Continuous Empirical Characteristic Function Estimation of Mixtures of Normal Parameters

Dinghai Xu* John Knight†

This paper develops an efficient method for estimating the discrete mixtures of normal family based on the continuous empirical characteristic function (CECF). An iterated estimation procedure based on the closed form objective distance function is proposed to improve the estimation efficiency. The results from the Monte Carlo simulation reveal that the CECF estimator produces good finite sample properties. In particular, it outperforms the discrete type of methods when the maximum likelihood estimation fails to converge. An empirical example is provided for illustrative purposes.

Keywords Empirical characteristic function; Mixtures of normal.
JEL Classification C13; C15; C16

1 Introduction

Finite mixture models, in particular, discrete mixture of normal (MN) family, have been of considerable interest in recent years across different areas such as biology, engineering, economics, finance, medicine, genetics etc. Excellent examples are documented in Everitt and Hand (1981), Titterington, Smith and Makov (1985) and McLachlan and Peel (2000). One attractive property for mixture models compared to the stationary Gaussian models is that they can capture the leptokurtic, skewed and multimodal characteristics of the empirical data. Furthermore, as a note, any continuous distribution can be approximated arbitrarily well by an appropriate finite MN.

In a general set up of discrete MN, we define an independently and identically distributed (iid) random variable r drawn from K different normal

*Department of Economics, University of Waterloo, Waterloo, Ontario, N2L 3G1, dhxu@uwaterloo.ca

†Department of Economics, University of Western Ontario, London, Ontario, N6A 5C2, jknight@uwo.ca

distributions with probability α_k , i.e.

$$pdf(r_j) = \sum_{k=1}^{K} \alpha_k N(\mu_k, \sigma_k^2) \tag{1}$$

where $j = 1, 2, ..., n$. Defining $\theta = (\alpha', \mu', \sigma')'$, where α, μ, σ are each K $\times$ 1 vectors, we have (3K-1) unknown parameters.

Various methods have been proposed and used for estimating unknown parameters in the MN models. Maximum Likelihood (ML) estimation is one of the most popular methods because of its attractive asymptotic statistical properties. To implement the ML estimation, one of the conditions is that the likelihood function should be bounded in its parameter space. It is well-known that this condition is not always satisfied in the MN case[1] and thus ML estimation may break down in practice. The difficulties in the ML approach have sparked considerable interest in searching for alternative estimation methods such as Method of Moments (MOM) in Cohen (1967) and Day (1969), method of Moment Generating Function (MGF) in Quant and Ramsey (1978) and Schmidt (1982), method of Discrete Empirical Characteristic Function (DECF) in Tran (1998). This class of estimation methods, in essence, minimize the distance between the theoretical components (Moment, MGF or CF) from the model and their empirical counterparts from the data over a set of fixed grid points. However, two major problems arise: one is the choice of the size of the grid points; the other is the "optimal" (if it exists) distance among those grid points[2]. As Tran (1998) and Knight and Yu (2002) documented, these two problems are difficult to handle in practice.

In this paper, we use an iterated procedure based on the continuous ECF (CECF) to efficiently estimate the parameters in the MN models. The proposed estimation method does not suffer from the two problems associated with the grid points since the theoretical CF is continuously matched with its empirical component. We show that with a particular class of weighting functions, a general closed form solution for the objective distance function is available. The existence of the closed form objective function simplifies the parameter estimation and reduces the computational cost. Furthermore, the iterated procedure improves the efficiency of the estimation. Monte Carlo evidence shows that the CECF estimator produces good finite sample properties and works reasonable well even in the cases where the ML method fails.

[1]Please refer to Chapter 2 of Quant (1998) for more discussion on the likelihood function in the MN models .

[2]In the Generalized Method of Moments (GMM), these two problems correspond to: how many moments should be chosen and which moments should be chosen?

The paper is organized as follows. Section 2 presents the CECF estimator in the MN context and outlines the iterated procedure based on the closed form objective distance function. In Section 3, some Monte Carlo experiments are conducted to compare the finite sample performance of the CECF estimator against other alternatives. Section 4 discusses some empirical applications with stock market data. Section 5 concludes the paper.

2 CECF Estimator and Its Asymptotic Properties

2.1 Closed Form Objective Distance Measure Based on the CECF

There are several advantages of using the CF-based estimator. The CF is always uniformly bounded and has a one-to-one correspondence with the distribution function. Hence, inference based on the CF should be theoretically equivalent to that based on the distribution function. As mentioned, the likelihood function is not well behaved in the MN models. In this section, we present an estimator based on the CECF.

Given the MN model defined in (1), the corresponding CF is of the following form,

$$C(t,\theta) = E(e^{itr}) = \sum_{k=1}^{K} \alpha_k \exp(i\mu_k t - \frac{1}{2}\sigma_k^2 t^2) \tag{2}$$

where $i = \sqrt{-1}$.

Noting that $\exp(itx) = \cos(tx) + i\sin(tx)$, (2) can be rewritten as:

$$C(t,\theta) = \sum_{k=1}^{K} \alpha_k \cos(\mu_k t)\exp(-\frac{1}{2}\sigma_k^2 t^2) + i\sum_{k=1}^{K} \alpha_k \sin(\mu_k t)\exp(-\frac{1}{2}\sigma_k^2 t^2) \tag{3}$$

Defining the empirical counterpart (ECF) by the following:

$$C_n(t) = \frac{1}{n}\sum_{j=1}^{n} \exp(itr_j) \tag{4}$$

where r_j, $j = 1, 2, ..., n$, constitutes a random sample from the distribution in (1).

It can also be written as,

$$C_n(t) = \frac{1}{n}\sum_{j=1}^{n}\cos(tr_j) + i[\frac{1}{n}\sum_{j=1}^{n}\sin(tr_j)] \tag{5}$$

By the Law of Large Numbers (LLN), $C_n(t) \xrightarrow{P} C(t,\theta)$. We define the CECF estimator as the minimizer of the following objective distance measure,

$$D(\theta; r) = \int_{-\infty}^{+\infty} |C_n(t) - C(t,\theta)|^2 w(t)dt \tag{6}$$

where $w(t)$ is some weighting function ensuring the convergence of the integral in (6). In this paper, we use the exponential weighting function, $w(t) = \exp(-bt^2)$, where b is a non-negative real number. This weighting function retains certain computational properties of the Gaussian kernels, which enable us to achieve a general closed form for (6).

Proposition 1 If a random sample is generated from the process (1) and the distance measure is defined as in (6), then the integral can be solved analytically and is given by:

$$\begin{aligned} D(\theta; r) &= \frac{1}{n^2}\sqrt{\frac{\pi}{b}}\sum_{i=1}^{n}\sum_{j=1}^{n}\exp(-\frac{1}{4b}(r_i - r_j)^2) \\ &+ \sum_{k=1}^{K}\alpha_k^2\sqrt{\frac{\pi}{b+\sigma_k^2}} \\ &+ 2\sum_{k=1}^{K}\sum_{h\neq k}^{K}\alpha_k\alpha_h\sqrt{\frac{\pi}{b+0.5(\sigma_k^2+\sigma_h^2)}}\exp(-\frac{(\mu_k-\mu_h)^2}{4b+2(\sigma_k^2+\sigma_h^2)}) \\ &- \frac{2}{n}\sum_{k=1}^{K}[\alpha_k\sqrt{\frac{\pi}{0.5\sigma_k^2+b}}\sum_{j=1}^{n}\exp(-\frac{(r_j-\mu_k)^2}{4b+2\sigma_k^2})] \end{aligned} \tag{7}$$

Proof: See the Appendix.

2.2 Asymptotic Properties

The asymptotic properties of the CF based estimator have been established in Heathcote(1977) and Knight and Yu (2002). In this section, we provide the calculation of the asymptotic covariance matrix of the CECF estimator in the MN model, which will be used in the iterated procedure later.

Proposition 2 Let $\hat{\theta} = \text{argmin}[D(\theta; r)]$, where $D(\theta; r)$ is defined in (7), then,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Lambda^{-1}\Omega\Lambda^{-1}) \tag{8}$$

where Λ and Ω are $(3K-1) \times (3K-1)$ matrices with the $ijth$ element[3],

$$\Lambda_{ij} = E\left(\frac{\partial D^2(\theta; r)}{\partial \theta_i \partial \theta_j}\right) \tag{9}$$

and

$$\Omega_{ij} = E\left(\frac{\partial D(\theta; r)}{\partial \theta_i}\frac{\partial D(\theta; r)}{\partial \theta_j}\right) \tag{10}$$

Proof is available in Heathcote (1977).

