Revisiting the Finite Mixture of Gaussian Distributions with Application to Futures Markets

Thierry Ané*
Chiraz Labidi

We present a new estimation method for Gaussian mixture modeling, namely, the kurtosis-controlled expectation-maximization (EM) algorithm, which overcomes the limitations of the usual estimation techniques via kurtosis control and kernel splitting. Our simulation study shows that the dynamic allocation of kernels according to the value of the total kurtosis measure makes the proposed kurtosis-controlled EM algorithm an efficient method for Gaussian mixture density estimation. This algorithm yielded considerable improvements over the classical EM algorithm. We then used the discrete Gaussian mixture framework to account for the observed thick-tailed distributions of futures returns and applied the kurtosis-controlled EM algorithm to estimate the distributions of real (agricultural, metal, and energy) and financial (stock index and currency) futures returns. We proved that this framework is perfectly adapted to capturing the departures from normality of the observed return distributions. Unlike

*Correspondence author, HEC Lausanne, Institute of Banking and Financial Management, CH-1015 Lausanne, Dorigny, Switzerland; e-mail: thierry.ane@hec.unil/ch

Received January 2000; Accepted June 2000

- *Thierry Ané is an Assistant Professor at HEC Lausanne, Institute of Banking and Financial Management in Dorigny, Switzerland.*
- *Chiraz Labidi is a Ph.D. Student at the University Paris IX Dauphine, CEREG and ESSEC in Cergy Pontoise Cedex, France.*

The Journal of Futures Markets, Vol. 21, No. 4, 347–376 (2001)

in previous studies, we found that a two-component Gaussian mixture is too poor a model to accurately capture the distributional properties of returns. Similar results have been obtained for stocks, indexes, currencies, interest rates, and commodities. This has important implications in many financial studies using Gaussian mixtures to incorporate the thickness of the tails of the distributions in the computation of the value at risk or to infer implied risk-neutral densities from option prices, to name but a few. © 2001 John Wiley & Sons, Inc. Jrl Fut Mark 21: 347–376, 2001

INTRODUCTION

For many years, both financial economists and statisticians have been concerned with the description of asset returns with a special interest in forecasting issues. Until recently, point and interval estimates seemed adequate to most users' needs, and no particular attention was given to constructing density estimates. In recent years, quantitative finance has emphasized the importance of accurate density estimation. Indeed, the form of asset return distributions has important implications for many financial models, including the mean-variance paradigm, the theoretical models of asset prices, and the pricing of contingent claims. In addition, the booming area of financial risk management is also effectively dedicated to providing density estimates of portfolio values and to tracking certain aspects of the densities, such as the value at risk.

Studies as early as Fama (1965) showed that asset returns are leptokurtic, that is, "peaked" and "fat-tailed" relative to the normal, and frequently positively skewed. Numerous models that incorporate these features in the distribution of asset returns have been tested so far with conflicting and often discouraging results. Although many sophisticated mathematical tools have been developed to present new distributions, far fewer efforts have been made to improve the estimation procedure of the more conventional framework of discrete Gaussian mixtures.

Kon (1984) carried out the most extensive study of discrete Gaussian mixtures but did not have the actual mathematical tools to accurately estimate the parameters of the model. He maximized the log likelihood of the normal mixture using the modified quadratic hill-climbing algorithm described in Goldfeld and Quandt (1972). Very recently, Lekkos (1999) used another well-known estimation technique with the test for the number of components already presented by Kon. Nevertheless, many estimation problems inherent to the finite mixture framework remain overlooked in the financial literature. Kon's algorithm often fails to con-

verge to a solution within the feasible parameter space (i.e., it produces a number of negative probabilities for π_j), especially when the number of Gaussians in the mixture is increasing. In addition, both Kon and Lekkos wrongly used the chi-square distribution of the generalized likelihood ratio to determine the number of components in the mixture. It is, therefore, very difficult to draw convincing conclusions from these empirical studies about both the validity of the Gaussian mixture model and the optimal number of components in the mixture.

Other authors such as Bahra (1996) and Söderlind (1997) suggested the use of a two-component discrete mixture to infer information from option prices without testing or at least discussing the choice of the number of mixing distributions. The discrete mixture model suffers from its ad hoc use in the quantitative finance literature. The main contribution of this article is to revisit the estimation techniques related to the discrete mixture models, using the algorithm introduced by Vlassis and Likas (1999) to prove that the Gaussian mixture densities are rich enough to capture the characteristics of financial asset returns. Hence, the family of discrete Gaussian mixtures represents a good candidate for the modeling of asset returns.

The article is organized as follows. In the Discrete Mixture Distribution section, we present the Gaussian mixture framework and discuss the different estimation problems inherent to this framework. The Kurtosis-Controlled EM Algorithm section introduces new advanced techniques to select and estimate the parameters of the model. The simulation results, summarized in the Simulation Study section, prove the validity of the kurtosis-controlled algorithm developed in the previous section. The results of an empirical study carried out on 24 futures contracts is reported in the Empirical Investigation section. It is shown that the mixture of Gaussian distributions is perfectly adapted to the futures markets. The resulting densities for futures returns have a high degree of fit to historical data, simple mathematical tractability for modeling, and flexibility to describe many of the diverse patterns of observed returns. An example of the benefit of this density estimation method for margin setting in futures markets concludes the section. The last section gives concluding remarks.

DISCRETE MIXTURE DISTRIBUTION

The most natural derivation of the mixture model arises when one samples from a population that consists of several homogeneous subpopulations indexed by $j = 1,2 \ldots q$ without having access to the index label.

One of the interests lies in ascertaining the number of components q in the mixture. However, mathematically the problem is ill posed unless one constrains the component densities to lie in some parametric family. In this article, the component densities are assumed to be normal because of the mathematical tractability of this distribution and its widespread use in finance.

We denote $n(x;\mu_j,\sigma_j^2)$ the Gaussian density of the jth component, represented by the vector of the parameters, $\theta_j = (\mu_j,\sigma_j^2)$. The proportion of the total population in the jth component, also called the *component weight*, is denoted by π_j. In this context, the observed data are a sample from the mixture density:

$$g(x;\, Q) = \int f(x;\, \theta)\, dQ(\theta) = \sum_{j=1}^{q} \pi_j\, n(x;\, \mu_j,\, \sigma_j^2) \tag{1}$$

The mixing distribution Q is assumed to be discrete, having a small number of points of support, $\theta_1, \theta_2 \ldots \theta_q$, with unknown support weights, $\pi_1, \pi_2 \ldots \pi_q$. If there are q components in the mixture, it is called a *q-component finite mixture model.*

When the number of components q is known, the expectation-maximization (EM) algorithm (Dempster, Laird, Rubin, 1977) is the most commonly accepted method for estimating the remaining parameters. The EM provides iterative formulae to estimate the parameters and can be proven to monotonically increase in each step the likelihood function. The EM estimates of the parameters are solutions to the system of nonlinear equations:

$$\pi_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} P(j/x_i) \tag{2}$$

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{n} P(j/x_i)x_i}{\sum_{i=1}^{n} P(j/x_i)} \tag{3}$$

$$\sigma_j^{2(t+1)} = \frac{\sum_{i=1}^{n} P(j/x_i)(x_i - \mu_j^{(t+1)})^2}{\sum_{i=1}^{n} P(j/x_i)} \tag{4}$$

$$P(j/x_i) = \frac{\pi_j^{(t)}\, n(x_i, \mu_j^{(t)}, \sigma_j^{2(t)})}{\sum_{k=1}^{q} \pi_k^{(t)}\, n(x_j, \mu_k^{(t)}, \sigma_k^{2(t)})} \tag{5}$$

Despite its simplicity, the EM algorithm presents many drawbacks that are overlooked in the financial literature. First, the EM requires an initialization of the parameter vector near the solution, and the algorithm is very sensitive to the choice of these starting values. In addition, the Gaussian densities in the mixture present nonlinearities with respect to their parameters μ_j and σ_j^2. Thus, the likelihood function will almost surely exhibit local maxima or saddle points. Although the EM monotonically increases the likelihood of the observations in each step, it cannot ensure a convergence of the parameters to a global maximum satisfying appropriate nonsingularity conditions and may easily get stuck in a local maximum or saddle point. Moreover, in financial problems, the number of components q is unknown and the EM algorithm cannot be used unless it is coupled with some estimation techniques for the parameter q.

When the number of components q is unknown, the objective is to find the mixture with the fewest components that provides a satisfactory fit to the data. The problem of estimating q also has a long history, but no definitive methods have been obtained (see McLachlan & Basford, 1988). First, it is known that considering the multimodality of the empirical density can be an insensitive approach to component estimation.[1] Moreover, the estimation of the number of normal distributions cannot be performed by maximization of the likelihood with respect to the number of kernels (i.e., the number of Gaussian distributions).[2]

In the financial literature, the problem has been answered with the generalized likelihood ratio test and its asymptotic χ^2 distribution (see Kon, 1984; Lekkos, 1999). Given a proposed mixture model with assumed parametric forms for the component densities, a hypothesized mixture with fewer components may be regarded as the imposition of a null hypothesis on the original model framework. The problem of comparing the two mixture models may, therefore, seem to be that of testing between nested hypotheses. Hence, the use of the generalized likelihood ratio test. However, this traditional test does not apply in this context.[3]

[1]For instance, it can be shown that in a two-component normal mixture, the resulting density is not bimodal unless the means are separated by at least 2 standard deviations.

[2]Such an approach would lead to using one kernel for each input sample of X, with the kernel mean equal to the sample. This solution would unambiguously give rise to overfitting.

