-0.003	0.0010	0.070	0.0187	-0.168	0.0424
	(.0010)	(.0003)	(.0037)	(.0022)	(.0038)	(.0012)
NAV3	-0.004	0.0009	0.082	0.0231	-0.165	0.0416
	(.0010)	(.0003)	(.0040)	(.0023)	(.0038)	(.0011)
OPT	-0.001	0.0004	-0.011	0.0037	-0.030	0.0103
	(.0006)	(.0000)	(.0020)	(.0002)	(.0029)	(.0006)
	$T = 1000,\ \delta = 0.6,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2.$					
MLE	-0.001	0.0006	0.002	0.0033	-0.005	0.0052
	(.0008)	(.0000)	(.0018)	(.0002)	(.0018)	(.0001)
QMLE	-0.001	0.0008	-0.001	0.0068	-0.005	0.0081
	(.0010)	(.0001)	(.0026)	(.0005)	(.0024)	(.0003)
NAV1	-0.001	0.0018	0.023	0.0107	-0.077	0.0127
	(.0013)	(.0001)	(.0032)	(.0008)	(.0026)	(.0006)
NAV2	-0.001	0.0010	0.007	0.0081	-0.079	0.0125
	(.0010)	(.0001)	(.0028)	(.0005)	(.0025)	(.0004)
NAV3	-0.001	0.0009	0.013	0.0088	-0.076	0.0119
	(.0010)	(.0001)	(.0029)	(.0005)	(.0025)	(.0005)
OPT	-0.001	0.0007	-0.010	0.0060	-0.028	0.0061
	(.0009)	(.0000)	(.0024)	(.0003)	(.0020)	(.0002)

* Simulation standard errors are in parenthesis

Table 2. Biases and MSE of estimators

Model B:
$$y_t = \beta x_t + \varepsilon_t$$
$$x_t = \lambda x_{t-1} + \omega_t,\ \omega_t \text{ i.i.d } \mathcal{N}(0,1)$$
$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2,\ t = 1,2,\ldots T$$
$$\varepsilon_t = \xi_t h_t^{1/2},\quad \xi_t \text{ i.i.d. } t(5)$$

	β		α_0		α_1	
	Bias	MSE	Bias	MSE	Bias	MSE
	T = 500, β = 1.0, α_0 = 0.6, α_1 = 0.4, λ = 0.8.					
MLE	-0.000	0.0005	0.004	0.0045	-0.005	0.0133
	(.0008)	(.0000)	(.0021)	(.0002)	(.0039)	(.0011)
QMLE	0.001	0.0007	-0.002	0.0079	-0.020	0.0266
	(.0009)	(.0000)	(.0031)	(.0009)	(.0043)	(.0016)
NAV1	-0.000	0.0017	0.028	0.0237	-0.180	0.0591
	(.0010)	(.0000)	(.0048)	(.0013)	(.0052)	(.0019)
NAV2	0.000	0.0010	0.009	0.0192	-0.197	0.0635
	(.0010)	(.0000)	(.0044)	(.0011)	(.0050)	(.0018)
NAV3	0.000	0.0006	0.065	0.0316	-0.194	0.0650
	(.0008)	(.0000)	(.0052)	(.0021)	(.0052)	(.0019)
OPT	0.000	0.0005	-0.026	0.0070	-0.040	0.0191
	(.0007)	(.0000)	(.0025)	(.0004)	(.0039)	(.0011)
	T = 500, β = 1.5, α_0 = 0.8, α_1 = 0.2, λ = 0.8.					
MLE	-0.001	0.0005	0.004	0.0067	-0.005	0.0089
	(.0007)	(.0000)	(.0026)	(.0004)	(.0030)	(.0005)
QMLE	-0.000	0.0006	-0.002	0.0118	-0.015	0.0143
	(.0009)	(.0000)	(.0038)	(.0008)	(.0029)	(.0009)
NAV1	0.000	0.0010	-0.032	0.0238	-0.084	0.0283
	(.0010)	(.0000)	(.0048)	(.0014)	(.0046)	(.0014)
NAV2	0.000	0.0010	-0.057	0.0225	-0.096	0.0280
	(.0010)	(.0000)	(.0044)	(.0012)	(.0043)	(.0012)
NAV3	0.000	0.0006	0.002	0.0282	-0.099	0.0366
	(.0008)	(.0000)	(.0053)	(.0019)	(.0051)	(.0020)
OPT	0.000	0.0006	-0.032	0.0106	-0.030	0.0104
	(.0008)	(.0000)	(.0030)	(.0005)	(.0026)	(.0005)

Table 2. Continue ...

$T = 1000,\ \beta = 1.0,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4,\ \lambda = 0.8.$

MLE	-0.000	0.0002	0.002	0.0022	-0.002	0.0065
	(.0001)	(.0000)	(.0015)	(.0001)	(.0026)	(.0003)
QMLE	-0.002	0.0003	0.001	0.0058	-0.001	0.0098
	(.0006)	(.0000)	(.0029)	(.0006)	(.0033)	(.0009)
NAV1	-0.000	0.0006	0.052	0.0220	-0.154	0.0454
	(.0007)	(.0000)	(.0042)	(.0015)	(.0047)	(.0017)
NAV2	-0.000	0.0004	0.037	0.0163	-0.168	0.0475
	(.0007)	(.0000)	(.0039)	(.0011)	(.0044)	(.0015)
NAV3	-0.000	0.0003	0.078	0.0277	-0.164	0.0504
	(.0006)	(.0000)	(.0046)	(.0017)	(.0048)	(.0019)
OPT	-0.000	0.0003	-0.018	0.0040	-0.018	0.0099
	(.0005)	(.0000)	(.0020)	(.0002)	(.0031)	(.0005)

$T = 1000,\ \beta = 1.5,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2,\ \lambda = 0.8.$

MLE	-0.000	0.0002	0.001	0.0033	-0.001	0.0041
	(.0005)	(.0000)	(.0013)	(.0001)	(.0020)	(.0002)
QMLE	-0.000	0.0003	-0.005	0.0059	-0.002	0.0070
	(.0007)	(.0000)	(.0020)	(.0005)	(.0024)	(.0006)
NAV1	-0.000	0.0005	-0.007	0.0143	-0.069	0.0186
	(.0007)	(.0000)	(.0038)	(.0008)	(.0037)	(.0008)
NAV2	-0.000	0.0005	-0.023	0.0128	-0.079	0.0186
	(.0007)	(.0000)	(.0035)	(.0008)	(.0035)	(.0009)
NAV3	-0.000	0.0003	0.019	0.0174	-0.080	0.0220
	(.0006)	(.0000)	(.0041)	(.0012)	(.0039)	(.0011)
OPT	-0.000	0.0003	-0.022	0.0059	-0.012	0.0058
	(.0006)	(.0000)	(.0024)	(.0003)	(.0022)	(.0003)

* Simulation standard errors are in parenthesis.

