**Fixed Income**
**Quantitative Credit Research**

May 2006

# Tail behavior of systematic risk factors in the Lehman Brothers Global Risk Model

(Extracted from the QCR Quarterly, vol. 2006-Q1/Q2)

Adam Purzitsky

LEHMAN BROTHERS

# Tail behavior of systematic risk factors in the Lehman Brothers Global Risk Model[1]

**Adam Purzitsky**
+44 020 7102 9023
apurzits@lehman.com

*The Lehman Brothers Global Risk Model is a tool for quantifying and managing the risk of global fixed income portfolios. As part of its continuing evolution, we are extending the model beyond the calculation of tracking error volatility to include other summary statistics for risk, particularly estimates of tail risk, such as Value-at-Risk. In the present study we introduce empirical evidence on the distributions of systematic risk factors that illustrates the need for such statistics. We then conduct an analysis of some possible models for describing the factor distributions. The results of the study are that the Normal distribution fails to accurately describe the tails for a large number of systematic risk factors but this failing can be rectified by more general distributions, such as a mixture of Normals or the Student's t distribution. Moreover, we show that in a heavy tailed world, tracking error volatility is not sufficient as a summary statistic for risk. Thus risk measures that describe portfolio exposures to tail events are indispensable for effective risk management and allocation.*

## INTRODUCTION

The portfolios analyzed by the Lehman Brothers Global Risk Model comprise a broad market on the order of several tens of thousands of securities. Extensive empirical study has allowed the Global Risk Model to capture the bulk of the randomness in this large market with a few hundred common risk factors. In this paper we study the distributions of these systematic risk factors and find that a significant number of the most important ones have distributions that are poorly described by the Normal distribution and, in particular, that the distributions tend to have tails that can be much heavier than those implied by the Normal distribution. We then explore alternatives to the Normal distribution that remedy its defects while remaining parsimonious and tractable.

The rest of the paper proceeds as follows. First we examine a few factors relative to the Normal distribution to highlight the inadequacy of the Normal assumption in matching the empirical distribution of these factors. We pay particular attention to the failure of the Normal assumption for the purposes of tail risk calculations. Along the way we try to illustrate the interaction between modeling and estimation, showing how an improved distributional specification can lead to more robust estimates of risk parameters.

We then introduce the mixture of Normals model as a generalization of the Normal distribution that can rectify its failings and allow transparent calculations to illustrate the basic ideas involved. Finally, the t distribution is introduced as a generalization of the mixture model. We then conduct a comprehensive study covering virtually all of the factors. This demonstrates that both the mixture of Normals and the t distribution provide a statistically good fit to the data and, in particular, are superior to the Normally distributed model for a large number of factors.

## ANALYSIS OF THE EMPIRICAL DISTRIBUTION OF SYSTEMATIC FACTORS

In this section we examine two basic empirical facts that are pervasive across asset markets and thus affect a large number risk factors.

1. The volatility of factors changes through time.

[1] *I would like to thank Jeremy Rosten, Vasant Naik and Anthony Lazanas for comments and suggestions.*

2. Even taking as given the best estimate of current volatility, the factors are *not* Normally distributed and in fact exhibit much heavier tails than the Normal.

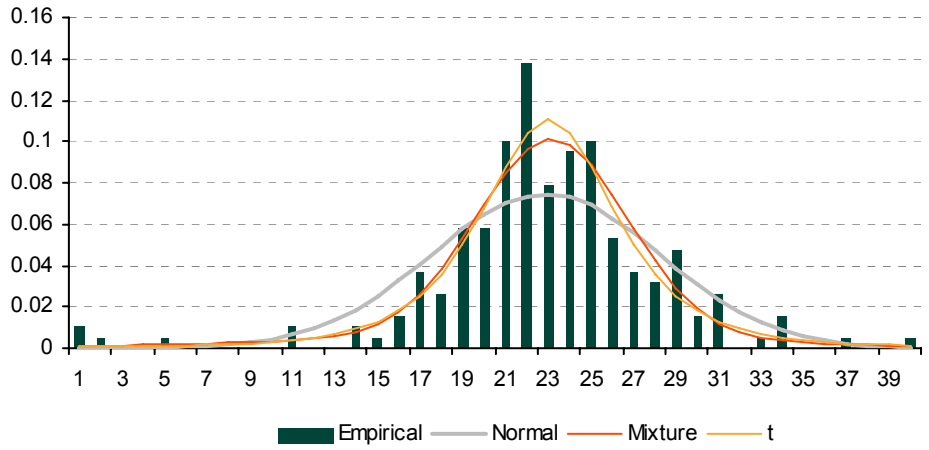
The first point is already well understood in the current risk model methodology by the use of a weighted covariance matrix in which variances and covariances are estimated from data that give geometrically declining weight to the data the further it is in the past. However, we mention changing volatility here because we will show that changing volatility alone is not sufficient to account for the non-normality of the data.

## Understanding the normal assumption

One very important issue that needs to be discussed before we confront the Normal distribution with the evidence in the data is that of estimation error. One possible criticism of point 2 might be that the problem is simply that the estimate of current volatility is a poor one and that the data is Normally distributed given the "true" value of the current volatility. We will argue that this is not the case; in any event, we can rebut this with the basic observation that since "true" volatility is unobservable we will always need to use some estimate of it in any practical implementation. Thus, modeling the factor distribution given the best *estimate* of current volatility is really the only interesting problem and that is the one we are trying to address. Nonetheless, in the following sections we will present evidence showing that, in fact, estimation error is not the root cause of the apparent non-normality and that the fat tails are a genuine feature of the data.

## The Normal distribution vs the data

**Figure 1. USD 5Yr swap spread factor**




*Source: Lehman Brothers*
*Histogram and fitted Normal probability density function for (the negative of) monthly changes in the USD 5 year swap spread from January 1990 through September 2005.*

Figure 1 plots a histogram of the 5 year USD swap spread factor[2] along with the Normal distribution evaluated with the sample mean and variance of the factor. The dataset has 189 monthly observations on the changes in the spread. The standard deviation of the sample is 5.7 basis points. This example shows very clearly the inadequacy of the Normal distribution for making tail risk calculations. A 3 standard deviation event in this sample is any move greater than 17.1 basis points (bp), according to the Normal distribution the probability of a 3 standard deviation event is about 2.7 in a thousand (.0027) and yet in our sample we see 5

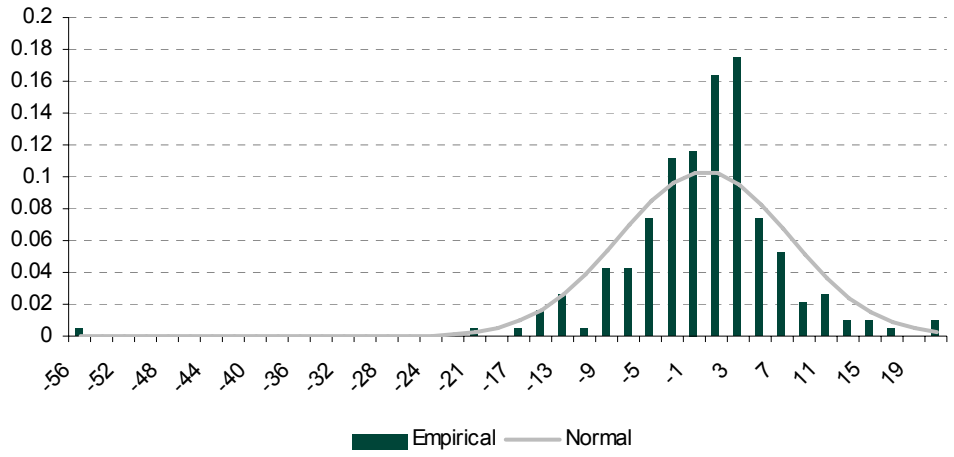
[2] *The factor is the negative of the monthly changes in the USD 5yr swap spread. The negative is so the factor loading (which is the spread duration) can be expressed as a positive number.*

such events for an empirical frequency of 5/185 = .027. This is 10 times too many tail events! Inspection of Figure 1 shows another feature of the data relative to the Normal: very small moves are also too common. For a distribution that is roughly symmetric this of course must be the case since the extra occurrences of large observations must be balanced by extra small ones ("extra" meaning relative to the Normal) in order to keep the standard deviation fixed. The point is simply that in a non-normal world the entire interpretation of standard deviation is different.

As mentioned above, one possible explanation for the factor distributions displaying heavier tails and a sharper peak in a histogram is simply time varying volatility. The idea is that some of the observations were collected during a time when the volatility was higher than its average for the whole sample and so we find more large moves than would be implied by the full sample volatility estimate. Other observations were collected when volatility was smaller than its full sample average and so we appear to get too many small moves as well.

To establish that, even conditional on the current volatility estimate, factor changes display heavier tails than the Normal distribution, we need to examine the factor realizations and volatilities dynamically. For our second example, we take a look at the USD Energy A credit factor, which is essentially[3] the negative[4] of the average spread change on USD denominated A-rated energy sector bonds. Its histogram and fitted Normal density is plotted in Figure 2 where one very large tail event stands out, this observation is September 2001. Clearly the Normal density is assigning virtually no probability to an event of that magnitude.

**Figure 2. USD energy A credit spread factor**

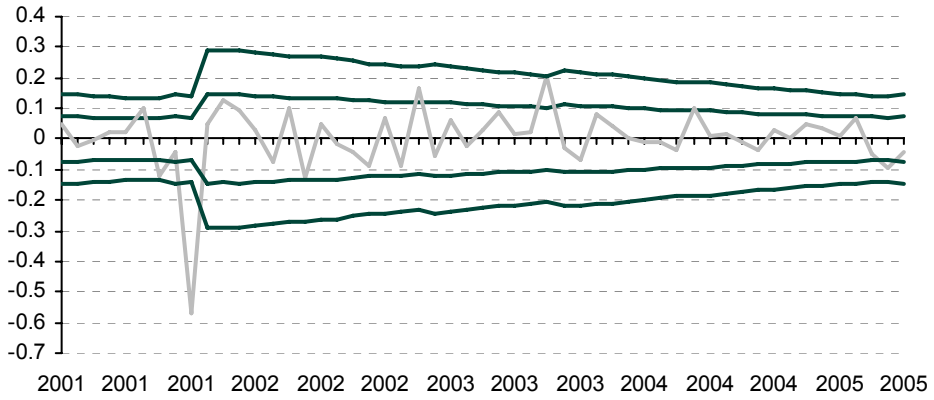



*Source: Lehman Brothers*
*Histogram and fitted Normal PDF for the USD energy A credit spread risk factor. The data sample is monthly from January 1990 through September 2005.*

In Figure 3 we plot the time series of factor realizations along with one and two standard deviation bands from the beginning of 2001 until the present. The volatility estimates at each point in time are the weighted estimates using data prior to that time point, the weighting makes the estimate determined only by the most recent data.