From the closed form of the objective function we are sometimes able to solve explicitly the expectations [4] in (9) and (10). Lemma 1 presents an example where the number of mixture components is 2.

Lemma 1 If a random sample is generated from the process (1) when K=2, and the distance measure via the CECF is defined as in (7), then the closed form expressions for Λ and Ω are given by:

$$\Lambda_{ij} = U_{ij} + V_{ij} \times \exp\left(-\frac{(\mu_1 - \mu_2)^2}{4b + 2\sigma_1^2 + 2\sigma_2^2}\right)$$

$$\Omega_{ij} = I_i' W_{ij} I_j$$

where the expressions for I, W, U and V are given in the proof.
Proof: See the Appendix.

2.3 Iterated Procedure

In the exponential weighting function, the parameter b plays an important role with the efficiency of the estimator. In most of the literature b is set to be 1, resulting in $w(t) = \exp(-t^2)$. But this may lead to poor efficiency. To see this, we present a simple example where the number of mixture components is 1, i.e., $r \sim N(\mu, \sigma^2)$. To estimate the two unknowns, μ and σ^2 using the

[3] θ_i and θ_j corresponds to the ith or jth element in the vector θ.

[4] The expectations in (9) and (10) can be also simply approximated by using some numerical methods. The results should be asymptotically equivalent to the results of the closed form solutions.

CECF estimator we apply Proposition 1 with K = 1. Then the closed form distance measure is given by,

$$\begin{aligned} D(\theta; r) &= \frac{1}{n^2}\sqrt{\frac{\pi}{b}}\sum_{i=1}^{n}\sum_{j=1}^{n}\exp(-\frac{1}{4b}(r_i - r_j)^2) \\ &+ \sqrt{\frac{\pi}{b+\sigma^2}} - \frac{2}{n}\sqrt{\frac{\pi}{\frac{1}{2}\sigma^2 + b}}\sum_{j=1}^{n}\exp(-\frac{(r_j-\mu)^2}{4b+2\sigma^2}) \end{aligned} \tag{11}$$

With Proposition 2, it is straightforward to derive the analytical form of the asymptotic covariance for the CECF estimator in the above case [5].

$$var(\hat{\mu}) = \frac{\sigma^2}{n}\left(\frac{b^2 + 2b\sigma^2 + \sigma^4}{b^2 + 2b\sigma^2 + \frac{3}{4}\sigma^4}\right)^{\frac{3}{2}} \tag{12}$$

$$var(\hat{\sigma}^2) = \frac{16(b+\sigma^2)^2}{9n}\left(\frac{16(b+\sigma^2)^3(2b^2 + 4b\sigma^2 + 3\sigma^4)}{(4b^2 + 8b\sigma^2 + 3\sigma^4)^{\frac{5}{2}}} - 1\right) \tag{13}$$

The Asymptotic Relative Efficiency (ARE)[6] is generally less than 1. As Heathcote (1977) and Yu (2004) mentioned, the larger the true variance, the less efficient the CECF estimator using weighting function $\exp(-t^2)$. For instance, when $\sigma^2 = 10$, the ARE is about 61.54%. But instead of fixing $b = 1$, if we adjust the b value, for instance, we increase b value to 20, the ARE is improved significantly to 94.36%. Therefore, with our general exponential weighting function, we may be able to improve the efficiency of the CECF estimator by changing the b value.

In practice, how can we choose b efficiently? We propose the following iterated procedure:

1. Start with an initial value of b, say b_0 ;

2. Given $b = b_0$, perform the CECF procedure to estimate θ_0 , i.e. $\hat{\theta_0} = argmin[D(\theta; r)]$;

3. Use $\hat{\theta_0}$ to calculate the covariance matrix S, say S_0, defined in Proposition 2, each element of which is a function of b;

4. Update b through the minimization of a certain measure (MS), such as the trace or the determinant, of the estimated covariance matrix in step 3 , i.e. $b_1 \in argmin[MS(S_0)]$;

5. Given the updated value of b in 4, repeat the steps from 2 to 4 until $|b_j - b_{j-1}|$ is small enough.

[5](12) and (13) are identical as in Yu (1998). Please refer to Yu (1998) for more derivation details.

[6]Here, the ARE is defined as the trace of the inverse of the information matrix (MLE) over the trace of the asymptotic covariance matrix for the CECF estimator.

3 Monte Carlo Simulation

In this section, we conduct several Monte Carlo experiments to compare the finite sample performance of the CECF procedure against that of the alternatives, including the MLE, MGF and DECF methods. For simplicity, the number of mixture components is set to be 2 [7]. We examined 8 cases (in 4 groups) and each case was replicated 1000 times. Most of the cases have previously been studied by Quandt (1972), Quandt and Ramsey (1978), Schmidt (1982) and Tran (1998). The experimental designs are shown in Table 1 and the corresponding density plots are shown in Figure 1. All the experiments were performed using Matlab version 7.0 on a Pentium IV PC (CPU: 2.40 G HZ; 512 MB of RAM).

In the Monte Carlo environment, the true parameter values are known. Hence we minimize the trace (or the determinant) measure of the asymptotic covariance matrix to determine the optimal b value in each estimation[8]. We report the optimal b values and the corresponding variances of the CECF estimators in Table 2 (Table 2a uses the trace measure and Table 2b uses the determinant measure). Figure 2 correspondingly plots the trace and the determinant over a certain range of b values in each simulation case. We find that the optimal b values are very similar for both the trace and determinant measures.

The CECF estimation procedure is also compared to the methods of MGF, DECF and MLE. The comparison results are shown in the Table 3.1 (Bias) and Table 3.2 (Root of Mean Square Error (RMSE))[9].

Table 3.1 and 3.2 show that, in general, the CECF estimator performs as efficiently as the MLE compared to the MGF and DECF when the MLE does not suffer from the unboundedness problem (see group A, B and C). However, conducting a direct comparison of the RMSEs between the CECF and MLE in group A, B and C, we found that RMSEs of MLE are slightly smaller than those of the CECF estimator. The reason is obvious in that theoretically the MLE is the most efficient method under the regularity conditions and the ARE (MLE/CECF) is generally less than 1. But, we also expect the difference would decrease as the number of the replications and the sample

[7]Since the general expression for the objective distance for any K is available, the extension (for cases $K > 2$) is straightforward. In the empirical application, we increase the mixture components up to 4.

[8]The iterated procedure in 2.3 was also applied in each case. The b values are very close to the optimal ones.

[9]The Bias measure captures the average distance from the true parameter value and RMSE is often considered as a good statistical measure of efficiency. In our working paper version, other measures were constructed for more detailed comparisons, such as range, mean, standard deviation and coverage rate. To save space, those results are not reported.

size increase. In the cases of group D where the MLE fails, the CECF estimator still performs very well and dominates the other two methods in terms of the bias and RMSE measure.

Table 3.1 and 3.2 also indicate similar results as shown in Quandt and Ramsey (1978), Schmidt (1982) and Tran (1998). (i) Group A: Case A1 and A2 only differ in the sample size in each replication. We found that the bias and RMSE of the five estimates in case A2 (with larger sample size) are, in general, smaller than those in case A1 (with smaller sample size). In other words, increasing sample size will improve the quality of the estimates. As expected, the differences of RMSE between the MLE and CECF decreased. (ii) Group B: Case B1 and B2 only differ in the mixture weights. We found that the bias and RMSE of the five estimates in case B1 (with low asymmetry) are, in general, higher than those in case B2 (with high asymmetry). That is, the quality of estimates gets worse as α increases. (iii) Group C: Case C1 and C2 only differ in the mean of the second mixture component. We found that the bias and RMSE of five estimates in case C1 (with low separability) are, in general, higher than those in case C2 (with high separability). That is, the quality of estimates is improved as the mixture components become separable. As Schmidt (1982) mentioned, it is hard to estimate accurately when two distributions are similar (Case B2, C1 and D1). There is also a severe problem of nonconvergence for the MGF, DECF and MLE in those cases. For instance, in Case D1, the failure rate for the MGF is about 64%; the failure rate for DECF is about 53%; the failure rate for the MLE is about 90% . For this reason, the results are not presented for the MLE. (iv) Group D: Case D1 and D2 only differ in the variance of the second mixture component. In this group, the MLE fails to converge due to the unboundedness of the likelihood. For instance, the failure rate in case D2 is about 86%. Furthermore, we found that the bias and RMSE of the five estimates in case D2 (with low variance) are, in general, slightly lower than those in case D3 (with high variance). In other words, we expect increasing one component's variance will deteriorate the quality of the estimates. Comparison of the results from group D also reveals that the CECF procedure outperforms the alternative discrete-type methods when the MLE fails to converge.