Table 3. Ratios of estimated standard errors to standard deviations

Model A: $y_t = \delta y_{t-1} + \varepsilon_t$

$h_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1}$, $t = 1,2,\ldots,T$

$\varepsilon_t = \xi_t h_t^{1/2}$, ξ_t i.i.d. $t(5)$

	δ	α_0	α_1
	$T = 500,\ \delta = 0.8,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4.$		
MLE	0.953	1.041	1.000
QMLE	0.962	1.008	0.927
GMM1	0.653	0.761	0.612
GMM2	0.700	0.712	0.570
GMM3	0.722	0.765	0.600
GMM*	0.948	0.956	0.952
	$T = 500,\ \delta = 0.6,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2.$		
MLE	1.010	1.089	0.969
QMLE	0.991	1.042	1.001
GMM1	0.811	0.860	0.630
GMM2	0.850	0.830	0.600
GMM3	0.856	0.840	0.688
GMM*	0.971	0.982	0.940
	$T = 1000,\ \delta = 0.6,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4.$		
MLE	0.988	1.026	0.978
QMLE	1.028	1.009	0.954
GMM1	0.708	0.674	0.585
GMM2	0.672	0.681	0.567
GMM3	0.722	0.765	0.600
GMM*	1.000	1.030	0.978
	$T = 1000,\ \delta = 0.6,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2.$		
MLE	1.019	1.032	0.950
QMLE	1.104	1.030	0.958
GMM1	0.676	0.779	0.596
GMM2	0.912	0.858	0.657
GMM3	0.714	0.671	0.600
GMM*	1.018	1.008	0.933

Table 4. Ratios of estimated standard errors to standard deviations

Model B: $y_t = \beta x_t + \varepsilon_t$

$x_t = \lambda x_{t-1} + \omega_t$, ω_t i.i.d $\mathcal{N}(0.1)$

$h_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1}$, $t = 1,2,\ldots,T$

$\varepsilon_t = \xi_t h_t^{1/2}$, ξ_t i.i.d. $t(5)$

	β	α_0	α_1
$T = 500$, $\beta = 1.0$, $\alpha_0 = 0.6$, $\alpha_1 = 0.4$, $\lambda = 0.8$.			
MLE	0.985	1.033	0.987
QMLE	1.009	0.998	0.900
GMM1	0.976	0.786	0.699
GMM2	0.955	0.766	0.622
GMM3	1.040	0.899	0.848
GMM*	0.986	0.983	0.938
$T = 500$, $\beta = 1.5$, $\alpha_0 = 0.8$, $\alpha_1 = 0.2$, $\lambda = 0.8$.			
MLE	1.010	1.033	0.934
QMLE	1.095	1.035	1.037
GMM1	0.992	0.859	0.790
GMM2	0.976	0.841	0.748
GMM3	1.018	0.985	0.936
GMM*	0.989	0.984	0.937

Table 4. continue

	$T = 1000,\ \beta = 1.0,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4,\ \lambda = 0.8.$		
MLE	1.037	1.027	0.978
QMLE	1.026	0.931	1.014
GMM1	0.989	0.745	0.684
GMM2	0.985	0.741	0.634
GMM3	0.985	0.859	0.791
GMM*	0.971	0.977	0.974
	$T = 1000,\ \beta = 1.5,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2,\ \lambda = 0.8.$		
MLE	1.020	1.027	0.953
QMLE	1.029	1.036	0.968
GMM1	0.988	0.870	0.797
GMM2	0.982	0.849	0.750
GMM3	0.980	0.959	0.908
GMM*	0.976	0.984	0.955

Table 5. Coverage of confidence intervals [a]

Model A: $y_t = \delta y_{t-1} + \varepsilon_t$

$$h_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1}, \quad t = 1,\dots,T$$

$$\varepsilon_t = \xi_t h_t^{1/2}, \quad \xi_t \text{ i.i.d. } t(5)$$

	δ	α_0	α_1
	$T = 500,\ \delta = 0.8,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4.$		
MLE	0.934	0.958	0.934
QMLE	0.929	0.927	0.893
GMM1	0.883	0.872	0.341
GMM2	0.884	0.864	0.320
GMM3	0.897	0.862	0.327
GMM*	0.934	0.904	0.880
	$T = 500,\ \delta = 0.6,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2.$		
MLE	0.952	0.972	0.924
QMLE	0.944	0.937	0.868
GMM1	0.900	0.914	0.477
GMM2	0.896	0.888	0.466
GMM3	0.898	0.897	0.502
GMM*	0.953	0.911	0.861

Table 5. continue

	$T = 1000,\ \delta = 0.8,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4.$		
MLE	0.951	0.960	0.954
QMLE	0.950	0.934	0.910
GMM1	0.911	0.776	0.335
GMM2	0.907	0.769	0.312
GMM3	0.910	0.755	0.348
GMM*	0.960	0.929	0.913
	$T = 1000,\ \delta = 0.6,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2.$		
MLE	0.951	0.964	0.932
QMLE	0.954	0.938	0.897
GMM1	0.924	0.928	0.476
GMM2	0.930	0.923	0.486
GMM3	0.939	0.918	0.531
GMM*	0.955	0.926	0.881

[a]This table shows the number of times in 1000 replications that the 95% confidence intervals of estimators include the true value of parameters

Table 6. Coverage of confidence intervals [b]

Model B: $y_t = \beta x_t + \varepsilon_t$

$$x_t = \lambda x_{t-1} + \omega_t, \quad \omega_t \text{ i.i.d } \mathcal{N}(0,1)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1}, \quad t = 1,\ldots,T$$

$$\varepsilon_t = \xi_t h_t^{1/2}, \quad \xi_t \text{ i.i.d. } t(5)$$

	β	α_0	α_1
	$T = 500,\ \beta = 1.0,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4,\ \lambda = 0.8.$		
MLE	0.955	0.955	0.933
QMLE	0.955	0.916	0.877
GMM1	0.943	0.871	0.486
GMM2	0.949	0.857	0.402
GMM3	0.952	0.887	0.514
GMM*	0.950	0.879	0.882
	$T = 500,\ \beta = 1.5,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2,\ \lambda = 0.8.$		
MLE	0.946	0.959	0.912
QMLE	0.961	0.929	0.851
GMM1	0.950	0.906	0.681
GMM2	0.948	0.858	0.634
GMM3	0.949	0.951	0.726
GMM*	0.948	0.882	0.877

