MODELING FINANCIAL MARKET RETURNS WITH A LOGNORMALLY SCALED STABLE DISTRIBUTION

RIMMER, ROBERT H., United States, mathestate@gmail.com

BROWN, ROGER J., United States

Abstract

A stable mixture distribution is presented as a model for intermediate range financial logarithmic returns. The model is developed from the observation of high frequency one minute market returns, which can be well modeled by random noise generated by a stable distribution multiplied by a non-random market parameter which is a measure of market volatility. The stable distribution has an α parameter of approximately 1.8, for the actively traded ETF, SPY. The daily time series of the scale factor shows strong serial dependence. Nevertheless the daily scale factor over periods of months to years is well fit by a lognormal distribution. Thus intermediate term market simulation and risk modeling can be accomplished with the product of a lognormal random variable and a standardized stable random variable. Although there is not a closed formula for the stable distribution, the mixture distribution and density functions can be approximated by numerical integration. Where ϕ is the stable characteristic function and λ is a lognormal density, the mixture characteristic function can be given by mcf.

$$\phi(t, \alpha, \beta) = e^{-|t|^\alpha \left(1 - i\,\beta\,\mathrm{sgn}(t)\tan\left(\frac{\pi\alpha}{2}\right)\right)}$$

$$\lambda(x, \mu, \sigma) = \frac{e^{-\frac{(\log(x)-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\, x\,\sigma}$$

$\mathrm{mcf}(t, \alpha, \beta, \gamma, \sigma, \delta) = e^{i\,\delta\,t} \int_0^\infty \lambda(s, \log(\gamma), \sigma)\,\phi(s\,t, \alpha, \beta)\,ds$, where α is the shape parameter of the stable distribution, β is the stable skewness parameter, γ is the median of the scale factor distribution, δ is a location parameter and σ is the shape parameter of the lognormal distribution fitting the varying scale factor. Numerically it is difficult to fit these parameters to data, but with the large sample sizes provided by intraday minute data, α can be approximated using the generalized extreme value distribution, and maxima of partitioned data. α can also be approximated by sequentially fitting each day's data; this value is surprisingly consistent, or by rescaling each day's data by the stable γ for the day and doing a stable fit to the rescaled data. The parameters for lower frequency daily returns can be approximated by taking advantage of the serial dependence, estimating the scale factor for partitioned data and rescaling the partitions.

The presentation shows evidence for the model with one minute returns of the SPY ETF collected since July 2007. This time frame includes a rather remarkable variation in market volatility, yet the model seems to remain valid. Calculations of the functions are demonstrated with *Mathematica*, and John Nolan's program, STABLE. A web resource of programs in *Mathematica* will be made available.

The model is attractive since it can account for all the stylized facts about financial returns and be explained as arising from the behavior of a continuous double auction market model that has limit order book return distributions with heavy power-tails, which over very short times measured in seconds yield independent returns obeying the generalized central limit theorem. The varying scale factor or volatility accounts for the serial dependence seen in the absolute value of market returns. The density of the mixture distribution has a higher peak than a stable distribution with the same parameters, α, β, γ, but on the tails it asymptotically approaches a stable distribution. Thus it is different from a truncated stable distribution. Sums of independent random variables from this distribution will converge to a stable distribution, but such behavior may not be observed in financial data because the scaling variables are not independent.

Introduction

We propose a model for financial markets that consists of random noise generated by a stable distribution multiplied by a scale factor. The scale factor variable of the model is not random and has a structure with strong serial dependence; it is a signal given off by the market, reflecting the market's volatility. The simple description in the first two sentences can account for all the stylized facts about financial returns (differences of logarithms of prices) which have been nicely laid out in *Quantitative Risk Management* as follows[1]:

"(1) Return series are not i.i.d. although they show little serial correlation.
(2) Series of absolute or squared returns show profound serial correlation.
(3) Conditional expected returns are close to zero.
(4) Volatility appears to vary over time.
(5) Return series are leptokurtic or heavy-tailed.
(6) Extreme returns appear in clusters."

As we proceed we will point out the reasons the stylized facts must be a consequence of such a model. We will also attempt to tie the model to the continuous double auction mechanism of price formation. The model is data driven. We have studied market returns with stable distributions for more than eight years and have found as have others that the fit is reasonably good except at the tails. When the tails from market return data are studied in isolation they consistently show a tail exponent that is too high for a stable regime. Looking closely at the stable fits to financial logarithmic returns, we find that market returns have more density in the central part of the distribution than the stable fit and that the tail exponent found by a stable fit is lower than slope found on the log-log plot of the empirical distribution function. Thus the assumption of a stationary stable distribution for market returns results in finding a set of parameters that overestimates extreme events. However, if we rescale the data by the non-random volatility component, we come up with an estimate of the tail exponent that is considerably higher than that found by the stationary stable assumption; such a model will be less likely to overestimate extreme returns.

We have observed that for time intervals of months to years, the histogram of the scaling variable is well fit by a lognormal distribution. We develop the lognormally scaled stable distribution as a stable mixture distribution. Since we know the scaling variable is not random, this is something of a thought experiment to explore the non-stationary stable behavior, but it may possibly be useful to simulate risk. The lognormal histogram is likely a consequence of the pattern of decay of the serial dependent structure that we will demonstrate.

Stable Distributions

For the analysis, we chose to use stable distributions, because they include the heavy tails consistent with the stylized fact (5), and our experience has been that although the stable fit is not ideal, it is better than that for most other commonly used continuous distributions. Stable distributions as a class have the attractive property that the distribution of sums of random variables from a stable distribution retain the same shape and skewness, although the summed distribution will change its scale and location parameters. Further they are the only class of statistical distributions with this property. The Normal distribution is one special member of the class as are the Cauchy and Levy distributions. These three forms, unfortunately, are the only members of the group that have simple mathematical formulas.