Several reliability experiments were carried out further to examine the asymptotic properties of the CECF estimator. One standard evaluation is to examine the asymptotic distributions of the CECF estimates. Table 4 summarizes the results from the Kolmogorov - Smirnov (KS) normality test. The KS test statistic shows that 25 out of 40 cases normality is not rejected at either 5% or 1% significance level. We expect more robust results if we increase the sample sizes. A direct comparison between case A1 and A2 in Table 4 confirms this expectation. At 5% (or 1%) significant level, the normality is not rejected in 3 out of 5 cases in case A1 (with smaller samples) while the normality is not rejected in 4 out of 5 cases in case A2 (with larger samples).

4 Empirical Application

The data contains 18.5-years (July 2, 1962 to December 31, 1980) of daily returns for 4 stocks[10] in the Dow-Jones Industrial Average and the Standard and Poor's 500 (S& P 500) Composite stock market index[11]. We also update the data set to December 31, 2004 for further empirical illustration. All the data sets are available at the Wharton Research Data Services (WRDS) website.

4.1 Empirical Estimation via the Iterated CECF Procedure

The first sample data set (1962 to 1980) was previously examined by Kon (1984). The estimation technique adopted in Kon (1984) was mainly the likelihood-based procedure. As mentioned, the regularity conditions for the ML estimation procedure do not always hold for the MN models. Instability of the estimates is one of the serious drawbacks. The estimation results from Kon (1984) also confirmed this point. He noted that given a three mixture normal model, only 15 out of 30 stocks in the data were successful in reaching an interior optimum and with mixtures of four only 10 out of 30 stocks reached an interior optimum. The poor performance of the ML estimation procedure indicates that alternative procedures should be adopted. We estimate the MN models (with two up to four mixture components) based on the CECF iterated procedure. Kon (1984) also conducted some model specification experiments for the data set. Results showed that four-mixture components should be flexible enough to capture the empirical characteristics. However, in practice, given any mixture component in the model, our closed form distance measure based on the CECF should be able to generate stable estimates. A discussion on specifications of the number of mixture

[10]The 4 stocks include: Allied Chemical Corp (ACD); Eastman Kodak Co. (EK); General Electric Co. (GE) and Union Carbide Corp. (UK).

[11]In the working paper version, the data set consists of daily returns on 30 stocks and 3 market indices. To save space, we do not report all estimation results in this paper. Please refer to our working paper for more empirical results.

components is provided in the next section. In our estimation procedure, we assume the number of mixture components is given.

To simplify the computational program, we redefine the distance measure in Proposition 1. Realizing that the first part in (7), $\frac{1}{n^2}\sqrt{\frac{\pi}{b}}\sum_{i=1}^{n}\sum_{j=1}^{n}\exp(-\frac{1}{4b}(r_i-r_j)^2)$, does not contain any unknown parameter, ignoring this term will not affect the optimization results and will speed up the computation by avoiding an n by n loop. Therefore, the distance measure is redefined as,

$$\begin{aligned} D(\theta;r)^* &= \sum_{k=1}^{K}\alpha_k^2\sqrt{\frac{\pi}{b+\sigma_k^2}} \\ &+ 2\sum_{k=1}^{K}\sum_{h\neq k}^{K}\alpha_k\alpha_h\sqrt{\frac{\pi}{b+0.5(\sigma_k^2+\sigma_h^2)}}\exp(-\frac{(\mu_k-\mu_h)^2}{4b+2(\sigma_k^2+\sigma_h^2)}) \\ &- \frac{2}{n}\sum_{k=1}^{K}[\alpha_k\sqrt{\frac{\pi}{0.5\sigma_k^2+b}}\sum_{j=1}^{n}\exp(-\frac{(r_j-\mu_k)^2}{4b+2\sigma_k^2})] \end{aligned} \quad (14)$$

The iterated procedure was implemented to generate the CECF estimates for each discrete MN model. All the stocks and market indices were successful reaching an optimum within the parameter space with 2 to 4 mixture components. A few experiments on sensitivity of initial guesses were conducted. The results revealed that the CECF estimator was insensitive for a reasonable range of initial values [12]. The initial guess for b value is 1 for all the CECF experiments. All the empirical parameter estimates and the associated optimal b values are reported in Table 5 (MN2), 6 (MN3) and 7 (MN4) [13]. As mentioned, an updated data set is also used for further empirical illustration. To save space, we only report the empirical results for the MN2 in Table 8. There are several possible interesting interpretations from these empirical estimates. One finding is consistent with Kon (1984) in that there is, in most of the cases, at least one statistically negative mean parameter

[12] For instance, we started with two different groups of initial values for the ACD data with an MN3 model, where the unknown $\theta = (\alpha_1, \alpha_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2)$. The CECF estimates were $\hat{\theta}^1$ = (0.5206, 0.0872, 0.0629, 0.4591, -0.0955, 2.8605, 12.6226, 0.7746) and $\hat{\theta}^2$ = (0.0860, 0.4165, 0.4413, -0.0945, 0.0727, 12.6940, 0.8198, 2.9377). These two sets of estimates can be viewed as the same optimums except the order of the MN components is different. In Kon (1982), the MLE failed to converge in this case.

[13] Compared to Kon 's ML estimates, the CECF did converge to a similar set of optimum when we make comparison for mixtures of two normal model. However, the MLE suffered from some convergence problems when the mixture components are beyond 2, which makes it very difficult for comparisons between these two procedures. The results of ML estimates are only available in the working paper version (Table 6 - 10), see Kon (1982).

estimate. This could be explained by the so-called Monday Effect [14] that there might be one Monday information component in the stock returns. We also find another interesting point similar to Kon (1984) and Tran (1998) that in general, a large proportion of returns are drawn from one or more small-variance distributions, while a relative small proportion of returns are drawn from high-variance distributions. The two findings are also carried through to the updated data set (1962 - 2004). Via a direct comparison of Table 5 and 8 in the case of MN2, we find that the means, in general, are shifted to the left slightly. This might be because of the inclusion of 1987's data (stock market crush) in the updated sample.

4.2 Model Specification - Assessing the Number of Mixture Components

In the CECF estimation procedure, we assume that the number of mixture components, K, is known. In fact, with empirical data, we never know the true data generating process. The information of mixture components must be inferred from the real data. Unfortunately, according to many authors, the determination of the number of clusters remains an unsolved issue. See for example, McLanchan (1987), Thode, Finch and Mendell (1988), Bozdogan (1992), Feng and McCulloch (1996), Polymedis and Titterington (1998) etc and reference therein.

Kon (1984) used an inappropriate Likelihood Ratio (LR) test in the model specification since, with the discrete MN models, the regularity conditions do not hold for - $2log(\lambda)$ [15] to have the usual asymptotic Chi-squared distribution. As Bozdogan (1994) claimed, *" ... To insist to use the LR test in determining the number of component clusters is fruitless, and moreover, it seems to be wrong, since the null hypothesis tested corresponds to the boundary rather than interior of the parameter space..."* A number of alternative procedures have been proposed. One group of procedures are bootstrapping or simulating the LR statistics, for example Mclanchan (1987), Thode, Finch and Mendell (1988), Bozdogan (1992), Feng and McCulloch (1996). Another type of measure is based on the Information Criteria (IC), attempting to balance the increase of fitting property against the increase of unknown parameters with more mixture components. For example, Akaike (1974), Basford and McLachlan (1988). Our purpose in this paper is not focused on testing the model misspecification. We do not present any model selection decision.

[14] French (1980) found that the stock returns on Monday have significant different behavior compared to those on other days during the week.

[15] λ is the likelihood ratio statistics from the two candidate model specifications.

However, we provide some useful information both from empirical data and theoretical model and let readers of interest investigate the model selection procedure.

