

5-1-2005

Using Scale Mixtures Of Normals To Model Continuously Compounded Returns

Hasan Hamdan
James Madison University

John Nolan
American University

Melanie Wilson
Allegheny College

Kristen Dardia
James Madison University

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Hamdan, Hasan; Nolan, John; Wilson, Melanie; and Dardia, Kristen (2005) "Using Scale Mixtures Of Normals To Model Continuously Compounded Returns," *Journal of Modern Applied Statistical Methods*: Vol. 4 : Iss. 1 , Article 20.
DOI: 10.22237/jmasm/1114906800

Using Scale Mixtures Of Normals To Model Continuously Compounded Returns

Hasan Hamdan

Department of Mathematics and Statistics
James Madison University

John Nolan

Department of Mathematics and Statistics
American University

Melanie Wilson

Department of Mathematics
Allegheny College

Kristen Dardia

Department of Mathematics and Statistics
James Madison University

A new method for estimating the parameters of scale mixtures of normals (SMN) is introduced and evaluated. The new method is called UNMIX and is based on minimizing the weighted square distance between exact values of the density of the scale mixture and estimated values using kernel smoothing techniques over a specified grid of x -values and a grid of potential scale values. Applications of the method are made in modeling the continuously compounded return, CCR, of stock prices. Modeling this ratio with UNMIX proves promising in comparison with other existing techniques that use only one normal component, or those that use more than one component based on the EM algorithm as the method of estimation.

Key words: Expectation-Maximization algorithm, UNMIX, kernel density smoothing, expected return

Introduction

The study of univariate scale mixtures of normals, SMN, has long been of interest to statisticians continuously hunting for better methods to model probability density functions. Modeling using these mixtures has many applications from genetics and medicine to economic and populations studies. More specifically, one can use SMN to model any data that is seemingly normally distributed and has a high kurtosis. Using SMN allows for the tails of

the density to be heavier than those in the normal density, giving a better coverage for data that varies greatly from the mean.

The most common estimation of the parameters of the mixtures is the EM algorithm of by Dempster, Larid, and Rubin (1977). This method is based on finding the maximum likelihood estimate of the parameters of a given data set. The EM algorithm performs well in cases where the distance between means of the components is relatively large. However, when estimating the parameters of a mixture of normals where all of the components have the same mean but different variances, the EM algorithm gives a poor estimation when these variances are small and close.

In this article, we elaborate on a new approach of estimation, UNMIX, proposed by Hamdan and Nolan (2004). The UNMIX program uses kernel smoothing techniques to get an empirical estimate of the density of the data. It then estimates the parameters of the mixture based on minimizing the weighted least squares of the distance between the values from the empirical density and the new scale mixture density over a pre-specified grid of x -values, and

Hasan Hamdan is Assistant Professor at James Madison University. His research interests are in mixture models, sampling and mathematical statistics. Email: hamdanhx@jmu.edu. John Nolan is a Professor at American University. His research interests are in probability, stochastic processes and mathematical statistics. Melanie Wilson is a graduate student in statistics at Duke and Kristen Dardia is graduating from James Madison University. This research was partially supported by NSF grant number NSF-DMS 0243845.

potential grid of σ values called r-grid. The UNMIX method will be used to estimate the density of the continuously compound return, CCR. The estimation of the density function is pertinent to knowing the probability that the closing stock price will stay within a certain interval during a given time period. The density function was first estimated simply by using a normal curve.

However, Mandelbrot (1963) and Fama (1965) showed that the normal estimation did not model market returns appropriately due to the excess kurtosis and volatility clustering that characterize returns in financial markets. Clark (1973) then tested the use of the lognormal distribution to estimate the density of the stock returns. This is analogous to using the normal distribution to estimate the density of the natural log of the stock returns (also called the continuously compounded return). Following Clark's estimation, Epps and Epps (1976) found that a better estimation is obtained when using a mixture of distributions.

However, their assumption used the transaction volume as the mixing variable thus introducing excess error. Another popular method evolved recently when Zangari (1996), Wilson (1998), and Glasserman, Heidelberger and Shahabuddin (2000), used the multivariate t distribution to estimate the stock return. Unfortunately, Glasserman, Heidelberger, and Shahabuddin (2000), pointed out that since most stock returns have equally fat tails, this model frequently comes up short. Additionally, the method involves solving non-linear equations to derive a numerical approximation of an input covariance matrix and requires the consuming and difficult job of inverting marginal distributions.

As proposed by Clark (1973) and Epps and Epps (1976), we look deeper into modeling the CCR (the natural log of the stock returns), we find that modeling the distribution using a simple normal curve should be avoided due to the fact that the CCR of most stock prices are mound shaped but have a high kurtosis (also known as a high volatility). Therefore, these ratios can be modeled using a SMN with mean zero, since the mean of the CCR of the prices is close to zero. A brief explanation of the concept of a random variable X having a density

function of the form of a SMN is introduced in Section 2. Next, in Section 3, techniques of estimation of SMN are listed and brief background on the common EM algorithm is also presented. In Section 4, the density of CCR is estimated for different stocks with SMN using the UNMIX program and using a single normal. Also, the density is also estimated using the EM algorithm and the results are compared. Finally, some suggestions for improving this method are made in the conclusion section.