[3] *It is actually the average spread change after the influence of some other factors has been controlled for.*
[4] *Again, the factor is defined as the negative of the monthly spread change so that the factor loading (which is the spread duration) can be expressed as a positive number.*

**Figure 3. Time series of USD energy a credit spread factor**

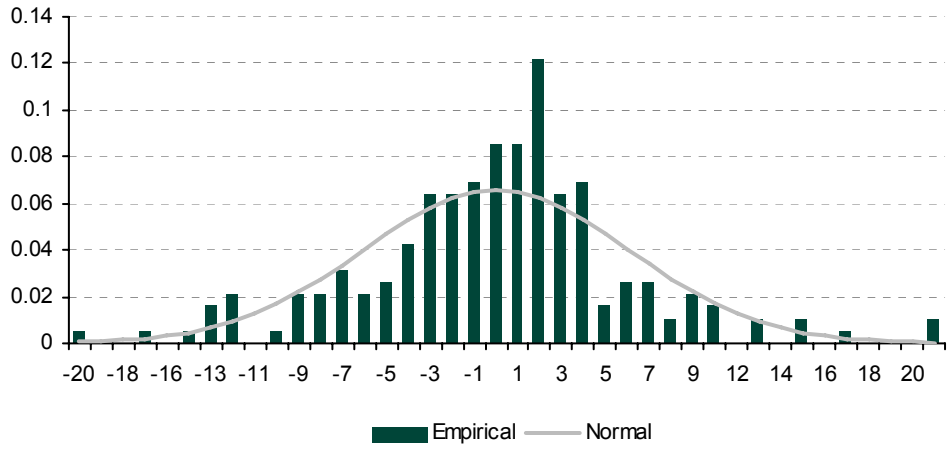



*Source: Lehman Brothers*
*This figure plots the monthly realizations of the USD energy A credit spread factor (the gray line) along with the one and two standard deviation bands (green lines) estimated according to the weighted estimation scheme. That is, at each point the standard deviation is estimated from all data prior to that point but with weights that decline geometrically with a one year half life.*

At the beginning of September 2001 the volatility estimate was about 7bp. During the month the factor exhibits a 57bp move – an eight standard deviation event.! *The probability of an eight standard deviation event under the Normal distribution is about* $10^{-15}$. Even once the volatility estimate is updated to include September, it is still only about 14.6bp. Thus, the 57bp move would still be about a four standard deviation event. This has a probability in the Normal distribution of about 3 in a 100,000. Moreover, in the following 43 months we would expect, under Normality, to see about 14 events of one standard deviation or more, but in the data we have only four, even though the estimated standard deviation gradually declines as the large event receives progressively less weight in the estimation. This shows quite clearly that simple changing volatility is insufficient to explain the data. Even at the highest volatility estimate we could possibly get from this data, under the Normal distribution the tail event is still exceedingly unlikely. Furthermore, the time series data show that the tail event has caused the volatility estimates to overestimate subsequent volatility and that even with an estimate that puts geometrically declining weight on the large event as it goes farther into the past this overestimation persists for a period of several *years*.

Before we introduce some models that can provide us with an acceptable fit to the data, we pause to make clear that for our workhorse example (the USD Energy A credit spread factor), although the September 2001 observation was a particularly sharp illustration of the issues involved, the non-Normality of the factor is not driven entirely by this one event. We display the histogram with fitted Normal density, omitting the September 2001 observation in Figure 4. The tails still appear fatter than those implied by the Normal, moreover, while the factor's sample kurtosis including September 2001 is a whopping 30, when we omit the Sept. 2001 observation we still get a decidedly non-Normal kurtosis of 4.3 (the Normal has a kurtosis of 3). Moreover, in a subsequent section we conduct a formal statistical test of the Normal assumption along with the models introduced in the sequel. The test is conducted both with and without the Sept. 2001 observation with similar results. In both cases, for this factor, we reject the Normal distribution.

**Figure 4. USD energy A credit spread factor omitting Sep. 01**




*Source: Lehman Brothers*
*This figure plots the histogram and fitted Normal PDF for the USD energy A credit spread factor from monthly data covering January 1990 through September 2005, but with the September 2001 observation omitted.*

To sum up, the Normal distribution fails rather spectacularly to describe the distribution of many systematic risk factors. Moreover, this means that volatility is not quite enough as a summary statistic for risk exposures. In a non-Normal world knowing the value of volatility gives precious little information on the risk associated with tail events.

## POSSIBLE MODELS FOR THE DISTRIBUTION OF THE SYSTEMATIC FACTORS

### The mixture of Normals

The preceding section makes clear that the Normal distribution is grossly inadequate as a model for tail risk calculations. Even using a standard deviation estimate that includes the event (an impossibility in reality) the event would still be given nearly zero probability in the Normal model. In this section we present a first pass at modeling the fat tails in the data. Motivated by the examples of the previous section we postulate a model in which there is a random mixture of two Normal distributions, one that describes the distribution of the factors in the "usual" case and another that describes the distribution of the factors in the "extreme" case along with an estimate of the probability of an "extreme event". This mixture of Normals model is formally specified as follows:

- In any month with probability $\lambda$ the factor distribution is $N(0,\sigma^2)$.
- With probability $1-\lambda$ the factor distribution is $N(0, \sigma^2 + \gamma^2)$.

The idea is that the distribution with volatility $\sigma$ is realized every month. Thus, in the absence of a large event the factor distribution is Normal with standard deviation $\sigma$. However, each month there is probability $1-\lambda$ that additionally a "large event" will occur. A large event is a one time shock with volatility $\gamma$, thus in any month with a large event the factor change is actually the sum of two Normals with standard deviations $\sigma$ and $\gamma$. Since the sum of two Normals is again Normal we conclude that in any month where a large event occurs the factor distribution is Normal with volatility $\sqrt{\sigma^2 + \gamma^2}$ .

#### *A Tail Probabilities Example*

Maximum likelihood parameter estimates for the mixture and Normal models fit to our example from the last section, the USD Energy A credit factor, are in Figure 5.

**Figure 5. Parameter estimates for USD energy A credit spread factor**

| **Normal Model** | | **Mixture Model** | |
|---|---|---|---|
| Mean | -2bp | Mean | -.29bp |
| Volatility | 7.6bp | Usual Volatility | 5.6bp |
| | | Jump Probability | 6% |
| | | Jump Volatility | 21bp |
| | | Total Volatility | 7.6bp |
| Prob. of 3 std event 0.27% | | Prob. of 3 std event | 1.76% |

*Source: Lehman Brothers*
*This table reports parameter estimates for the Normal and mixture of Normals model for the USD energy A credit spread factor. Parameters are estimated by maximum likelihood using monthly data from January 1990 through September 2005.*

The estimates imply that the probability of a large event in each month is about 6%. This means that in about 19 out of every 20 months we expect the volatility to be 5.6 basis points but in about 1 in 20 months we expect it to be 21.7 basis points ( $\sqrt{\sigma^2 + \gamma^2}$ ). Thus, although the true volatility is 7.6 basis points a month we realize a volatility of only 5.6 basis points most of the time.

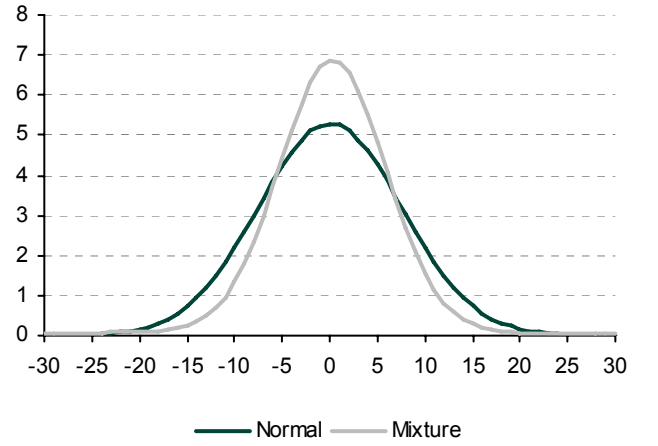
Moreover, as noted before, the meaning of volatility is different from the Normal world. The total volatility is 7.6 basis points per month so under Normality the probability of a move of 22.8 basis points (a 3 standard deviation event) is about 2 in a thousand. However in the mixture model we have a 6% probability that our factor is the realization of a Normal with variance 21.7 basis points. Given that we realize a large event, the probability of a move of at least 22.8 basis points is about .29 (it's close to a 1 standard deviation event). Thus, since the large event has a probability of .06, the total probability of a move greater than 22.87 basis points is about .06*(.29) = .017, *it's 6.5 times bigger*. Thus, in 189 observations (the size of our sample) the Normal model would be very surprised to see a move of 22 basis points, whereas the mixture model fully expects to see at least one observation of that magnitude.

The estimates in Figure 5 also illustrate the differences in the estimation that accounting for the fat tails can have. The mean of the distribution estimated from the Normal assumption (which is just the sample mean) is about 10 times as big as the mean estimated from the mixture model even though the model is estimated with the mean of the "large event" distribution set to zero! We discuss in more detail the interaction between properly modeling the tails of the distribution and robust estimation of parameters in the next section on volatility estimation.

In Figure 6 we plot the Normal distribution and the mixture distribution for our example. The right panel displays a comparison of the tails from about 3 the standard deviation mark outwards. It is easy to see that the mixture displays both a higher frequency of small moves and much heavier tails.

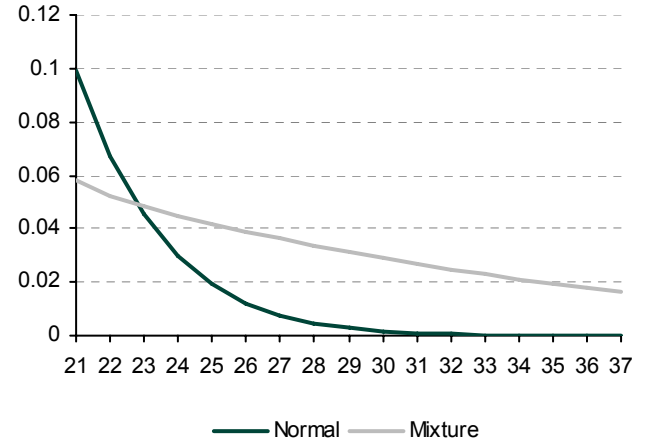
**Figure 6. Mixture and normal distributions for USD Energy A credit spread factor**

**Mixture and normal distributions for USD Energy A credit spread factor**




*Source: Lehman Brothers*

**Tail of mixture and normal distributions for USD Energy A credit spread factor**




*Source: Lehman Brothers*
*The left panel shows the Normal and Mixture PDFs with the parameters as reported in Figure 5. The right panel shows the right tail of the same distributions from 2.8 to 5 standard deviations.*

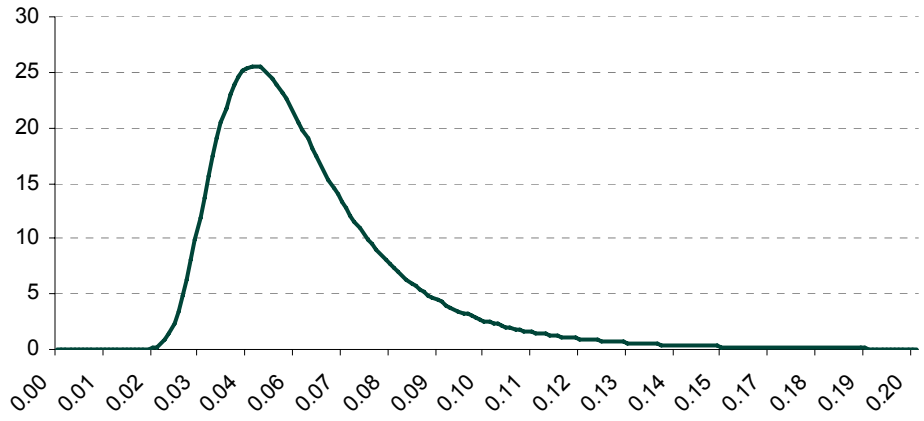
## The Student's t distribution

It turns out that a continuous mixture of Normal distributions, with the right weighting of the mixing distribution, yields a Student t distribution. In this sense, the t distribution can be understood as a direct generalization of the mixture model.

To be specific, a more formal representation of the mixture model from the previous section is as the product of two random variables. Let $X$ be a standard Normal random variable and let $B(p)$ be a Binomially distributed random variable (that is, $B$ is 1 with probability $p$ and 0 with probability 1-$p$). Then the mixture model can be written as $\left((1-B)\sigma + B\sqrt{\sigma^2+\gamma^2}\right)X$. When $B$ takes the value 0 we are left with $\sigma X$, which has a Normal distribution with volatility $\sigma$. When $B$ takes the value 1 we get a Normally distributed variable with volatility $\sqrt{\sigma^2+\gamma^2}$. But of course, letting $Z = \left((1-B)\sigma + B\sqrt{\sigma^2+\gamma^2}\right)$ then we obtain a compact representation of the mixture variable as $Y=Z\cdot X$.

Recall that if we take a random variable $X$ to be distributed as a standard Normal and a second variable $Y$ to have a Chi-squared distribution with $\nu$ degrees of freedom then a Student t distributed random variable is obtained as $\frac{\sigma\sqrt{\nu}}{\sqrt{Y}}X$. Letting $Z = \frac{\sigma\sqrt{\nu}}{\sqrt{Y}}$ we can then represent a t distributed random variable by the product $Z\cdot X$. Thus, the t distribution is also a mixture of Normals where each observation is normally distributed with volatility randomly chosen by the continuously distributed random variable $\frac{\sigma\cdot\sqrt{\nu}}{\sqrt{Y}}$. In Figure 7 we plot the distribution of the random variable $\frac{\sigma\cdot\sqrt{\nu}}{\sqrt{Y}}$ for $\sigma=.05$ and $\nu=5$.