[3]Indeed, to give an example, consider the apparent simple problem of testing between a Gaussian model $[H_0\colon\ g(x,Q) = n(x;\mu_1,\sigma_1^2)]$ and a two-component Gaussian mixture $[H_1\colon\ g(x,Q) = \pi_1 n(x;\mu_1,\sigma_1^2) + (1 - \pi_1)n(x;\mu_2,\sigma_2^2)]$. The traditional approach for testing between nested hypotheses-

To overcome these limitations, we present in the next section a new algorithm for Gaussian mixture modeling, introduced by Vlassis and Likas (1999), that does not require starting values, introduces higher moments to increase the chance of escaping from local maxima or saddle points, and dynamically finds the number of components that provides the best fit to the data.

KURTOSIS-CONTROLLED EM ALGORITHM

Capturing the Information Content beyond the Second Moment

If we set aside for some time the problems discussed in the previous section, for a given number of kernels q in the Gaussian mixture, the EM algorithm provides a maximization of the likelihood function. However, the value of the likelihood at the maximum does not constitute an indicator of the fit to the unknown distribution. Indeed, two densities may produce similar maximum likelihood values but may differ significantly according to other distance measures. The key idea of Vlassis and Likas (1999) is to integrate the EM algorithm into a new algorithm that also monitors the goodness of fit of an actual model to the unknown distribution.

To shed light on how the EM works, let us first note that the posterior probability $P(j/x_i)$ can be regarded as the probability that a sample x_i belongs to a kernel j. According to these probabilities, the EM algorithm tries to optimally distribute the q kernels over the input space while capturing the main information about the unknown distribution. Whenever the number of kernels is too small to accurately describe the unknown distribution in some parts, the EM will provide an estimated density that underestimates the unknown density in these parts. In this situation, the parameters of each kernel (μ_j and σ_j^2) are well estimated by Equations 3 and 4, although the fit remains unsatisfactory.

Therefore, the first two moments of the kernels cannot be used to assess the adequacy of the estimated model to the unknown distribution. We need to resort to higher moments to decide whether a particular ker-

would use the generalized likelihood ratio test, referring for significance to a table of the χ_α^2 distribution, where α is the number of constraints imposed on H_1 to produce H_0. The difficulty here arises from the fact that there is not a unique way of obtaining H_0 from H_1. Indeed, we could impose $\pi_1 = 0$ (one constraint) or $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$ (two constraints). What are the appropriate degrees of freedom for the chi-square distribution in this context? Chapter 5 of Titterington, Smith, and Makov (1985) shows that the generalized likelihood ratio test cannot be applied at all in this context.

nel j adequately fits the samples x_i lying in its vicinity. The most natural tool to use in the context of Gaussian distribution is undoubtedly the kurtosis of the distribution. Under the hypothesis of Gaussianity, the following equation holds for each kernel:

$$\int_{-\infty}^{+\infty} \left(\frac{x - \mu_j}{\sigma_j}\right)^4 n(x; \mu_j, \sigma_j^2)\, dx = 3 \tag{6}$$

At this stage, we can use Bayes formula to rewrite the Gaussian kernel $n(x;\mu_j,\sigma_j^2)$ in terms of the density $g(x,Q)$ of the mixture:

$$n(x; \mu_j, \sigma_j^2) = \frac{P(j/x_n)}{\pi_j} g(x, Q) \tag{7}$$

The insertion of Equation 7 into Equation 6 and the approximation of the integral yield the following:

$$\frac{1}{n\pi_j} \sum_{i=1}^{n} \left(\frac{x_i - \mu_j}{\sigma_j}\right)^4 P(j/x_i) = 3 \tag{8}$$

With Equation 2, this can be restated as

$$\frac{\sum_{i=1}^{n} \left(\frac{x_i - \mu_j}{\sigma_j}\right)^4 P(j/x_i)}{\sum_{i=1}^{n} P(j/x_i)} = 3 \tag{9}$$

On the basis of this result, we can define the so-called *weighted kurtosis* k_j of the jth kernel of the Gaussian mixture:

$$k_j = \frac{\sum_{i=1}^{n} \left(\frac{x_i - \mu_j}{\sigma_j}\right)^4 P(j/x_i)}{\sum_{i=1}^{n} P(j/x_i)} - 3 \tag{10}$$

This is the weighted equivalent of the kurtosis of a distribution, with weights equal to the posterior probabilities $P(j/x_i)$, in the same way as Equations 3 and 4 represent the weighted sample mean and variance of the random variable X.

In a Gaussian mixture, the weighted kurtosis k_j should be zero for all components $j = 1, 2 \ldots q$. For some kernel j, the departure from zero of the weighted kurtosis k_j clearly indicates that the distribution of

the samples in its vicinity is non-Gaussian. On the basis of this assertion, we can construct a measure of the goodness of fit of an estimated q-component Gaussian mixture by a weighted average of the individual weighted kurtoses k_j of the kernels of the mixture:

$$K_T = \sum_{j=1}^{q} \pi_j |k_j| \tag{11}$$

This new measure, introduced by Vlassis and Likas (1999), was called the *total kurtosis* of the mixture.

Unlike for the maximum likelihood, which has no reference value, we know that the lower bound for the total kurtosis K_T is zero. As a result, a low value indicates that each kernel correctly fits the samples in its vicinity, whereas a large value of the total kurtosis indicates that at least one kernel fails to provide an acceptable approximation of the unknown density in its vicinity.

Toward a New Algorithm (Vlassis & Likas, 1999): The Kurtosis-Controlled EM

The basic idea of the new algorithm, hereafter called the *kurtosis-controlled EM algorithm*, is that the maximization of the likelihood by the EM method should go together with a decrease in the total kurtosis value of the mixture. Starting with a small number q of kernels, we carry out EM steps to adjust the parameters of these q kernels so that the likelihood is increasing and the total kurtosis is decreasing. This procedure is continued until either a local maximum is reached or the total kurtosis reaches a minimum and starts increasing.

In the first situation, the local maximum of the likelihood inevitably corresponds to a local minimum of the total kurtosis because no further EM steps would update the parameters. Either the total kurtosis value is considered small enough to accept the solution (i.e., the estimated model fits well the unknown distribution) and the algorithm is stopped, or the total kurtosis is too high to consider that the estimated model correctly fits the data and a Gaussian mixture with more components is needed. Then, the current local maximum parameters will be used to create an initial parameter vector for the EM algorithm with $q + 1$ kernels. The number of kernels in the mixture is increased by one of the q previous kernels being split according to a rule described later in this section.

In the second situation, the algorithm is stopped by an increase in the total kurtosis, although no maximum of the likelihood has been

reached yet. It is considered that the EM algorithm has provided the best possible fit with q kernels and that additional kernels are needed to improve the fit to the unknown density. An increase in the total kurtosis value is thus also regarded as a signal for kernel splitting. At least $q + 1$ kernels will be required to better the approximation.

In both cases, kernel splitting is triggered according to the value of the total kurtosis. We now need to explain the mechanism of the split and particularly which of the previous q kernels will be split into two new kernels for the next EM step. A reasonable selection criterion is to split the kernel j that contributes most significantly to total kurtosis K_T. Hence, the selected kernel j will be the kernel with the higher value of $\pi_j|k_j|$. This kernel will be split into two kernels with means $\mu_j - \sigma_j$ and $\mu_j + \sigma_j$, a similar variance σ_j^2, and prior probabilities $\pi_j/2$, so that Equation 2 still holds with the new $q + 1$ kernels. After splitting, the $q + 1$ kernels continue to be updated with the EM algorithm.

To end this algorithm, we still have to specify a termination criterion. Given that the EM converges for a fixed number of kernels, producing a termination criterion is tantamount to disabling kernel splitting at some stage when a condition is satisfied. The EM will then inevitably converge to a maximum value of the likelihood with the current number of kernels.

The total kurtosis is our benchmark for estimating the degree of goodness of fit. At each kernel split, the value of the total kurtosis just after the split is thus stored and compared with the value of the total kurtosis at the next split. When the difference becomes small, it is thought that splitting is no longer effective in the sense that it does not bring a better adequacy to the unknown distribution. From that time on, we keep the number of kernels fixed and perform EM steps to reach the local maximum corresponding to this number of kernels and accept this solution.

Advantages of the Kurtosis-Controlled EM Algorithm

Vlassis and Likas' (1999) algorithm, described in the previous subsection, dynamically adds kernels according to the value of the total kurtosis. With this method, no initial knowledge of the number of components q of the mixture is required. The selection of q is made by an assessment of the goodness of fit, and the problems related to the applicability of the generalized likelihood ratio test are avoided.

Further, for $q = 1$, the maximum likelihood estimates of the parameters are easy to compute:

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{12}$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \tag{13}$$

We can initialize the previous algorithm with $q = 1$ and the values in Equations 12 and 13, compute the total kurtosis K_T, and perform the first split if the kurtosis is not sufficiently small. With this method, no initial parameter vector is needed. We no longer have the usual problems of the sensitivity of the EM to the starting values nor the necessity to provide initial parameters near the solution. Last but not least, this algorithm is also capable of potentially escaping from local maxima of the likelihood function by splitting one kernel.

In conclusion, via kurtosis control and kernel splitting, ad hoc estimation techniques for the number of kernels are no longer needed. The kurtosis-controlled EM algorithm avoids the usual shortcomings of the EM while preserving its remarkable property of monotonic convergence.