Methodology

A random variable X is a scale mixture of normals or SMN if $X \stackrel{d}{=} AZ$, where $Z \sim N(0,1)$, $A > 0$, A and Z independent. Here $N(0,1)$ is the standard normal variable with mean 0 and standard deviation 1. Therefore, X has a probability density function

$$f(x) = \int_0^{\infty} \frac{1}{\sigma} \phi(x/\sigma) \pi(d\sigma), \tag{1}$$

where ϕ is the standard normal density and the mixing measure π is the distribution of A .

An SMN can either be an infinite or a finite mixture, depending upon the mixing measure π . If our mixing measure is discrete and A takes on a finite number of values, say $\sigma_1, \dots, \sigma_m$ with respective probabilities π_1, \dots, π_m then the probability density function can be rewritten as

$$f(x) = \sum_{j=1}^m \frac{1}{\sigma_j} \phi(x/\sigma_j) \pi_j \tag{2}$$

A common finite mixture, called the contaminated normal, occurs when A takes on two values, with $\sigma_1 < \sigma_2$ and $\pi_1 > \pi_2$. In this case our density function can be simplified to

$$f(x) = \pi_1 \phi(x/\sigma_1) + (1 - \pi_1) \phi(x/\sigma_2).$$

Some common examples of infinite SMN are the Generalized t distribution, Exponential power family, and Sub-Gaussian distributions. The following theorem gives the characteristics necessary for a distribution to be SMN with mean zero.

Theorem: (Schoenberg, 1938) Given any random variable X with density $f(x)$, X is a scale mixture of normals if and only if $h(x) = f(\sqrt{x})$ is a completely monotone function. See Feller (1971) for definition of completely monotone function. As we have seen above when A takes on a finite number of values, the density of X can be written more simply in the same manner as equation (2). When π is not concentrated at a finite number of points, Hamdan and Nolan (2004) give a constructive method on how to discretize π so that equation (2) is uniformly close to equation (1).

Estimating Scale Mixtures of Normals

In estimating SMN one needs to find the following: number of components, estimated parameters of each component, and estimated weights of each component. We highlight some of the important developments in this area.

This problem of estimating SMN has been the subject of a large diverse body of literature. Dempster, Larid, and Rubin (1977) introduced the EM algorithm for approximating the maximum likelihood estimates. Because other methods have been developed based on the EM algorithm. A robust powerful approach based on minimizing distance estimation is analyzed by Beran (1977) and Donoho and Liu (1988). Zhang (1990) used Fourier methods to derive kernel estimators and provided lower and upper bounds for the optimal rate of convergence. Priebe (1994) developed a nonparametric maximum likelihood technique from related methods of kernel estimation and finite mixtures.

EM algorithm

The EM algorithm developed by Dempster, Larid, and Rubin (1977), is based on finding the maximum likelihood estimate of the components, parameters, and weights of a

mixture of normals. It should be noted that though we will only use the EM algorithm for a mixture of normals, it can be generalized for other mixtures. However, differentiation problems become more complicated in the M step of the algorithm for non-normal mixtures. The EM algorithm does not assume that we are dealing with SMN and allows each density function to have a different mean. Therefore, given the data points, x_1, \dots, x_n , from the finite normal mixture of k components

$$\sum_{j=1}^k \frac{1}{\sigma_j} \phi\left(\frac{x - \mu_j}{\sigma_j}\right) \pi_j,$$

the data are completed by letting each x_i correspond to a y_i . The new y_i is a vector giving the initial value x_i and also a sequence of values z_1, \dots, z_k which tells the location of the x value as follows:

$$y_i = (x_i, z_{i1}, \dots, z_{ik})$$

$$\text{where } z_{ij} = \begin{cases} 1 & \text{if } x_i \text{ is generated by} \\ & \text{the } j\text{th component;} \\ 0 & \text{otherwise} \end{cases}$$

Therefore the only missing values are the labels, z_{i1}, \dots, z_{ik} . Next the maximum likelihood estimate of each y_i is found in the in the Expectation Step of the EM algorithm. An initial guesses for the parameters

$$\hat{\pi}_1, \dots, \hat{\pi}_k, \hat{\mu}_1, \dots, \hat{\mu}_k, \hat{\sigma}_1, \dots, \hat{\sigma}_k$$

are taken. Then an estimate of probability of category membership of the i^{th} observation, conditional on x_i is found based on using the parameter estimate

$$(\hat{\pi}_1, \dots, \hat{\pi}_k, \hat{\mu}_1, \dots, \hat{\mu}_k, \hat{\sigma}_1, \dots, \hat{\sigma}_k).$$

This estimation is noted by

$$\hat{\gamma}_{ji} = \frac{\hat{\pi}_j \frac{1}{\hat{\sigma}_j} \phi(x_i, \hat{\mu}_j, \hat{\sigma}_j)}{\sum_{j=1}^k \hat{\pi}_j \frac{1}{\hat{\sigma}_j} \phi(x_i, \hat{\mu}_j, \hat{\sigma}_j)}$$

$i = 1, \dots, n$ and $j = 1, \dots, k$.

The next step is to compute the weighted means and variances in the Maximization Step of the EM algorithm for mixtures of normals. Then the E and the M steps are iterated until the parameters converge, and the final values are used as the parameters estimates of the mixture of normals. The EM algorithm works well in modeling SMN where the variance of the components are relatively large, but as the variances approach zero, the algorithm shows a poor performance. In general, as shown in many simulation studies, when the components are not well-separated, estimation based on maximum likelihood is poor (Dick & Bowden, 1973).