**Figure 7. Density of mixing distribution in Student's t model**




*Source: Lehman Brothers*
*This figure plots the PDF of the mixing distribution that plays the role of the weighting function when the Student's t distribution in represented as a continuous mixture of Normals.*

For these parameter values, the mean of the distribution is .0595 and its mode (the point at which the PDF attains its maximum value) is .0456. It will always be the case that σ is between the mean and mode of the distribution of $\frac{\sigma \cdot \sqrt{\nu}}{\sqrt{Y}}$ and thus we can interpret σ as measuring something very similar to the "usual" volatility in the mixture model.

The t distribution is specified by only three parameters: its mean[5] (or location parameter), the dispersion parameter σ and its degrees-of-freedom parameter. It generally displays fat tails relative to the Normal just as the mixture distribution does. In fact, the tails of the t distribution decay even more slowly than those of the mixture. Moreover, like the mixture distribution the t also nests the Normal as a limiting case. Indeed, the Normal distribution is a t distribution with an infinite degrees-of-freedom.

One somewhat subtle advantage of the t distribution over the mixture is that it has only three parameters whereas the mixture distribution has five. Having fewer parameters is an advantage in reducing the estimation error in the parameters. To see this point consider how the thickness of the tails is controlled in each of the two models. In the t distribution, the thickness of the tails is controlled by the degrees of freedom parameter, whereas in the mixture model it is controlled jointly by the volatility of large events and the probability of large events. Now, suppose we estimate the parameters from a dataset. In the t distribution the degrees of freedom controls not only the thickness of the tails but also, jointly with the dispersion parameter, the height of the peak of the distribution. Thus every observation contains some information on the degrees of freedom parameter. On the other hand, in the mixture distribution the tails are controlled jointly by the large event probability and the "large event" variance, the peak of the distribution is controlled jointly by the large event probability and the "usual" volatility. Thus, while every observation contains information on the large event probability and so we expect it to be reasonably well estimated, we have very little information on the large event variance and thus would expect it to be estimated relatively poorly.

In Figure 9, we display both the mixture and the t distributions estimated for our canonical example from the previous sections, the USD Energy A credit spread factor. The right tail is also displayed. As the figure shows, both models give about the same probability weight to a shock of 57bp (the largest observed move in our sample) but the tail of the t distribution is

[5] *In the case that the degrees of freedom is less than or equal to 1, the mean does not exist and this parameter is interpreted simply as a location parameter.*

decaying much more slowly and so gives more probability to the possibility of a shock of, say, 90bp. Of course, this implies that the t must give correspondingly more weight to small moves as well. Figure 8 presents parameter estimates for the t distribution on this factor. Notice that the probability of a three standard deviation event is actually slightly higher in the mixture model, despite the fact that far in the tail the t density is higher. The reason for this is that the t density does not become larger than the mixture density until about six standard deviations into the tail and the total mass in the tail outside of six standard deviations is much smaller than the mass from 3-6 standard deviations. . A similar phenomenon occurs with value at risk (VaR) statistics. Consider a hypothetical bond with a 5 year spread duration. The Normal distributon implies a 95% VaR for such a bond of 63 basis points which is higher than the VaR that either of our heavy tailed alternatives yields. On the other hand the 99% VaR for the heavy tailed distributions is larger than the 99% VaR from the Normal distribution (see Figure 8). This is simply because the Normal density doesn't cross below the heavy tailed densities until farther into the tail than the 95% VaR point. Figure 8 also reports the 95% and 99% expected shortfalls (ES) for both models for the same hypothetical bond with a 5 year spread duration.

**Figure 8. Parameter estimates for USD Energy A credit spread factor**

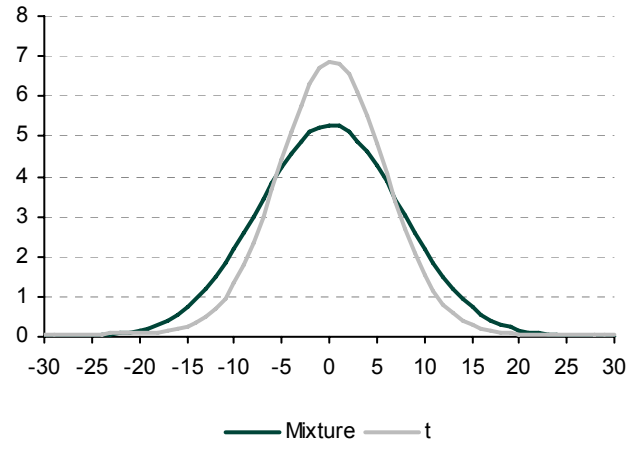
| **Student's t Model** | | **Mixture Model** | |
|---|---|---|---|
| Mean | .23bp | Mean | -.029bp |
| Dispersion | 4.4bp | Usual Volatility | 5.6bp |
| Degrees of Freedom | 2.87 | Jump Probability | 6% |
| | | Jump Volatility | 21bp |
| | | Total Volatility | 7.6bp |
| Prob. of 3 std event | 1.4% | Prob. of 3 std event | 1.76% |
| 95% VaR | 53 bp | 95% VaR | 52 bp |
| 99% VaR | 104 bp | 99% VaR | 106 bp |
| 95% ES | 88 bp | 95% ES | 85 bp |
| 99% ES | 163 bp | 99% ES | 163 bp |

*For Normal distribution with volatility of 7.6 bp the 95% ES is 78 bp and the 99% ES is 101 bp. The 95% VaR is 63 bp and the 99% VaR is 88 bp.

*Source: Lehman Brothers*
*This table reports parameter estimates for both the Student's t distribution and the mixture of Normals distribution fit to the USD energy A credit spread factor. The parameters are estimated by maximum likelihood. The sample period for the estimates is January 1990 through September 2005. The expected shortfall (ES) numbers assume a bond with a 5 year spread duration.*

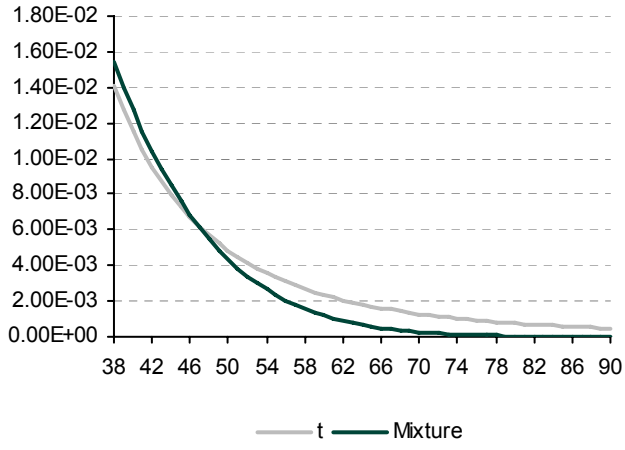
**Figure 9. Densities of mixture and t distributions for USD Energy A credit spread factor**

**Densities of mixture and t distributions for USD Energy A credit spread factor**




*Source: Lehman Brothers*

**Tail of densities of mixture and t distributions for USD Energy A credit spread factor**




*Source: Lehman Brothers*
*The left panel is the probability density function of both the mixture and t distributions with estimated parameters fit to the USD energy A credit spread factor. The right panel is the right tail of the density functions from about 5 to 12 standard deviations. The sample period for the estimation is January 1990 through September 2005.*

## Time varying volatility with fat-tailed conditional distributions

Having introduced fat tailed distributions as mixtures of Normals with stochastically chosen volatilities, it might be asked if this really differs from the time varying volatility explanation of the fat tails in the factor distributions. As a practical matter, the major difference is in the perceived *persistence* of the volatility. We will see below that in the mixture model the usual volatility is itself time varying but shocks to it are highly persistent. On the other hand, as we have seen in the example in Figure 3, large events can be quite isolated. A more subtle distinction is found in *when* the two sources of randomness resolve themselves. In a stochastic volatility model the volatility of the time t factor realization is understood to be *determined* (though not necessarily observed) at time t-1. In the mixture model and t distributed model the volatility and the realization of the Normal variable are chosen *simultaneously*. Thus we refer to the conditional distribution as the distribution of the factor *conditional* on the current value of volatility and we model the conditional distribution as t or mixture. In the following sections we make the distinction concrete as we formally model stochastically varying volatility in our heavy-tailed models.

### *Modeling stochastic volatility in the Mixture distribution*

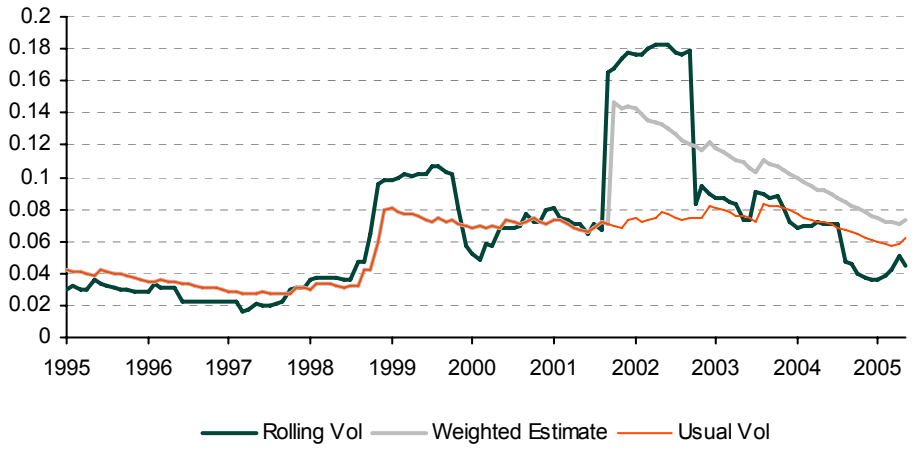
In the mixture model time varying volatility is understood to mean time variation in the sigma parameter. Thus we obtain a model specified as follows:

- In any month with probability $\lambda$ the factor distribution is $N(0,\sigma_t^2)$.
- With probability $1-\lambda$ the factor distribution is $N(0,\sigma_t^2 + \gamma^2)$.

Note that this is identical to the original specification except for the time subscript on $\sigma$. We assume that the variance of the large event distribution is constant.

To illustrate the distinction in Figure 10 we plot the time series of realized 12 month volatilities for our example USD Energy A credit spread factor along with the weighted estimate and the usual volatility from the mixture model.

**Figure 10. Time series of volatility estimates for USD Energy A credit spread factor**




*Source: Lehman Brothers*
*Rolling Vol estimates volatility with the most recent 12 months of data equally weighted. The 'Weighted Estimate' estimates standard deviation with all data prior to that point but with geometrically declining weights with a one-year half life. The 'Usual Vol' is the estimate from the mixture model where parameters are estimated using the two-stage procedure described in the text.*

In this plot, the usual volatility is estimated in a weighted manner with the jump parameters held fixed at their full sample values. Clearly, for most of the sample the usual volatility estimate tracks well the realized volatility. Moreover, what we mean by time varying volatility can be seen from the rolling volatility estimates. From 1995 until 1998, the realized volatility is almost always between 2 and 4bp per month. However, after the Russian/LTCM crisis, the realized volatility shifts upwards and stays around 6-8bp/month (except after Sept. 2001) until the beginning of 2005 when it again drops. In September 2001 the rolling volatility estimate jumps from about 7 basis points to 17 from this one event, stays around 17-18bp until the extreme event drops out of the estimation window and then falls back. However, we saw in Figure 3 that this is not at all an accurate description of actual volatility dynamics. The realizations subsequent to the shock of September 2001 are not that much more volatile than the ones leading up to it. Thus, it is clear that Sept. 2001 was a one time event and not a shift in volatility per se. To rationalize the data in the simple stochastic volatility model we would need to have volatility jump by a large amount *twice*. We would need a large increase in volatility to rationalize the large event observation followed immediately by a large decrease in volatility to rationalize the subsequent data. Moreover, the second (downward) change in volatility is highly persistent, lasting for years, while the upward jump lasted only a month. It would require a very complicated model of volatility dynamics to generate this sort of behavior in the basic stochastic volatility model. On the other hand, the mixture model with a persistent time varying usual volatility generates this type of dynamics by construction.