Table 6. continue

	$T = 1000, \beta = 1.0, \alpha_0 = 0.6, \alpha_1 = 0.4, \lambda = 0.8.$		
MLE	0.966	0.958	0.947
QMLE	0.960	0.930	0.900
GMM1	0.952	0.898	0.472
GMM2	0.944	0.834	0.413
GMM3	0.955	0.808	0.517
GMM*	0.948	0.897	0.914
	$T = 1000, \beta = 1.5, \alpha_0 = 0.8, \alpha_1 = 0.2, \lambda = 0.8.$		
MLE	0.950	0.960	0.936
QMLE	0.957	0.925	0.901
GMM1	0.945	0.918	0.671
GMM2	0.934	0.900	0.639
GMM3	0.956	0.941	0.710
GMM*	0.949	0.897	0.910

[b] See note in table 6.

Table 7. Wilcoxon Matched-pairs Signed-ranks Test.

Model A: $y_t = \delta y_{t-1} + \varepsilon_t$

$h_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1}$, $t = 1,2,\ldots,T$

$\varepsilon_t = \xi_t h_t^{1/2}$, ξ_t i.i.d $t(5)$

	δ		α_0		α_1	
	Rank(-)[a]	Z[b]	Rank(-)	Z	Rank(-)	Z
	$T = 500,\ \delta = 0.8,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4$					
MLE-QMLE	535.7	-7.32 (0.00)	550.1	-9.76 (0.00)	527.4	-7.42 (0.00)
MLE-NAV1	554.3	-10.71 (0.00)	572.0	-16.69 (0.00)	538.2	-21.57 (0.00)
MLE-OPT	517.5	-4.50 (0.00)	543.9	-7.79 (0.00)	529.5	-12.95 (0.00)
QMLE-NAV1	531.6	-5.48 (0.00)	550.9	-10.54 (0.00)	533.2	-18.42 (0.00)
QMLE-OPT	473.0	-2.49 (0.01)	474.8	-2.40 (0.02)	517.3	-8.34 (0.00)
	$T = 500,\ \delta = 0.6,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2$					
MLE-QMLE	549.3	-8.75 (0.00)	558.9	-11.08 (0.00)	528.2	-7.88 (0.00)
MLE-NAV1	538.5	-8.09 (0.00)	546.2	-9.79 (0.00)	532.6	-14.46 (0.00)
MLE-OPT	518.8	-3.73 (0.00)	538.3	-8.43 (0.00)	515.6	-9.85 (0.00)
QMLE-NAV1	493.1	-0.08 (0.94)	484.2	-0.47 (0.64)	500.9	-8.03 (0.00)
QMLE-OPT	452.5	-5.89 (0.00)	470.4	-3.86 (0.00)	476.9	-3.56 (0.00)

Table 7. continue

	$T = 1000,\ \delta = 0.8,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4$					
MLE-QMLE	539.8	-7.59	548.0	-11.60	535.4	-7.42
		(0.00)		(0.00)		(0.00)
MLE-NAV1	556.3	-12.39	570.8	-21.15	552.6	-24.50
		(0.00)		(0.00)		(0.00)
MLF-OPT	544.7	-6.77	541.0	-11.28	541.6	-14.29
		(0.00)		(0.00)		(0.00)
QMLE-NAV1	526.2	-6.94	564.4	-16.90	534.1	-22.32
		(0.00)		(0.00)		(0.00)
QMLE-OPT	490.0	-1.32	504.4	-0.55	509.1	-9.05
		(0.19)		(0.59)		(0.00)
	$T = 1000,\ \delta = 0.6,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2$					
MLE-QMLE	531.0	-7.42	539.7	-12.25	539.3	-7.02
		(0.00)		(0.00)		(0.00)
MLE-NAV1	539.2	-9.44	558.7	-12.48	547.7	-18.77
		(0.00)		(0.00)		(0.00)
MLE-OPT	537.1	-4.88	536.4	-10.89	537.4	-12.20
		(0.00)		(0.00)		(0.00)
QMLE-NAV1	513.5	-3.35	523.8	-3.23	532.1	-14.25
		(0.00)		(0.00)		(0.00)
QMLE-OPT	460.9	-3.18	491.2	-1.42	500.1	-6.94
		(0.00)		(0.16)		(0.00)

[a] Rank(-) denotes the negative mean rank between a pair of estimators, e.g. for MLE-QMLE, Rank(-) = 535.7 records the mean rank for which MLE has smaller absolute bias than QMLE.

[b] Z denotes Wilcoxon statistics for testing H_0: the two estimators are equivalent. 2-tailed p-values are given in the parenthesis.

Table 8. Wilcoxon Matched-pairs Signed-ranks Test

Model B: $y_t = \beta x_t + \varepsilon_t$

$x_t = \lambda x_{t-1} + \omega_t, \quad \omega_t$ i.i.d $N(0,1)$

$h_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1}, \quad t = 1,\ldots,T$

$\varepsilon_t = \xi_t h_t^{1/2}, \quad \xi_t$ i.i.d $t(5)$

	β		α_0		α_1	
	Rank(-)[a]	Z[b]	Rank(-)	Z	Rank(-)	Z
$T = 500,\ \beta = 1.0,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4,\ \lambda = 0.8$						
MLE-QMLE	526.6	-2.35 (0.02)	534.9	-7.74 (0.00)	535.6	-7.49 (0.00)
MLE-NAV3	509.4	-0.13 (0.90)	573.3	-19.61 (0.00)	555.9	-21.23 (0.00)
MLE-OPT	494.6	-1.13 (0.26)	538.9	-5.93 (0.00)	552.9	-12.19 (0.00)
QMLE-NAV3	478.4	-1.94 (0.05)	553.6	-14.66 (0.00)	544.9	-17.23 (0.00)
QMLE-OPT	474.6	-2.87 (0.00)	481.9	-2.28 (0.03)	510.8	-6.77 (0.00)
$T = 500,\ \beta = 1.5,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2,\ \lambda = 0.8$						
MLE-QMLE	538.0	-7.88 (0.00)	545.0	-7.86 (0.00)	533.6	-5.84 (0.00)
MLE-NAV3	520.0	-4.71 (0.00)	569.0	-15.59 (0.00)	558.7	-17.99 (0.00)
MLE-OPT	526.5	-4.99 (0.00)	539.7	-5.87 (0.00)	533.4	-9.22 (0.00)
QMLE-NAV3	473.2	-3.56 (0.00)	530.9	-9.45 (0.00)	537.6	-13.21 (0.00)
QMLE-OPT	467.0	-3.32 (0.00)	468.3	-2.32 (0.00)	492.8	-3.68 (0.00)