There are also many practical difficulties in estimating SMN using the EM. Some of these are computationally difficult and intractable. For example, when the MLE of the mixing measure in the finite case is found, a large local maxima might be found that occurs as a consequence of a fitted component having a very small (but nonzero) variance. Moreover, it is not clear how to initialize the estimates, especially when the mixture is a scale mixture. Though, methods have recently been developed by Biernacki, Celeux, & Govaert (2003) in order to find the most efficient initializing conditions. Another key problem in finite mixture models is determining the number of components in the mixture. Several criteria based on the penalized log-likelihood, such as Akaike Information Criterion, AIC, the Bayesian Information Criterion, BIC, and the Information Complexity Criterion introduced by Bozdogan (1993), have been used.

UNMIX

The next approach, UNMIX, uses kernel smoothing techniques to estimate the empirical density of a sample. It then minimizes the weighted square distance between the kernel smoothing estimate and the density computed by

discretizing the mixture over a pre-specified grid of x-values and potential grid of sigma values. Given a sample of size n from the mixture, we fix a grid of possible sigma values (called the σ -grid), and possible x values (called the x-grid), $x_1 \dots x_k$, where $k \leq m$.

In order to obtain an estimate $\hat{f}(x)$ of $f(x)$ for each x in the x -grid, we use kernel smoothing techniques discussed briefly at the end of the section. Our model is

$$y_i = \sum_{j=1}^m \frac{1}{\sigma_j} \phi(x_i / \sigma_j) \pi_j + \varepsilon_i,$$

where ε_i are independent with mean 0. That is solved for π_j by minimizing

$S(\pi)$ where $\pi^T = (\pi_1, \dots, \pi_m)$ and

$$S(\pi) = \sum_{i=1}^k \left(w_i \left(y_i - \sum_{j=1}^m \phi_{ij} \pi_j \right) \right)^2,$$

with $\phi_{ij} = \frac{1}{\sigma_j} \phi(x_i / \sigma_j)$ and w_i are weights. We

will use $w_i=1$ throughout. However, if the data are heavy-tailed then one can try different weights until he finds a good fit (in the heavy-tailed case, a good strategy might be weighting the points that are close to the mean of the x-grid less than those that are far from the mean of the x-grid. Next consider the problem as a quadratic programming problem with two constraints: $\sum \pi_j = 1$ and $\pi_j \geq 0$ for all j. Expanding $S(\pi)$:

$$\begin{aligned} S(\pi) &= \sum_{i=1}^k \left[w_i^2 y_i^2 - 2w_i y_i \sum_{j=1}^m \phi_{ij} \pi_j + w_i^2 \left(\sum_{j=1}^k \phi_{ij} \pi_j \right)^2 \right]; \\ &= \sum_{i=1}^k w_i^2 y_i^2 - 2 \sum_{i=1}^k \left(w_i y_i \sum_{j=1}^k \phi_{ij} \pi_j \right) + \sum_{i=1}^k \left(\sum_{j=1}^k \phi_{ij} \pi_j \right)^2; \\ &= \sum_{i=1}^k w_i^2 y_i^2 - 2 \sum_{j=1}^m \left(\sum_{i=1}^k w_i y_i \phi_{ij} \right) \pi_j \\ &\quad + \sum_{i=1}^k \sum_{j=1}^m \sum_{l=1}^m (w_j \phi_{ij} \pi_j)(w_l \phi_{il} \pi_l). \end{aligned}$$

Because $\sum_{i=1}^k w_i^2 y_i^2$ is independent of π , it is a constant. Reformulating the problem in a matrix environment, we let g be the $(m \times 1)$ vector defined as

$$g = - \left(\sum_{j=1}^k w_j y_j \phi_{j1}, \dots, \sum_{j=1}^k w_j y_j \phi_{jm} \right)^T,$$

and let \mathbf{H} be the $(m \times m)$ matrix defined as $\mathbf{H} =$

$$\begin{pmatrix} \sum_{i=1}^k w_i^2 \phi_{i1} \phi_{i1} & \sum_{i=1}^k w_i^2 \phi_{i1} \phi_{i2} & \cdots & \sum_{i=1}^k w_i^2 \phi_{i1} \phi_{im} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^k w_i^2 \phi_{im} \phi_{i1} & \sum_{i=1}^k w_i^2 \phi_{im} \phi_{i2} & \cdots & \sum_{i=1}^k w_i^2 \phi_{im} \phi_{im} \end{pmatrix}$$

To simplify, let $c = \frac{1}{2} \sum_{i=1}^k w_i^2 y_i^2$ be a constant, resulting in the following formula for $S(\pi)$:

$$S(\pi) = 2 \left[c + g^T \pi + \frac{1}{2} \pi^T \mathbf{H} \pi \right].$$

Therefore π can be found by minimizing $\left[g^T \pi + \frac{1}{2} \pi^T \mathbf{H} \pi \right]$ subject to our two quadratic programming constraints $\sum_{j=1}^m \pi_j = 1$ and $\pi_j \geq 0$, this latter constraint can be rewritten in matrix form as $\mathbf{A} \pi \geq b$ where

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

of order $(m \times m)$ and $b^T = (0 \cdots 0)$ of order $(m \times 1)$. A quadratic programming routine, QPSOLVE which is a Fortran subroutine, is used to solve this problem. UNMIX is a Splus program that takes the sample, x-grid, r-grid, and a vector of weights as the input and calls

QPSOLVE. The program's output is a vector of estimated weights over the given r-grid.