Furthermore, the estimation problem caused by the heavy tails is also plainly evident in Figure 10. Let us look at the weighted volatility estimate. It also jumps upwards after September 2001 and then, as the weight assigned to the large shock, subsides the weighted estimate gradually declines. Yet in the middle of 2005, nearly four years after the event, the weighted estimate is still overestimating volatility. The reason for this overestimation is that the model does not expect to see any events of that magnitude and thus, after the event, has no choice but to conclude that volatility must actually be much higher. In contrast the

estimate of "usual" volatility from the mixture model doesn't change at all when the shock hits. This is because the mixture model fully expects to see the odd large event and so sees no need to change its estimates.

*Modeling stochastic volatility with the t distribution*

Consider now a factor that we model with the t distribution, with some degrees of freedom $\nu$ and a dispersion parameter $\sigma$. The standard deviation, or volatility, of the factor is calculated from the parameters as $\sigma \cdot \text{sqrt}(\nu/\nu\text{-}2)$. In the current risk model estimation methodology we use sample standard deviation calculated from data that overweight recent data to match the time variation in volatility. However, specifying the factors as t-distributed and with tail risk calculations in we would prefer to estimate the degrees of freedom parameter with all the data receiving full weight. The reason is that since the degrees of freedom control the tails of the distribution and we already do not get very many observations there we want to use all the data we have to estimate the degrees of freedom. After all, it is fairly common to go more than five years between extreme events, and if the degrees of freedom are estimated with weighted data then the large observations will have virtually no weight in the estimation. Our estimates would imply that we do not think an extreme event is possible in the future just because we haven't seen one in a while!

Furthermore, there is a correspondence between the dispersion parameter in the t distribution and the standard deviation in the normal. In both cases one starts with a standard distribution, either standard normal or standard t, and then defines the parameters $\mu$ and $\sigma$ to be such that $(x\text{-}\mu)/\sigma$ has that standard distribution. It appears that in both cases $\sigma$ is acting in the same role as a dispersion parameter and it's just a coincidence that under normality $\sigma$ also happens to equal the standard deviation.

Thus, with the t distribution we propose to model time variation in volatility by letting $\sigma$ vary, thus at time t volatility will by $\sigma_t \cdot \text{sqrt}(\nu/\nu\text{-}2)$. We estimate the degrees of freedom with all available data given full weight. We then re-estimate the dispersion parameter $\sigma_t$ using weighted data and keeping the degrees of freedom fixed. Since the degrees of freedom parameter is fixed in the second step the estimate of $\sigma_t$ will track variation in volatility even though it is not numerically equal to volatility.

Above it was shown how the mixture model also leads to more robust volatility estimates as the model didn't need to radically increase its volatility in order to digest a large event. Of course since the t distribution also has fat tails relative to the normal its volatility estimate will also have this robustness property as the occasional large event is "expected" and thus doesn't warrant a large re-evaluation of parameter values.

## HEAVY TAILS OF SYSTEMATIC RISK FACTORS

In this section we explore in detail the tail behavior of some of the more important factors to get an overview of how the data actually look. For this purpose we use the t distributed model as the tail behavior of the t distribution is entirely characterized by a single parameter, the degrees of freedom (DoF). This allows a more concise summary of tail behavior since with the mixture model a higher arrival rate of large events combined with a smaller large event volatility makes the tail relatively thicker in some places but relatively thinner in others and this makes direct comparisons among the factors slightly more difficult.

### Yield curve factors in USD and EUR

*Monthly frequency*

We begin with by far the most important risk factors for the USD and EUR markets: the treasury yield curve factors. Figure 11 presents parameter estimates using monthly data from

the t distribution as well as sample standard deviation and kurtosis to give an another measure of the degree of tail thickness. The kurtosis estimates presented are not "excess kurtosis" and so a value of 3 corresponds to a Normally distributed variable. Since the DoF of a random variable that is actually Normally distributed is infinity we must put some arbitrary upper bound on the admissible degrees of freedom in order to get a maximum of the likelihood. We choose 60, which is significantly higher than the point at which the Normal and t become completely indistinguishable in a sample of the size we have.

**Figure 11. Yield curve factors with monthly data**

| | Sample Moments | | t-Parameters | |
|---|---|---|---|---|
| | **Std Dev** | **Kurtosis** | **Sigma** | **DoF** |
| EUR 6m Par Yield | 15.7 | 4 | 13.2 | 6.8 |
| EUR 2Y Par Yield | 22.7 | 2.7 | 22.3 | 60+ |
| EUR 10Y Par Yield | 20.5 | 2.9 | 20.1 | 60+ |
| EUR 30Y Par Yield | 18.6 | 3 | 18.2 | 60+ |
| USD 6m Par Yield | 23.7 | 4.7 | 18 | 4.4 |
| USD 2Y Par Yield | 29.1 | 2.7 | 28.6 | 60+ |
| USD 10Y Par Yield | 27.1 | 3.7 | 24.3 | 10.2 |
| USD 30Y Par Yield | 22.2 | 3.8 | 19.6 | 9.1 |

*Source: Lehman Brothers*
*Sample moments and parameter estimates for USD and EUR yield curve factors are estimated from monthly data. The factors are the negative of the change in the par rates over the month. The sample is from January 1990 through September 2005 for USD and from Feb. 1994 to Sep. 2005 for EUR.*

As can be seen from Figure 11, at the monthly frequency several of the yield curve factors do appear to have tails that are well described by the Normal distribution. However, in both markets the short end has thick tails and in USD the longer rates also appear to have thick tails. Moreover, if we look at data at the daily or weekly frequencies we see that curve factors have thicker tails than the normal across the yield curve.

### *Daily and weekly frequencies*

Figures 12 and 13 present t distribution parameter estimates and sample moments for yield curve factors using daily and weekly data. Although the Global Risk Model uses only monthly data it is still of interest to note that at higher frequencies even yield curve factors are uniformly heavy-tailed.

**Figure 12. USD Yield Curve Factors at Daily and Weekly Frequency**

| | Sample Moments | | t-Parameters | |
|---|---|---|---|---|
| **Daily** | **Std Dev** | **Kurtosis** | **Sigma** | **DoF** |
| USD 2Y Par Yield | 5.8 | 6.7 | 4.2 | 4 |
| USD 5Y Par Yield | 6.1 | 5.2 | 4.8 | 4.8 |
| USD 10Y Par Yield | 5.8 | 4.8 | 4.7 | 5.6 |
| USD 30Y Par Yield | 4.9 | 4.7 | 4.1 | 6.1 |
| Weekly | | | | |
| USD 2Y Par Yield | 12.8 | 4 | 11.6 | 12.1 |
| USD 5Y Par Yield | 13.6 | 3.3 | 12.9 | 19.6 |
| USD 10Y Par Yield | 12.7 | 3.4 | 12 | 18.9 |
| USD 30Y Par Yield | 10.9 | 3.6 | 10.2 | 15.9 |

*Source: Lehman Brothers*
*Sample moments and parameter estimates for USD yield curve factors estimated from daily and weekly data. The factors are the negative of the change in the par rates over the week/day. The sample is from Jul. 1986 through Nov. 2005.*

**Figure 13. EUR yield curve factors at daily and weekly frequency**

| | Sample moments | | t-Parameters | |
|---|---|---|---|---|
| **Daily** | **Std Dev** | **Kurtosis** | **Sigma** | **DoF** |
| EUR 6m Par Yield | 4.5 | 6 | 3.2 | 3.8 |
| EUR 2Y Par Yield | 4.8 | 4.9 | 3.6 | 4.4 |
| EUR 10Y Par Yield | 4.5 | 5.6 | 3.4 | 4.3 |
| EUR 30Y Par Yield | 4.3 | 5.6 | 3.3 | 4.6 |
| Weekly | | | | |
| EUR 6m Par Yield | 10.2 | 4 | 8.8 | 7.5 |
| EUR 2Y Par Yield | 10.7 | 3.9 | 9.4 | 8.6 |
| EUR 10Y Par Yield | 10 | 4.1 | 8.7 | 7.7 |
| EUR 30Y Par Yield | 9.7 | 4.3 | 8 | 6.2 |

*Source: Lehman Brothers*
*Sample moments and parameter estimates for EUR yield curve factors estimated from daily and weekly data. The factors are the negative of the change in the par rates over the week/day. The sample is from January 1994 through Nov. 2005.*

## Global credit spread factors

Figure 14 presents estimates of average parameter values over classes of credit factors. Clearly global credit factors are almost uniformly non-normal.

**Figure 14. Average parameter estimates for global credit factors with monthly data**

| | Sample Moments | | t-Parameters | |
|---|---|---|---|---|
| | **Std Dev** | **Kurtosis** | **Sigma** | **DoF** |
| USD Credit | 9.1 | 14.2 | 4.1 | 2.2 |
| Global High Yield | 84.6 | 8.7 | 37.4 | 2.3 |
| USD C/MBS | 6 | 9.1 | 3.9 | 3 |
| EUR Credit | 5 | 7.5 | 2.4 | 7.9 |
| GBP Credit | 6.5 | 6.7 | 4.6 | 4.4 |
| Global EMG Credit | 118.7 | 21.5 | 42.5 | 2 |

*Source: Lehman Brothers*
*Average sample moments and parameter estimates for global credit spread factors are estimated from monthly data. For example, the DoF entry for USD Crd is the average DoF estimate over the 31 USD credit spread factors. Data sample for USD covers January 1990 to September 2005, sample period for global high yield is August1997 to September 2005, for USD C/MBS sample period is January 1998 to September 2005. For EUR Crd sample period is from 1999 (with various beginning dates) until September 2005. GBP credit sample covers Mar.1999 to September 2005. Finally, global emerging market sample periods vary but most begin in 1997 and extend to September 2005.*

## Global FX factors

In this section, we present the relevant estimates for some major foreign exchange crosses. We report numbers estimated from both monthly and weekly data. It is interesting to note that unlike the yield curve factors there is very little tendency for the FX factors to have thinner tails at the monthly frequency than the weekly frequency.

**Figure 15. Parameter Estimates for Global FX Factors with Monthly Data**

| | Sample Moments | | t-Parameters | |
|---|---|---|---|---|
| | **Std Dev** | **Kurtosis** | **Sigma** | **DoF** |
| EUR | 2.8 | 3.9 | 2.5 | 10.3 |
| GBP | 2.7 | 7.5 | 2.1 | 4.9 |
| JPY | 3.2 | 5.3 | 2.6 | 6.1 |
| CHF | 3.2 | 3.5 | 3 | 20.9 |
| DKK | 2.9 | 4.1 | 2.5 | 8.3 |
| NOK | 3 | 4.1 | 2.5 | 6.8 |
| SEK | 3.2 | 6.7 | 2.6 | 6.3 |
| AUD | 2.6 | 2.7 | 2.6 | 60+ |
| CAD | 1.6 | 3.3 | 1.5 | 12.4 |
| NZD | 2.5 | 3.5 | 2.2 | 7.4 |

*Source: Lehman Brothers*
*Sample moments and parameter estimates for global FX factors estimated from monthly data. The factors are the percentage changes in the exchange rate for the indicated currency against USD. USD is the base currency in all cases. The sample is from January 1990 through December 2005.*

**Figure 16. Parameter Estimates for Global FX Factors with Weekly Data**

| | Sample Moments | | t-Parameters | |
|---|---|---|---|---|
| | **Std Dev** | **Kurtosis** | **Sigma** | **DoF** |
| EUR | 1.4 | 4.3 | 1.2 | 9.2 |
| GBP | 1.3 | 6.6 | 1 | 5.3 |
| JPY | 1.5 | 6.2 | 1.2 | 5.8 |
| CHF | 1.6 | 4 | 1.4 | 9.7 |
| DKK | 1.4 | 3.6 | 1.3 | 11.9 |
| NOK | 1.4 | 4.1 | 1.2 | 6.8 |
| SEK | 1.5 | 6.2 | 1.3 | 7.3 |
| AUD | 0.8 | 4.1 | 0.7 | 7.7 |
| CAD | 1.3 | 5.7 | 1.1 | 5.7 |
| NZD | 1.4 | 4.3 | 1.2 | 9.2 |