We present in Table 9 the measures on Akaike Information Criterion (AIC), which is defined as,

$$AIC = -2Log(L) + 2K \tag{15}$$

where L is the likelihood function and K is the number of mixture components.

Kass and Raftery (1995) suggested that the Bayesian Information Criterion (BIC) provides a more reasonable indicator of the number of mixture components, and mentioned that the comparison between two competing models is not significant with a BIC difference less than 2. A BIC difference between 2 and 6 implies a positive significance, while a BIC difference between 6 and 10 (or greater than 10) provides strong evidence in model selections. Table 9 also displays the BIC measures, which is defined as,

$$BIC = -2Log(L) + KLog(N) \tag{16}$$

where L is the likelihood function, K is the number of mixture components and N is the sample size.

Furthermore, Table 9 provides the moment comparisons of the real data (1962 - 1980) across different number of mixture-component specifications. In the mixtures of K normal distribution, it is straightforward to calculate the first four moments: mean, variance, skewness and kurtosis. The formulas are given as follows:

$$\mu = \sum_{k=1}^{K} \alpha_k \mu_k \tag{17}$$

$$\sigma^2 = \sum_{k=1}^{K} \alpha_k (\mu_k^2 + \sigma_k^2) - \mu^2 \tag{18}$$

$$\text{Skewness} = \frac{1}{\sigma^3} \sum_{k=1}^{K} \alpha_k (\mu_k - \mu)[3\sigma_k^3 + (\mu_k - \mu)^2] \tag{19}$$

$$\text{Kurtosis} = \frac{1}{\sigma^4} \sum_{k=1}^{K} \alpha_k [3\sigma_k^4 + 6(\mu_k - \mu)^2 \sigma_k^2 + (\mu_k - \mu)^4] \tag{20}$$

Table 9 indicates that the MN model is very flexible to capture various shapes of the return distributions, especially the heavy tail characteristics. The estimated moments values are close to the corresponding empirical ones. However, the choices of the mixture-component number are not consistent from the moment comparisons, the AIC and BIC measures. We will leave this for further investigation.

5 Conclusion

This paper proposed an iterated CECF procedure for the estimation of the discrete MN models. As there is a general closed-form expression for the objective distance measure, the estimation procedure is relatively easily implemented. The results of the Monte Carlo experiment reveal that the CECF estimator produces good finite sample properties and is a comparable estimator to the standard MLE. In particular, the CECF procedure performs very well against discrete-type methods when the MLE fails to converge. This is also consistent with the findings from the empirical study.

The CF based approaches have been widely used in financial modelling in recent years, see for examples Knight and Satchell (1996), Jiang and Knight (2002), Knight and Yu (2002) etc. One other possible application of the CECF procedure in financial econometrics is to identify daily returns data with different information components. For instance, Kon (1984) stated that stock returns might be drawing from three mixture regimes - a non-information distribution, a firm-specific information distribution and a market-wide information distribution. With our CECF procedure proposed in this paper, we can efficiently estimate any finite mixture-component normal distributions and thus provide a better explanation of the financial mechanism from the empirical side.

Appendix

Proof of Proposition 1 With (3) and (5), we can write,

$$|C_n(t) - C(t,\theta)|^2 = A^2 + B^2$$

where $A = \frac{1}{n}\sum_{j=1}^{n}\cos(tr_j) - \sum_{k=1}^{K}\alpha_k\cos(\mu_k t)\exp(-\frac{1}{2}\sigma_k^2 t^2)$ and $B = \frac{1}{n}\sum_{j=1}^{n}\sin(tr_j) - \sum_{k=1}^{K}\alpha_k\sin(\mu_k t)\exp(-\frac{1}{2}\sigma_k^2 t^2)$.

Then, we decompose $A^2 + B^2$ into the following four parts,

$$\begin{aligned}
A^2 + B^2 &= \frac{1}{n^2}[\sum_{j=1}^{n}\cos(tr_j)]^2 + [\sum_{j=1}^{n}\sin(tr_j)]^2 \\
&+ \sum_{k=1}^{K}\alpha_k^2\exp(-\sigma_k^2 t^2) \\
&+ 2\sum_{k=1}^{K}\sum_{h\neq k}^{K}\alpha_k\alpha_h\exp(-\frac{1}{2}t^2(\sigma_k^2+\sigma_h^2))\cos(t(\mu_k-\mu_h)) \\
&- \frac{2}{n}\sum_{k=1}^{K}\alpha_k\exp(-\frac{1}{2}\sigma_k^2t^2)[\cos(\mu_k t)\sum_{j=1}^{n}\cos(tr_j)+\sin(\mu_k t)\sum_{j=1}^{n}\sin(tr_j)]
\end{aligned}$$

We evaluate each part in the integral with the exponential weighting function.

$$\begin{aligned}
\text{part 1} &= \int \frac{1}{n^2}\left([\sum_{j=1}^{n}\cos(tr_j)]^2 + [\sum_{j=1}^{n}\sin(tr_j)]^2\right)\exp(-bt^2)dt \\
&= \frac{1}{n^2}\int\left(\sum_{j=1}^{n}\cos(tr_j)^2 + 2\sum_{i\neq j}\cos(tr_i)\cos(tr_j) + \sum_{j=1}^{n}\sin(tr_j)^2 + 2\sum_{i\neq j}\sin(tr_i)\sin(tr_j)\right)dt \\
&= \frac{1}{n^2}\sqrt{\frac{\pi}{b}}\sum_{i=1}^{n}\sum_{j=1}^{n}\exp(-\frac{1}{4b}(r_i-r_j)^2)
\end{aligned}$$

$$\begin{aligned}
\text{part 2} &= \int\sum_{k=1}^{K}\alpha_k^2\exp(-\sigma_k^2t^2)\exp(-bt^2)dt = \sum_{k=1}^{K}\int\alpha_k^2\exp(-(\sigma_k^2+b)t^2)dt \\
&= \sum_{k=1}^{K}\alpha_k^2\sqrt{\frac{\pi}{b+\sigma_k^2}}
\end{aligned}$$

$$\begin{aligned}
\text{part 3} &= \int 2\sum_{k=1}^{K}\sum_{h\neq k}^{K}\alpha_k\alpha_h\exp(-\frac{1}{2}t^2(\sigma_k^2+\sigma_h^2))\cos(t(\mu_k-\mu_h))\exp(-bt^2) \\
&= 2\sum_{k=1}^{K}\sum_{h\neq k}^{K}\int\alpha_k\alpha_h\exp(-\frac{1}{2}t^2(\sigma_k^2+\sigma_h^2))\frac{\exp[it(\mu_k-\mu_h)+\exp[-it(\mu_k-\mu_h)]}{2}\exp(-bt^2)dt \\
&= 2\sum_{k=1}^{K}\sum_{h\neq k}^{K}\alpha_k\alpha_h\sqrt{\frac{\pi}{b+0.5(\sigma_k^2+\sigma_h^2)}}\exp(-\frac{(\mu_k-\mu_h)^2}{4b+2(\sigma_k^2+\sigma_h^2)})
\end{aligned}$$

$$
\begin{aligned}
\text{part 4} &= -\frac{2}{n}\int\sum_{k=1}^{K}\alpha_k \exp(-\frac{1}{2}\sigma_k^2 t^2)[\cos(\mu_k t)\sum_{j=1}^{n}\cos(tr_j)+\sin(\mu_k t)\sum_{j=1}^{n}\sin(tr_j)]\exp(-bt^2)dt \\
&= -\frac{2}{n}\sum_{k=1}^{K}\int\alpha_k \exp(-\frac{1}{2}\sigma_k^2 t^2)\sum_{j=1}^{n}\cos(t(r_j-\mu_k))\exp(-bt^2)dt \\
&= -\frac{2}{n}\sum_{k=1}^{K}[\alpha_k\sqrt{\frac{\pi}{0.5\sigma_k^2+b}}\sum_{j=1}^{n}\exp(-\frac{(r_j-\mu_k)^2}{4b+2\sigma_k^2})]
\end{aligned}
$$

Combining the results will yield the solution in Proposition 1.