In obtaining an estimate for $f(x)$, kernel smoothing techniques were used. One important variable in density estimates using kernel smoothing techniques is the bandwidth. In general, using a large bandwidth over-smoothes the density curve, and small bandwidths can under-smooth the density curve. In essence, the bandwidth controls how wide the kernel function is spread about the point of interest. If there are a large number of values, x_i near x , then the weight of x is relatively large and the estimation of the density at x will also be large.

There are four sources of variability involved when using UNMIX to estimate a SMN. The first is the sampling variability, the second is due to the method of density estimation and bandwidth used. The third variability is the choice of the x-grid and finally, the fourth is the choice of the r-grid.

Controlling sampling variability can be done by increasing the sample size. However, controlling the variability introduced by the method of density estimation requires care and investigation of the sample and bandwidth used. For example, we can weight the observations by using their distance from the center. There is considerable literature on how to pick the most effective bandwidth including articles by Hardle and Marron (1985) and Muller (1985). For the purposes of this article, when using the UNMIX program, the default bandwidth based on the literature given in R-Software is used.

UNMIX performs well for estimating distributions with a high kurtosis but losses accuracy for data that is extremely concentrated about the mean. However, these difficulties can be overcome due to the flexibility of the program in terms of fitting the data. In particular, the r-grid can be changed and the weights interactively in a systematic way until a good fit is found. We have found that the most useful x-grid is evenly distributed and symmetric about the mode, where the distance from the mode on both sides is the absolute maximum of the sample data because the mode is 0 in this case. This allows the x-grid to cover all data points. Also in creating the σ -grid, a

simple guideline is to make it evenly distributed from a point close to zero to a point at least three sample standard deviations away from zero. This again allows for the σ -grid to cover a large percentage of potential sigma values no matter what the original distribution of sigma is. Here, we assume that the values in the x-grid and the σ -grid are the only possible values for each x and σ , therefore it is important to pick them within in the range of the sample.

Results

Estimating the density of stock returns has been important to statisticians and those interested in finance since the stock market opened. Fama (1965) and many others model stock prices based on simple random walk assumption. In other words the actual price of a stock will be a good estimate of its intrinsic value. The standard assumption is that the percentage changes in the stock price in a short period of time are normally distributed with parameters μ , expected return of the stock, and σ which is the volatility of the stock price. The expected return is estimated by

$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and $x_i = \ln\left(\frac{S_i}{S_{i-1}}\right)$, where S_T is the current stock price. Therefore, the 1 period volatility is estimated by

$$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} .$$

The continuously compounded return, CCR, can now be estimated as follows with $S_{T-\tau}$ as the stock price τ time units earlier:

$$\ln\left(\frac{S_T}{S_{T-\tau}}\right) \sim \phi\left[\left(\mu - \frac{\sigma^2}{2}\right)\tau, \sigma\sqrt{\tau}\right].$$

Comparing Normal Estimate to UNMIX Estimate

We now estimate and compare the density of the CCR using a single normal curve and a scale mixture of normals. Taking advantage of Yahoo's (an internet search engine) intensive finance sources, three stocks were found whose price quotes showed

relatively high volatility: Ciber Inc, ExxonMobil, and Continental Airlines. For each of the stocks, we sampled the weekly closing prices over the past four years, from July 14, 2000 to July 14, 2004. The natural log of the return was taken to find the CCR for each stock.

Modeling with the single normal method described above and the UNMIX program, their performances were compared against the empirical density found using kernel smoothing techniques. The empirical density is then used to estimate the density over an x-grid of 51 equally-spaced points between $-4S$ and $4S$, where S is the sample standard deviation. Because the empirical density can be made very close to the true density at any given point, it is considered as the true density in each of the following error calculations which are presented in Table 1, Table 2 and Table 3.

Example 1: In this example, the density of the CCR, of Ciber Inc. stock, is estimated. The normal estimate based on the random walk assumption has a mean of $-.00686$ and standard deviation of $.09041$. The estimated SMN was found using the UNMIX program and has 4 components with weight vector of $(.52951, .07374, .39415, .00260)$ and an estimated σ -vector of $(.12266, .06048, .03885, .03750)$. The estimated densities were evaluated on the same x-grid and the results are shown in Figure 1.

The maximum and average error between each estimate and the empirical density can be seen in Table 1. In Figure 2, the three density estimates were found for an x-grid located in the tail of distribution of CCR and it consists 25 equally-spaced points between $.2$ and $.45$. Using the normal assumption, the probability of any sample point falling in such range is approximately $.012$ and approximately $.035$ when the scale mixture assumption is used. Though this probability is not high, most density estimation techniques do not recover the tails well where the most extreme occurrences can be found. This could be very problematic in finance and risk analysis.

Figure 1: Estimated Density of CCR for Ciber Inc stock.