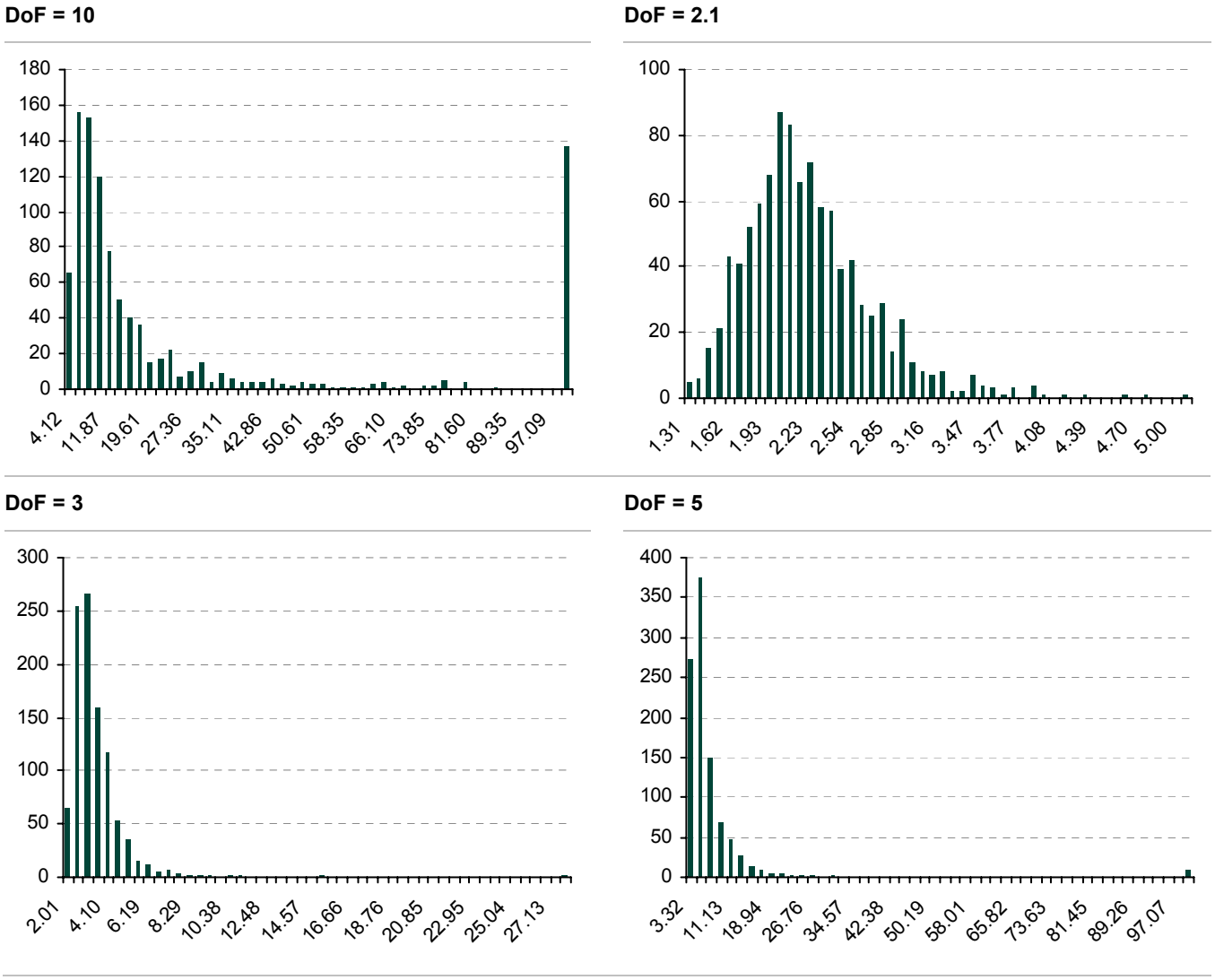
*Source: Lehman Brothers*
*Sample moments and parameter estimates for global FX factors are estimated from weekly data. The factors are the percentage changes in the exchange rate for the indicated currency against USD. USD is the base currency in all cases. The sample is from January 1990 through December 2005.*

## Understanding the parameter estimates

In interpreting the DoF estimates it is important to keep in mind that our sample sizes are all less than 200 observations and thus there is a lot of uncertainty in the estimates. In this section we attempt to get a handle on the distribution and implications of this uncertainty.

We perform a Monte Carlo experiment in which we simulate 1,000 samples, each with 190 observations, from a Student's t distribution. Then, from each sample we estimate the parameters by maximum likelihood. In these simulations the location parameter is always zero and the dispersion parameter is .25. We repeat the experiment for DoF values of 10, 5, 3 and 2.1. Histograms of the estimated DoF parameters are presented in Figure 17 and summary statistics are reported in Figure 18. We report medians instead of means because the extreme right skew makes the mean a fairly meaningless statistic.

**Figure 17. Histograms of DoF estimates form simulated data samples**




*Source: Lehman Brothers*
*This figure shows histograms of the estimated DoF parameters for 1000 simulated data samples each with 190 observations. In each sample the true value of the location parameter is 0 and the true dispersion parameter is .25.*

**Figure 18. Summary statistics for DoF estimates from simulated samples**

| True DoF | Median Estimate | Maximum Estimate | Minimum Estimate | 1st percentile | 5th percentile | % of Estimates > 20 |
|---|---|---|---|---|---|---|
| 10 | 11.15 | 100+ | 3.15 | 4.12 | 4.81 | 29.2% |
| 5 | 5.31 | 100+ | 2.34 | 2.87 | 3.22 | 3.1% |
| 3 | 3.12 | 27.92 | 1.75 | 1.88 | 2.22 | 0.1% |
| 2.1 | 2.14 | 5.12 | 1.27 | 1.42 | 1.58 | 0% |

*Source: Lehman Brothers*
*This table gives some descriptive statistics of the estimated DoF parameters for 1000 simulated data samples each with 190 observations. In each sample the true value of the location parameter is 0 and the true dispersion parameter is .25.*

In looking at the histograms the first thing that jumps out is the long right tail of the distributions. For example, for the case with a true DoF of 10 we still have 29.2 % of the estimates being greater than 20 and 13.7% of our estimates hit the upper bound of 100. On the other hand, for true DoF = 10 the smallest estimate is 3.15 and for true DoF=5 the smallest is 2.34. Clearly, extreme overestimation is fairly common but extreme underestimation is rarer. This has implications for our interpretation of the estimates obtained in the real data. For example, it may seem slightly odd that the USD yield curve factors have such disparate DoF estimates, with the 20-year and 30-year points having estimates of about 10 and the others having estimates on the upper bound. Our Monte Carlo results imply that the estimates of 10 are far more believable. That is, if the truth were even as high as, say, 25 then the probability of getting an estimate of 10 from the data is tiny while, on the other hand, if the truth is 10 we still have a reasonable probability of getting an estimate on the upper bound of 100. Motivated by these considerations in our POINT implementation we estimate parameters by maximizing the likelihood under the constraint that the DoF be less than or equal to 20.

On the other hand the very low DoF estimates are quite a bit more believable. For example, recall that for the case that the true DoF was 5, the smallest estimate (out of 1000) was 2.34, while in the data the *average* USD credit spread factor estimate was 2.2. Given an estimated DoF of 2.2 there is basically no chance that the true DoF was as high as 5. Moreover, for the case of the true DoF = 3 the proportion of estimates less than 2 was only 1.3%. Thus, for those factors whose estimated DoF is less than 2 there is less than a 1.3% chance that the truth was even as high as 3. However, even though the low DoF estimates are generally more credible, when the estimate is less than two there is a further complication.

### *Infinite variance?*

Some spread factors, such as credit or swap spread factors, have estimated DoF parameters of less than or equal to two. This implies that the factor volatilities are infinite. What should we make of such estimates? There is no theoretical reason that market spread factors *have* to have finite variances, so one possibility is to take the estimates at face value as our simulation study seems to suggest we do. After all, these are the best guesses given the data. However we can argue quite conclusively that in fact these spread factors do not have infinite variances. We simply estimate the t parameters of the changes in the swap yields themselves. When we do this we find that the DoF estimates are well above two (and generally above five) and thus it appears that the swap yield changes have finite variance. But since the swap yield is just the sum of the treasury yield and the spread, it can only have finite variance if the spread has finite variance. We solve this in our POINT implementation simpy by putting a lower bound on the DoF in our likelihood maximization, just as we imposed an upper bound. At the time of writing, this lower bound is 2.5.

As a practical matter the very low DoF estimates cause another problem in the implementation of tail risk calculations. Although a full explanation of the implementation details of our tail risk calculations is beyond the scope of this paper, we feel compelled to mention one implementation issue. The current implementation of tail risk calculations in POINT is simulation based. Very low DoF parameters and the resulting slow decay of the tails of the implied distribution can cause the occasional simulated value to be large beyond reason (see Figure 25 and the discussion immediately after the figure). Although such values have extremely small probability in any given sample, the number of samples drawn by our simulation algorithm is very large and this will tend to generate one of these “ultra-extreme” values fairly regularly. Our solution is straightforward: we simply truncate the simulated distribution so that these ultra-extreme values are deleted. This allows us to keep the good fit of our model over most of the relevant range without trying to come up with a different and possibly less parsimonious model just because of a few outliers. We mention this here only

because it serves to illustrate that there are, of course, limitations to any parametric model. That said, since we have available simple fixes to these issues we certainly think the advantages of our parametric modeling strategy far outweigh the difficulties.

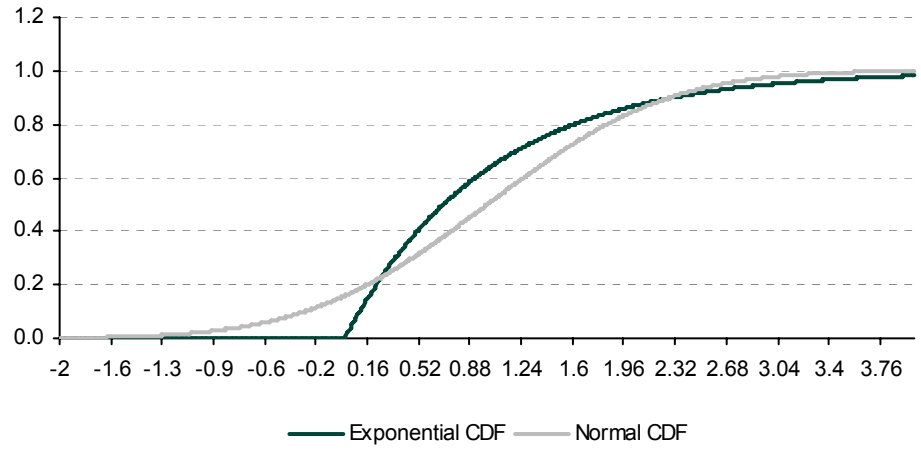
## SELECTING A SUITABLE MODEL FOR THE DISTRIBUTION OF THE SYSTEMATIC FACTORS

To assess the general quality of the fit of the models, we need to employ a specification test. The basic idea of the test is to see if the model provides an acceptably accurate description of reality. To be clear, we do not attempt to establish whether or not the model is literally "true" but instead ask only if it is an acceptable approximation of the truth. As detailed in Appendix A, we estimate parameters by the method of maximum likelihood. This procedure, as its name implies, chooses the parameter estimates that make the data as likely as possible to have been generated by the given model. The main reason we favor this method is exactly that it makes sense without requiring that the model be perfectly specified. When the model is only an approximation to the literal truth then there is no "true" parameter to estimate. However, it still makes just as much sense to choose the parameter that makes the data as likely as possible to have come from the given model. Moreover, it is a fact from the theory of mathematical statistics that this procedure yields a consistent and optimal estimate of the parameter values that make the data's *population* distribution as likely as possible to have come from the model's population distribution.