Proof of Lemma 1 When K=2, the closed form objective in (7) (by ignoring the constant term) is:

$$
\begin{aligned}
D(\theta;r) &= \alpha^2\sqrt{\frac{\pi}{b+\sigma_1^2}}+(1-\alpha^2)\sqrt{\frac{\pi}{b+\sigma_2^2}} \\
&+ 2\alpha(1-\alpha)\sqrt{\frac{\pi}{b+0.5(\sigma_1^2+\sigma_2^2)}}\exp(-\frac{(\mu_1-\mu_2)^2}{4b+2(\sigma_1^2+\sigma_2^2)}) \\
&- \frac{2}{n}[\alpha\sqrt{\frac{\pi}{0.5\sigma_1^2+b}}\sum_{j=1}^{n}\exp(-\frac{(r_j-\mu_1)^2}{4b+2\sigma_1^2})] \\
&- \frac{2}{n}[(1-\alpha)\sqrt{\frac{\pi}{0.5\sigma_2^2+b}}\sum_{j=1}^{n}\exp(-\frac{(r_j-\mu_2)^2}{4b+2\sigma_2^2})]
\end{aligned}
$$

where $\theta=(\alpha,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)$. Referring to Heathcote (1977),

$$
\Lambda=\int\left(\frac{\partial Re[C(t,\theta)]}{\partial\theta}\frac{\partial Re[C(t,\theta)]}{\partial\theta'}+\frac{\partial Im[C(t,\theta)]}{\partial\theta}\frac{\partial Im[C(t,\theta)]}{\partial\theta'}\right)\exp(-bt^2)dt
$$

Re[.] and Im[.] stand for the real and imaginary part of the function [.]. In the case where K=2,

$$
\begin{aligned}
Re[C(t,\theta)] &= \alpha\cos(\mu_1 t)\exp(-\frac{1}{2}\sigma_1^2t^2)+(1-\alpha)\cos(\mu_2 t)\exp(-\frac{1}{2}\sigma_2^2t^2) \\
Im[C(t,\theta)] &= \alpha\sin(\mu_1 t)\exp(-\frac{1}{2}\sigma_1^2t^2)+(1-\alpha)\sin(\mu_2 t)\exp(-\frac{1}{2}\sigma_2^2t^2)
\end{aligned}
$$

It is straightforward to solve the one-dimensional integral in the Λ expression. To display the result compactly, we define the following two symmetric matrices.

$$
U=\begin{bmatrix}
\frac{\sqrt{\pi}}{\sqrt{b+\sigma_1^2}}+\frac{\sqrt{\pi}}{\sqrt{b+\sigma_1^2}} & 0 & 0 & -\frac{\alpha\sqrt{\pi}}{4(b+\sigma_1^2)^{\frac{3}{2}}} & \frac{(1-\alpha)\sqrt{\pi}}{4(b+\sigma_1^2)^{\frac{3}{2}}} \\
 & \frac{\alpha^2\sqrt{\pi}}{2(b+\sigma_1^2)^{\frac{3}{2}}} & 0 & 0 & 0 \\
 & & \frac{(1-\alpha)^2\sqrt{\pi}}{2(b+\sigma_2^2)^{\frac{3}{2}}} & 0 & 0 \\
 & & & \frac{3\alpha^2\sqrt{\pi}}{16(b+\sigma_1^2)^{\frac{5}{2}}} & 0 \\
 & & & & \frac{3(1-\alpha)^2\sqrt{\pi}}{16(b+\sigma_2^2)^{\frac{5}{2}}}
\end{bmatrix}
$$

$$V = \begin{bmatrix} V11 & V12 & V13 & V14 & V15 \\ & V22 & V23 & V24 & V25 \\ & & V33 & V34 & V35 \\ & & & V44 & V45 \\ & & & & V55 \end{bmatrix}$$

With

$$\begin{aligned}
V11 &= \frac{2\sqrt{\pi}}{\sqrt{b+0.5\sigma_1^2+0.5\sigma_2^2}}; \quad V12 = \frac{\alpha\sqrt{\pi}(\mu_1-\mu_2)}{2(b+0.5\sigma_1^2+0.5\sigma_2^2)^{\frac{3}{2}}}; \quad V13 = \frac{(1-\alpha)\sqrt{\pi}(\mu_1-\mu_2)}{2(b+0.5\sigma_1^2+0.5\sigma_2^2)^{\frac{3}{2}}} \\
V14 &= \frac{\alpha\sqrt{\pi}(1-\frac{(\mu_1-\mu_2)^2}{2b+\sigma_1^2+\sigma_2^2})}{4(b+0.5\sigma_1^2+0.5\sigma_2^2)^{\frac{3}{2}}}; \quad V15 = -\frac{(1-\alpha)\sqrt{\pi}(1-\frac{(\mu_1-\mu_2)^2}{2b+\sigma_1^2+\sigma_2^2})}{4(b+0.5\sigma_1^2+0.5\sigma_2^2)^{\frac{3}{2}}}; \quad V22 = 0 \\
V23 &= \frac{\alpha(1-\alpha)\sqrt{\pi}(1-\frac{(\mu_1-\mu_2)^2}{2b+\sigma_1^2+\sigma_2^2})}{(2b+0.5\sigma_1^2+0.5\sigma_2^2)^{\frac{3}{2}}}; \quad V24 = 0; \quad V25 = \frac{\alpha(1-\alpha)\sqrt{\pi}(\mu_1-\mu_2)(1-\frac{(\mu_1-\mu_2)^2}{6b+3\sigma_1^2+3\sigma_2^2})}{8(b+0.5\sigma_1^2+0.5\sigma_2^2)^{\frac{5}{2}}} \\
V33 &= 0; \quad V34 = \frac{\alpha(1-\alpha)\sqrt{\pi}(\mu_1-\mu_2)(1-\frac{(\mu_1-\mu_2)^2}{6b+3\sigma_1^2+3\sigma_2^2})}{8(b+0.5\sigma_1^2+0.5\sigma_2^2)^{\frac{5}{2}}}; \quad V35 = 0 \\
V44 &= 0; \quad V45 = \frac{3\alpha(1-\alpha)\sqrt{\pi}((\mu_1-\mu_2)(1-\frac{(\mu_1-\mu_2)^2}{b+0.5\sigma_1^2+0.5\sigma_2^2}+\frac{(\mu_1-\mu_2)^4}{12(b+0.5\sigma_1^2+0.5\sigma_2^2)^2})}{8(b+0.5\sigma_1^2+0.5\sigma_2^2)^{\frac{5}{2}}}; \quad V55 = 0
\end{aligned}$$

Hence, we can express Λ as,

$$\Lambda_{ij} = U_{ij} + V_{ij} \times \exp\left(-\frac{(\mu_1-\mu_2)^2}{4b+2\sigma_1^2+2\sigma_2^2}\right)$$

Similarly, it is straightforward to derive the first order derivative with respect to the unknown parameters respectively. Recall that, $\Omega_{ij} = \int \left(\frac{\partial D(\theta;r)}{\partial \theta_i}\frac{\partial D(\theta;r)}{\partial \theta_j}\right)\exp(-bt^2)dt$. Through a tedious derivation, the integrals can be solved explicitly. We express the formula in the matrix form, i.e., $\Omega_{ij} = I_i' W_{ij} I_j$. To save space, we do not provide the expressions for I and W. But in our working paper version, we have all the detailed calculations and analytical forms.

References

Basford, K. E. and McLachlan, G.(1988). *Mixture Models - Inference and Applications to Clustering*, Marcel Dekker, INC.

Bozdogan. H.,(1992). Choosing the Number of Component Clusters in the Mixture-Model Using a New Information Complexity Criterion of the Inverse-Fisher Information Matrix, *Information and Classification: Concepts, Methods and Applications*: 44 - 54.

Bozdogan. H., (1994). Mixture- Model Cluster Analysis Using Model Selection Criteria and a New Information Measure of Complexity, In Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach, Kluwer Academic Publishers, 69 - 113.

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.

Everitt, B. S. and Hand, D. J. (1981). *Finite Mixture Distributions*, Chapman and Hall.

Feng, Z. D and McCulloch, C.E. (1995). Using Bootstrap Likelihood Ratios in Finite Mixture Models, *Journal of the Royal Statistics Society*, B (58): 609 - 617.

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

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

French,K. R., (1980). Stock Returns and the Weekend Effect, *Journal of Financial Economics*, 8 (1): 55-70.

Jiang, G. J. and Knight, J., (2002).Estimation of Continuous - Time Processes via the Empirical Characteristic Function,*Journal of Business & Economic Statistics*, 20: 198 - 212.