Stable distributions are the limiting distributions of sums of independent and identically distributed (i.i.d.) random variables. When the identically distributed random variables arise from a distribution with light tails, where the variance exists, the Normal distribution is the limiting distribution. This is the classical central limit theorem. When tails of the starting distribution are heavy, variance is infinite. In this case we observe the generalized central limit theorem for sums of random variables and the limiting distribution is a general stable distribution with shape parameter, $\alpha < 2$. The Normal distribution has tails that are light. All other members of the class have heavier tails ranging from slightly heavier than to Normal to extreme. The Normal distribution is the only distribution of the class for which the second moment and variance exist. For all other stable distributions variance is infinite. The characteristic function, $\phi(t)$, for the general case is given in equation (1). In the analysis of financial data we are not concerned with the case where $\alpha = 1$.

$$\begin{aligned} \phi(t) &= \exp\left(i\, t\, \delta - \gamma^{\alpha}\, |t|^{\alpha} \left(1 - i\, \beta\, \text{sgn}(t) \tan\left(\frac{\pi\, \alpha}{2}\right)\right)\right); \qquad \alpha \neq 1 \\ \phi(t) &= \exp\left(i\, t\, \delta - \gamma\, |t| \left(1 + \frac{2\, i\, \beta\, \text{sgn}(t) \log(|t|)}{\pi}\right)\right); \qquad \alpha = 1 \end{aligned} \tag{1}$$

Stable distributions are characterized by four parameters.

α is the shape parameter with a domain (0, 2], when it is 2, the distribution is the Normal distribution. When $1 < \alpha \leq 2$, the mean or expectation of the distribution exists. All financial market return data appear to be in this range of α.

β is the skewness parameter, domain, [-1, 1], defining the asymmetry of the distribution. When $\beta = 1$, it is maximally skewed to the right with a heavy tail, and the left tail is extremely light. The opposite is true when $\beta = -1$. At $\beta = 0$, the distribution is symmetric. As α increases toward 2, the effect of the skewness parameter diminishes and at $\alpha = 2$, the distribution is the symmetric Normal distribution.

γ, domain $(0, \infty)$, is the scale parameter. We will be suggesting this as a measure of volatility if α is constant or confined to a narrow range.

$\delta \in$ Reals, is the location parameter, and in the parameterization designated by the characteristic function $\phi(t)$ above, it is the expectation of the distribution, when $\alpha > 1$.

The summation-stability property has the interesting feature that it is possible to calculate the parameters of the distribution for any number of sums of random variables, if one knows the original stable distribution parameters. Thus the distributions are attractive for financial logarithmic returns, where the sum of a series of returns is the return for the series interval. Stable distributions thus are scalable in the sense that if a process arises from the sum of many small events, the summed event has a similar distribution. For financial market prices where price changes may occur thousands of times in a minute, it would be very convenient if the process were stable and stationary, in which case probabilities for events could be calculated across many intervals. Unfortunately there is no reason to believe that financial markets should output data in a distribution with stationary parameters.

Continuous Double Auction

Since we can easily obtain a large collection of financial return events, we tend to think of these as a random sample. But financial events are generated sequentially by a continuous double auction (CDA) system; so our model also needs to be consistent with this process of price formation[2].   In the past human market makers matched buy and sell orders and kept tract of the limit order books. Now market participants have access to data faster than the human brain can process the information, but they also have computers that can be programmed with algorithms to act faster than the information arrives; most orders are matched by computer algorithmicals. We know some of the rules, but we don't necessarily know exactly how the CDA is programmed, so we will consider it a black box and try to make very few assumptions. The graphic shows what the structure of orders in the market black box might look like at an instant when the market price is 50.

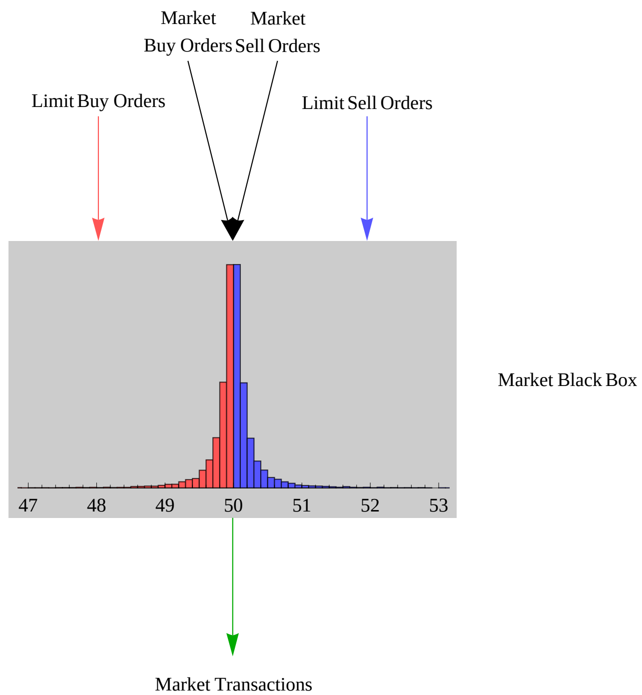

Figure 1

Orders flow into the system in two types, market orders which are executed as soon as possible within a very short time interval and limit orders which are sorted by price and time into an order database. The limit orders are filled on a first come first serve basis when the price limit is reached in the market. When numbers of buy and sell market orders are equally matched, the orders are executed at the current market price and each party is charged a commission on top of the market price. The market price does not change and the return over this interval is zero. If the market orders are not equally matched, then the unmatched buy or sell orders are filled in the respective sell or buy order books. The price changes, generating positive or negative returns. If the implied return structure of the order books has heavy power tails, then the continuous double auction will produce heavy tailed returns.

There is a lag in transmission of the executed price information, across networks, and prices may be executed in different markets and on different servers. So for some brief interval which now may be measured in hundreds of milliseconds to several seconds, one trader is not aware of the actions of other traders; there is also a lag in the execution of orders. Over some brief interval actions of traders are independent because of the information lag. The CDA process theoretically also involves both a buyer and a seller; over some defined interval, as the price changes, one of the parties will have made an inferior choice. The probability of making a better or poorer choice in the transaction is almost surely very close to 1/2, adding another element of randomness to the transactions. Thus over very short time frames of several seconds, market prices and the implied returns are *independent*.