SIMULATION STUDY

In this section, we present some simulation results to examine the effectiveness of this new approach in locating the maximum likelihood estimates of a Gaussian mixture. We first drew observations from several known Gaussian mixtures that we then tried to approximate using both the kurtosis-controlled EM algorithm and the traditional EM algorithm. In every estimation, we started the kurtosis-controlled EM algorithm with one Gaussian component with the traditional maximum likelihood parameter estimates in Equations 12 and 13. Then, splits were performed according to the level of total kurtosis, as previously described. The EM algorithm was performed with the right number of kernels because no reliable method is available to estimate this parameter in the traditional framework. For the EM starting values, the same probability was affected for each kernel, the variance of each kernel was set equal to the empirical variance of the series with Equation 13, and the means were randomly selected within the range of the data. We generated 20 mean vectors and launched the EM algorithm for each corresponding global parameter vector. Among the 20 resulting solutions, we kept the solution that maximized the log likelihood.

We considered 40[4] simulated data sets of sample size $n = 10{,}000$. For brevity, we only report the results from four typical simulations in Table I. Although the log likelihood is generally not a good indicator of the fit for an empirical investigation, as we have already argued, it conveys more information in a simulation study. Indeed, because the original density $g^{\text{theoretical}}(x,Q)$ is known, we can use the series $\{x_i\}_{i=1,2\ldots n}$ of the simulated data and the true parameters to compute the theoretically optimal log likelihood $L^{\text{theoretical}} = \Sigma_{i=1}^{n} \ln [g^{\text{theoretical}}(x_i,Q)]$, which can be compared to the log likelihoods obtained with the kurtosis-controlled EM algorithm ($L^{\text{K-EM}}$) and the conventional EM algorithm (L^{EM}). In the same manner, a theoretical total kurtosis measure can be computed with the simulated data and the true distribution parameters.

Simulation 1

In this first sample, the random variable follows a three-component mixture with density:

$$g_1^{\text{theoretical}}(x, Q) = 0.4n(x; -2, 1) + 0.3\, n(x; 1, 0.64) + 0.3n(x; 5, 2.25) \quad (14)$$

The main feature of this Gaussian mixture is to clearly exhibit multimodality. The value of the theoretical log likelihood is $L^{\text{theoretical}} = 0.0255 = -24{,}071.02$. The theoretical total kurtosis is $K_T^{\text{theoretical}} = 0.0155$. Figure 1(a) displays the original density as well as the estimated densities obtained with both the kurtosis-controlled EM approach and the traditional EM algorithm. The resulting log likelihood for the EM method is $L^{\text{EM}} = -24{,}086.31$ with a total kurtosis of $K_T^{EM} = 0.1543$. In comparison, the kurtosis-controlled EM algorithm was able to assess the right number of kernels and produced a log likelihood of $L^{\text{K-EM}} = -24{,}071.13$ for a total kurtosis of $K_T^{K\text{-}EM} = 0.0263$.

It is apparent in Figure 1(a) that the kurtosis-controlled EM algorithm provided a better fit than the EM. In addition, let us insist on the fact that this algorithm was also able to estimate the good number of Gaussian densities in the mixture, a variable that remains difficult to obtain with the traditional approach. Moreover, the departure from the true density is measured by the total kurtosis. The level of K_T^{EM} and $K_T^{K\text{-}EM}$ can be used to form an opinion about the degree of misfitting.

[4]We drew simulations of mixtures with the number of kernels ranging from 1 to 10. For each number of kernels, four simulations were drawn, two exhibiting multimodality and two within the unimodal region. For all simulations, the parameters were chosen randomly.

TABLE I

EM Versus Kurtosis-Controlled EM Parameter Estimates

Simulation 1

Kurtosis-Controlled EM Algorithm			*EM Algorithm*		
μ	σ^2	π	μ	σ^2	π
−2.0279	0.9619	0.3988	−2.0259	0.9816	0.3867
0.9907	0.6754	0.3066	0.9828	0.6433	0.3164
5.0382	2.1737	0.2946	5.0288	2.1566	0.2969
	$L^{K\text{-}EM}$ −24071.13			L^{EM} −24086.31	
	$K_T^{K\text{-}EM}$ 0.0263			K_T^{EM} 0.1543	

Simulation 2

Kurtosis-Controlled EM Algorithm			*EM Algorithm*		
μ	σ^2	π	μ	σ^2	π
−0.0610	0.0015	0.0105	−0.0356	0.0082	0.0144
−0.0013	0.0087	0.4118	−0.0083	0.0002	0.3682
0.0007	0.0003	0.5360	0.0010	0.0002	0.5654
0.0589	0.0026	0.0417	0.0503	0.0063	0.0520
	$L^{K\text{-}EM}$ 15570.35			L^{EM} 11431.55	
	$K_T^{K\text{-}EM}$ 0.0512			K_T^{EM} 1.5354	

Simulation 3

Kurtosis-Controlled EM Algorithm			*EM Algorithm*		
μ	σ^2	π	μ	σ^2	π
−0.0021	3.2369E-04	0.0406	−0.0017	3.1444E-04	0.0412
−0.0001	1.3800E-06	0.6182	−0.0001	1.3800E-06	0.6070
0.0002	2.6350E-05	0.3412	0.0001	2.3330E-05	0.3518
	$L^{K\text{-}EM}$ 43733.32			L^{EM} 43729.27	
	$K_T^{K\text{-}EM}$ 0.0270			K_T^{EM} 0.0367	

Simulation 4

Kurtosis-Controlled EM Algorithm			*EM Algorithm*		
μ	σ^2	π	μ	σ^2	π
−0.2002	2.5820E-03	0.0801	−0.1976	2.5306E-03	0.1180
−0.0501	4.8180E-05	0.3499	−0.0501	5.0400E-05	0.3497
0.1499	4.9290E-05	0.4500	0.1498	4.9180E-05	0.3200
0.2501	6.1980E-05	0.1200	0.2498	6.5060E-05	0.2200
	$L^{K\text{-}EM}$ 21807.89			L^{EM} 21375.83	
	$K_T^{K\text{-}EM}$ 0.1307			K_T^{EM} 0.8503	

Note: The parameters were obtained both with the EM algorithm and the kurtosis-controlled EM algorithm for the simulated data sets drawn from the densities in Equations 14 to 17. The table is divided into four parts corresponding to the different simulations. Each part compares the effectiveness of both methods. We give the estimated parameters, the log likelihood, and the level of total kurtosis. Each simulation confirms the superiority of the kurtosis-controlled EM algorithm. The theoretical log likelihoods for the different simulations are −24071.02, 15571.92, 43736.49, and 21913.58, respectively. The theoretical total kurtosis values are 0.0155, 0.0345, 0.0114, and 0.1116, respectively.