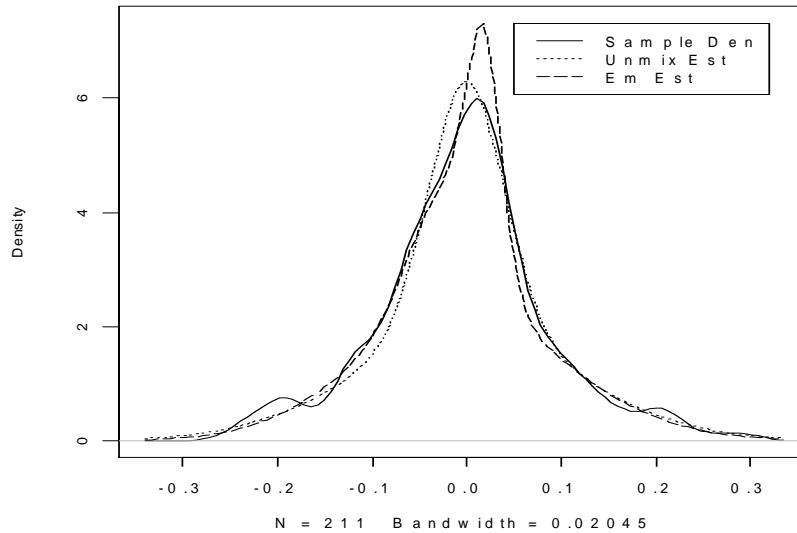
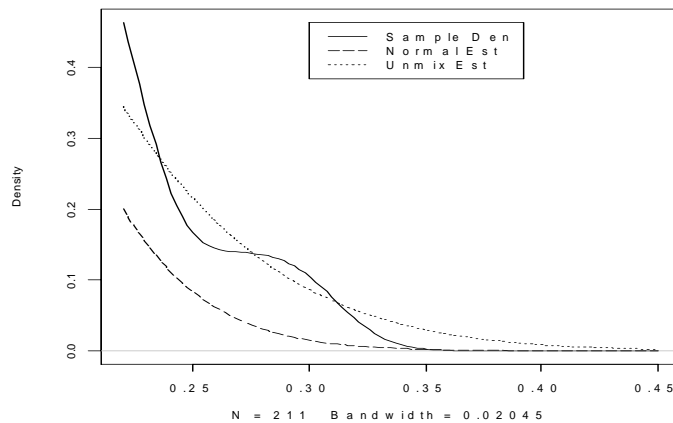


Figure 2: Estimated Density in the right tail of CCR of Ciber Inc.



Notice in Figure 2 that estimating with SMN produces a better fit in the tails. In contrast to overestimating the rate of return in the body, where around 95% of the data are located, the normal curve tends to underestimate the density in the tails. As in our examples, the distributions for the CCR tend to have fatter tails than the proposed normal has. Because the tails of the data are heavy, the scale mixture estimation will produce a better fit than the normal.

Under the single normal assumption, the 95% confidence interval for the mean of the CCR is $(-.1767, .1767)$. Equivalently and by exponentiation, the interval for the mean rate of return is $(.8381, 1.1933)$. The corresponding UNMIX estimate is found to be $(.8469, 1.1808)$. In comparison to UNMIX, the normal curve tends to overestimate the rate of return in the body of the density. Though the gap does not seem large when investing a small amount, for

big time investors 1 penny off per dollar can translate to thousands of dollars lost when investing millions. In Table 4, we summarize the bounds for the middle 95% probability of the distribution for all three examples.

Examples 2 & 3: In these two examples, estimate the densities for the CCR of ExxonMobil and Continental Airline stocks are found based on a single normal and using UNMIX. The single normal and UNMIX estimates are plotted against the empirical density, as described in the previous example, the results are shown in Figure 3 and Figure 4 respectively. The single normal for the ExxonMobil case has an estimated mean of .000248 and an estimated standard σ of 0.038. The scale mixture for the ExxonMobile case has 4 components with a weight vector of (.04734,.47208,.4412,.03938) and the corresponding estimated σ -vector of (.091,.02587,.02542,.01276).

However, the single normal for the Continental Airline case has an estimated mean of -0.0076 and an estimated σ of 0.0930. Finally, the scale mixture for the Continental Airline case has 5 components with a weight vector of (.07006,.01495,.44468,.32486,.14544) and the corresponding estimated σ -vector of (.24215,.24074,.08372,.08232,.02343).

Notice in Figures 3 and 5 the empirical density tends to be negatively skewed. This is common in the densities of CCR since there is a greater probability of the stock market to produce large downward movements than large upward movements. This can be explained by the public's tendency to pull-out of a falling market thus causing prices to drop even further.

In the following tables, the maximum absolute difference between the empirical density and the estimated density over the selected grid using a single normal is indicated by Max. Norm., and the average value is indicated by Avg. Norm. Similarly, Max. UNMIX and Avg. UNMIX are the corresponding values when a scale mixture, with UNMIX as a method of estimation, is used rather than one single normal as a model for CCR.

Next, the performance of the UNMIX method is compared to the EM algorithm in estimating the density of the CCR of the same three stocks. The number of components to be used with the EM is also unknown, and there are many ways that can be used to estimate it. Here, we tried two, three, four and five component mixture.