The structure of the limit order books is also dynamic with orders being added and removed as the price changes. The structure of the limit order book, especially in the tails will likely remain relatively constant over some interval of time. Over the interval that this structure remains constant, returns will be *identically distributed*.

Sums of independent identically distributed random variables converge to a stable distribution. But over periods of time longer than a few seconds trading behavior is not independent and it is not likely identically distributed over long periods of time. Therefore it is not surprising that we find the parameters of the convergent stable distributions are not stationary. On a minute by minute basis, we expect the skewness parameter, β, to be varying wildly as this will directly follow the ratios of orders executed in the buy and sell market order books over the interval. δ, the mean, is expected to be close to zero by the stylized fact (3); so in looking at stable fits to the daily data we will be most interested in the structure of the α and γ parameters of the stable distribution.

Data

For our investigation we began to collect prospectively one-minute price data on the SPY exchange traded fund (ETF) since July 2007. We also collected at the same time data on 50 other securities with similar results, but will present here only the SPY data. The SPY ETF is ideal for study since it is heavily traded and has a relatively large price to the tick size of one cent. These properties assure that there will be enough high resolution returns to analyze and the calculated returns will not have too much granularity created by the discrete tick size.

The data points are at one minute intervals during the regular trading day of New York markets. The data do not include the before or after market trading sessions. Each day's data are concatenated and the log returns are calculated for the entire series. Concatenation in this manner creates a wider step in price and the calculated return at each day's open than is seen from minute to minute during the rest of the day, but no price information is lost. The Figure 2 shows the closing price each day for the data set.

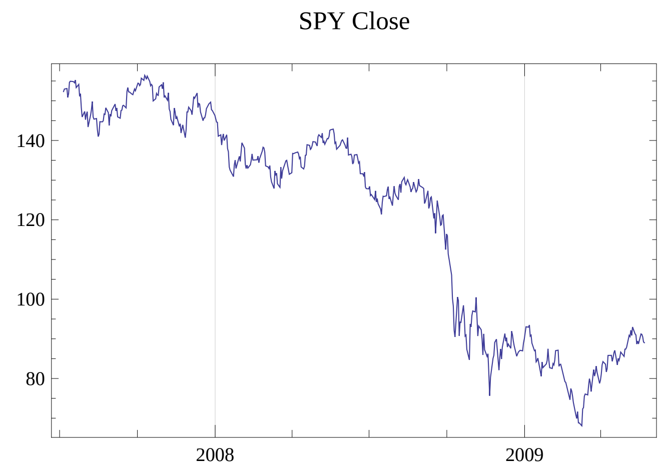

Figure 2

Data from Thu 5 Jul 2007 through Fri 22 May 2009.

The one minute logarithmic returns in Figure 3 show the clustering and variation of volatility that is typical of other financial data as noted in stylized facts (4) and (6).

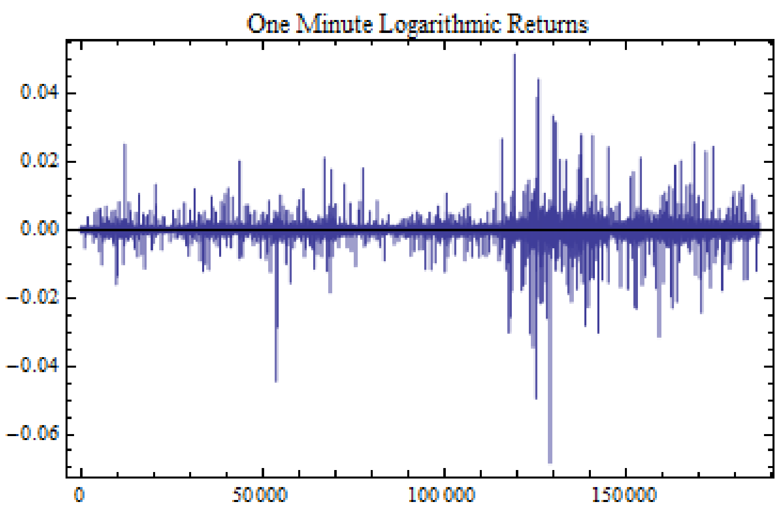

Figure 3

Figure 4 shows the autocorrelation function of the returns. The blue plot is the autocorrelation of the raw returns, there is no serial correlation, consistent with the stylized fact (1). The red plot is the autocorrelation plot of the absolute value of the log returns, which reveals a very interesting structure. The plot covers a lag of 15 days of one-minute returns. The tall spikes represent the larger returns experienced from the close of one market day to the open of the next. But there is also an intraday variation, which reflects higher volatility at the open and close of each market day. The daily variations are superimposed on a slowly decay of the autocorrelation function that persists for many months. Clearly there is significant serial dependence to the absolute value of the returns, stylized fact (2). When we take the absolute value of returns, we are in a sense looking at volatility at its most elementary level. Since there is a clear daily cycle in the data, we decided to partition the data into days to see what we could learn from analysis of the days individually.

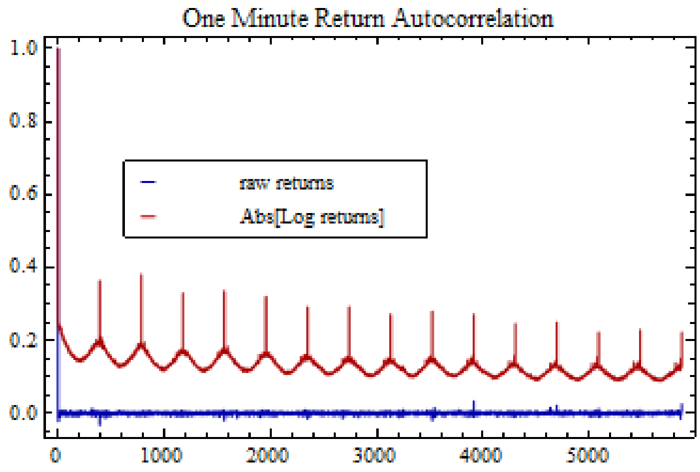

Figure 4

Daily Data Structure

To examine this large data set more closely we, partitioned the data into days and fit each day's returns to a stable distribution[9]. The mean of the parameters is shown. The average fit to each day's data gives significantly higher α than we obtain, taking the data set as a whole.