Table 8. continue

$T = 1000,\ \beta = 1.0,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4,\ \lambda = 0.8$

MLE-QMLE	545.0	-7.63	550.5	-8.70	538.2	-5.83
		(0.00)		(0.00)		(0.00)
MLE-NAV3	529.2	-6.61	559.0	-23.58	557.9	-23.79
		(0.00)		(0.00)		(0.00)
MLE-OPT	518.0	-4.59	548.1	-9.20	540.6	-11.37
		(0.00)		(0.00)		(0.00)
QMLE-NAV3	483.9	-1.86	558.2	-20.09	553.7	-21.34
		(0.06)		(0.00)		(0.00)
QMLE-OPT	485.1	-4.24	501.9	-1.12	515.3	-7.92
		(0.00)		(0.26)		(0.00)

$T = 1000,\ \beta = 1.5,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2,\ \lambda = 0.8$

MLE-QMLE	525.5	-3.04	554.7	-9.82	539.6	-6.28
		(0.00)		(0.00)		(0.00)
MLE-NAV3	514.3	-3.85	572.4	-18.73	560.4	-20.21
		(0.00)		(0.00)		(0.00)
MLE-OPT	535.3	-2.84	536.0	-9.28	542.8	-10.10
		(0.01)		(0.00)		(0.00)
QMLE-NAV3	503.1	-1.08	552.2	-11.29	544.1	-15.49
		(0.28)		(0.00)		(0.00)
QMLE-OPT	488.4	-0.36	489.0	-1.60	515.8	-5.01
		(0.72)		(0.11)		(0.00)

[a,b] See footnote in table 7.

Table 9. Kolmogorov-Smirnov Test of Normality

Model A: $y_t = \delta y_{t-1} + \varepsilon_t$

$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \quad t = 1,\ldots,T$

$\varepsilon_t = \xi_t h_t^{1/2}, \quad \xi_t$ i.i.d $t(5)$

	δ		α_0		α_1	
	K-S	P-value	K-S	P-value	K-S	P-value
	$T = 500,\ \delta = 0.8,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4$					
MLE	1.229	0.098	1.351	0.052	1.291	0.071
QMLE	2.842	0.000	2.906	0.000	2.770	0.000
NAV1	2.462	0.000	1.734	0.005	2.309	0.000
OPT	1.555	0.016	1.870	0.002	0.910	0.379
	$T = 500,\ \delta = 0.6,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2$					
MLE	1.168	0.131	1.319	0.062	0.957	0.319
QMLE	3.996	0.000	5.234	0.000	5.739	0.000
NAV1	1.017	0.253	1.174	0.127	2.280	0.000
OPT	0.791	0.802	0.003	0.300	1.244	0.091
	$T = 1000,\ \delta = 0.6,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4$					
MLE	0.972	0.301	0.720	0.678	0.987	0.284
QMLE	1.630	0.010	4.638	0.000	4.793	0.000
NAV1	2.304	0.000	2.017	0.001	2.589	0.000
OPT	0.817	0.516	1.472	0.026	1.088	0.187
	$T = 1000,\ \delta = 0.6,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2$					
MLE	0.781	0.576	0.862	0.447	0.843	0.477
QMLE	3.354	0.000	4.314	0.000	2.547	0.000
NAV1	2.834	0.000	1.737	0.004	2.554	0.000
OPT	0.638	0.810	1.461	0.028	1.534	0.018

Table 10. Kolmogorov-Smirnov Test of Normality

Model B: $y_t = \beta x_t + \varepsilon_t$

$x_t = \lambda x_{t-1} + \omega_t$, ω_t i.i.d $N(0,1)$

$h_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1}$, $t = 1,\ldots,T$

$\varepsilon_t = \xi_t h_t^{1/2}$, ξ_t i.i.d $t(5)$

	β		α_0		α_1	
	K-S	P-value	K-S	P-value	K-S	P-value
$T = 500$, $\beta = 1.0$, $\alpha_0 = 0.6$, $\alpha_1 = 0.4$, $\lambda = 0.8$						
MLE	1.438	0.032	1.196	0.114	0.377	0.999
QMLE	5.829	0.000	4.530	0.000	3.540	0.000
NAV3	0.567	0.905	1.196	0.115	2.322	0.000
OPT	0.335	1.000	1.433	0.033	0.944	0.277
$T = 500$, $\beta = 1.5$, $\alpha_0 = 0.8$, $\alpha_1 = 0.2$, $\lambda = 0.8$						
MLE	0.624	0.832	1.186	0.120	0.935	0.346
QMLE	4.359	0.000	9.394	0.000	6.198	0.000
NAV3	0.617	0.841	1.504	0.022	1.952	0.001
OPT	0.358	0.998	1.096	0.181	2.013	0.001
$T = 1000$, $\beta = 1.0$, $\alpha_0 = 0.8$, $\alpha_1 = 0.2$, $\lambda = 0.8$						
MLE	0.660	0.777	0.670	0.761	0.813	0.524
QMLE	2.839	0.000	2.101	0.000	2.217	0.000
NAV3	0.705	0.703	1.563	0.015	2.656	0.000
OPT	0.859	0.452	1.235	0.095	1.013	0.256
$T = 1000$, $\beta = 1.5$, $\alpha_0 = 0.8$, $\alpha_1 = 0.2$, $\lambda = 0.8$						
MLE	0.633	0.817	0.841	0.479	0.793	0.556
QMLE	8.172	0.000	9.078	0.000	4.921	0.000
NAV3	0.750	0.627	0.997	0.273	2.204	0.000
OPT	0.865	0.443	0.816	0.518	1.619	0.011

Table 11. Ratios of MSE and Ratios of Asymptotic variances [a]

Model A: $y_t = \delta y_{t-1} + \varepsilon_t$

$h_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1}, \quad t = 1,\ldots,T$