There was no noticeable difference between the four-component mixture and the five-component mixture. Therefore, the four-component mixture was used for our examples. The parameters were then estimated using the EM algorithm and it was compared to that found using the UNMIX estimation. The initialization of the parameters was somewhat arbitrary because our goal is to find the best density fit and not to investigate the speed or the convergence of these estimation methods. The π 's were initialized such that each component has an equal weight of .25, and the μ 's were initialized such that = the mean of the sample, μ_1 and μ_2, μ_3 and $\mu_4 = .2, .4$, and $.8$ times the mean of the sample respectively.

Then, the σ 's were initialized for each component in the same manner as the μ 's. For each of the three examples, the process was repeated 50 times and the mean of the parameter estimate was taken as the final EM estimate. The estimated densities of the stocks are shown in Figure 7.

Notice that the EM estimate tends to overestimate the mean of the empirical density which is a consequence of the fitted component having a very small variance. The EM captures the skewness of the density better but in general, UNMIX outperforms it. This is seen by the fact that in the three examples, the EM algorithm produces both a greater maximum and average error as summarized in Table 5.

Figure 3: Estimated Density of CCR for Exxon Mobile stock.

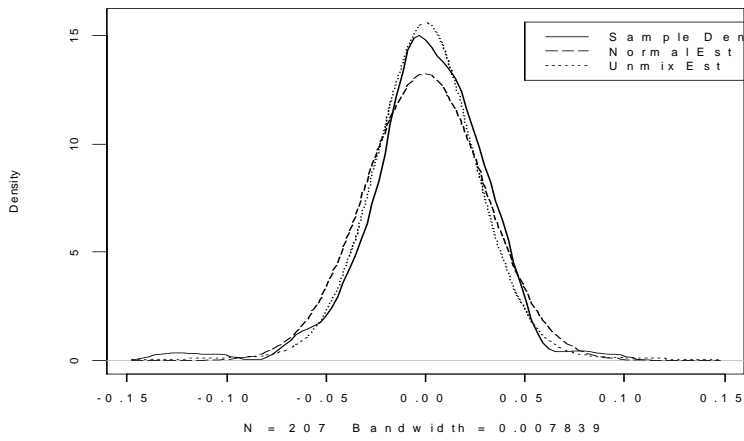


Figure 4: Estimated Density of CCR in the right tail of Exxon Mobile stock. Probability of being in the tail is approximately .0372.

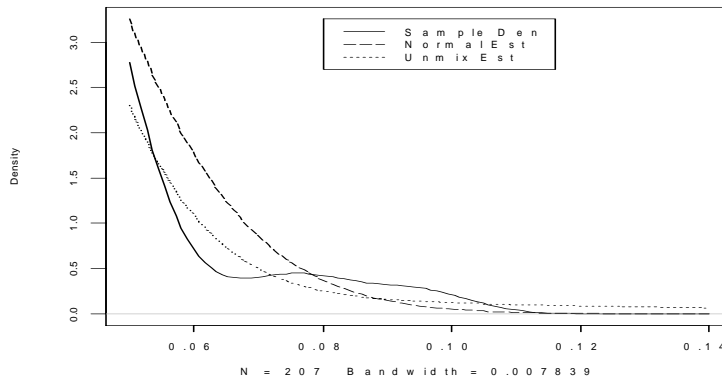


Figure 5: Estimated Density of CCR for Continental Airline.

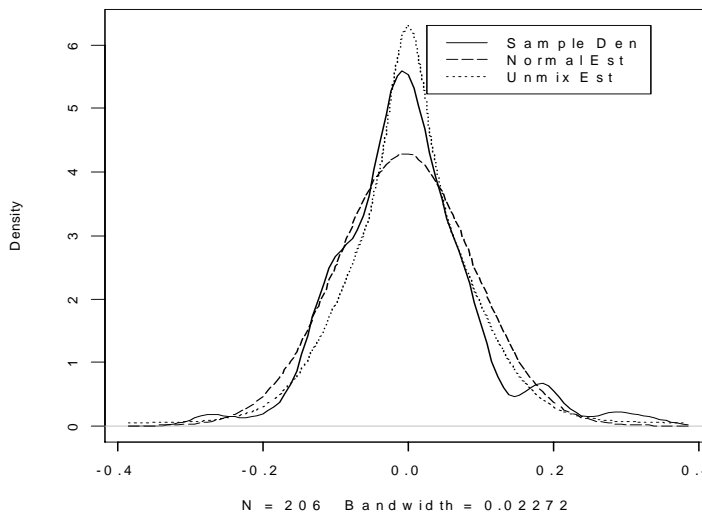


Figure 6: Estimated Density in the right tail of CCR for Continental Airlines stock.
Probability of being in this tail is .0186.

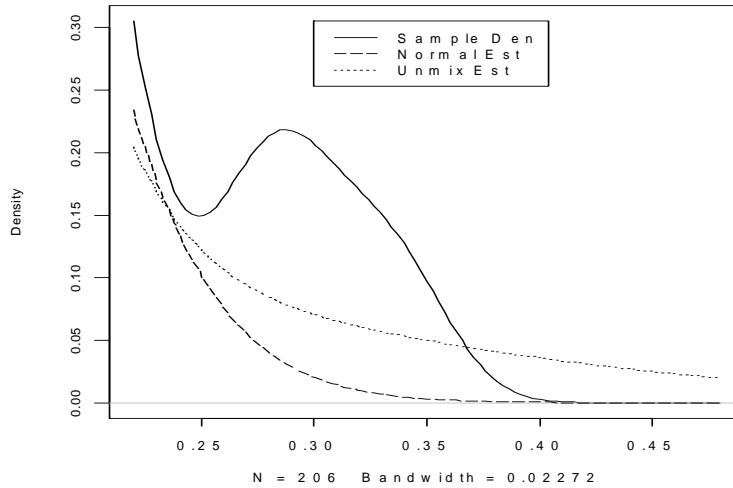


Table 1: Maximum and average errors of the Normal and UNMIX estimates of CCR for Ciber Inc.