α	β	γ	δ
1.81028	0.0379846	0.000487981	-1.28519×10^{-7}

The plot, Figure 5, shows that α on a daily basis clearly clustering to significantly higher than the value found when the data are evaluated as a whole. The red lines are the 95 % confidence intervals for the maximum likelihood fit method, based on a usual sample size of 391. (Note : the days before some holidays are shorter; the sample size is considerably smaller on these days.)

SPY Alpha

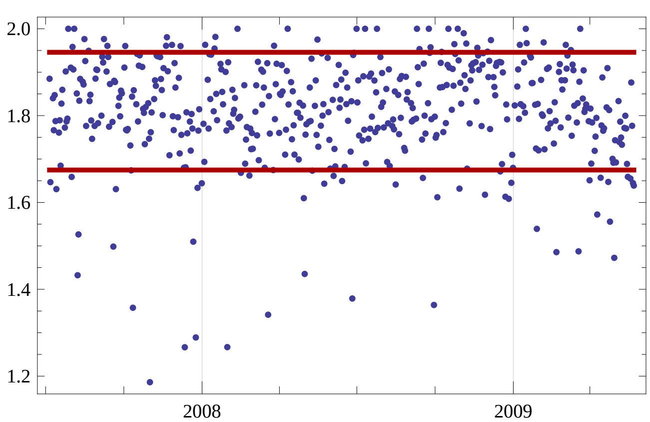

Figure 5

The daily structure of the stable scale factor, γ, in Figure 6 is very interesting; used in this way γ can be thought of as a measure of volatility. Clearly volatility varies over time, stylized fact (4). The plot shows the scale measured two ways, in blue is the maximum likelihood fit and in red is a very fast stable characteristic function method, which will be used throughout the presentation. The results are very close by both methods. The fast characteristic method is shown in the appendix.

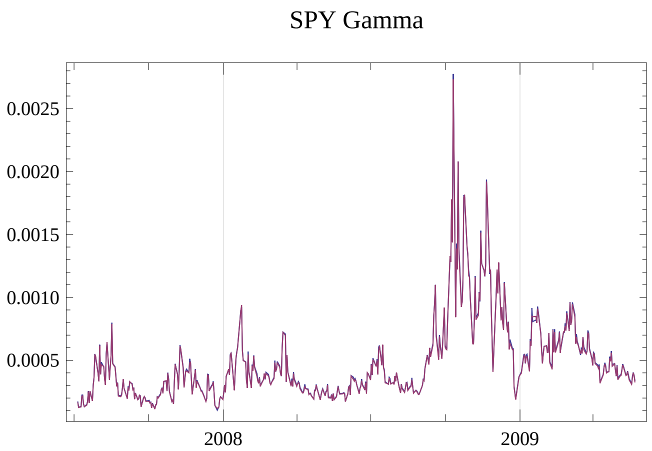

Figure 6

The structure of the β parameter appears to be fairly random over a wide range. δ picks up same pattern of volatility as the scale factor γ. At this point we can say that the logarithmic returns are not independent, and most of the serial dependency seems to be carried in the γ parameter; the stable scale factor. Figure 8 shows the very strong serial dependence in the daily γ values; since γ is the scale factor of the stable distribution, the very strong serial dependent structure accounts for stylized fact (6), clustering of volatility. α seems to be confined to a narrow range, mostly within the confidence intervals of the method. Since γ is the scale factor of the distribution, we can standardize the distribution to a stable distribution with $\gamma = 1$, by dividing by each day's return data by the scale factor for the day. When we rescale in this fashion, we will also remove the volatility seen in the plot of the δ parameter in Figure 7.