### A specification test

Specification testing takes the next step and asks, given that the parameter estimates make the data as likely as possible to have come from the model, does the model actually imply that the data is very likely? For example, suppose the data are drawn from an exponential distribution with mean and variance both equal 1 but we model it as normally distributed. We can still estimate the mean and variance of the normal model and let us suppose that our estimates also give 1 so there is no estimation error in the parameters. A normal distribution with mean and variance both equal to one assigns 84% probability of any observation being positive, which is quite high. However, if we have 100 observations, then the probability according to the Normal of all of them being positive is $.84^{100}$. which is on the order of $10^{-8}$, or basically zero. Since the data came from an exponential distribution, all of the observations will be positive and so it would be virtually a zero probability event that the normal distribution had generated them. The intuition for this example can be restated by saying that the CDF of the normal distribution takes the value .16 at zero (1-.84). The CDF of the exponential distribution is zero at zero. Thus, for values near zero the two CDFs are quite far from each other (see Figure 19 which plots the two CDFs).

**Figure 19. Normal and exponential distribution CDFs**




*Source: Lehman Brothers*
*This figure plots the exponential and normal CDFs with the mean and variance of both distributions set equal to one.*

Consider now an arbitrary random variable $X$ with CDF $F(x)$. A well known test of the specification of the CDF is the Kolmogorv-Smirnov test which, following the logic of the example in the preceding paragraph, is based on the distance between the CDF implied by the model and the empirical CDF of the data.

To be precise, let $F(x) = \text{Prob}(X < x; \textit{Model})$ and let $G(x) = \frac{\#obs < x}{\#obs} = \Pr ob(X < x; Empirical)$ . The logic of the preceding example implies that if these two functions are sufficiently different at any point then the data are very unlikely, given the model. The Kolmogorv-Smirnov test thus looks at the test statistic $T = \max\{|F(x) - G(x)|\}$. The only problem left is to decide on a critical value for how large the statistic must be before we reject the model as unacceptable. This critical value is generally chosen to make the test have a certain significance level[6]. Below we implement the test with significance levels of 5% and 1%. The distribution of the statistic is known, and this distribution is used to determine the critical value in the standard K-S test.

There is, however, one further complication that should be addressed. The problem is that the distribution of the test statistic does not account for the fact that $F$ is itself only an approximation of what the true CDF would be even if the model is literal truth. This is because $F$ is evaluated at parameter values that are only estimates and thus differ from the "true" (large sample) values. Parameters are estimated by fitting the distribution of the data as well as possible and we had a whole family of distributions, indexed by (at least 3) free parameters, available in fitting the data. This means that, to whatever extent the estimated parameters differ from the "true" values, the distribution evaluated at the *estimates* will fit *better in sample* than will the true parameter values. Of course, future observations will be drawn from the true distribution and not the estimated one. Thus, in order not to reject we should actually require a tighter in sample fit than we would for testing just a single fixed distribution. Moreover, because the mixture model has five free parameters and the t only three, we expect the mixture to fit slightly better than the t in-sample. Thus, for the test to have the desired significance level we should require an even tighter in-sample fit for the mixture than the t in order not to reject it.

Our solution to this problem is to use the K-S test statistic but to determine the critical values from a simulation study that will give us the true distribution of the statistic accounting for

*6 A significance level of 5% means, by definition, that if the model is the literal truth then we have only a 5% probability of falsely rejecting it.*

the error in the parameter estimates. The result is a far more powerful test and thus tougher test for the models. The simulation study is described in an appendix.

### The specification test on real data

We test the specification of the normal distribution, the mixture distribution and the t distribution on two sets of factors. One is the set of all factors for which we have at least 185 observations and the other is all factors with at least 40 or more observations (including the ones with 185+ observations). We conduct two tests, the first is the standard K-S test which tests whether the in-sample fit is acceptably good, this test treats the estimated parameter values as the truth. The reason that this version of the test is reported is because it is a particularly tough test of the model. Since the parameter estimates are chosen to make the model fit as well as it possibly can then a rejection by this version of the test is especially strong evidence against the models ability to describe the data. The results of the first test are reported as "Test 1" in Figures 20 and 21. We then repeat the test on the mixture and t models using the critical values from calculated by Monte Carlo to adjust for the fact that the model with more parameters should fit better in sample. Results are summarized as "Test 2" in Figures 22 and 23. When reading the results we must bear in mind that the test with significance level 5% has, by construction, a 5% probability of rejecting a true model, thus it appears quite clear that there is very little evidence against either the t or the mixture model and both provide an acceptable fit. These test results omit the convexity factors which have distributions with support on the positive real numbers and so a priori can't really be described by any of the candidate models. As a robustness check the second number in the cells reporting the results against the normal distribution also report the test result omitting the observation from Sept. 2001[7]. Although there are slightly fewer rejections of the normal the basic conclusions are unchanged.

**Figure 20. Kolmogorov-Smirnov Test 1: Results (185+ observations)**

| Model | Number of rejections at 5% significance level (76 tests) | Number of rejections at 1% significance level (76 tests) |
|---|---|---|
| Normal | 42 (55%) / 36 (47%)[8] | 32 (42%) / 25 (33%)[8] |
| Mixture | 0 (0%) | 0 (0%) |
| t | 0 (0%) | 0 (0%) |

*Source: Lehman Brothers*
*This table reports the results of the Kolmogorov-Smirnov specification test applied to those factors with more than 185 observations. This version of the test uses the exact critical values of the test treating the estimated parameters as the truth.*

**Figure 21. Kolmogorov-Smirnov Test 1: Results (40+ observations)**

| Model | Number of rejections at 5% significance level (334 tests) | Number of rejections at 1% significance level (334 tests) |
|---|---|---|
| Normal | 107 (32%) / 94 (28%)[8] | 67 (20%) / 55 (16%)[8] |
| Mixture | 2 (1%) | 0 (0%) |
| t | 2 (1%) | 1 (0%) |

*Source: Lehman Brothers*
*This table reports the results of the Kolmogorov-Smirnov specification test applied to those factors with more than 40 observations. This version of the test uses the exact critical values of the test treating the estimated parameters as the truth*

*7 Recall that both the t and the mixture distributions nest the normal distribution as special cases. This explains why for those factors where the normal is not rejected the t and mixture models aren't rejected either.*
*8 Results with Sept. 2001 observation omitted.*

**Figure 22. Kolmogorov-Smirnov Test 2: Results (185+ observations)**

| Model | Number of rejections at 5% significance level (76 tests) | Number of rejections at 1% significance level (76 tests) |
|---|---|---|
| Mixture | 12 (16%) | 4 (5%) |
| t | 6 (8%) | 2 (3%) |

*Source: Lehman Brothers*
*This table reports the results of the Kolmogorov-Smirnov specification test applied to those factors with more than 185 observations. This version of the test uses the critical values of the test calculated by Monte Carlo to correct for the effects of parameter uncertainty.*

**Figure 23. Kolmogorov-Smirnov test 2: Results (40+ observations)**

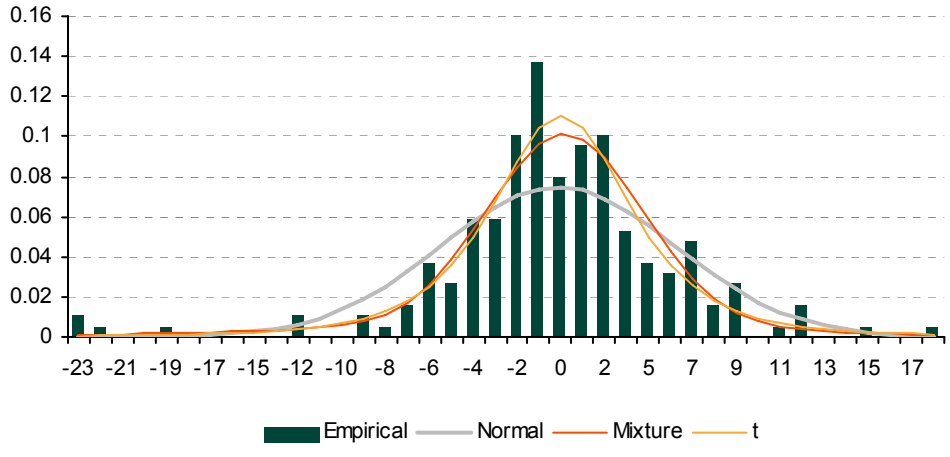
| Model | Number of rejections at 5% significance level (334 tests) | Number of rejections at 1% significance level (334 tests) |
|---|---|---|
| Mixture | 29 (9%) | 12 (4%) |
| t | 15 (5%) | 7 (2%) |

*Source: Lehman Brothers*
*This table reports the results of the Kolmogorov-Smirnov specification test applied to those factors with more than 40 observations. This version of the test uses the critical values of the test calculated by Monte Carlo to correct for the effects of parameter uncertainty.*

Let us examine in some detail a couple of examples to illustrate the specification test in action on real factors. We first revisit the example of the USD 5-year maturity swap spread. In our specification test neither the t distribution nor the mixture model is rejected for this factor. Way back in Figure 1 we had plotted the histogram of this factor against the normal density to show the gross inadequacy of the normal model in describing its distribution. Now in Figure 24 we plot its histogram again, this time we add the estimated densities of both the t distribution and the mixture model. We include close ups of both tails. All the pictorial evidence appears to support the claim that both models give a much improved description of this factor's distribution.

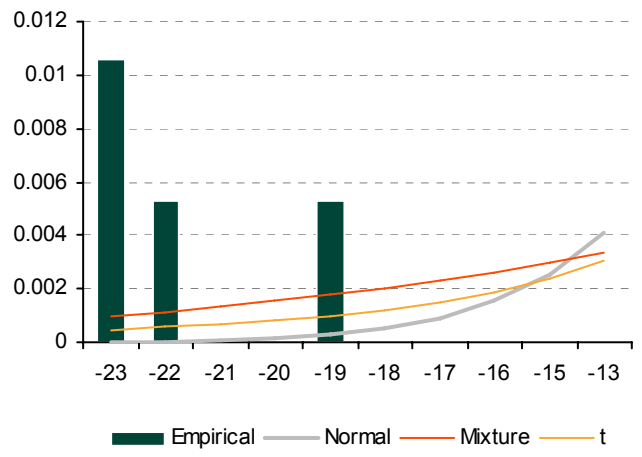
Figure 24 shows the histogram and fitted PDFs of the Normal, mixture and t distributions for the (negative of) the monthly changes in the USD 5 year swap spread. The parameters are estimated by maximum likelihood using data from January 1990 through September 2005. The lower panels show the left and right tails of the same distributions.

**Figure 24. Empirical, Normal, Mixture and t PDFs for USD 5yr Swap Spread Factor**



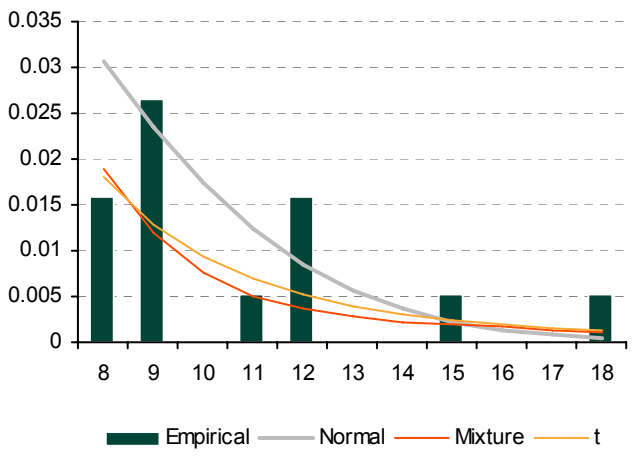

*Source: Lehman Brothers.*

**Left tail of Empirical, Normal, Mixture and t PDFs for USD 5yr Swap Spread Factor**




*Source: Lehman Brothers*

**Right tail of Empirical, Normal, Mixture and t PDFs for USD 5yr Swap Spread Factor**

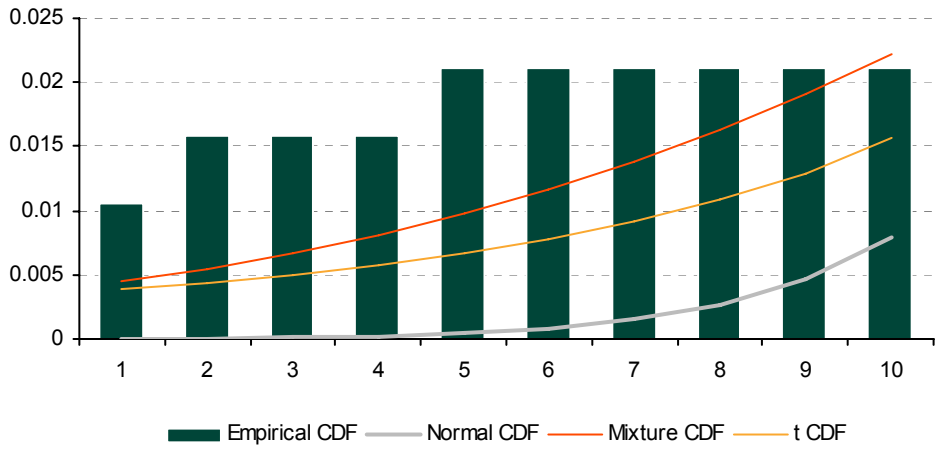



*Source: Lehman Brothers*
*The top panel shows the histogram and fitted PDFs of the Normal, mixture and t distributions for the (negative of) the monthly changes in the USD 5 year swap spread. The parameters are estimated by maximum likelihood using data from Jan. 1990 through Sep. 2005. The lower panels show the left and right tails of the same distributions.*