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

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors, *Journal of American Statistical Association*, 90: 773 - 795 .

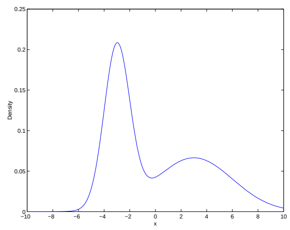
Knight, J. L. and Yu, J., (2002). Empirical Characteristic Function in Time Series Estimation, *Econometric Theory*, 18, 691 - 721.

Kon, S. (1984). Models of Stock Returns - A Comparison, *The Journal of Finance*, 39: 147 - 165.

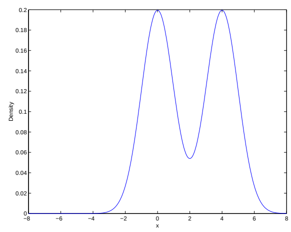
McLachlan, G.(1987). On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in a Normal Mixture, *Applied Statistics*, 38: 318 - 324.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*, Wiley.
Paulson, A. S. , Holcomb, E. W. and Leitch, R.A. (1975). The Estimation of the Parameters of the Stable Laws, *Biometrika*, 62: 163-170.
Polymedis, A. and Titterington, D. M (1998). On the Determinants of the Number of Components in a Mixture, *Statistics & Probability Letters*, 38: 295 - 298.
Quandt, R. E. and Ramsey, J. B. (1978). Estimating Mixtures of Normal Distributions and Switching Regressions, *Journal of the American Statistical Association*, 73: 730 - 738.
Quandt, R. E. (1988). *The Econometrics of Disequilibrium*, Basil Blackwell.
Schmidt, P. (1982). An Improved Version of the Quandt-Ramsey MGF Estimator for Mixtures of Normal Distributions and Switching Regressions, *Econometrica*, 50: 501 - 516.
Thode. H, Finch, S. and Mendell, N.R. (1988). Simulated Percentage Points for the Null Distribution of the Likelihood Ratio Test for a Mixture of Two Normals. *Biometrics*. 44(4): 1195-1201.
Titterington, D. M. , Smith, A. F. M. and Makov, U. E. (1985). *Statistical Analysis of Finite of Mixture Distributions*, John Willey & Son Ltd.
Tran, K. (1998). Estimating Mixtures of Normal Distributions via Empirical Characteristic Function, *Econometric Reviews*, 17(2): 167-183.
Yu, J. (1998). Empirical Characteristic Function in Time Series Estimation and a Test Statistic in Financial Modelling, Ph. D Thesis, The University of Western Ontario.
Yu, J. (2004). Empirical Characteristic Function Estimation and Its Applications, *Econometric Reviews*, 2004: 23, 93-123.

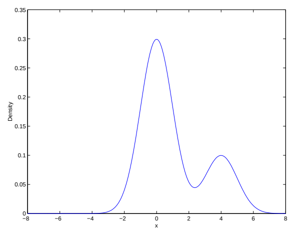
Figure 1. Densities of all the Simulation Cases



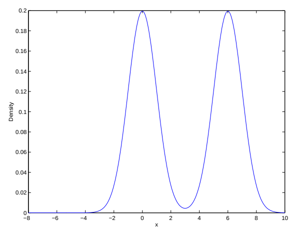
Case A1 and A2



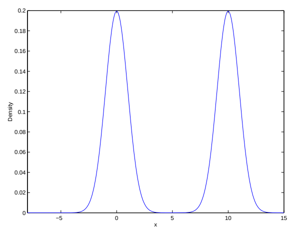
Case B1



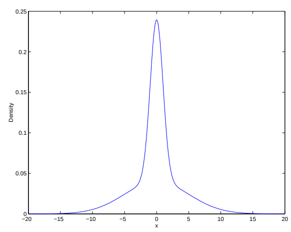
Case B2



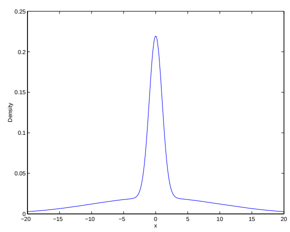
Case C1



Case C2



Case D1



Case D2

Table 1 Monte Carlo Simulation Design

Case	α	μ_1	μ_2	σ_1^2	σ_2^2	n
1 (A1)	0.5	-3	3	1	3	100
2 (A2)	0.5	-3	3	1	3	200
3 (B1)	0.5	0	4	1	1	100
4 (B2)	0.75	0	4	1	1	100
5 (C1)	0.5	0	6	1	1	100
6 (C2)	0.5	0	10	1	1	100
7 (D1)	0.5	0	0	1	5	100
8 (D2)	0.5	0	0	1	10	100

Table 2a Asymptotic Variances of the CECF estimator under Optimal b values (by Trace Minimization)

Case	var(α)	var(μ_1)	var(μ_2)	var(σ_1^2)	var(σ_2^2)	$b*$
1 (A1)	0.0030	0.0284	0.0942	0.0764	0.7205	2.4075
2 (A2)	0.0015	0.0142	0.0471	0.0382	0.3603	2.4075
3 (B1)	0.0033	0.0338	0.0338	0.0821	0.0821	1.2581
4 (B2)	0.0027	0.0206	0.0846	0.0501	0.2053	1.2520
5 (C1)	0.0025	0.0219	0.0219	0.0502	0.0502	1.9609
6 (C2)	0.0025	0.0203	0.0203	0.0417	0.0417	4.6289
7 (D1)	0.0806	0.0572	0.1444	0.4604	3.9221	1.7577
8 (D2)	0.0259	0.0494	0.2505	0.3388	9.9057	2.2387

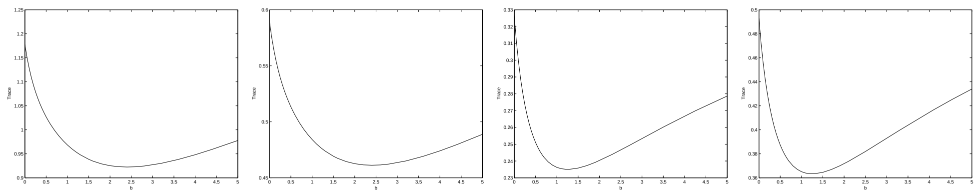
Table 2b Asymptotic Variances of the CECF estimator under Optimal b values (by Determinant Minimization)

Case	var(α)	var(μ_1)	var(μ_2)	var(σ_1^2)	var(σ_2^2)	$b*$
1 (A1)	0.0030	0.0279	0.0933	0.0753	0.7240	2.1492
2 (A2)	0.0015	0.0140	0.0467	0.0377	0.3620	2.1492
3 (B1)	0.0033	0.0343	0.0343	0.0820	0.0820	1.5173
4 (B2)	0.0027	0.0208	0.0867	0.0502	0.2044	1.5247
5 (C1)	0.0025	0.0219	0.0219	0.0503	0.0503	2.1992
6 (C2)	0.0025	0.0203	0.0203	0.0417	0.0417	4.9260
7 (D1)	0.0799	0.0570	0.1458	0.4527	3.9362	1.6229
8 (D2)	0.0232	0.0460	0.2726	0.2627	10.5002	1.2223

Figure 2 Optimal b Values

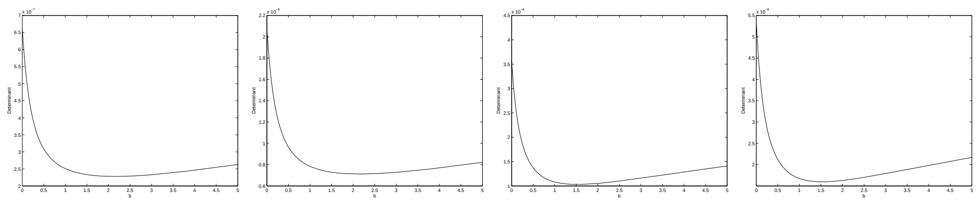
By Trace Minimization:

A1 A2 B1 B2



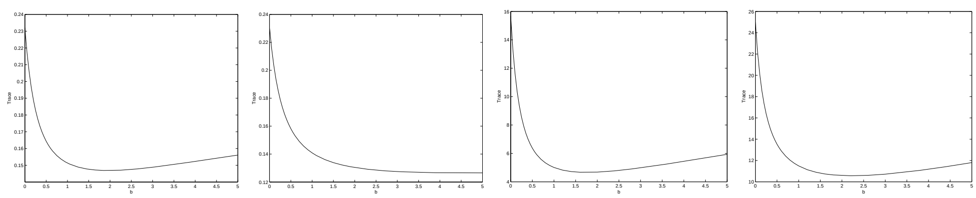
By Determinant Minimization:

A1 A2 B1 B2



By Trace Minimization:

C1 C2 D1 D2



By Determinant Minimization:

C1 C2 D1 D2

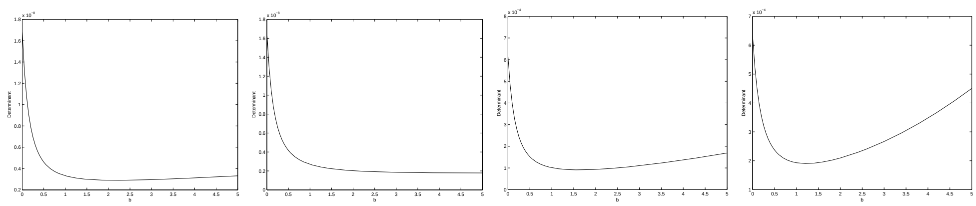


Table 3.1 Bias Comparison of CECF, DECF, MGF and MLE Estimates
("-" means that MLE fails in the case.)

	α	μ_1	μ_2	σ_1^2	σ_2^2
1(A1)					
CECF	0.0028	-0.0036	-0.0114	-0.0121	-0.0004
DECF	-0.0084	-0.0279	0.0472	0.0620	0.6814
MGF	0.0219	0.0793	0.0860	0.2332	0.2180
MLE	0.0034	-0.0025	-0.0083	-0.0237	-0.0408
2(A2)					
CECF	-0.0030	0.0022	0.0082	-0.0235	0.0450
DECF	-0.0149	-0.0244	0.0633	0.0356	0.7910
MGF	0.0165	0.0708	0.1044	0.2228	0.1836
MLE	-0.0020	0.0058	0.0122	-0.0269	-0.0137
3(B1)					
CECF	-0.0016	0.0049	0.0098	-0.0003	-0.0086
DECF	-0.0027	-0.0564	0.0919	0.4644	0.4865
MGF	0.0195	0.0676	0.0833	0.2590	0.2585
MLE	-0.0016	0.0063	0.0096	-0.0122	-0.0186
4(B2)					
CECF	-0.0029	-0.0118	-0.0286	-0.0124	0.0514
DECF	-0.0322	-0.0654	-0.1115	0.1704	1.1304
MGF	0.0469	0.1708	0.2685	0.3950	1.4732
MLE	-0.0023	-0.0096	-0.0252	-0.0185	0.0181
5(C1)					
CECF	0.0006	-0.0070	0.0063	0.0025	-0.0103
DECF	-0.0043	-0.0383	0.0552	0.2162	0.2022
MGF	-0.0012	-0.0115	-0.0082	0.0904	0.1087
MLE	0.0005	-0.0067	0.0061	-0.0076	-0.0134
6(C2)					
CECF	-0.0022	0.0014	0.0069	0.0005	-0.0189
DECF	-0.0094	0.0362	-0.0936	0.2562	0.2653
MGF	-0.0074	-0.0207	-0.0764	0.0719	0.1387
MLE	-0.0022	0.0007	0.0060	-0.0030	-0.0220
7(D1)					
CECF	0.0605	0.0001	0.0173	0.2585	0.0105
DECF	0.0630	0.0004	0.0335	0.0291	2.1875
MGF	0.0769	0.0067	0.0241	0.2287	0.7363
MLE	-	-	-	-	-
8(D2)					
CECF	0.0442	0.0128	-0.0014	0.2788	0.1941
DECF	-0.0087	0.0095	-0.0679	-0.0104	2.0466
MGF	0.0573	0.0098	0.0098	0.3782	0.5114
MLE	-	-	-	-	-

Table 3.2 RMSE Comparison of CECF, DECF, MGF and MLE Estimates

	α	μ_1	μ_2	σ_1^2	σ_2^2
1(A1)					
CECF	0.1739	0.5506	0.9932	0.9386	2.9606
DECF	0.1901	0.5896	1.0716	1.1702	4.7028
MGF	0.2019	0.6969	1.0677	1.6131	3.2438
MLE	0.1651	0.5143	0.9045	0.8061	2.5867
2(A2)					
CECF	0.1727	0.5446	1.0081	0.8734	2.8875
DECF	0.1917	0.5759	1.0946	1.0565	4.8631
MGF	0.1905	0.6613	1.0931	1.4920	2.8478
MLE	0.1660	0.5036	0.9267	0.7875	2.4525
3(B1)					
CECF	0.1880	0.6371	0.6033	0.9406	0.9228
DECF	0.2482	0.9678	1.1277	2.6268	2.7071
MGF	0.2928	1.0993	1.0832	4.2026	5.6895
MLE	0.1869	0.6258	0.5922	0.8892	0.9040
4(B2)					
CECF	0.1757	0.4654	1.0733	0.7189	1.8494
DECF	0.3395	0.6183	2.3995	1.4892	6.4183
MGF	0.2877	1.0308	2.4973	2.1365	15.8561
MLE	0.1716	0.4497	1.0544	0.7136	1.6518
5(C1)					
CECF	0.1587	0.4807	0.4588	0.7441	0.7223
DECF	0.1729	0.5639	0.5684	1.3674	1.3159
MGF	0.1510	0.4995	0.5409	2.2495	1.3553
MLE	0.1577	0.4673	0.4404	0.6932	0.6629
6(C2)					
CECF	0.1596	0.4527	0.4649	0.6433	0.6565
DECF	0.1817	0.5602	0.4805	1.4895	1.5112
MGF	0.1581	0.5138	0.5042	1.3494	1.4460
MLE	0.1597	0.4509	0.4594	0.6302	0.6379
7(D1)					
CECF	0.2575	0.6143	0.9268	1.4060	1.1538
DECF	0.7049	0.8092	4.8173	1.9240	12.7466
MGF	0.4171	0.6984	2.0667	1.7885	8.6132
MLE	-	-	-	-	-
8(D2)					
CECF	0.2095	0.7094	1.4376	1.5437	1.8368
DECF	0.4922	0.8462	4.5140	1.8310	13.9995
MGF	0.3184	0.8357	2.1330	2.8875	6.2428
MLE	-	-	-	-	-

Table 4 Kolmogorov-Smirnov Normality Test

P-values are given in parentheses. * Normality is not rejected at 5% significant level (cut-off value is 0.0428); ** Normality is not rejected at 1% significant level (cut-off value is 0.0513).

Case	α	μ_1	μ_2	σ_1^2	σ_2^2
1 (A1)	0.0154*	0.0221*	0.0271*	0.0552	0.0699
	(0.9710)	(0.7081)	(0.4497)	(0.0043)	(1.0454e-004)
2 (A2)	0.0158*	0.0316*	0.0276*	0.0403*	0.0810
	(0.9627)	(0.2662)	(0.4261)	(0.0755)	(3.6570e-006)
3 (B1)	0.0184*	0.0251*	0.0384*	0.0647	0.0618
	(0.8860)	(0.5505)	(0.1029)	(4.2789e-004)	(9.1855e-004)
4 (B2)	0.0212*	0.0187*	0.0839	0.0629	0.1258
	(0.7571)	(0.8714)	(1.3654e-006)	(6.9300e-004)	(2.8219e-014)
5 (C1)	0.0217*	0.0216*	0.0174*	0.0597	0.0461**
	(0.7290)	(0.7347)	(0.9204)	(0.0015)	(0.0276)
6 (C2)	0.0360*	0.0203*	0.0179*	0.0314*	0.0313*
	(0.1465)	(0.7985)	(0.9030)	(0.2723)	(0.2765)
7 (D1)	0.0595	0.0143*	0.0323*	0.1390	0.1124
	(0.0016)	(0.9862)	(0.2450)	(2.4779e-017)	(1.7499e-011)
8 (D2)	0.0709	0.0278*	0.0352*	0.1307	0.0739
	(7.9795e-005)	(0.4181)	(0.1643)	(2.2266e-015)	(3.2983e-005)

Table 5 Empirical Parameter Estimates for MN(2)

(The statistics in the parentheses are t-values).