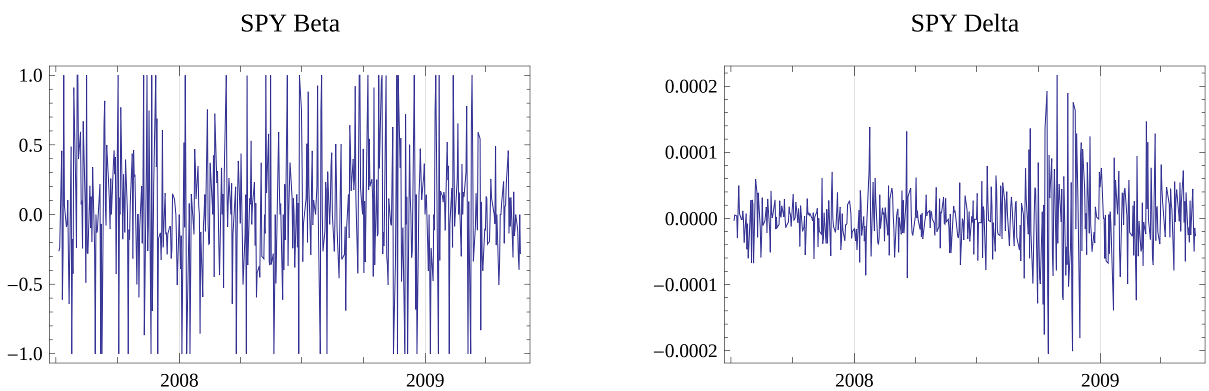

Figure 7

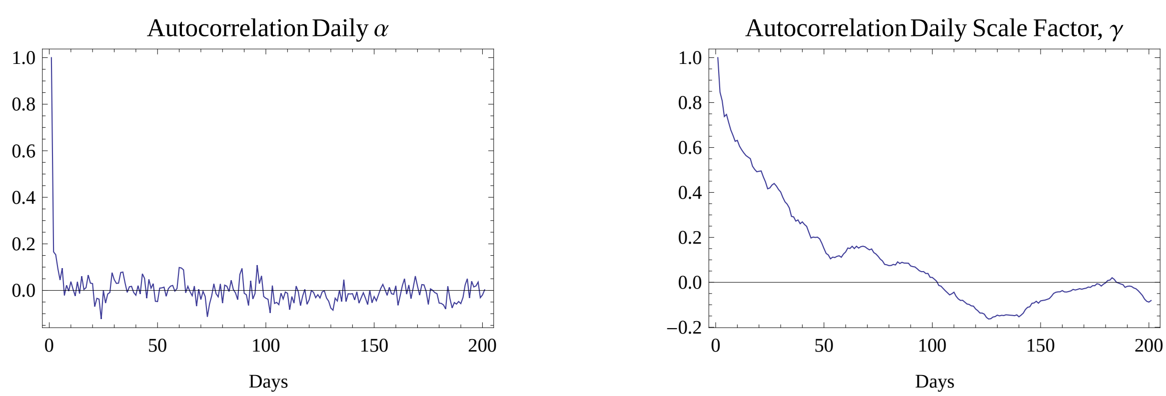

Figure 8

Rescaled Data

The finding that α for the daily data is relatively confined to a narrow range and the finding that γ clearly shows serial dependence and a wide range of variation suggests the idea that markets might be modeled by a non-stationary stable distribution with a varying γ parameter. Since γ is the scale parameter of the distribution, the data can be rescaled simply by dividing by the gamma for each day. Figure 9 shows the autocorrelation of the raw absolute log returns in blue and same returns rescaled (in red) by dividing each day's returns by the stable scale factor, γ, calculated from the stable fit to each days returns. This removes the sequential daily serial dependence in the data, leaving the intraday cycle and the inter-day jumps. We could remove these as well, but the effects are smaller and they are cyclical for each day. It is important to remember that the scale factor is constantly changing; we are not dealing with a stationary stable distribution even within a day.

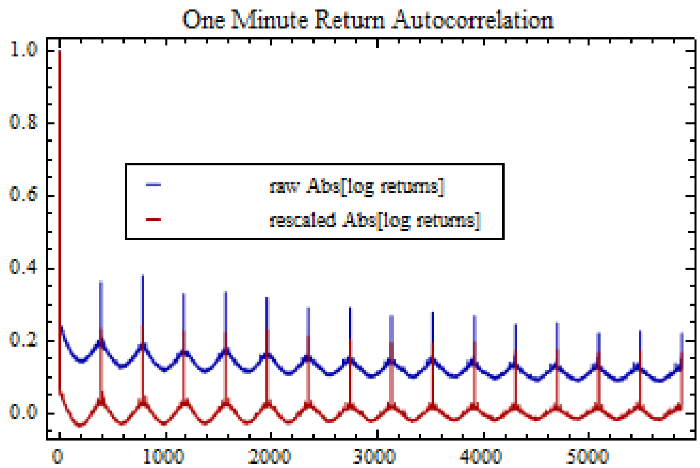

Figure 9

Here is a stable fit to the rescaled data. α is significantly higher than in the fit to the raw minute data. γ is approximately 1.0 as should be expected since each day's data was divided by the scale factor calculated for that day. It is not one exactly because we left the intra and inter day cycles alone.

Rescaled Return Data

α	β	γ	δ
1.79297	2.97653×10^{-9}	1.01272	−0.00331071

Raw Return Data

α	β	γ	δ
1.41505	1.92315×10^{-9}	0.000389648	-2.4746×10^{-6}

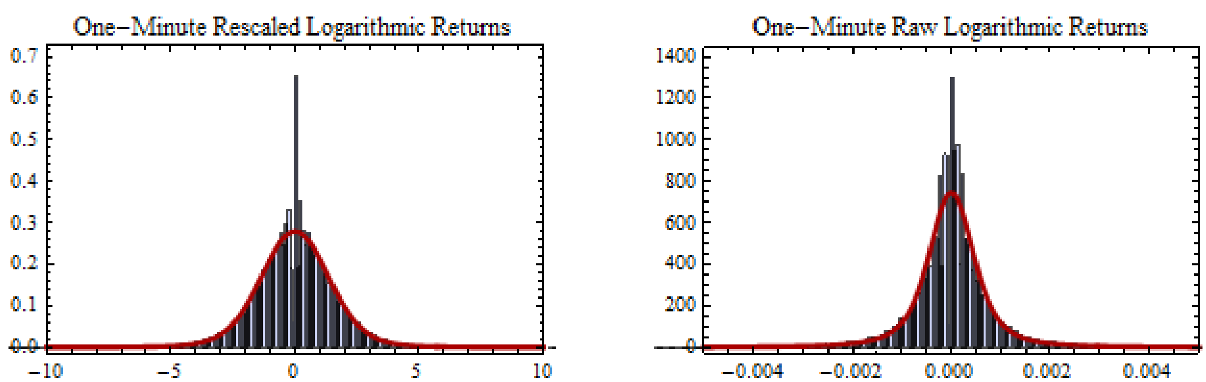

Figure 10

The stable distribution fit to the rescaled data is excellent. Figure 11 shows a log-log plot (rescaled data on left), which fits the data to the stable distribution function with parameters from the maximum likelihood fit. There are five plots in this illustration, three of which are lines representing the tails of distributions implied by parameters and two of which are plots of actual data. To align the tails, the absolute value of the left tail is shown in blue; the right tail is shown as 1 - probability in red. Displayed this way a stable distribution with $\alpha < 2$, reflects linear parallel tails with the slope of minus α. If β is zero the tails are superimposed. The tail of a Normal distribution, defined by its first two moments calculated from the sample, is shown in green and does not have linear tails. The dots show the data points of the two respective tails. In this representation, stable distributions have a characteristic shape with parallel tails that become linear when $\alpha \neq 2$; except when $\beta = \pm 1$, in which case the lighter tail is not linear. The data in the tails matches the tail exponent very closely for the rescaled data, but in the fit to the raw data, the calculated α is smaller than the slope of the data in the tails. There is still some excess central mass in the histogram caused by an excess of zero returns in the minute data. The phenomenon may be either due to very low volume on those minutes so that the price does not change, or it may be due to the granularity of the tick size, showing prices that changed but returned to the same value by the end of the minute, because of trades being at one cent price intervals.