It appears from these graphs that the primary sources of the rejection of the normal are the left tail and the peak of the distribution. Of course, as previously noted, for symmetric distributions these are both reflections of the same problem. In examining the left tail plot it may appear at first glance that the t and mixture densities aren't that far above the normal density and the absolute difference between them is small, however if we look at the ratio of the t and mixture densities to the normal density we find that the t and mixture densities are between 30 and 50 times as large in the left tail and this can make a substantial difference. In Figure 25 we take a look at the empirical CDF compared with the normal, mixture and t model CDFs in the left tail. The source of rejection of the normal appears quite clear. At the most extreme point on the graph, corresponding to a 23bpwidening, the empirical CDF has a value of .01 (2/189). The normal CDF takes a value very close to zero (it's .00002) while the mixture and t CDF's both have a value of about .005.

To appreciate just how large the improvement in fit is let us revisit the basic logic of our test procedure. We have 189 observations on this factor, and 2 observations that are less than or equal to -23bp. The normal CDF says that the probability of any observation being less than -23bp is .00002 which implies that the probability of 2 (or more) out of 189 being less than -23bp is on the order of $10^{-6}$. On the other hand, both the mixture and t models give about a 24% probability to the event that 2 or more out of 189 observations are less than -23bp and we certainly would not rule a model out based on that result.

**Figure 25. Left tail of Empirical, Normal, Mixture and t CDFs for USD 5yr swap spread factor**




*Source: Lehman Brothers*
*This figure shows the empirical and fitted CDFs of the Normal, mixture and t distributions for the (negative of) the monthly changes in the USD 5 year swap spread. The parameters are estimated by maximum likelihood using data from January 1990 through September 2005.*

Figure 25 illustrates the potential need for truncation of the parametric distribution. While both the t distribution and mixture model CDFs appear to fit well over the range where data is observed they both are declining slowly and thus leave substantial probability to the left of the smallest observed factor realization. By contrast, the empirical distribution stops abruptly. On one hand this behavior of the parametric models is desirable. After all, just because we have not yet seen any smaller realizations does not mean we cannot ever, and our sample size is still small. One of the main reasons we employ a parametric model is to extrapolate outside the range of observed data, especially when the sample isn't large. The problem though is how slowly the tails are declining. This is necessary to match the empirical distribution, but the parametric models will extrapolate this slow decline far into the left tail while simple common sense tells us that in reality the factor distribution probably will eventually start to decline more quickly. This is really just a more extreme example of the phenomenon that most parametric models (including the Normal) will allow some probability of unboundedly large and small realizations that we know have a true probability of zero (like the fact that the Normal model allows for the possibility of negative interest rates). We accept this small defect as the cost of having a simple and tractable model that can describe the data well over the range that is likely to occur in practice. And so it is here that we accept this limitation of the models and employ our (admittedly *ad hoc*) truncation of the distribution to ensure we get reasonable output from a parsimonious model that fits the data well over the range most relevant in reality.

## ALTERNATIVE VOLATILITY SPECIFICATIONS

### DTS

Ben Dor *et al* (2005) find that the spread volatility of credit securities tends to be proportional to the spread level across a wide range of spreads. This implies that return volatility is proportional to duration times spread (DTS).

Incorporating the DTS result into our heavy-tailed models is straightforward. In the mixture model we simply specify the functional form of $\sigma_t$ as $\sigma_t = \sigma \cdot S_t$ where $S_t$ is the spread level and the sigma on the right side of the equation (without a t subscript) is the proportionality constant (it is a standard abuse of notation to use sigma in both meanings). Alternatively we

could simply specify that instead of the change in spread $\Delta S_t$ having the mixture distribution, the proportional change $\frac{\Delta S_t}{S_t}$ has the mixture distribution. This is tantamount to assuming that both the usual and large event volatilities are proportional to spread level.

For the t distributed model things are even simpler. We can proceed in a similar manner and model the dispersion parameter as $\sigma_t = \sigma \cdot S_t$ or simply model $\frac{\Delta S_t}{S_t}$ as t distributed but in the case of the t distribution both approaches are equivalent.

As we've argued throughout, accounting for stochastic volatility does not eliminate the need for heavy tailed models. To document this we formulate some examples of spread factors where $\frac{\Delta S_t}{S_t}$ is modeled as t distributed and report the estimates for the degrees of freedom parameter along with the absolute estimate (the model where $\Delta S_t$ is t distributed). It is clear from these estimates that even with the DTS formulation the factors still have distributions with far heavier tails than the Normal.

**Figure 26. DoF Estimates with Monthly Data**

| Factor | DTS Estimate | Absolute Estimate |
|---|---|---|
| USD Non-Cyclical A | 4.34 | 2.22 |
| USD Energy BAA | 3.29 | 2.27 |
| USD 2 yr Swap Spread | 3.45 | 3.33 |
| USD 5 yr Swap Spread | 3.73 | 2.9 |
| USD 10 yr Swap Spread | 2.86 | 2.4 |

*Source: Lehman Brothers*
*This table reports degrees of freedom estimates where either the factor normalized by its initial value is t distributed (DTS estimate) or the factor itself is t distributed (absolute estimate). In all cases the sample covers January 1990 through September 2005.*

## GARCH models

Another way to handle time varying volatility is to write down an explicit time series model for its evolution. By far the most common specification for this process is the GARCH (1,1) specification:

$$\sigma_t^2 = \omega + \alpha \cdot f_{t-1}^2 + \beta \cdot \sigma_{t-1}^2$$

In this specification, the current level of volatility is a linear combination of last period's volatility and last period's squared factor realization. In general $\alpha + \beta < 1$ and so volatility reverts to a long run mean. However, since $\alpha + \beta > 0$ and is generally estimated to be very close to 1, volatility is persistent. This model is designed to capture the tendency of volatility to cluster: a large shock today increases the level of volatility tomorrow and so large shocks tend to be followed by more large shocks. On the other hand, the model, by construction, is not consistent with large shocks to the factors that do not cause large increases in subsequent volatility and such shocks are in fact present in the data (see Figure 3). Moreover, as with the DTS formulation above, estimated models with t distributed factors and a GARCH volatility specification also generate fat tails with degrees of freedom estimates commonly well below 10.

## MODELING DEPENDENCE: CORRELATION AND COPULAS

The first three sections of this paper were concerned with modeling the marginal distributions of the systematic factors. For many, we soundly rejected the Normal distribution in favor of distributions with heavy tails, such as the Student's t distribution. However, the factors display a wide variety of different values for their degrees of freedom estimates, including many that are Normally distributed. Ultimately, it is the distribution of portfolios that we are actually interested in and the central lesson of portfolio theory is that the distribution of a portfolio depends as much on the dependence structure among the factors as it does on the distributions of the factors themselves.

Once we admit that we are not in a normally distributed world, the modeling of dependence becomes considerably more complicated. For example, the sum of two independent t distributed random variables is not t distributed (unless the variables happen to be Normal). The sum of two jointly t distributed random variables is again t but two variables can only be jointly t if they have the same degrees of freedom and it's clearly not the case that all systematic factors can be modeled with the same degrees of freedom parameter. Thus we need a way to express dependence between variables and thus calculate the distributions of linear combinations without constraining the factor's marginal distributions.

### Copulas

Consider a random variable $X$ with cdf $F(x)$. By definition $F(x)$=Prob($X<x$), now let's look at the cdf of the random variable $Y=F(X)$. Prob($F(X)<x$)=Prob($X<F\text{-}1(x)$)

$= F(F^{-1}(x)) = x$ and so we see that $Y$ is uniformly distributed. What's important in the preceding derivation is that its truth doesn't require any statements about the possible dependence between $X$ and any other variables. It only requires that $F$ be the marginal cdf of $X$. Moreover, the reasoning clearly goes the other way as well. Given a uniform random variable $U$ and a cdf $F$ we can form a random variable $X$ that, by construction, will have cdf $F$ by defining $X= F^{-1}(U)$. This appears to give hope that we might be able to define dependence in terms of variables with uniformly distributed marginals and then transform them to variables with specific marginal distributions. This in fact turns out to be the case.

We refer to the cumulative distribution function of any set of N variables with uniformly distributed marginals as a copula. The value of this can be seen from the following theorem, gleaned from Embrechts *et al*. (2001):

***Sklar's Theorem:*** Let $H$ be an n-dimensional cumulative distribution function with continuous[9] margins $F_1,...,F_n$. Then there exists a unique n-dimensional copula $C$ such that $H(x_1,...,x_n) = C(F_1(x_1),...,F_n(x_n))$. Conversely, if $C$ is an n-dimensional copula and $F_1,\ldots,F_n$ are distribution functions then the function $H$ defined by $H(x_1,...,x_n) = C(F_1(x_1),...,F_n(x_n))$ is an n-dimensional distribution function with margins $F_1,\ldots,F_n$.

If we suppose further that the distribution functions are *strictly* increasing then they will have well defined inverses and we get the following corollary, which allows us to represent copulas in terms of functions we already know:

***Corollary:*** Let $H$ be an n-dimensional cumulative distribution function with continuous margins $F_1,...,F_n$ and associated copula C. Then for $u_1,...,u_n$ each in [0,1], we have

$$C(u_1,...,u_n) = H\left(F_1^{-1}(u_1),...,F_n^{-1}(u_n)\right).$$

[9] *The most general statement of the theorem does not require the margins to be continuous. However, if the margins are not continuous then the copula is not unique.*

### Modeling the dependence of the systematic factors

As stated above, in a non-Normal world, correlation concepts are different than in the Normally distributed world. With jointly Normally distributed random variables, a correlation of zero is equivalent to independence, with jointly t distributed random variables this is no longer the case. The reason for this is as follows. To construct a pair of random variables $X_1$ and $X_2$ that are *jointly* t distributed with zero correlation we start with a pair of random variables, $Z_1$ and $Z_2$, that are independently Normally distributed and we form the joint t variables as $(X_1, X_2) = \frac{\sqrt{\nu}}{\sqrt{Y}}(Z_1, Z_2)$ where $Y$ has a Chi-square distribution with ν degrees of freedom. The key is to notice that for $X_1$ and $X_2$ to be jointly t we use the *same* Chi-square variable to construct them. Thus, even though $X_1$ and $X_2$ have zero correlation they are quite strongly dependent. In particular they are strongly dependent in the extreme tails. The reason is simply that what drives the really extreme events (say 4+ standard deviation moves) are the very small realizations of the Chi-square variable. This must be the case because such events are given virtually no probability by the Normal distribution. However, when the Chi-square variable has a small realization this applies to both $X_1$ and $X_2$ simultaneously and so they will tend to have extreme moves at the same time, a property referred to as tail dependence.

Our situation at first appears somewhat difficult. We have established that the Normal distribution is a poor model for the marginal distribution of many of our systematic factors. On the other hand, we do not want to model them all as jointly t since this would imply they would all have the same degrees of freedom and that is equally counterfactual. Yet, clearly we need some model of dependence to calculate portfolio level distributions. The answer of course is to use a copula to specify the dependence structure and then we are free to model each marginal distribution appropriately.