$\varepsilon_t = \xi_t h_t^{1/2}, \quad \xi_t$ i.i.d. $t(5)$

	δ		α_0		α_1	
	MSE	AV	MSE	AV	MSE	AV
	$T = 500,\ \delta = 0.8,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4.$					
MLE/QMLE	0.743	0.731	0.526	0.556	0.505	0.591
MLE/NAV1	0.331	0.706	0.205	0.463	0.270	0.444
MLE/OPT	0.824	0.834	0.562	0.697	0.705	0.857
QMLE/NAV1	0.447	0.967	0.390	0.834	0.534	0.751
QMLE/OPT	1.110	1.141	1.069	1.254	1.395	1.689
	$T = 500,\ \delta = 0.6,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2.$					
MLE/QMLE	0.688	0.714	0.509	0.548	0.566	0.535
MLE/NAV1	0.604	0.840	0.441	0.734	0.537	0.517
MLE/OPT	0.790	0.855	0.554	0.698	0.846	0.955
QMLE/NAV1	0.825	1.230	0.895	1.321	0.944	1.034
QMLE/OPT	1.149	1.196	1.088	1.274	1.494	1.959

Table 11. continue

$T = 1000,\ \delta = 0.8,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4.$

MLE/QMLE	0.699	0.684	0.503	0.519	0.488	0.514
MLE/NAV1	0.300	0.583	0.104	0.334	0.157	0.250
MLE/OPT	0.806	0.788	0.589	0.603	0.635	0.697
QMLE/NAV1	0.429	0.901	0.207	0.642	0.321	0.534
QMLE/OPT	1.444	1.153	1.171	1.140	1.300	1.489

$T = 1000,\ \delta = 0.6,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2.$

MLE/QMLE	0.753	0.642	0.483	0.485	0.643	0.506
MLE/NAV1	0.345	0.785	0.307	0.565	0.411	0.522
MLE/OPT	0.832	0.834	0.548	0.585	0.855	0.813
QMLE/NAV1	0.458	1.223	0.634	1.165	0.639	1.033
QMLE/OPT	1.106	1.300	1.134	1.206	1.331	1.609

[a] MSE denotes the ratios of mean square errors of estimators and AV denotes the ratios of asymptotic variances of estimators.

Table 12. Ratios of MSE and ratios of asymptotic variances [b]

Model B: $y_t = \beta x_t + \varepsilon_t$

$x_t = \lambda x_{t-1} + \omega_t$, ω_t i.i.d $N(0,1)$

$h_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1}$, $t = 1,\ldots,T$

$\varepsilon_t = \xi_t h_t^{1/2}$, ξ_t i.i.d. $t(5)$

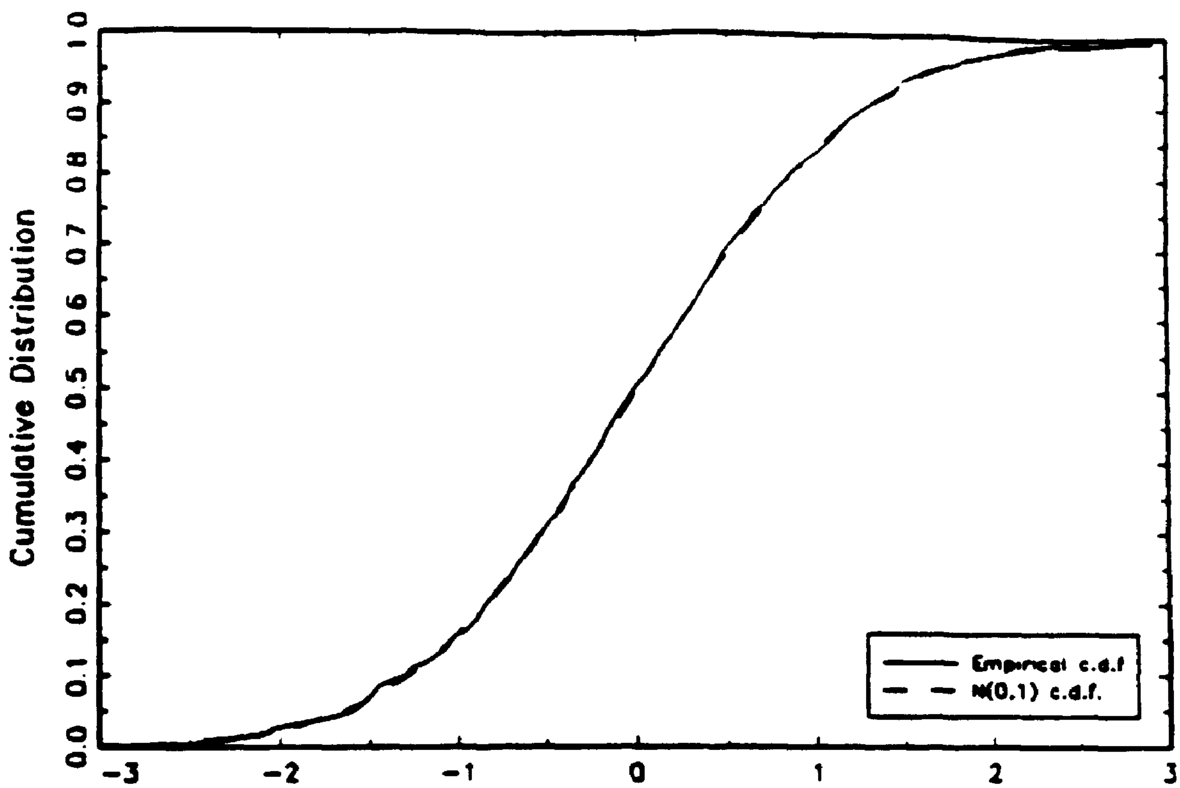
	β		α_0		α_1	
	MSE	AV	MSE	AV	MSE	AV
	$T = 500$, $\beta = 1.0$, $\alpha_0 = 0.6$, $\alpha_1 = 0.4$, $\lambda = 0.8$.					
MLE/QMLE	0.690	0.659	0.564	0.602	0.499	0.608
MLE/NAV3	0.833	0.747	0.141	0.215	0.204	0.656
MLE/OPT	0.914	0.912	0.637	0.774	0.694	0.880
QMLE/NAV3	1.208	1.134	0.251	0.357	0.409	1.078
QMLE/OPT	1.325	1.384	1.130	1.286	1.390	1.376
	$T = 500$, $\beta = 1.5$, $\alpha_0 = 0.8$, $\alpha_1 = 0.2$, $\lambda = 0.8$.					
MLE/QMLE	0.834	0.709	0.569	0.565	0.610	0.513
MLE/NAV3	0.347	0.833	0.238	0.261	0.244	0.330
MLE/OPT	0.845	0.880	0.636	0.775	0.857	0.930
QMLE/NAV3	1.016	1.175	0.418	0.461	0.391	0.643
QMLE/OPT	1.014	1.242	1.117	1.371	1.371	1.812