Error	Body. Den.	Tail Den.
Max. Norm.	1.6440	.2630
Max. UNMIX	.7874	.1188
Avg. Norm.	.3645	.0499
Avg. UNMIX	.1559	.0203

Table 2: Maximum and average errors of the Normal and UNMIX estimates of CCR for ExxonMobile stock.

Error	Body. Den.	Tail Den.
Max. Norm.	1.9545	1 .0582
Max. UNMIX	1.7341	.4712
Avg. Norm.	.5838	.2268
Avg. UNMIX	.4015	.1283

Table 3: Maximum and average errors of the Normal and UNMIX estimates of CCR for Continental Airline stock.

Error	Body. Den.	Tail Den.
Max. Norm.	1.3104	.1897
Max. UNMIX	.8375	.1410
Avg. Norm.	.26911	.0699
Avg. UNMIX	.02090	.0579

Table 4: Bounds for the middle 95% probability of the distribution for the CCR of Ciber Inc., ExxonMobil, and Continental Airlines in both the normal and UNMIX estimates.

Stock	Normal	UNMIX
Ciber Inc.	(.8381, 1.1933)	(.8469, 1.1808)
ExxonMobile	(.9428, 1.0607)	(.9462, 1.0569)
Continental	(.8334, 1.200)	(.8416, 1.1882)

Figure 7: Estimated Density of CCR using the UNMIX program and the EM algorithm for (a) Ciber Inc.; (b) ExxonMobil; (c) Continental Airlines.

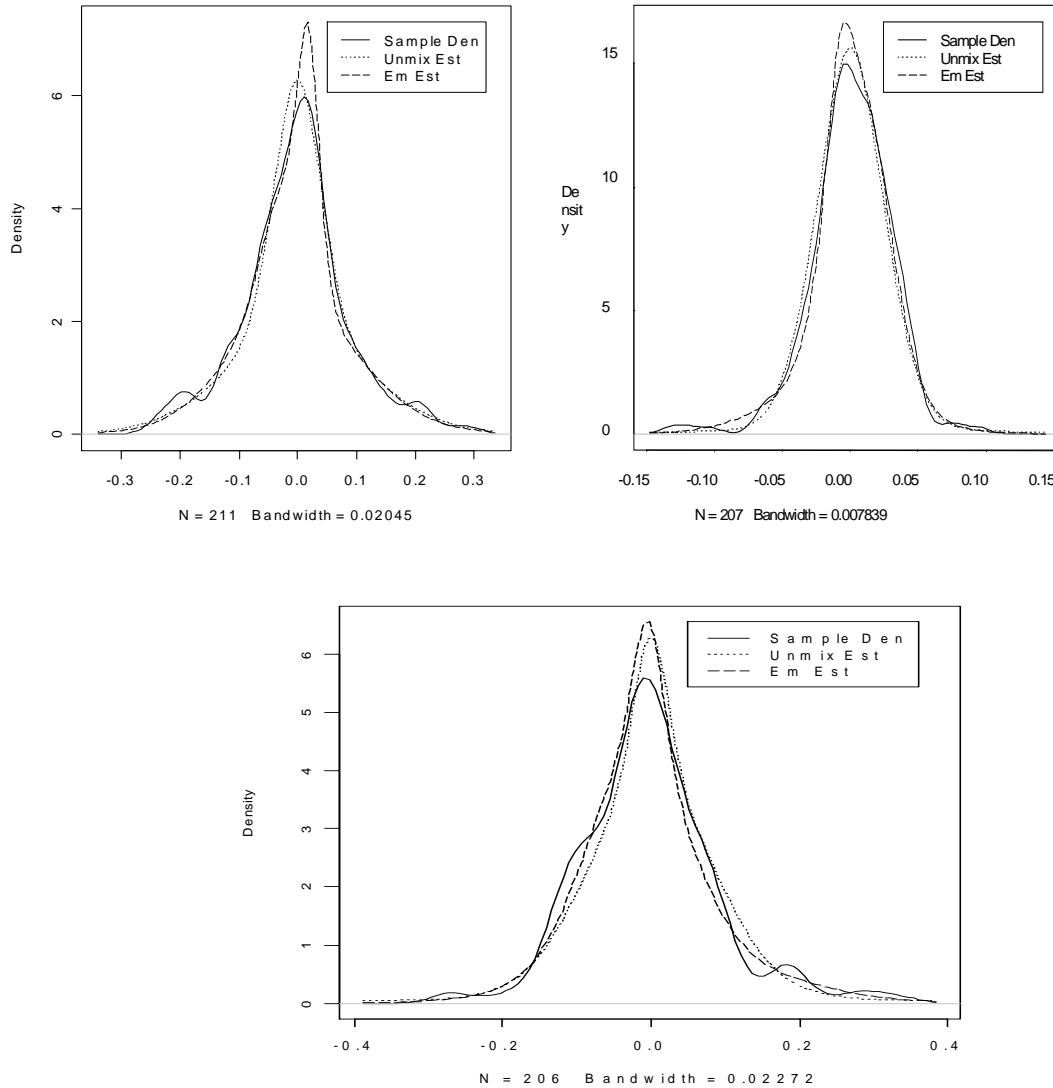


Table 2: Maximum and average errors of the UNMIX and EM estimates of CCR for all examples.