In the Risk Model, at the time of writing, we choose to model dependence by a Normal copula. This has the advantage that the meaning of correlation is preserved from the jointly Normal world and yet we are still free to model the marginal distributions as t distributed, with each factor having its own estimated degrees of freedom.

## DIRECTIONS FOR FUTURE RESEARCH

### Conditional skewness

Most spread type factors, most notably credit factors, are precluded by arbitrage requirements from ever having negative levels. The risk model factors model changes in spreads and this implies that such factors should have skewed distributions where the amount of skewness depends directly on the spread level. This is because to keep the spread level positive the spread change can never be larger in the left tail than the current spread level. This truncation does not apply to the right tail and thus the distribution must be asymmetric even if it appears to be roughly symmetric around the mode in normal markets.

The empirical modeling of conditional skewness is, however, a very difficult task. For example, it is not obvious how the right tail of the factor distribution should depend on the spread level. Moreover, since large events are rare to begin with and thus large events occurring when spreads are initially tight are even rarer, we have precious little data on systematic factors to guide our modeling. One way to mitigate this might be to use bond level data but this confounds systematic risk with idiosyncratic and it's not entirely obvious that the pattern of conditional skewness in the bond level data necessarily translates in the same way to the distribution of systematic factors.

### Tail dependence

We alluded to the concept of tail dependence in the section on copulas. In particular, the student t copula is a dependence structure that exhibits tail dependence whereas the Normal distribution does not. Again, the relative rarity of large events in general makes it difficult to statistically detect the presence of tail dependence in the data. After all, tail dependence between two random variables means only that extreme events have a *tendency* to occur together. It does not imply that if one variable has an extreme event then the other *must* also. A further difficulty in potentially modeling tail dependence is that we might believe that it only applies to one tail and thus how we might want to introduce it to the model is intimately related to how we choose to model conditional skewness. Thus, for the purposes of the present study we content ourselves with leaving these two issues for further work.

## CONCLUSIONS

Quantifying the risk exposure of bond portfolios with potential exposure to a myriad of markets and risks is challenging. Based on intense and thorough empirical study, the Lehman Brothers Global Risk model has been providing tracking error volatility calculations for bond portfolios of virtually any composition. However, in a non-Normally distributed world, volatility is an insufficient statistic for a portfolio's risk. Thus, providing measure of a portfolio's tail risk is of great interest. The first step towards accurately modeling portfolio tail risk is to study the entire distribution of the systematic risk factors instead of only their volatilities. In particular we focus on accurately modeling the tails of the systematic factor distributions.

In the present paper we have conducted a comprehensive empirical analysis of the systematic factor distributions, showing that, for many the Normal distribution is a grossly inadequate description of reality. We proceed to demonstrate that heavy-tailed alternatives to the Normal model can rectify its failings while retaining much of its tractability. Moreover, we show how the heavy-tailed models can accommodate changes in volatility regimes without difficulty and provide a parsimonious yet accurate description of the factor distributions. Finally, we have explained how we can model the dependence among the factors using a copula that allows complete flexibility in specifying the dependence structure while keeping the freedom to appropriately model each factor's marginal distribution.

In conclusion, we believe that the proper modeling of the complete distribution of our systematic risk factors and the associated ability to provide accurate tail risk measures is a large step forward in the natural evolution of the model and we hope this development will make the model even more valuable to managers of fixed income portfolios.

## REFERENCES

Ben Dor, A., Dynkin L., Houweling, P., Hyman J., van Leeuwen E. and Penninga O.(2005): "A New Measure of Spread Exposure in Credit Portfolios", Fixed Income Research, Lehman Brothers.

Embrechts P., Lindskog F. and McNeil A. (2001): "Modelling Dependence with Copulas and Applications to Risk Management", Department of Mathematics, ETH Zurich.

## APPENDIX

### Estimation of parameters of approximate models: the method of maximum likelihood

In this section we briefly review our chosen estimation method, the method of maximum likelihood. Although it is well known from basic statistics that for a perfectly specified model the maximum likelihood estimator is optimal in large samples, we wish to stress that we favor it because of its performance on models that are not perfectly specified; that is, approximate models. This seems to be the more important criterion because the assumption that *any* parametric model *perfectly* describes the randomness generating real world data would be beyond the limits of reason.

Formally, the maximum likelihood estimator is calculated as follows. Our model implies a probability density function (PDF) for the data that depends on some vector of parameters $\theta$. For example, if the model is the normal distribution then $\theta$ is the two element vector $(\mu,\sigma^2)$, if the model is the t distribution then $\theta$ is the three element vector $(\mu,\sigma,\nu)$. Let $p(x;\theta)$ be the PDF for the model with parameter vector $\theta$. As a PDF, this is a function where $x$ is the variable and $\theta$ is considered a fixed parameter. However, in a statistical estimation problem the data $x$ is fixed and it is over different parameter values that we need to search. That is, we treat $x$ as fixed and consider $p(x;\theta) = L(\theta;x)$ as function $\theta$. Viewed as a function of $\theta$, we refer to the PDF as the "likelihood" of the data and denote it by $L(\theta;x)$. The maximum likelihood estimator is then simply the value $\hat{\theta}$ that maximizes $L(\theta;x)$. In applications, the data are generally assumed to be a sample of independent draws from the specified distribution and so the PDF of the data is just the product of the PDFs of each observation. For this reason it is usually easier to work with the logarithm of the likelihood, defined by

$l(\theta;x)=\log(L(\theta;x))$. This is convenient because the log function turns products into sums and since the logarithm is an increasing function the value of $\theta$ that maximizes the log-likelihood will also maximize the likelihood.

The preceding discussion leads us to the question of what, if anything, are we estimating when the model is not perfectly specified? In elementary statistics the estimation problem is generally posed with the assumption that there is some value $\theta_0$ that is literally the true value in the sense that the data are draws from the distribution with PDF $p(x;\theta_0)$. Suppose now, more realistically, that the model is not perfectly specified. Let $q(x)$ denote the true PDF of the data, if the model is approximate then *there is no* $\theta_0$ *such that* $q(x) = p(x;\theta_0)$. So what are we estimating? The answer is that there is still a $\theta_0$ such that $p(x;\theta_0)$ is as close as possible to $q(x)$, for an appropriate measure of closeness.

Consider the quantity $M = -E\left[\log\left(\frac{p(x;\theta)}{q(x)}\right)\right] = -\int \log\left(\frac{p(x;\theta)}{q(x)}\right) q(x)dx$ which is referred to as the Kullback-Liebler divergence. We argue that this quantity provides a good measure of the distance between distributions. First, as the most minimal requirement, it clearly assigns zero distance if the two distributions are in fact the same. Moreover, $M$ is simply the average difference in the probability densities using the logarithmic scale and averaging over the true distribution. The logarithmic scale is appropriate because it explicitly corrects for the problem that in a region where both densities are small their absolute difference will also be small (no larger than the bigger of the two densities) and yet one density could still be thousands of times bigger than the other. For example, using the log scale has the desirable consequence that if the PDF $p$ assigns zero probability to any event that actually has positive probability then $M$ will equal infinity (using the convention that log(0) = -infinity). Thus limiting cases are handled appropriately. Moreover, it can be shown that $M$ is always positive

which is another obvious requirement of any distance measure. Using $M$ as our measure of the distance between distributions we thus define the "true" parameter to be the value $\theta_0$ that minimizes $M$. It can then be shown that the MLE is a consistent and asymptotically optimal estimator of this $\theta_0$.

### *Maximum likelihood estimation in practice*

To illustrate the maximum likelihood method we first observe that actually the usual sample mean and standard deviation *are* the maximum likelihood estimates if the model specifies the normal distribution. This explains why sample standard deviation is the best estimate of volatility in the normal model. To see this let us write out the likelihood for N independently and identically distributed (iid) observations from a normal distribution.

$$L(\mu,\sigma;x_1,...,x_n)=\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2\right)$$

which implies:

$$l(\mu,\sigma;x_1,...,x_n)=\log\left(L(\mu,\sigma;x_1,...,x_n)\right)=-\log\left(\sqrt{2\pi}\right)-N\log(\sigma)-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2$$

As usual, to maximize the log-likelihood we just differentiate with respect to each parameter and solve for the parameter values that simultaneously set the derivatives to zero. Differentiating the log-likelihood yields:

$$\frac{\partial l}{\partial \mu}=\frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i-\mu)$$

$$\frac{\partial l}{\partial \sigma}=-\frac{N}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{N}(x_i-\mu)^2$$

Setting the first of these equations to zero yields:

$$0=\frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i-\mu)\ \Rightarrow\ 0=\sum_{i=1}^{N}(x_i-\mu)$$

$$\Rightarrow\sum_{i=1}^{N}x_i=\sum_{i=1}^{N}\mu$$

$$\Rightarrow\sum_{i=1}^{N}x_i=N\cdot\mu\Rightarrow\mu=\frac{\sum_{i=1}^{N}x_i}{N}$$

Thus we see that the maximum likelihood estimate of $\mu$ is just the sample mean. In a similar fashion, we substitute the result for $\mu$ into the formula for the derivative with respect to $\sigma$ and set the result equal to zero to obtain:

$$\frac{N}{\sigma} = \frac{1}{\sigma^3}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 \quad \text{which implies } N = \frac{1}{\sigma^2}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2$$

$$\text{or } \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2$$

Thus, the sample mean and standard deviation are the maximum likelihood estimates in the normal model.

In the t distribution and mixture models we maximize the log-likelihood numerically.

### 10.2 Monte Carlo calculation of critical values of Kolmogorov-Smirnov Test

As discussed in the text, the standard K-S test simply tests whether a candidate CDF is likely to have generated a data set by comparing the candidate CDF to the empirical CDF. The test does not account for the fact that the CDF of the candidate model is itself subject to estimation error and thus does not properly penalize overfitting of the data. To correct for this we perform Monte Carlo experiments to determine the correct critical values so that the test will have the desired significance level.

Monte Carlo calculation of critical values proceeds as follows. For a given parameter vector θ and sample size N we draw K independent random samples, each with N observations, from the model. For each of the K samples we estimate the parameters by maximum likelihood and then, using the estimated parameter vector, we evaluate the test statistic. We now have a set of K values of the test statistic and the critical value for the 5% significance level is the $95^{th}$ percentile point of this set. Thus, by construction, the test will reject the true model exactly 5% of the time, which is the desired level.

At first glance it might be asked why it is not desirable to reject a true model as rarely as possible. The reason is that the model will, by chance, occasionally yield samples that are very unlikely. Any test that does not reject in those events will simply never reject – it will have no power to reject a false model. Thus, what we are actually doing in insisting that the significance level be exactly right is preserving the power of the test to discriminate against models that actually are poor approximations of reality.

In practice, we implement the Monte Carlo experiment in two steps. In the first, we choose K=1000 simulation runs and calculate the critical values for a variety of sample sizes and parameter vectors. Not surprisingly, we find that the critical values depend a lot on the sample size but do not change much with the parameter vector over the range of values relevant to us. Thus, in our second step we increase K to 5000 and calculate critical values for sample sizes 40 and 185 using a degrees of freedom parameter equal to 10, a dispersion parameter equal to 0.1 and a mean of 0. In the implementation of the test on real data we then get the critical value for a particular sample size by linearly interpolating (or extrapolating) the critical values for the sample sizes of 40 and 185

The views expressed in this report accurately reflect the personal views of Alexei Jiltsov, Arik Ben Dor and Vasanttilak Naik, the primary analysts responsible for this report, about the subject securities or issuers referred to herein, and no part of such analysts' compensation was, is or will be directly or indirectly related to the specific recommendations or views expressed herein.

Any reports referenced herein published after 14 April 2003 have been certified in accordance with Regulation AC. To obtain copies of these reports and their certifications, please contact Valerie Monchi (vmonchi@lehman.com; 212-526-3173).

Lehman Brothers Inc. and any affiliate may have a position in the instruments or the Company discussed in this report. The Firm's interests may conflict with the interests of an investor in those instruments.

The research analysts responsible for preparing this report receive compensation based upon various factors, including, among other things, the quality of their work, firm revenues, including trading, competitive factors and client feedback.