Figure 1a: Simulation 1

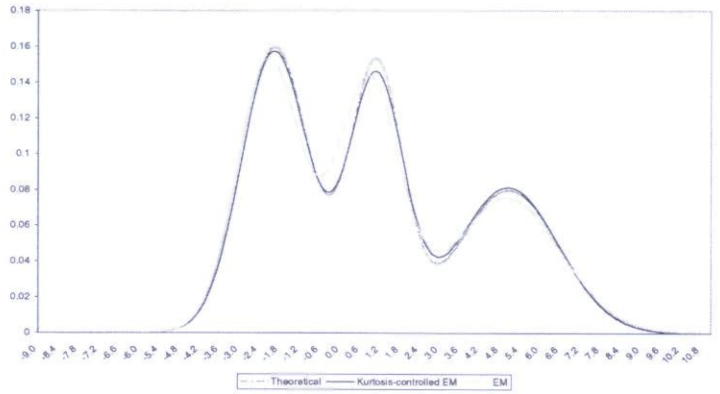

Figure 1b: Simulation 2

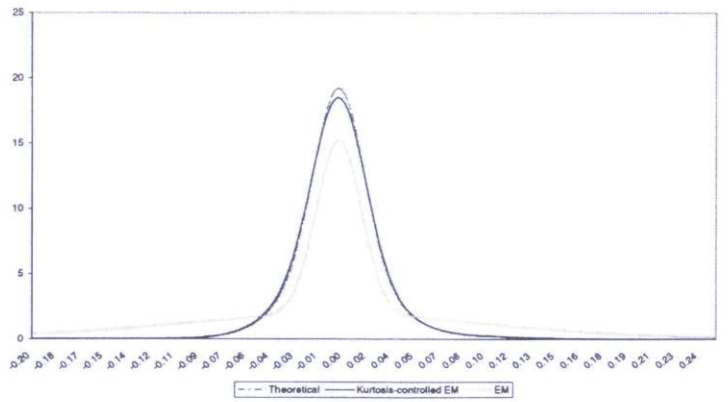

Figure 1c: Simulation 3

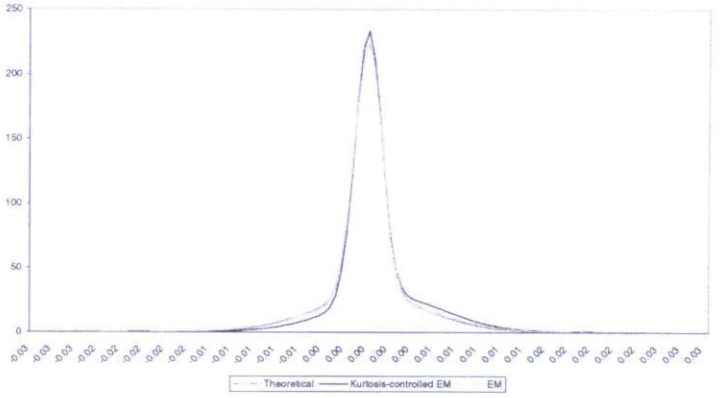

Figure 1d: Simultation 4

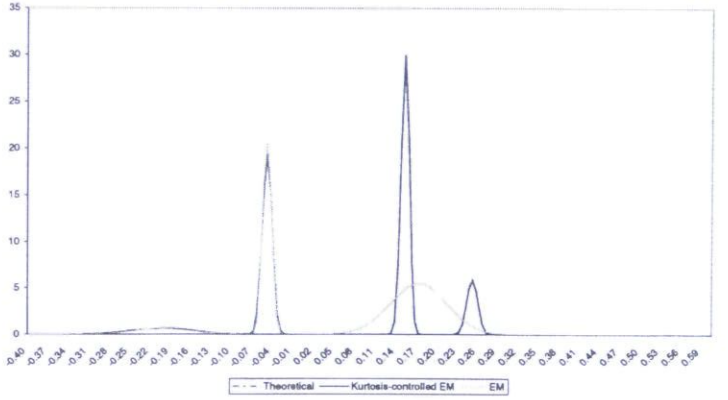

FIGURE 1

EM versus kurtosis-controlled EM Gaussian mixture densities: For each simulation data set, we used the EM and kurtosis-controlled EM estimated parameters to plot the corresponding return densities. On each chart is represented the graph of the theoretical Gaussian mixture density that has been used to generate the data set. The mere observation of the graphs reveals the superiority of the kurtosis-controlled EM algorithm to approximate the theoretical density. In addition, let us recall that the EM algorithm has been run with the right number of components as input (a very difficult parameter to estimate in this framework), whereas the kurtosis-controlled EM algorithm has found the exact number q of kernels.

Simulation 2

Observations for the second simulation were generated from the density:

$$g_2^{\text{theoretical}}(x,Q) = 0.01n(x; -0.06, 0.0144) + 0.4n(x; -0.001, 0.009) + 0.55n(x; 0.0007, 0.0003) + 0.04n(x; 0.06, 0.0025) \quad (15)$$

Unlike in the previous simulation, this mixed density is within the unimodal region described by Robertson and Fryer (1969). Hence, we would expect it to be more difficult to detect the different components in the mixture.

In practice, empirical return distributions are seldom multimodal. However, the usual simulation tests presented in the statistical literature to support the EM use multimodal densities similar to that of Simulation 1. This new simulation will give us some insights into the behavior of both algorithms when the mixture becomes more subtle and thus more difficult to detect.

Figure 1(b) displays the original density as well as the solutions obtained with both methods. The theoretical log likelihood is $L^{\text{theoretical}} = 15{,}571.92$. Although it was applied on the right number of kernels, the EM algorithm was stuck in a local maximum with $L^{\text{EM}} = 11{,}431.55$ and $K_T^{EM} = 1.5354$. Because we used the right number of kernels, no better fit has been found because of the starting values and local maxima or saddle points. None of the 20 vectors of the starting values enabled the EM algorithm to get close to the solution. We have already explained that the EM algorithm needs starting parameters near the solution. Performing the EM with 20 different initial vectors of the parameters was not enough to reach a suitable fit.

On the contrary, the kurtosis-controlled EM algorithm managed to determine the number of kernels in the mixture and provided a much better fit with $L^{\text{K-EM}} = 15{,}570.35$ and $K_T^{K\text{-}EM} = 0.0512$. Let us note that the theoretical kurtosis is $K_T^{\text{theoretical}} = 0.0345$.

Simulations 3 and 4

The estimation of the number of components and parameters of a finite mixture is more complex when the density exhibits no multimodality. Another problem arises when the parameters of the Gaussian densities in the mixture are small. This will undoubtedly occur with financial asset

returns. To test the behavior of the proposed algorithm in this situation, we generated two more series respectively from the following densities:

$$g_3^{\text{theoretical}}(x, Q) = 0.04n(x; -0.002, 3.24E-04) + 0.62n(x; -0.0001, 1.44E-06) + 0.34n(x; 0.0002, 2.5E-05) \quad (16)$$

and

$$g_4^{\text{theoretical}}(x, Q) = 0.08n(x; -0.2, 2.5E-03) + 0.35n(x; -0.05, 4.9E-05) + 0.45n(x; 0.15, 3.6E-05) + 0.12n(x; 0.25, 6.4E-05) \quad (17)$$

The parameters of these two simulations were chosen to approximately match the estimated parameters of two futures contracts presented in the following section.

In Simulation 3, the theoretical log likelihood was $L^{\text{theoretical}} = 43{,}736.49$ for a total kurtosis of $K_T^{\text{theoretical}} = 0.0114$. The EM algorithm ended up with a log likelihood value of $L^{\text{EM}} = 43{,}729.27$ and a total kurtosis of $K_T^{EM} = 0.0367$, indicating a satisfactory fit. Although the departure from the theoretical model is less important than what we found in the previous simulations, Figure 1(c) still displays a difference in the rate of convergence between the theoretical model and the mixture obtained with the EM algorithm.

Again, the kurtosis-controlled EM provided a better solution. The obtained parameters are very close to the true values, the log likelihood $L^{\text{K-EM}} = 43{,}733.32$ is improved, and the total kurtosis measure $K_T^{K\text{-}EM} = 0.0270$ is closer to what was found with the theoretical density.

Simulation 4 gave a total kurtosis of $K_T^{\text{theoretical}} = 0.1116$ and a theoretical log likelihood of $L^{\text{theoretical}} = 21{,}913.58$. The best local maximum found with the EM merely provided a log likelihood value of $L^{\text{EM}} = 21{,}375.83$. Figure 1(d) shows how badly this solution reflects the characteristics of the true density. The EM total kurtosis measure $K_T^{EM} = 0.8503$ indicates how high the total kurtosis is for a very bad goodness of fit.

The near perfect identity of the kurtosis-controlled EM estimated density and the true density reveals the effectiveness of this new method in estimating the parameters of a Gaussian mixture. Nevertheless, the total kurtosis $K_T^{K\text{-}EM} = 0.1307$ is the highest obtained with the kurtosis-controlled EM among the four simulations. However, the graph and the

estimated parameters compared with the true values clearly show that the fit is good. Let us note that this total kurtosis must be compared to the theoretical kurtosis $K_T^{\text{theoretical}} = 0.1116$, also very high in this simulation.

For a single Gaussian distribution, there exists a confidence interval for the empirical value of the kurtosis. As long as the kurtosis remains within the bounds of this confidence interval, the kurtosis is not statistically different from zero. The same idea applies to each kernel in the mixture: There exists a confidence interval for each kernel kurtosis k_j. If the kurtosis k_j remains in its confidence interval, it can be concluded that this kernel correctly fits the samples in its vicinity. Suppose that all the kernels produce kurtosis values k_j near the bounds of the intervals; then, the total kurtosis measure will be high even though the fit is good. In this respect, let us note that the theoretical total kurtosis in Simulation 4 is also higher than what we obtained with the other simulations. This supports our explanation for the level of $K_T^{K\text{-}EM}$. It thus seems that a total kurtosis value lower than 10^{-1} corresponds to a very good fit. However, to have another way of measuring the fit, we present in the empirical section another measure for the goodness of fit, namely, the Anderson–Darling (AD) statistic.

In conclusion, these simulations prove the superiority of the kurtosis-controlled EM algorithm in each situation. It enables us to estimate the number of kernels and then provides parameter values closer to the true parameters. In addition, the theoretical and estimated values of the total kurtosis measure provide us with a basis for comparison. At first, the theoretical values of the total kurtosis range from 0.0155 to 0.1116. Hence, a total kurtosis value of about 10^{-2} represents a perfect fit. However, the kurtosis of $K^{\text{EM}} = 0.8503$ obtained with the fourth simulation provides a reference value for a very badly estimated model. The different levels of kurtosis obtained for the kurtosis-controlled EM algorithm reflect the different levels of good fit, whereas the values of K^{EM} together with the graphs allow us to appreciate the degree of misfitting.

EMPIRICAL INVESTIGATION

Data and Descriptive Statistics

This study includes real (agricultural, metal, and energy) and financial (stock index and currency) futures contracts. Daily settlement prices of 24 contracts traded on nine different exchange markets were obtained from Datastream International for the period January 1, 1990 to Septem-

ber 29, 1999. Continuous series of nearby daily futures prices were constructed. During the maturity months, the nearby futures prices were rolled over by the daily prices of the first deferred contract. Classically, the rollover was made when the traded volume on the first deferred contract became greater than the traded volume on the maturing contract. Return series were computed with the natural logarithm of the ratio of two consecutive prices. Each data set contained 2,542 return observations.

Table II indicates the exchange market where each contract was traded as well as the usual descriptive statistics for the return series. The last column of Table II yields the Jarque–Bera statistic. For all the futures contracts, normality was rejected at a 5% confidence level (the critical value of the statistic is 7.38). The abnormality of returns is supported by the existence of high kurtosis levels, ranging from 4.65 to 70.55. These values are obviously outside the Gaussian 95% confidence interval, 3 ± 0.19. Eighteen futures contracts out of 24 also exhibit significant skewness at the 5% level (the absolute skewness values are greater than 0.10).

Empirical Results

For each future return series in the sample, the vectors of the parameters $(\mu_1, \mu_2 \ldots \mu_q)$, $(\sigma_1^2, \sigma_2^2 \ldots \sigma_q^2)$, and $(\pi_1, \pi_2 \ldots \pi_q)$ of the Gaussian mixture were estimated simultaneously with the value of q, the number of components, with the kurtosis-controlled EM algorithm. Table III summarizes the estimation results obtained when the algorithm was performed. It is organized into 24 subtables corresponding to the various futures contracts we considered. Each subtable provides the estimated number q of the components in the mixture as well as three columns reporting the vectors of the parameters. Four futures return series yielded a three-component mixture, 12 series were estimated by a four-component mixture, three return series were better approximated by a five-component mixture, another three series yielded a six-component mixture, and the remaining two series ended up with a seven-component mixture.