Error	Ciber Inc.	ExxonMobile	Continental
Max. EM	1.2832	2.143	1.2987
Max. UNMIX	.7971	1.7269	.8320
Av. EM	.1592	.4059	.2193
Avg. UNMIX	.1542	.3985	.2048

Conclusion

Estimation of the CCR of stocks has been an interest of both statisticians and financiers due to the importance of producing accurate models for the data. As evidenced by the previous examples, UNMIX allows for this analysis to occur with smaller error in comparison to the single normal assumption and the common methods based on the EM algorithm.

Although the EM algorithm is well developed and allows for different location and different scales, sometimes it has some practical difficulties. For example, when trying to find the MLE of the parameters, it might find a large local maxima that occurs as a consequences of a fitted component having a very small (but non-zero) variance. Also, there are still some problems associated with initializing the parameters including the number of components.

However, UNMIX fitted the data better than the EM. We believe that it will always fit the data well, because it is based on minimizing the weighted distance between empirical density and the mixture over a given grid. However, in terms of estimating the actual parameters, more work needs to be done because the EM still does a better job in estimating the actual values as we have seen in many simulated examples where the actual mixtures are known.

Here are some areas where we can improve UNMIX. First, make it most applicable is the possibility of handling not only scale, but location conditions. Also improvements to the program can be made by developing guidelines

to choose the most optimal x-grid and r-grid. Finally, we can improve the empirical density estimate by using optimal kernel functions and bandwidths. Implications of the UNMIX program can apply beyond the scope of the stock market. This program can be used to model distributions with relatively high possibilities of outlying events. Staying in the realm of finance the program can be used to estimate exchange rates.

However, there are also many examples outside of the finance field including fitting extreme data. For example, the UNMIX program was used to fit the density of some heavy-tailed data. These data were generated from the class of stable densities that have infinite variance and known to be infinite variance mixture of normals such as Cauchy density. Although more work needs to be done, but the UNMIX method looks promising in fitting such data.

References

Akaike, H. (1954). An approximation to the density function. *Annals of the Institute of Statistical Mathematics*, 6, 127-132.

Beran, R. (1977). Minimum hellinger distance estimates for parametric models. *Annals of Statistics*, 5, 445-463.

Biernacki, C., Celeux, G., & Govaert, G. (2003). Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics and Data Analysis*, 41, 561-575.

Bozdogan, H. (1993). Choosing the number of component clusters in the mixture-model using a new informational complexity criterion of the Inverse-Fisher Information Matrix, *Information and Classification*, 40-54.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41, 135-155.

Dempster, A. P., Larid, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39, 1-38.

Dick, N. P., & Bowden, D. C. (1973). Maximum likelihood estimation for mixtures of two Normal distributions. *Biometrics*, 29, 781-790.

Donoho, D. L., & Liu, R. C., (1988). The 'Automatic' robustness of minimum distance functional. *Annals of Statistics*, 16, 552-586.

Epps, T. W., & Epps, M. L. (1976). The stochastic dependence of security price changes and transaction volumes: implications for the mixture-of-distributions hypothesis. *Econometrica*, 44, 305-321.

Fama, E. F. (1965). The behavior of stock market prices. *Journal of Business*, 38, 34-105.

Feller, W. J. (1971). An introduction to probability and its applications. (2nd ed.). Vol. II. NY: Wiley.

Fix, E., & Hodges, J. L. (1989). Discriminatory data analysis - nonparametric discrimination: consistency properties., *International Statistics Review*, 57, 238-247.

Glasserman, P., Heidelberger, P., & Shahabuddin, P. (2000). Portfolio value-at-risk with heavy-tailed risk factors. *IBM Research Report*.

Hamdan, H., & Nolan, J. (2004). Approximating scale mixtures. *Stochastic Processes and Functional Analysis*, 283, 161-169.

Hamdan, H., & Nolan J (2004, in press). Estimating the parameters of infinite scale mixtures of normals. *Computing Science & Statistics*. 36, the proceeding of the 36th Symposium on the Interface.

Hardle, W., & Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function Estimation. *Annals of Statistics*, 13, 1465-1481.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36, 394-419.

Muller, H. G. (1985). Empirical bandwidth choice for nonparametric kernel regression by means of pilot estimators. *Statistical Decisions*, 2, 193-206.

Priebe, E. C. (1994). Adaptive mixtures. *Journal of the American Statistical Association*., 89, 796-806.

Schoenberg, I .J. (1938). Metric spaces and completely monotonic functions. *Annals of Math*, 39, 811-841.

Wilson, T. C. (1998). Value at risk. *Risk Management and Analysis*, 1, 61-124.

Zangari, P. (1996). An improved methodology for measuring VaR. *Risk Metrics Monitor*, 7-25.

Zhang, C. (1990). Fourier methods for estimating mixing densities and distributions. *Annals of Statistics*, 18, 806-831.