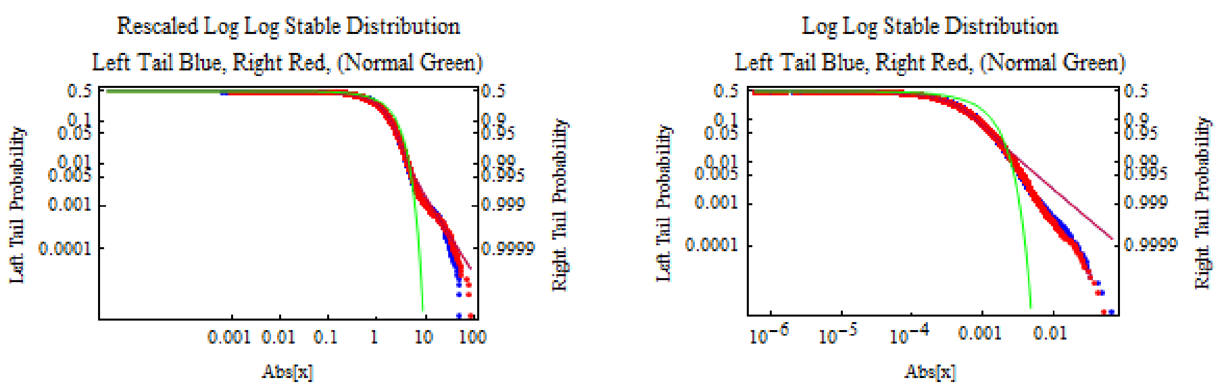

Figure 11

From these observations we form the hypothesis that financial market returns can be modeled by a non-stationary stable distribution with a scale factor which is continuously changing. The pattern of the change can itself be studied. The scale factor appears to account for almost all of the serial dependent structure seen in market return data. Over intervals as brief as a day, market behavior can be modeled by a volatility parameter *multiplied* by random α-stable noise. The volatility parameter, measured by stable γ, conveys non-random market information. The stable shape parameter, α, is relatively stationary with a value of about 1.8. Most other financial market models presume noise and a market signal are added together, the consequences of multiplying signal and noise are quite different. The heavy-tailed noise is multiplied by the volatility parameter, so events can be quite extreme and extreme events will cluster because of the persistence in the scaling volatility parameter, γ.

Structure of Volatility

The Figure 12 shows the serial daily volatility and the price of the SPY ETF along with the CBOE VIX volatility index. There is a general inverse relationship between the trend in price and that of volatility, both plots exhibit serial dependence, such that the best predictor of the next price is the last price and the same can be said of the volatility measure. The measure of stable γ closely parallels the popular volatility index, VIX. The volatility increases in late 2008 were dramatic and not predictable from the preceding data, yet the shape parameter of the stable distribution describing the daily returns remained relatively constant. We also note that the skewness parameter, β, and the location parameter, δ are small so that volatility can reasonably be measured by the stable scale parameter γ. We have shown that a stable distribution accurately describes data which has been rescaled by the value of γ. The structure of γ, can perhaps be studied by time-series analysis, but it is hard to imagine that any causal model could generate the rise in volatility that occurred in the Fall of 2008. Traditionally standard deviation has been used as a measure of volatility; we should point out that a similar graph could also be made using sample standard deviation for each day, but such a calculation implicitly assumes that the second moment of the distribution exists and that the distribution has light tails. In choosing a stable distribution scale factor, we have not excluded the possibility of finding a non-stationary Normal distribution in the analysis. The Normal assumption simply is not supported by the data, which are showing α of about 1.8.

Data from Thu 5 Jul 2007 through Fri 22 May 2009.

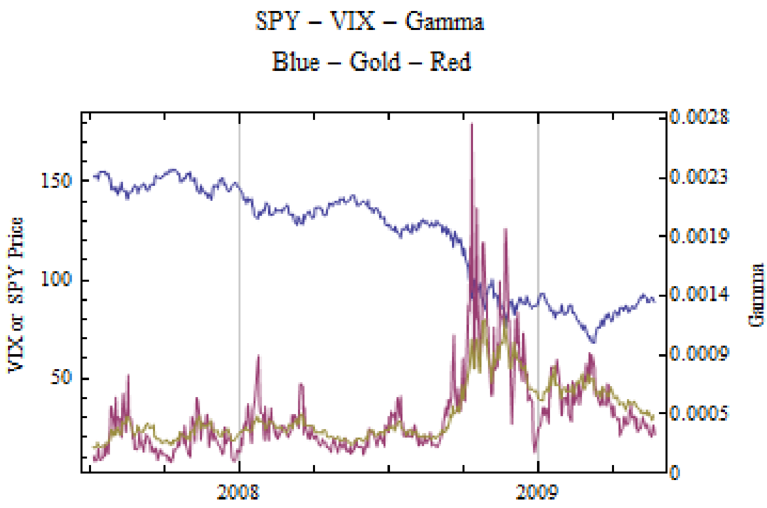

Figure 12

In Figure 13, we note a relationship with volume and stable γ, but we will not pursue that in this presentation.