Table 12. continue

	$T = 1000,\ \beta = 1.0,\ \alpha_0 = 0.6,\ \alpha_1 = 0.4,\ \lambda = 0.8.$					
MLE/QMLE	0.507	0.521	0.380	0.462	0.667	0.620
MLE/NAV3	0.565	0.626	0.079	0.145	0.130	0.425
MLE/OPT	0.669	0.764	0.553	0.662	0.661	0.686
QMLE/NAV3	1.114	1.200	0.209	0.315	0.194	0.686
QMLE/OPT	1.322	1.464	1.456	1.434	0.991	1.107
	$T = 1000,\ \beta = 1.5,\ \alpha_0 = 0.8,\ \alpha_1 = 0.2,\ \lambda = 0.8.$					
MLE/QMLE	0.722	0.709	0.557	0.550	0.586	0.568
MLE/NAV3	0.726	0.785	0.191	0.223	0.188	0.291
MLE/OPT	0.769	0.839	0.558	0.661	0.709	0.725
QMLE/NAV3	1.006	1.108	0.343	0.406	0.320	0.513
QMLE/OPT	1.066	1.184	1.001	1.203	1.211	1.277

[b] See note on table 11.

Fig. 1a. Empirical Distribution of beta (MLE)



GAUSS [illegible]

Fig. 1b. Empirical Distribution of alpha1 (MLE)

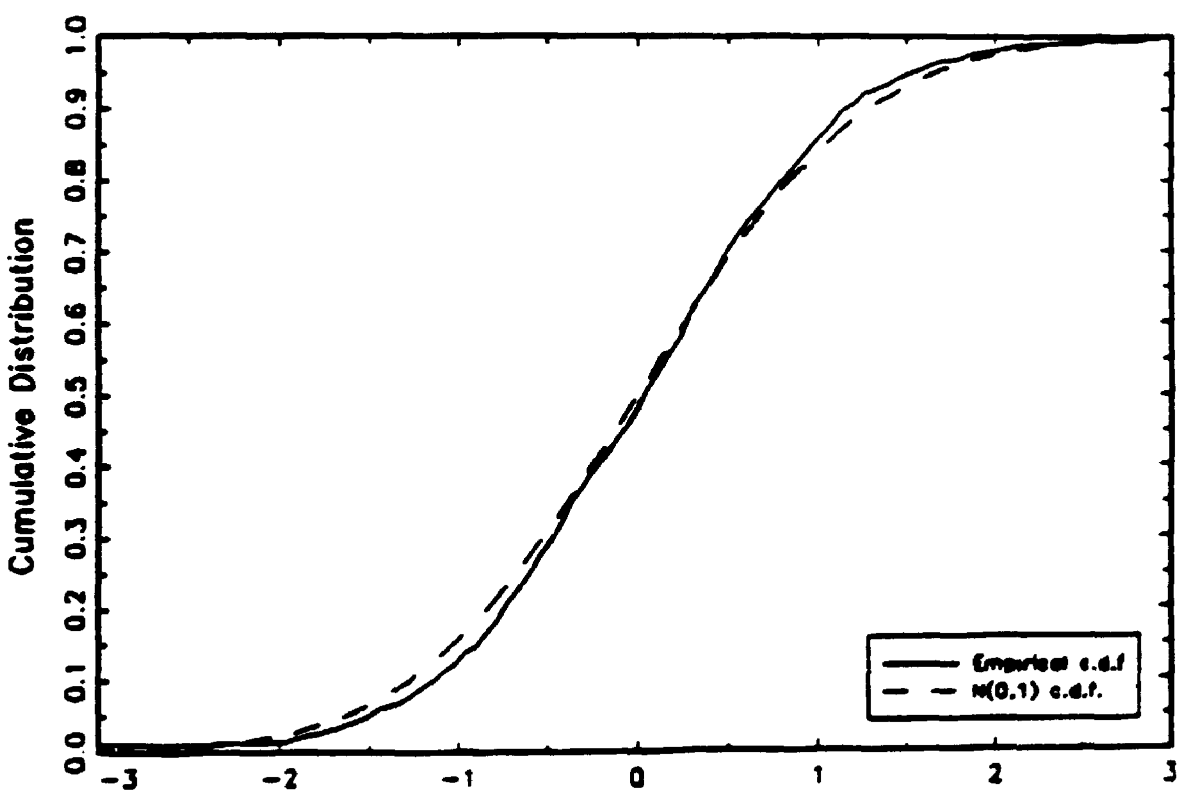


Fig. 2a. Empirical Distribution of beta (OMLE)

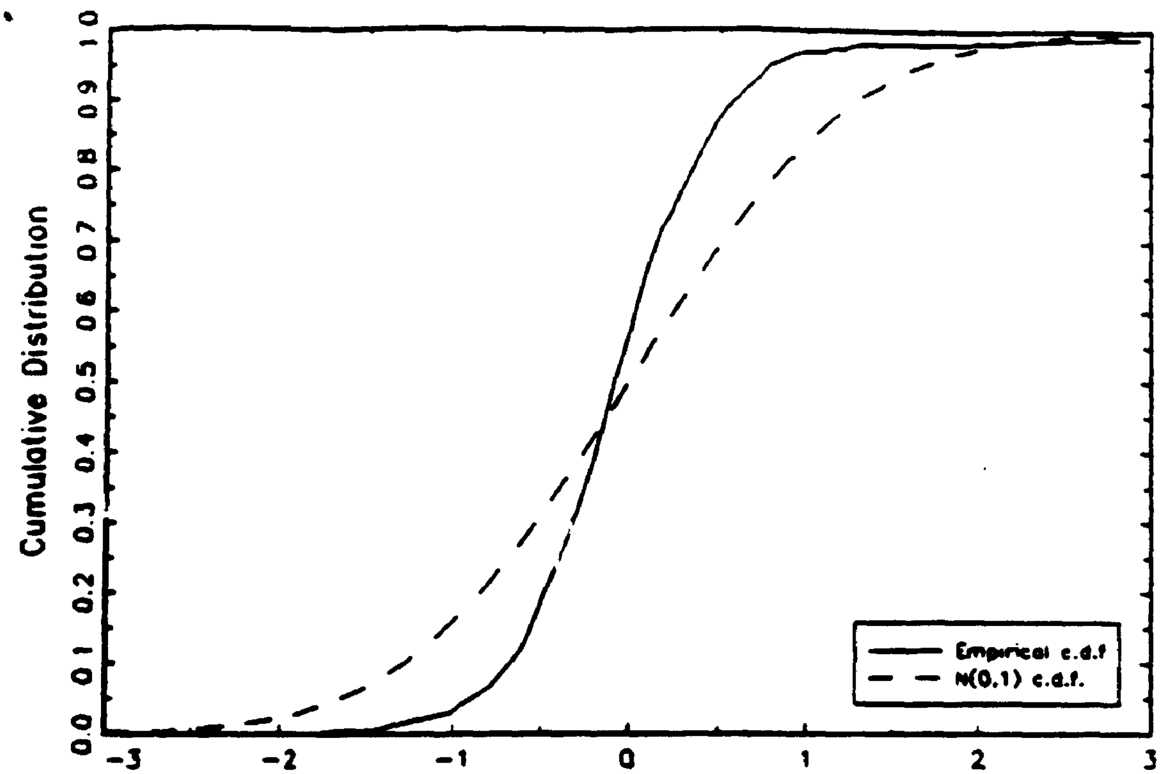


Fig. 2b. Empirical Distribution of alpha1 (OMLE)