At this stage, it is important to note that the results largely differ from previous studies. Lekkos (1999) set forth the superiority of the two-component Gaussian mixture for the modeling of spot and forward interest rate returns. In his study, the specification testing was performed by the commonly used likelihood ratio test, where the maximized likelihood values from both specifications are compared. Aïtkin and Wilson (1980) remarked, however, that for finite mixture problems, the likelihood ratio test is not valid because of the nonregularity of the model. The

TABLE II

Descriptive Statistics of the Data

Futures Contract	*Futures Exchange*	*M*	*SD*	*Skewness*	*Kurtosis*	*Jarque–Bera*
Agricultural						
Cocoa	NYCSCE	4.50E-05	1.81E-02	0.7071	6.3602	1407.73
Corn	CBT	−4.52E-05	1.35E-02	−1.4643	32.1323	90799.13
Cotton	NYCE	−1.14E-04	1.50E-02	−3.4893	70.5505	488463.33
Feeder cattle	CME	−2.36E-05	7.98E-03	−0.0736*	6.6642	1424.36
Live cattle	CME	−5.49E-05	9.44E-03	−0.4845	12.9845	10658.33
Orange juice	NYCE	−2.38E-04	2.20E-02	0.9516	17.8906	23868.56
Soybeans	CBT	−5.67E-05	1.26E-02	−0.7196	11.7758	8376.46
Sugar	NYCSCE	−6.02E-05	4.55E-03	−1.4949	53.3135	269068.90
Wheat	CBT	−1.43E-04	1.50E-02	−0.6121	12.0180	8772.30
Energy						
Crude oil	NYMEX	4.49E-05	2.43E-02	−1.8167	38.7198	121827.95
Heating oil	NYMEX	−1.95E-04	2.34E-02	−2.5454	36.0680	118564.19
Unleaded gasoline	NYMEX	5.79E-05	2.26E-02	−1.1665	21.5787	37135.40
Currency						
Australian dollar	CME	−6.76E-05	5.76E-03	0.0063*	8.1625	2822.81
British pound	CME	1.35E-05	6.56E-03	−0.3620	7.9128	2611.65
Deutsche mark	CME	−3.05E-05	6.79E-03	0.0098*	5.3268	573.46
Japanese yen	CME	1.20E-04	7.81E-03	0.6839	11.7215	8387.41
Metals						
Copper	COMEX	−1.14E-04	1.46E-02	−0.2661	8.0667	1026.09
Gold	COMEX	−1.17E-04	7.64E-03	−0.3819	21.8008	37500.03
Platinum	NYMEX	−7.06E-05	1.08E-02	−0.0242*	6.6614	1420.18
Silver	CBT	2.48E-05	1.51E-02	−0.1060	7.0731	1761.89
Stock index						
CAC 40 index	MATIF	3.26E-04	1.28E-02	−0.0395*	6.1944	1081.47
FTSE 100 index	LIFFE	3.53E-04	1.04E-02	0.0143*	4.6497	288.33
NIKKEI 225 index	OSE	−3.26E-04	1.49E-02	0.2551	5.6692	782.18
S&P 500 index	CME	5.03E-04	9.48E-03	−0.3128	9.6796	4767.12

Note: Descriptive statistics of the 24 futures returns over the period January 1, 1990 to September 29, 1999 are presented. The data collected from Datastream International are extracted from nine different exchange markets. The number of observations for each series is 2542. Population skewness and kurtosis for a normal distribution are 0 and 3, respectively variance coefficients for a size n sample extracted from a normal population are $6/n$ and $24/n$, respectively. Confidence intervals (95%) for a test of futures returns normality are given by ± 0.1 for the sample skewness and 3 ± 0.19 for the sample kurtosis. The statistic of the Jarque–Bera test of normality is chi square distributed with two degrees of freedom. The critical value of χ^2_2 at the 5% level is 7.38, * indicates nonsignificance at a 5% level. Six series display no asymmetry, but because of significant excess kurtosis, normality is rejected for each series. The Jarque–Bera statistic corroborates this conclusion.

mixing proportion is not identifiable in the general (nonmixed) specification (see the example developed in the Discrete Mixture Distribution section).

According to the values of the mixing proportions, π_j, the great majority of futures returns are attributed to the component distributions with means closer to zero and smaller variances, whereas small values of

TABLE III

Parameter Estimates

CAC 40			*Gold*			*Japanese Yen*		
$q = 3$			$q = 3$			$q = 3$		
μ	σ^2	π	μ	σ^2	π	μ	σ^2	π
−1.0815E-03	7.8984E-04	0.0529	−1.2336E-02	1.6540E-03	0.0081	−3.9852E-04	1.1830E-05	0.3972
−1.3296E-04	2.5596E-04	0.2700	−2.8499E-04	1.3870E-05	0.6284	9.1990E-05	6.5790E-06	0.6845
8.1783E-04	7.8470E-05	0.6771	4.4408E-04	9.6200E-05	0.3635	5.9089E-03	4.6509E-04	0.0383
Australian Dollar			*Cocoa*			*Crude Oil*		
$q = 4$			$q = 4$			$q = 4$		
μ	σ^2	π	μ	σ^2	π	μ	σ^2	π
2.2874E-04	3.7249E-04	0.0123	−1.5491E-02	3.3906E-04	0.1434	−5.7887E-02	1.0700E-02	0.0099
1.8319E-03	8.5400E-06	0.2058	−1.8083E-03	1.3830E-04	0.6886	−1.2095E-03	7.7070E-04	0.4057
−6.7063E-04	4.1400E-06	0.2129	1.4882E-02	2.8374E-04	0.1574	8.8961E-04	1.2027E-04	0.8712
−5.3514E-04	4.4680E-05	0.5690	3.8997E-02	1.0354E-03	0.0306	5.4745E-02	2.3055E-03	0.0132
Deutsche Mark			*Feeder Cattle*			*FTSE 100*		
$q = 4$			$q = 4$			$q = 4$		
μ	σ^2	π	μ	σ^2	π	μ	σ^2	π
−2.8735E-03	7.0400E-05	0.1090	−4.6352E-03	1.4431E-04	0.0886	−4.8106E-03	2.2044E-04	0.1280
−1.8324E-04	9.7700E-06	0.3705	1.4265E-04	2.2310E-05	0.6133	−3.2264E-03	4.5120E-05	0.3881
6.4167E-04	4.7830E-05	0.4498	6.6313E-04	6.6880E-05	0.2317	4.3483E-03	5.1490E-06	0.3838
1.0300E-03	1.7060E-04	0.0707	2.1317E-03	2.2077E-04	0.0664	5.5223E-03	2.5436E-04	0.1004
Live Cattle			*NIKKEI 225*			*Platinum*		
$q = 4$			$q = 4$			$q = 4$		
μ	σ^2	π	μ	σ^2	π	μ	σ^2	π
4.6680E-06	1.2615E-03	0.0215	−8.9264E-03	1.9476E-04	0.1451	−2.7724E-03	8.9810E-05	0.3681
−5.3878E-04	9.4720E-05	0.4502	5.7680E-05	4.5920E-05	0.3960	−4.9521E-04	4.6119E-04	0.1031
1.6248E-04	1.5910E-06	0.3955	1.8519E-03	2.9975E-04	0.4170	5.1764E-04	2.3800E-05	0.3556
1.5896E-03	0.7800E-06	0.1320	4.1647E-03	9.1620E-04	0.0419	4.7147E-03	1.1471E-04	0.1732

π_j are associated with Gaussian distributions exhibiting high absolute values of mean and variance estimates. A representative example of this feature is the estimated density of the Standard and Poor's 500 Index futures returns. Of the observations, 58% and 30% are generated from the distributions $n(1.4E-04,9.7E-05)$ and $n(2.5E-04,7.7E-06)$, re-

TABLE III (Continued)