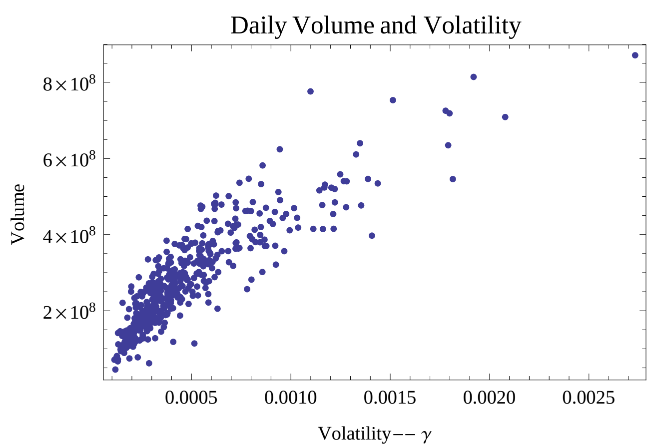

Figure 13

Although the volatility data clearly are not random, before September 2008, we found a remarkably good fit of the daily volatility to a lognormal distribution. This suggested that we might use a stable mixture distribution with the γ parameter distributed lognormally to give us a model distribution to represent financial returns over rather long time frames. The lognormal fit broke down with the marked increase in market volatility in the Fall of 2008. But subsequently it seems that we can divide the market data into two periods before and after the beginning of September 2008 and obtain good fits to a lognormal distribution--the difference is that the parameters have changed.

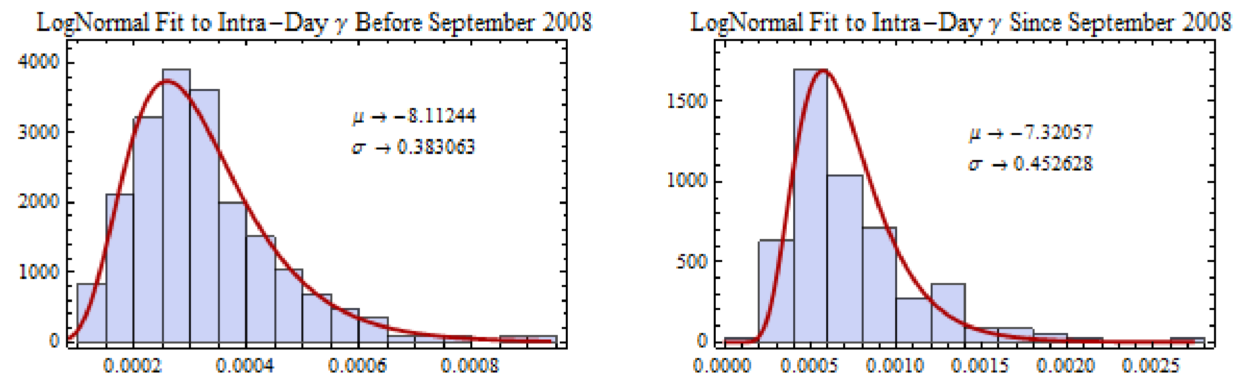

Figure 14

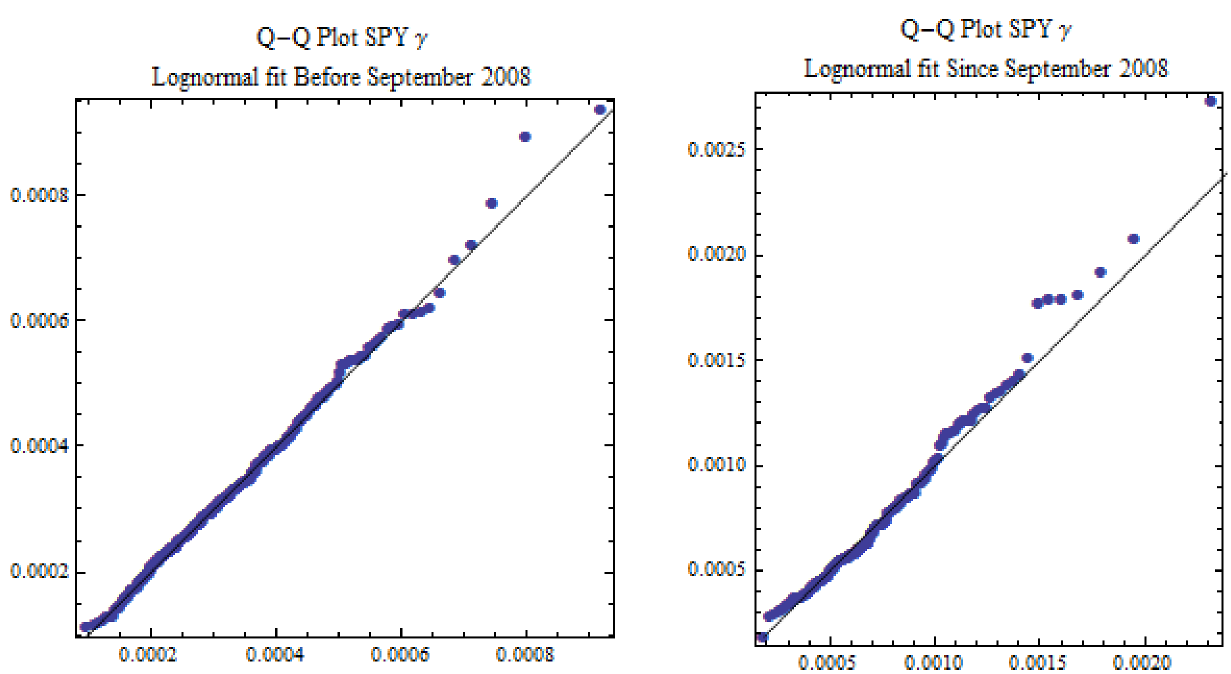

Figure 15

Although we know that the daily scale data are not random or independent, the histogram and fits to a lognormal distribution are very good in the partitioned samples. We believe this tells us something about the decay structure of volatility, which seems to be characterized by a series of rapid rises and slow decays that have a logarithmic relationship to γ. The fit to the lognormal distribution led us to investigate the behavior of a lognormally scaled stable mixture distribution with the hope that it might be useful over the long term to describe daily data. Our thought was that over a sufficiently long interval, the behavior of the scale factor γ, might appear more random.