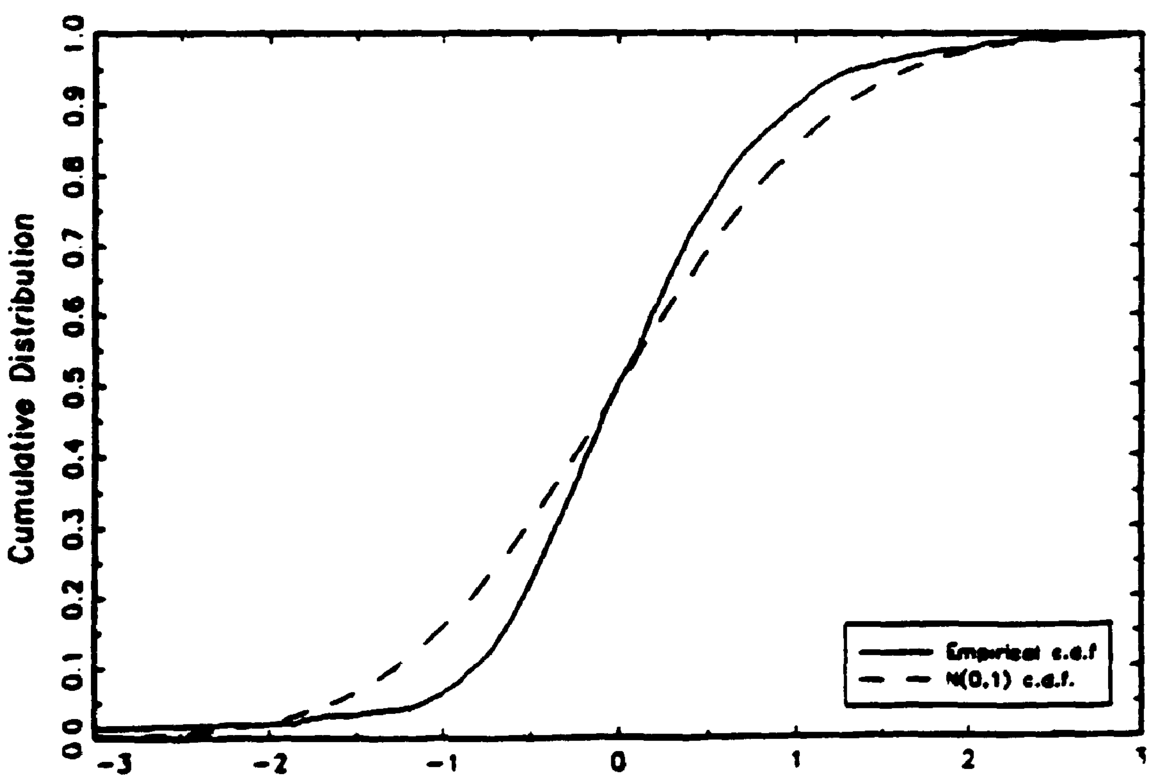


Fig. 3a. Empirical Distribution of beta (NAV3)

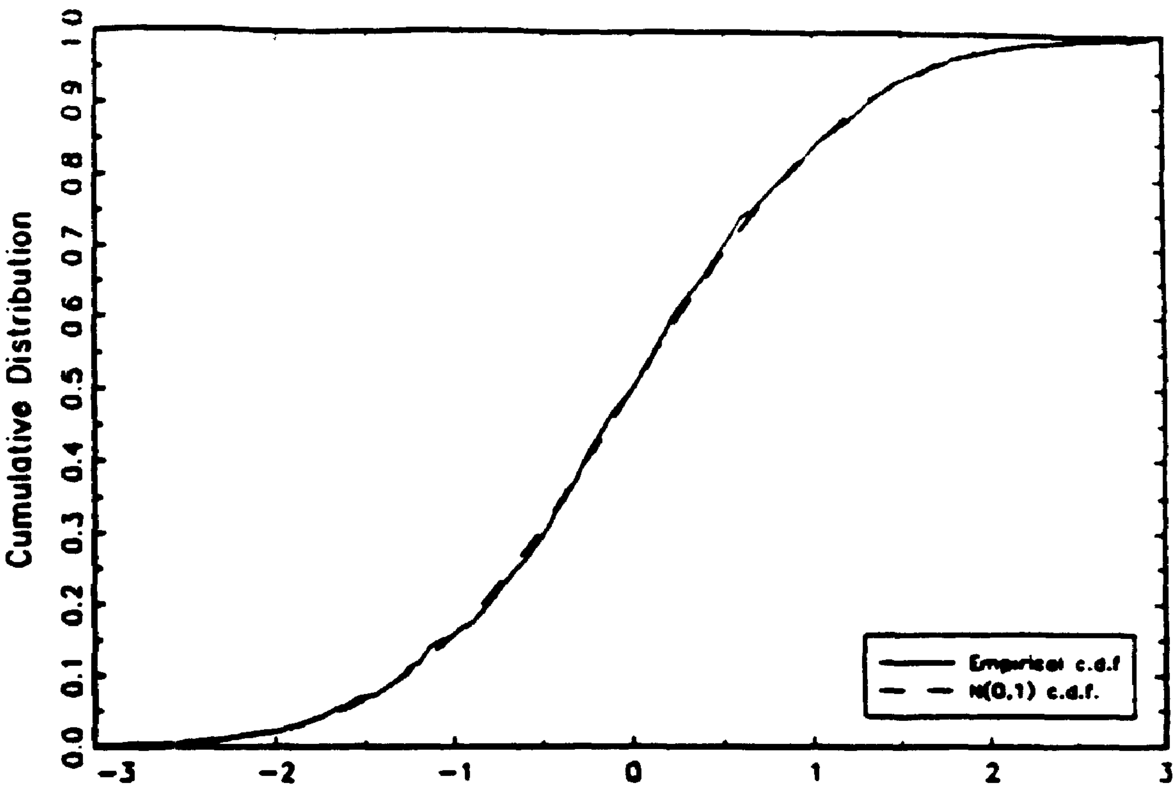


Fig. 3b. Empirical Distribution of alpha1 (NAV3)