Parameter Estimates

S&P 500			*Soybeans*			*Unleaded Gasoline*		
$q = 4$			$q = 4$			$q = 4$		
μ	σ^2	π	μ	σ^2	π	μ	σ^2	π
−1.1405E-03	6.7949E-04	0.0401	−1.2016E-01	1.5720E-05	0.0008	−3.2383E-02	1.4563E-02	0.0065
−2.5213E-04	7.7600E-06	0.3027	−3.1480E-03	4.8170E-05	0.3781	−4.7938E-03	2.1691E-04	0.4118
1.3921E-04	9.7620E-06	0.5768	−2.4874E-04	4.0884E-04	0.2527	−8.1214E-04	1.3222E-03	0.1696
8.7747E-03	8.6900E-06	0.0804	3.5012E-03	5.2660E-05	0.3684	5.6770E-03	2.1286E-04	0.4134
Copper			*British Pound*			*Orange Juice*		
$q = 5$			$q = 5$			$q = 5$		
μ	σ^2	π	μ	σ^2	π	μ	σ^2	π
−9.7900E-03	9.9949E-04	0.0416	−4.6593E-02	3.3600E-06	0.0006	−9.7584E-03	1.1714E-04	0.2420
−9.5586E-03	2.0574E-04	0.2050	−1.8493E-02	2.3030E-05	0.0252	−2.7635E-03	9.2929E-04	0.2668
−4.6540E-03	5.6130E-05	0.2653	−8.6970E-05	3.3700E-06	0.2343	7.4990E-04	4.0300E-05	0.3057
2.7025E-03	3.8770E-06	0.2752	2.5721E-04	3.1950E-05	0.6851	1.3634E-02	1.4087E-04	0.1734
1.2902E-02	1.6531E-04	0.2129	6.5923E-03	1.2767E-04	0.0546	2.2339E-02	8.8979E-03	0.0121
Corn			*Cotton*			*Heating Oil*		
$q = 6$			$q = 6$			$q = 6$		
μ	σ^2	π	μ	σ^2	π	μ	σ^2	π
−7.2218E-02	7.0501E-03	0.0019	−8.6950E-02	8.8489E-03	0.0030	−5.1532E-02	1.2140E-02	0.0121
−1.2000E-02	2.4162E-04	0.1000	−9.9853E-03	1.0503E-04	0.2061	−1.9476E-02	7.6524E-04	0.0930
−8.7807E-03	8.9740E-05	0.1591	−8.1443E-03	2.3892E-04	0.1321	−1.2080E-02	1.3955E-04	0.1788
1.0796E-03	2.2440E-05	0.2224	−1.0528E-03	2.1600E-05	0.2877	−5.6820E-04	4.6050E-05	0.2098
3.6474E-03	8.0520E-05	0.4334	9.5855E-03	9.2420E-05	0.2900	7.1743E-03	1.9042E-04	0.4163
1.1377E-02	5.1968E-04	0.0832	9.9255E-03	3.2729E-04	0.0611	1.7044E-02	7.7872E-04	0.0900
Silver			*Wheat*			*Sugar*		
$q = 7$			$q = 7$			$q = 3$		
μ	σ^2	π	μ	σ^2	π	μ	σ^2	π
−2.3983E-02	1.4585E-03	0.0104	−7.2582E-02	2.1120E-03	0.0039	−1.2847E-03	3.4855E-04	0.0388
−9.6195E-03	3.6347E-04	0.0300	−1.0758E-02	4.9640E-05	0.1401	−1.0620E-04	1.5000E-06	0.6249
−5.1187E-03	2.9857E-04	0.2537	−8.4565E-03	2.2873E-04	0.1734	1.8951E-04	1.8430E-06	0.3365
−3.3628E-03	3.0390E-05	0.1806	9.2655E-04	6.6170E-05	0.1587			
1.6094E-03	1.8840E-05	0.1869	3.3203E-03	7.1210E-05	0.2859			
4.9938E-03	1.6121E-04	0.2980	6.7521E-03	1.6775E-04	0.1946			
1.6256E-02	7.4556E-04	0.0424	1.5529E-02	6.6017E-04	0.0454			

Note: The estimated values of the parameters obtained when the kurtosis-controlled EM algorithm was run for the 24 futures return series are given. They are organized in subtables corresponding to each futures contract. Each subtable provides the estimated number q of components in the mixture as well as three rows reporting the means, μ_j, the variances, σ_j^2, and the probabilities, π_j. Four futures return series yield a three-component mixture, 12 series are estimated by a four-component mixture, three return series are better approximated by a five-component mixture, another three series yield a six-component mixture, and two series end up with a seven-component mixture.

spectively. The remaining two distributions, $n(6.7E-03,6.6E-06)$ and $n(1.1E-03,6.8E-04)$, characterized by a higher absolute mean, variance, or both, only account for 12% of the returns.

These results agree with the conventional idea of rare events. Most of the time, asset returns are small in magnitude (they have a mean close to zero and a small variance). Once in a while, unexpected events cause large (positive or negative) returns that are responsible for the observed excess kurtosis and skewness.

As a point of reference, we reconstructed the empirical distributions of the futures returns by the kernel method and compared these nonparametric distributions with the q-component Gaussian mixtures estimated with our algorithm. Figure 2 displays the resulting densities for four different futures contracts: soybeans ($q = 4$), Japanese yen ($q = 3$), heating oil ($q = 6$), and copper ($q = 5$). These density distributions are representative of the goodness of fit obtained within our sample of futures contracts. Whatever the number of kernels required to approximate the empirical densities, the kurtosis-controlled EM algorithm provides accurate solutions. To give another visual representation of the accuracy of the fit, we also plotted, in Figure 3, the empirical distribution function $F_e(\bullet)$, reconstructed through the kernel method, together with the Gaussian estimated mixture distribution function $F_m(\bullet)$. Four futures contracts are represented: FTSE 100 ($q = 4$), corn ($q = 6$), silver ($q = 7$), and British pound ($q = 5$). The quasiperfect graphical fit obtained in Figures 2 and 3 is further evidence of the accuracy of our estimation method.

Moreover, to confirm the goodness of fit of the estimated Gaussian mixtures, we introduced the AD test (Anderson & Darling, 1952) test, which statistically determines how close the empirical distribution function and the estimated Gaussian mixture distribution function are for the observed return series. The AD test emphasizes deviations between the distributions in the tails, an important area of the distribution of returns for risk management, by using the following weighted test statistic:

$$\mathrm{AD} = \max_{x \in \mathcal{R}} \frac{|F_e(x) - F_m(x)|}{\sqrt{F_m(x)(1 - F_m(x))}} \tag{18}$$

where $F_e(\bullet)$ is the empirical distribution function and $F_m(\bullet)$ is the distribution function of the mixture model. A small (large) value of the AD test statistic corresponds to a small (large) difference between the two distribution functions indicating a good (bad) fit. For all the return series, the AD test statistics were computed for the optimal mixture model spec-

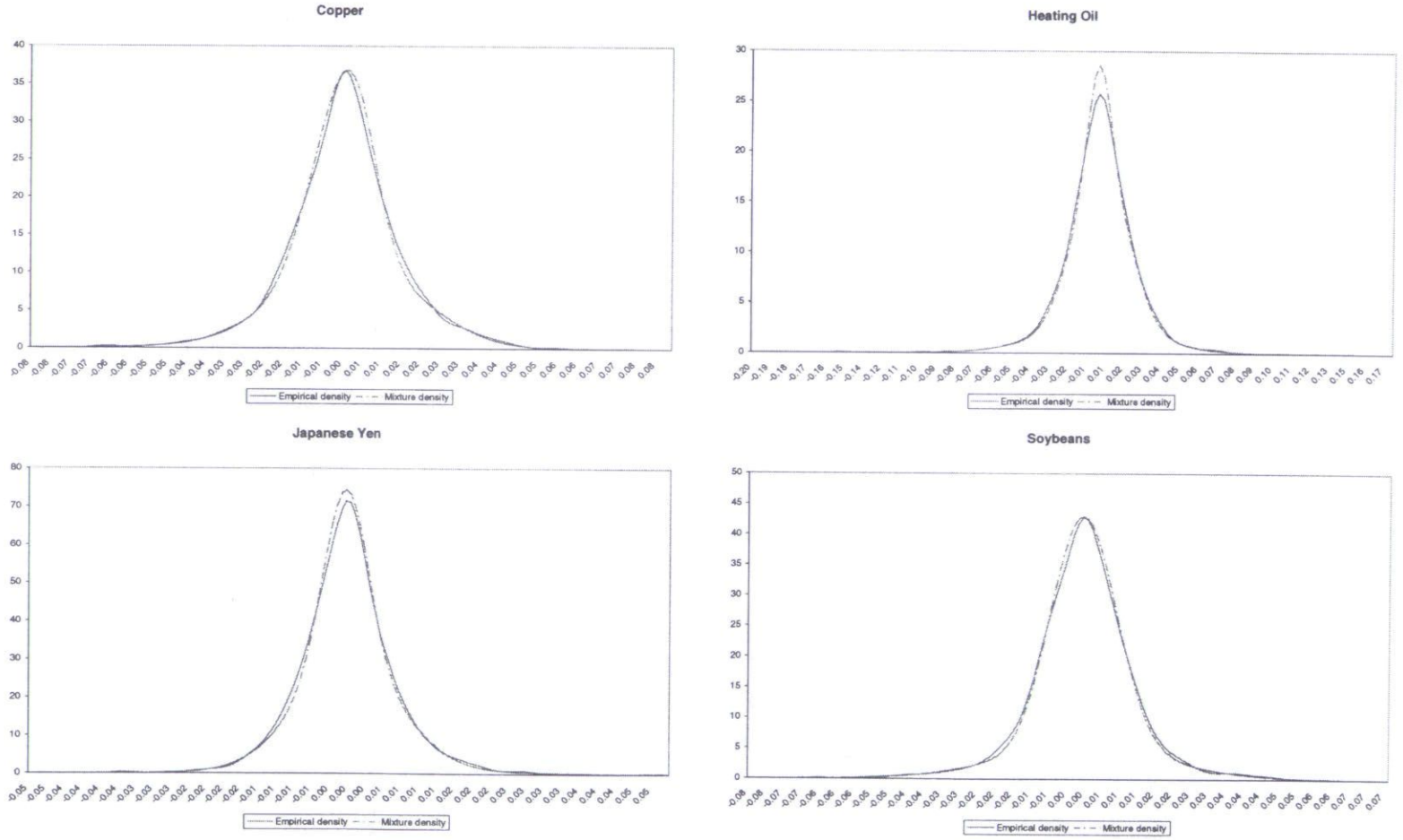

FIGURE 2

Empirical and Gaussian mixture densities: The density distributions of the futures returns have been reconstructed through the kernel method (solid lines), and the Gaussian mixture densities have been estimated by the kurtosis-controlled EM algorithm (dotted lines) for four different futures contracts: soybeans ($q = 4$), Japanese yen ($q = 3$), heating oil ($q = 6$), and copper ($q = 5$). These density distributions are representative of the goodness of fit obtained within our sample of futures contracts, regardless of the number of components.