Stock Codes	1.ACD	2.EK	3.GE	4.UK	5.S&P 500
$\hat{\alpha}$	0.7683	0.7200	0.7776	0.6675	0.4726
	(27.4893)	(18.9803)	(26.2464)	(18.1050)	(11.9926)
$\hat{\mu}_1$	-0.0451	-0.0549	0.0029	-0.0617	0.0656
	(-1.7600)	(-2.1300)	(0.1341)	(-2.6204)	(4.3890)
$\hat{\mu}_2$	0.2950	0.2764	0.1400	0.2002	-0.0452
	(2.8070)	(0.1341)	(1.5979)	(3.1016)	(-1.9483)
$\hat{\sigma}_1^2$	1.3653	1.0875	0.9753	0.8047	0.1640
	(18.5977)	(-2.8072)	(18.5655)	(14.2838)	(9.7373)
$\hat{\sigma}_2^2$	7.6156	5.0037	5.0334	3.9785	0.8835
	(11.4942)	(11.3784)	(13.1297)	(11.6148)	(17.6586)
$\hat{b}$	2.7768	2.0430	2.0860	1.4458	0.2294

Table 6 Empirical Parameter Estimates for MN(3)

Stock Codes	1.ACD	2.EK	3.GE	4.UK	5.S&P 500
$\hat{\alpha}_1$	0.5206 (1.869e2)	0.0598 (83.5030)	0.1461 (2.6229)	0.1454 (21.5250)	0.3270 (3.3363e7)
$\hat{\alpha}_2$	0.0872 (2.3614)	0.5755 (3.8032)	0.6946 (4.7433)	0.6250 (8.3025)	0.0285 (2.8938)
$\hat{\mu}_1$	0.0629 (1.1446)	0.9542 (3.0034)	0.2449 (1.7329)	0.4269 (2.3875)	-0.1466 (-2.4569)
$\hat{\mu}_2$	0.4591 (5.9096e3)	0.0314 (37.133)	-0.0084 (-0.2549)	-0.0192 (-7.7440)	1.1124 (6.5916e6)
$\hat{\mu}_3$	-0.0955 (-1.9928)	-0.0761 (0.1341)	0.0403 (0.5616)	-0.0810 (-2.0968)	0.0596 (3.2079)
$\hat{\sigma}_1^2$	2.8605 (6.6308)	6.0766 (-2.8072)	0.9753 (5.0153)	5.7035 (12.7160)	0.9882 (10.3450)
$\hat{\sigma}_2^2$	12.6230 (4.3795)	2.4761 (6.5220)	1.3707 (3.7875)	1.5743 (10.9570)	2.2435 (5.4204)
$\hat{\sigma}_3^2$	0.7746 (11.4051)	0.5961 (2.2330)	0.2885 (0.8236)	0.2901 (2.5972)	0.2497 (16.4490)
$\hat{b}$	2.8825	1.9801	2.0860	0.6821	4.8987

Table 7 Empirical Parameter Estimates for MN(4)

Stock Codes	1.ACD	2.EK	3.GE	4.UK	5.S&P 500
$\hat{\alpha}_1$	0.5092 (1.4518e5)	0.3783 (1.5517e2)	0.1886 (3.8410)	0.1497 (1.2171e4)	0.5648 (4.2869)
$\hat{\alpha}_2$	0.2511 (0.4804)	0.0990 (0.5580)	0.0813 (1.1625)	0.6158 (9.0737)	0.2120 (2.0606)
$\hat{\alpha}_3$	0.1440 (0.2805)	0.4714 (2.3951)	0.7293 (19.2600)	0.0012 (0.9410)	0.2168 (2.1414)
$\hat{\mu}_1$	0.1058 (1.0122)	0.0601 (0.7271)	0.1740 (0.9385)	0.4646 (2.6886)	0.0778 (4.7807)
$\hat{\mu}_2$	-0.2624 (-0.3104)	-0.1085 (-1.3729)	0.0604 (0.8684)	-0.0307 (-0.7206)	0.0274 (0.2776)
$\hat{\mu}_3$	0.0789 (8.7547e2)	-0.0396 (-1.6413e3)	-0.0066 (-2.2344)	-5.9359 (-1.9232)	-0.1934 (-5.7360e3)
$\hat{\mu}_4$	0.3493 (1.2429)	1.0572 (2.0307)	6.6409 (3.4665)	-0.0737 (-1.7313)	2.2974 (1.9996)
$\hat{\sigma}_1^2$	2.7534 (6.69295)	3.0016 (5.2987)	5.1679 (5.5598)	5.3235 (8.8234)	0.2221 (17.7830)
$\hat{\sigma}_2^2$	0.9481 (4.2815)	0.1552 (0.2794)	0.1102 (0.5521)	1.5792 (10.3920)	1.3068 (5.5835)
$\hat{\sigma}_3^2$	0.5205 (0.6918)	1.1759 (3.1943)	1.1861 (6.5135)	0.5788 (0.8830)	0.6149 (4.0215)
$\hat{\sigma}_4^2$	12.1480 (4.9056)	9.5961 (3.2794)	0.5443 (0.3120)	0.2944 (3.0774)	2.2305 (4.3608)
$\hat{b}$	1.2767	1.4856	0.4201	0.6740	0.8082

Table 8 Empirical Parameter Estimates for MN(2) – [1962 - 2004]

(The statistics in the parentheses are t-values).

Stock Codes	1.ACD	2.EK	3.GE	4.UK	5.S&P 500
$\hat{\alpha}$	0.6343	0.8110	0.7682	0.7296	0.6870
	(21.3312)	(44.6887)	(35.2605)	(37.2728)	(32.9421)
$\hat{\mu}_1$	-0.1067	-0.0448	-0.0019	-0.0670	0.0388
	(-4.5820)	(-2.7147)	(-0.1177)	(-3.7294)	(4.1417)
$\hat{\mu}_2$	0.3415	0.4783	0.2837	0.3843	0.0112
	(6.1790)	(0.1341)	(5.9146)	(5.8255)	(0.3995)
$\hat{\sigma}_1^2$	1.4619	1.4072	1.1843	1.2534	0.3322
	(18.8673)	(30.0482)	(26.4908)	(25.1290)	(23.9436)
$\hat{\sigma}_2^2$	6.3121	7.4265	5.7382	7.3022	1.8034
	(19.4443)	(15.0703)	(16.1080)	(18.4195)	(21.1692)
$\hat{b}$	2.5410	3.1241	2.3615	2.4957	0.6849

Table 9 Empirical Comparison across Different Model Specifications

	Data	MN2	MN3	MN4
ACD				
Mean	0.0406	0.0337	0.4181	0.0328
Variance	2.9124	2.8341	2.9160	2.9127
Skewness	0.4181	0.7402	1.0201	0.8365
Kurtosis	6.6346	5.6490	6.6299	6.5583
AIC	–	17578.34	17557.22	17564.98
BIC	–	17610.09	17608.67	17634.52
EK				
Mean	0.0459	0.0379	0.0474	0.0476
Variance	2.2702	2.2062	2.2565	2.2560
Skewness	0.3255	0.6160	1.3349	1.3538
Kurtosis	6.2845	4.9526	5.7889	5.7885
AIC	–	16571.16	16534.18	16544.76
BIC	–	16603.37	16585.72	16615.63
GE				
Mean	0.0370	0.0334	0.0364	0.0382
Variance	1.8952	1.8813	1.8940	1.8890
Skewness	0.1679	0.2848	0.4684	0.3900
Kurtosis	5.6936	5.4268	5.6827	5.5894
AIC	–	15719.61	15705.01	15712.36
BIC	–	15751.82	15756.55	15783.22
UK				
Mean	0.0305	0.0254	0.0315	0.0266
Variance	1.9369	1.8752	1.9069	1.9132
Skewness	0.3412	0.4904	0.8195	0.7383
Kurtosis	5.9529	4.9348	5.4159	5.4503
AIC	–	15793.56	15759.50	15764.83
BIC	–	15825.77	15811.04	15835.70
S&P 500				
Mean	0.0221	0.0072	0.0221	0.0220
Variance	0.5920	0.5465	0.5919	0.5937
Skewness	0.2133	-0.1567	0.4272	0.3476
Kurtosis	5.8488	4.2935	5.8805	6.0812
AIC	–	10335.87	10285.51	10281.48
BIC	–	10368.08	10337.05	10352.34