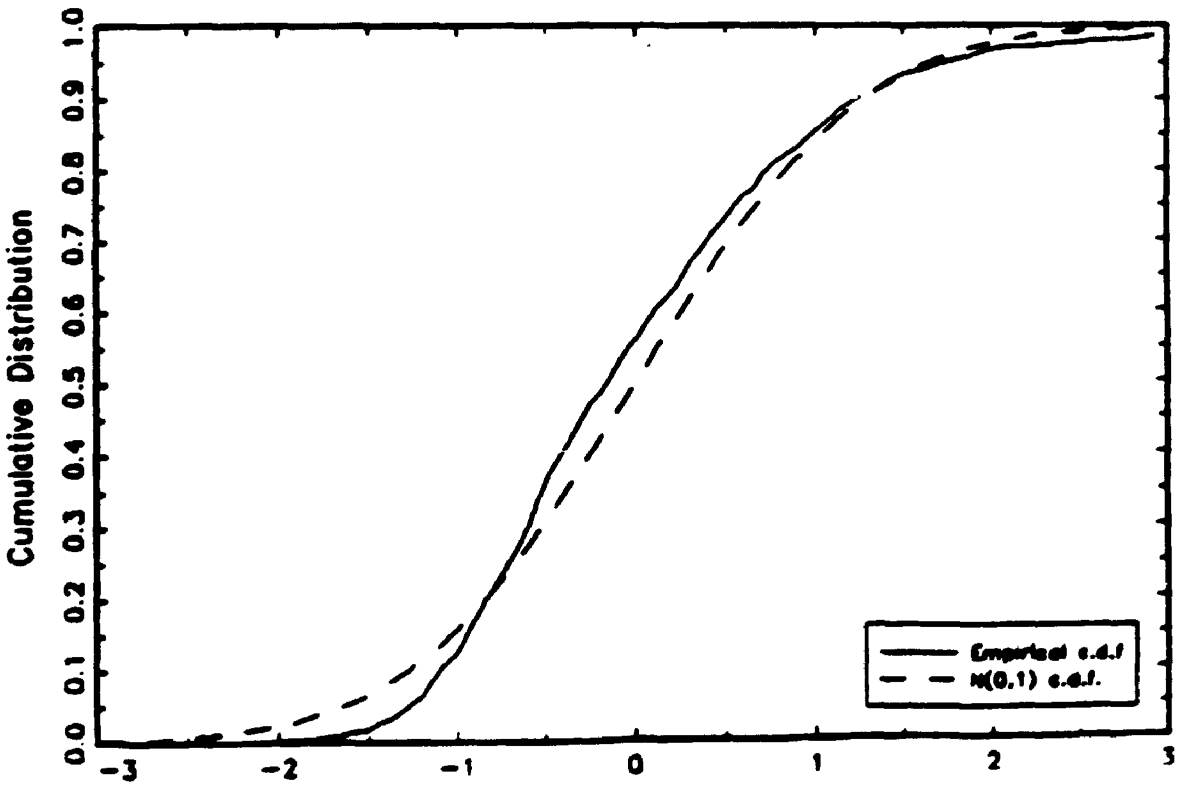


Fig. 4a. Empirical Distribution of beta (OPT)

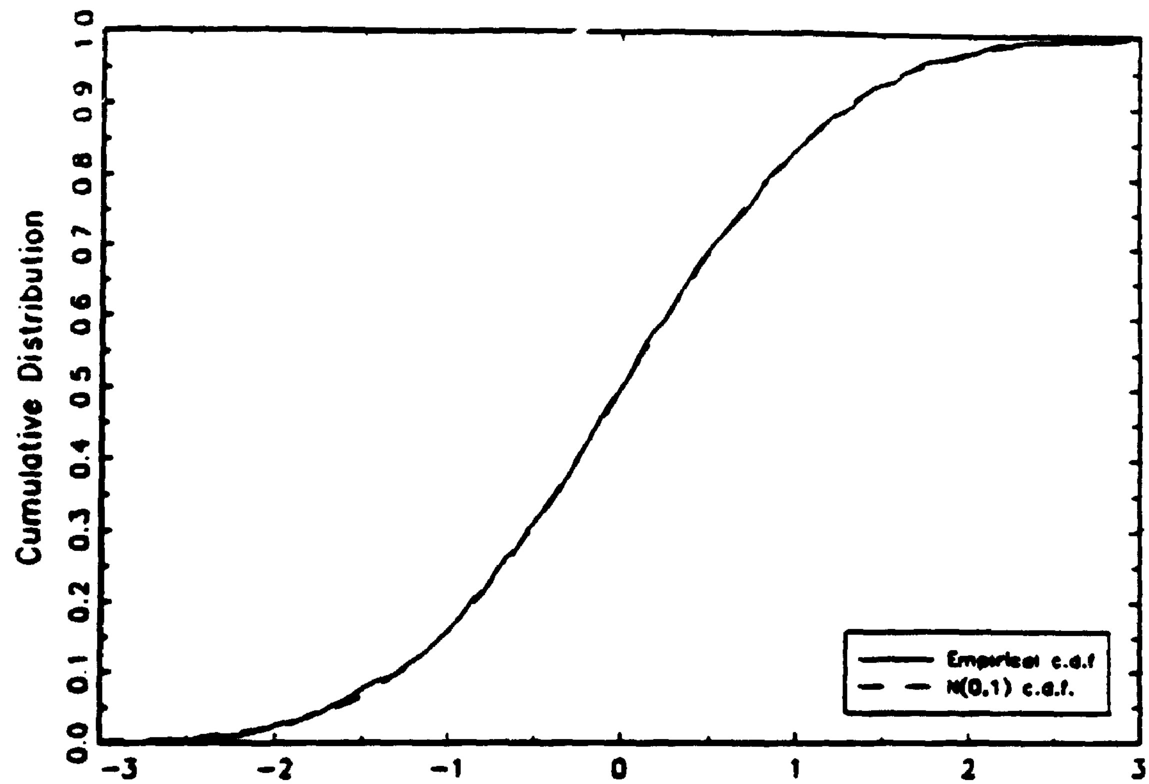


Fig. 4b. Empirical Distribution of alpha1 (OPT)