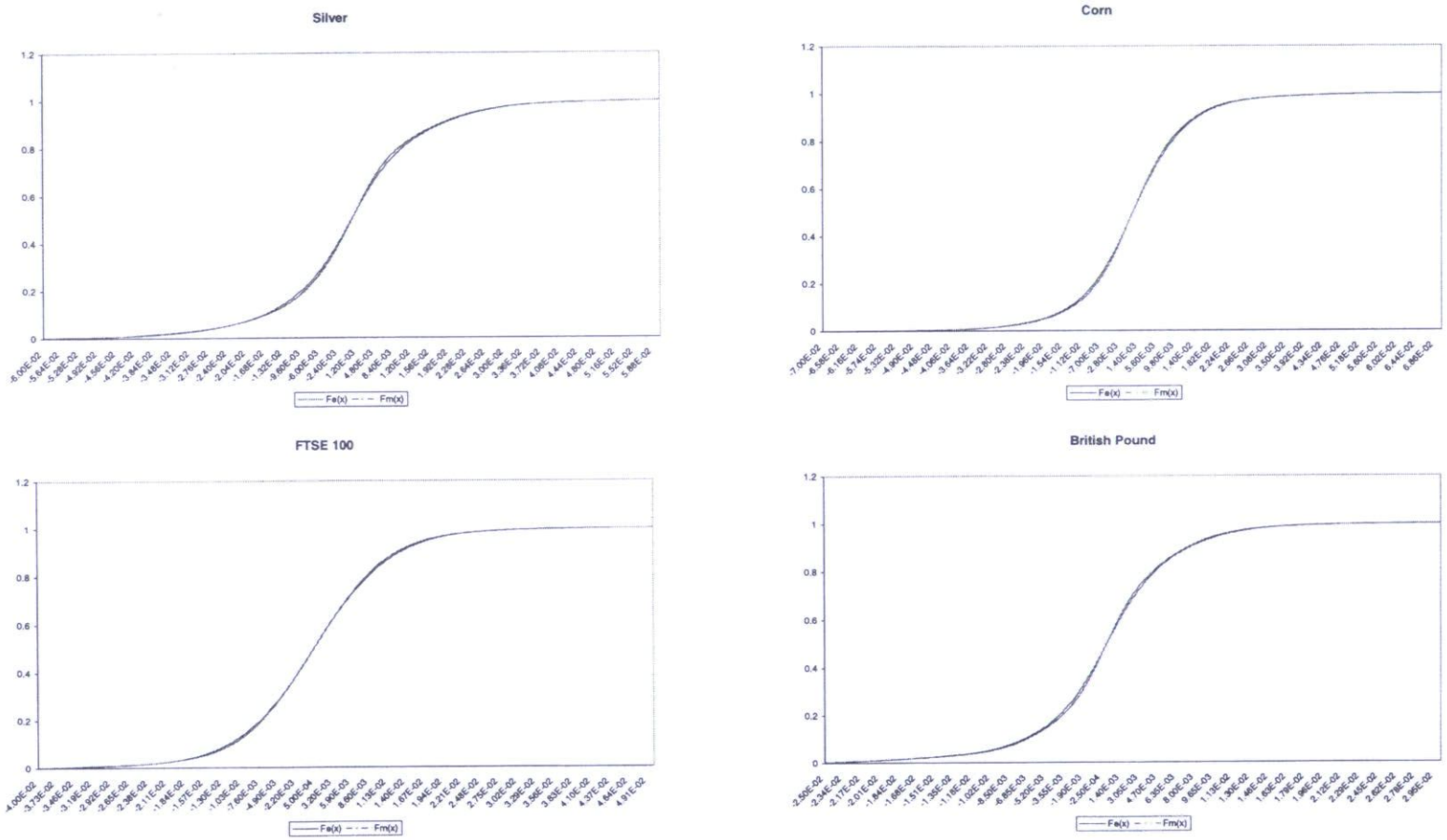

FIGURE 3

Empirical and Gaussian mixture distribution functions: The empirical distribution functions, $F_e(\bullet)$ (solid lines), have been reconstructed through the kernel method together with the Gaussian mixture distribution functions, $F_m(\bullet)$ (dotted lines), computed with the estimated parameters. Four futures contracts are represented: FTSE 100 ($q = 4$), corn ($q = 6$), silver ($q = 7$), and British pound ($q = 5$). The distance between the two curves is directly related to the AD statistic.

TABLE IV

Goodness-of-Fit Measures for the Estimated Densities

Futures Contract	*Log Likelihood*	$K_T^{K\text{-}EM}$	*AD*
Agricultural			
Cocoa	6738.73	0.0226	0.0278
Corn	7671.00	0.0425	0.0359
Cotton	7388.09	0.0326	0.0573
Feeder cattle	8828.25	0.1950	0.0464
Live cattle	8571.91	0.0559	0.0381
Orange juice	6484.70	0.0522	0.0438
Soybeans	7765.91	0.0399	0.0317
Sugar	11316.61	0.2020	0.2292
Wheat	7261.53	0.0154	0.0284
Energy			
Crude oil	6287.71	0.1026	0.1486
Heating oil	6436.59	0.0297	0.0458
Unleaded gasoline	6308.77	0.1016	0.0363
Currency			
Australian dollar	9674.30	0.0268	0.0385
British pound	9444.53	0.0352	0.0392
Deutsche mark	9202.88	0.0843	0.0412
Japanese yen	8973.90	0.0576	0.1072
Metals			
Copper	7271.58	0.0278	0.0345
Gold	9208.14	0.0974	0.0460
Platinum	8084.72	0.0429	0.0372
Silver	7291.29	0.0476	0.0394
Stock Index			
CAC 40 index	7592.32	0.0722	0.0395
FTSE 100 index	8068.27	0.0225	0.0291
NIKKEI 225 index	7201.89	0.0564	0.0992
S&P 500 index	8484.00	0.0765	0.0625

Note: Three different goodness-of-fit measures are given for the estimated densities. The log likelihood and the total kurtosis, K_T, are computed at each step of the kurtosis-controlled EM algorithm. The values obtained for the optimal estimated parameters are given. Having no reference value, the log likelihood is a poor indicator of the goodness of fit. On the contrary, the total kurtosis is decreasing to zero with the accuracy of the fit. In addition, our simulated results give us some insights into how close to zero the value should be to effectively capture the information content of the data set about the true distribution. The third row provides the AD statistic measuring the closeness of the estimated mixture distribution to the empirical distribution. The lower the value of the statistic is, the better the fit is.

ifications obtained by our estimation procedure and are reported in Table IV in addition to the log likelihood and total kurtosis values. The maximized values of the likelihood constitute poor indicators of the goodness of fit because we have no reference values. On the contrary, the total kurtosis and the AD statistic are both decreasing to zero when the estimated density comes closer to the unknown density. Moreover, the simulation investigation provides us with an indication of how close to zero

these two measures should be for a suitable estimation of the true distribution.

The theoretical total kurtosis measure obtained when various Gaussian mixture distributions were simulated ranges from 0.0178 to 0.0437. Twenty of the reported $K_T^{K\text{-}EM}$ values are about 10^{-2}, indicating a very good fit to the true distribution. The four remaining values range from 0.1016 to 0.2020. The densities estimated with the EM algorithm in Figures 1(a) ($K_T^{EM} = 0.1880$) and 1(c) ($K_T^{EM} = 0.1132$) roughly provide total kurtosis measures corresponding to the previous interval.[5] Comparing the EM estimated densities with the simulated densities represented in these graphs shows us that the degree of misfitting remains acceptable. Similarly to the total kurtosis measures, the AD statistics reported in the last column of Table IV reveal an excellent fit.

In conclusion, these tests show that the finite Gaussian mixtures together with the kurtosis-controlled EM algorithm provide a perfectly adapted framework to capture the departure from normality of the futures returns distributions. Many financial applications could be found that would undoubtedly benefit from this accurate estimation. In the following section, we give an example of the implications for margin setting in futures markets.

Application to Margin Setting in Futures Markets

Futures commission merchants as well as futures exchange clearinghouses attempt to protect themselves against losses by imposing margin requirements on their customers. In fixing required margin levels for futures, the desire for protection is weighed against the cost higher margins impose on traders. The existence of margins decreases the likelihood of customers' default. However, if the cost of maintaining sufficient margins is too high, market participants and, therefore, liquidity will be reduced. Futures markets officials are thus confronted with the difficult task of keeping margins at appropriate levels: high enough to maintain market integrity yet low enough to maintain market liquidity. All studies investigating how to set optimal margins rely on this trade-off.

In general, the margin is set so that it represents an amount that is expected to be exceeded by a price change over a specified period at an acceptable probability level. Let us denote by M and p the margin level and the probability of margin violation. The descriptive statistics pre-

[5]We would like to draw the reader's attention to the fact that the worst fits obtained when the futures return densities were estimated with the kurtosis-controlled EM algorithm correspond to the lowest levels of total kurtosis measures obtained with the EM algorithm on the simulated data.

sented in Table I indicate a significant skew for most of the futures contracts. Margin levels may thus be different for long and short positions. For a given value of the probability of margin violation p tolerated by the margin committee, we denote M^{long} and M^{short} the associated margin level for a long position and a short position, respectively. The margin level associated with a given value of the probability of margin violation p over 1 trading day for a long position is obtained with the following equation:

$$p = \text{Prob}(\Delta P < -M^{\text{long}}) = F(-M^{\text{long}}) \tag{19}$$

where F is the density function of the daily futures price changes, ΔP. Similarly, the margin level for a short position that is only exceeded with probability p over 1 trading day[6] is computed as follows:

$$p = \text{Prob}(\Delta P > M^{\text{short}}) = 1 - F(M^{\text{short}}) \tag{20}$$

Margin setting amounts to determining what is the margin level associated with a given value of the probability of margin violation over 1 trading day. For a given value of the probability of margin violation or potential default tolerated by the margin committee, the answer is given by Equations 19 and 20 for long and short positions. As the answer depends on the tails of the distribution of futures price changes, the choice of the distribution function F is crucial, and margin setting is very sensitive to the choice of a parametric distribution family.

Relying on the assumption that the stock index futures returns are generated from a normal distribution, Figlewski (1984) suggested that margin levels for futures contracts are set conservatively. Warshawsky (1989), however, showed that the normality assumption is inappropriate. Applying a nonparametric approach, he stated that margin levels are not set as conservatively as implied by the previous study. Booth, Broussard, Martikainen, and Puttonen (1997) suggested that the normal distribution may underestimate the probability of margin violation because methods that assume normality do not take into account the added risk contained in leptokurtotic data.

The problem of margin setting in futures markets is now addressed through the consideration of the base case of a margin level for a speculative account with a grace period of 1 day.[7] Two different methods for computing the margin levels are compared for a given probability of mar-

[6]Let us also note that although margin committees tend to express the margin level in terms of dollars per contract and not in terms of percentages, following Figlewski (1984) and Kofman (1993), we used the more convenient percentage price changes to compute the margin levels.

[7]Because futures traders are required to mark-to-market daily, the relevant time interval is 1 day.

gin violation.[8] The first method assumes that daily price changes are drawn from a normal distribution. The second method uses the finite Gaussian mixture framework and estimates the empirical distribution with the kurtosis-controlled EM algorithm. Optimal margin levels obtained with these two distributions are presented in Table V for long and short positions in the different futures contracts.

The remarkable result is that for conservative values of the probability of margin violation (1%), the appropriate margin levels obtained under the assumption of normality are well below those obtained with the finite mixture distributions. For example, for a long position on the crude oil futures contract, the appropriate margin level is equal to 5.65% under normality compared with 6.32% under the four-component mixture, that is, 10.67% higher. The same analysis applies to all the futures contract listed in Table V. Because the distributions of daily prices are slightly skewed, some asymmetry is found in setting margins for long and short positions when a non-Gaussian model is used.

Moreover, Gaussian mixture have also been used in the literature to infer implied risk-neutral densities from option prices and to assess how these densities are modified to incorporate major economic news (see Bahra, 1996; Söderlind, 1997). Such an analysis may provide important information for central banks. Nevertheless, following the ad hoc use of the two Gaussian mixture model in the literature, Bahra and Söderlind based their studies on a two-component mixture. Our empirical results prove that this is too poor a model to capture the distributional properties of returns.[9]

CONCLUDING REMARKS

In conclusion, our simulation study shows that the dynamic allocation of kernels according to the value of the total kurtosis measure makes the proposed kurtosis-controlled EM algorithm an efficient method for Gaussian mixture density estimation. This algorithm yields considerable improvements over the classical EM algorithm.

We used the mixture of normal models to account for the observed thick-tailed distributions of futures price returns and proved that these

[8]Like many margin committees, we considered the historical distribution of futures prices over the recent past and focused on the 99% quantile of the distribution as defined by Equations 19 and 20.
[9]In this article, we only report results relative to futures markets. However, we carried out the same empirical investigation on stocks, indexes, currencies, interest rates, and commodities. Although the average number of components in the mixture may vary from one market to another, it always remains greater than two. Hence, the two-component mixture density lacks the degree of freedom to accurately reproduce the shape of the distribution of returns.

TABLE V

Margin Levels with Gaussian Mixture Densities

Futures Contract	*Margin Level for a Long Position*			*Margin Level for a Short Position*		
	Normal	*Gaussian Mixture*	*Difference in Percentage*	*Normal*	*Gaussian Mixture*	*Difference in Percentage*
Agricultural						
Cocoa	4.21%	4.31%	2.36%	4.22%	5.48%	29.94%
Com	3.15%	3.50%	11.37%	3.14%	3.85%	22.77%
Cotton	3.47%	3.50%	0.81%	3.38%	3.48%	3.08%
Feeder cattle	1.86%	2.29%	23.40%	1.85%	2.19%	17.93%
Live cattle	2.20%	2.42%	9.75%	2.19%	2.29%	4.49%
Orange juice	5.14%	6.06%	17.87%	5.09%	5.89%	15.57%
Soybeans	2.94%	3.65%	24.36%	2.93%	3.53%	20.50%
Sugar	1.06%	1.37%	28.41%	1.05%	1.21%	15.37%
Wheat	3.50%	3.62%	3.39%	3.48%	3.92%	12.85%
Energy						
Crude oil	5.65%	6.32%	11.94%	5.66%	6.52%	15.33%
Heating oil	5.46%	6.61%	21.05%	5.42%	5.55%	2.29%
Unleaded gasoline	5.25%	6.22%	18.36%	5.26%	5.86%	11.25%
Currency						
Australian dollar	1.35%	1.54%	14.38%	1.33%	1.44%	8.23%
British pound	1.52%	2.04%	33.97%	1.53%	1.74%	14.22%
Deutsche mark	1.58%	1.83%	15.60%	1.58%	1.81%	14.87%
Japanese yen	1.80%	1.90%	5.26%	1.83%	2.26%	23.40%
Metals						
Copper	3.41%	4.06%	19.10%	3.39%	3.78%	11.72%
Gold	1.79%	2.01%	12.41%	1.77%	2.01%	13.60%
Platinum	2.52%	2.96%	17.52%	2.51%	2.97%	18.71%
Silver	3.51%	4.15%	18.27%	3.52%	4.09%	16.47%
Stock index						
CAC 40 index	2.95%	3.48%	18.26%	3.01%	3.40%	12.80%
FTSE 100 index	2.38%	2.76%	15.59%	2.45%	2.79%	13.73%
NIKKEI 225 index	3.50%	3.86%	10.24%	3.43%	4.05%	17.87%
S&P 500 index	2.16%	2.53%	17.32%	2.51%	2.97%	18.71%

Note: The computation of margins is given for long and short positions in the 24 different futures contracts with a probability of margin violation of $p = 1\%$. For each position, the first row gives the margin levels obtained with a Gaussian density, whereas the second row displays the margin levels resulting from the Gaussian mixture estimated via the kurtosis-controlled EM algorithm. The last row of each subtable (i.e., position) indicates the percentage increase in margin level when one moves from the normal distribution to the finite mixture framework.

models represent a perfectly adapted framework to capture the departure from normality of the return distributions. Unlike in previous studies, we showed that a two-component Gaussian mixture is too poor a model to accurately capture the distributional properties of futures returns. Similar results have been obtained for stocks, indexes, currencies, interest rates, and commodities.

Gaussian mixture distributions have been used in different areas of finance, such as for incorporating the abnormality of returns in the computation of the value at risk (see Hull & White, 1998) and inferring from option prices how the market reacts to economic news (see Bahra, 1996; Söderlind, 1997). Until now, these investigations were carried out arbitrarily with two-component Gaussian mixtures. Our findings have important implications in this context. Contradicting the common use of two kernels, they suggest that the results and conclusions of these studies should be updated.

BIBLIOGRAPHY

Anderson, T. W., & Darling, D. A. (1954). A test of goodness of fit. Journal of the American Statistical Association, 49, 765–769.

Aïtkin, M., & Wilson, G. T. (1980). Mixture models, outliers and the EM algorithm. Technometrics, 22, 325–331.

Bahra, B. (1996, August). Probability distribution of future asset prices implied by option prices. Bank of England Quarterly Bulletin, 36(3), 299–311.

Booth, G., Broussard, J. P., Martikainen, T., & Puttonen, V. (1997). Prudent margin levels in the Finnish stock index futures market. Management Science, 43, 1177–1187.

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

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

Figlewski, S. (1984). Margins and market integrity: Margin setting for stock index futures and options. Journal of Futures Markets, 4, 385–416.

Goldfeld, S. M., & Quandt, R. E. (1972). Nonlinear methods in econometrics. Amsterdam: North-Holland.

Hull, J., & White, A. (1998). Value-at-risk when daily market variables are not normally distributed. Journal of Derivatives, 5(3), 9–19.

Kofman, P. (1993). Optimizing futures margins with distribution tails. Advances in Futures and Options Research, 6, 263–278.

Kon, S. J. (1984). Models of stock returns—A comparison. Journal of Finance, 39, 147–165.

Lekkos, I. (1999). Distributional properties of spot and forward interest rates: USD, DEM, GBP and JPY. Journal of Fixed Income, (March) 35–54.

McLachlan, G. J.; Basford, K. E. (1988). Mixture models: Inference and applications to clustering. New York: Marcel Dekker.

Robertson, C. A., & Fryer, J. G. (1969). Some descriptive properties of normal mixtures. Skandinavian Aktuarietidskrift, 52, 137–146.

Silverman, B. W. (1986). Density estimation for statistics and data analysis. London: Chapman & Hall.

Söderlind, P. (2000). Market expectations in the UK before and after the ERM crisis. Economica, 67, 1–18.

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). Statistical analysis of finite mixture distributions. New York: Wiley.

Vlassis, N., & Likas, A. (1999). A kurtosis-based dynamic approach to Gaussian mixture modeling. IEEE Transactions on Systems, Man and Cybernetics, 29, 393–399.

Warshawsky, M. J. (1989). The adequacy and consistency of margin requirements: The cash, futures and options segments of the equity markets. Review of Futures Markets, 8, 420–437.