Lognormally Scaled Stable Distribution

The lognormally scaled stable mixture distribution (LNS) is developed to describe what we have found empirically. It can be generated as the product of a lognormal random variable and a stable random variable. Since the stable distribution does not have a closed density, we will first describe it as a characteristic function. This function will have five parameters, the usual α, β, γ, and δ of stable distributions plus σ, representing the standard deviation of the distribution of Log[γ], where γ is a random variable, a scale factor that multiplies the stable distribution. We have shown that the distribution of the scale factor is in fact not random: there is strong serial dependence, but over many months to years the events described by measuring this parameter are well fit by a lognormal distribution, so this distribution may be useful to model financial market risk over months to years.

Let λ **(x, μ, σ)** be the lognormal density function.

$$\lambda(x,\ \mu,\ \sigma) = \frac{e^{-\frac{(\log(x)-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\ x\sigma} \tag{2}$$

ϕ(t, α, β) is the standardized stable characteristic function. (1-parameterization will be used for financial work.)[3]

$$\phi(t,\ \alpha,\ \beta) = e^{-|t|^\alpha \left(1 - i\,\beta\, \text{sgn}(t) \tan\left(\frac{\pi\alpha}{2}\right)\right)} \tag{3}$$

A characteristic function for a mixed distribution of characteristic functions is given by[4]:

$$\sum_{i=1}^{n} p_i\, \omega_i(t)$$

Where ω_i(t) is a characteristic function and

$$\sum_{i=1}^{n} p_i = 1$$

Extending this to a single characteristic function with a varying scale factor, where $s_i > 0$.

$$\sum_{i=1}^{n} p_i\, \omega(\,s_i\, t)$$

We substitute the integral of the lognormal density function for the weighted sum, yielding the LNS characteristic function, lnscf, where s > 0; γ is the median of the scale factor distribution. (Note: the median of equation (2) is at x = e^{μ}.)

$$\text{lnscf}(t,\ \alpha,\ \beta,\ \gamma,\ \sigma) = \int_0^\infty \lambda(s,\ \log(\gamma),\ \sigma)\ \phi(s\,t,\ \alpha,\ \beta)\, ds \tag{4}$$

Adding a location parameter, we have a five parameter characteristic function.

$$\text{lnscf}(t,\ \alpha,\ \beta,\ \gamma,\ \sigma,\ \delta) = e^{i\,\delta\, t} \int_0^\infty \lambda(s,\ \log(\gamma),\ \sigma)\ \phi(s\,t,\ \alpha,\ \beta)\, ds \tag{5}$$

Since we have numerical approximations for the stable distribution, we can also compute by numerical integration the distribution function and density functions[5]. Where scdf(x, α, β, γ, δ) is the stable distribution function, the distribution function of our LNS becomes:

$$\text{lnscdf}(x,\ \alpha,\ \beta,\ \gamma,\ \sigma,\ \delta) = \int_0^\infty \text{scdf}(x,\ \alpha,\ \beta,\ s,\ \delta)\ \lambda(s,\ \log(\gamma),\ \sigma)\, ds \tag{6}$$

Likewise the LNS density is shown below, where spdf(x, α, β, γ, δ) is the stable density function.

$$\text{lnspdf}(x,\ \alpha,\ \beta,\ \gamma,\ \sigma,\ \delta) = \int_0^\infty \text{spdf}(x\,,\ \alpha,\ \beta,\ s,\ \delta)\ \lambda(s,\ \log(\gamma),\ \sigma)\, ds \tag{7}$$

Plots of the distribution function with parameters, $\{\alpha, \beta, \gamma, \sigma, \delta\}$.

Plots of the LNS and stable density functions with the same parameters, $\{\alpha, \beta, \gamma, (\sigma), \delta\}$. The stable functions are in the lighter colors. For the same γ, the LNS will have a taller mode, and when there is skewness the LNS will appear less skewed.

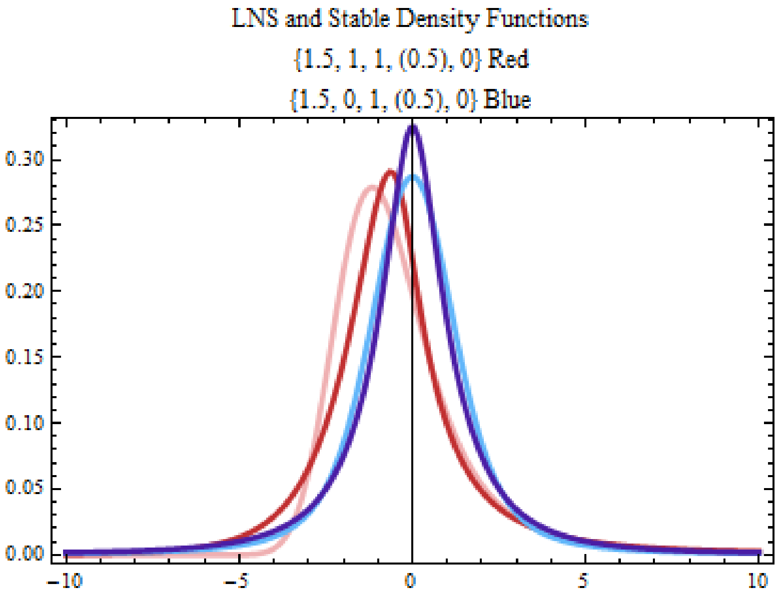

Figure 16

The three dimensional plot shows the LNS for a range of the parameter σ from 0.1 to 1, for parameters $\{\alpha, \beta, \gamma, \delta\} = \{1.5, 1, 1, 0\}$.

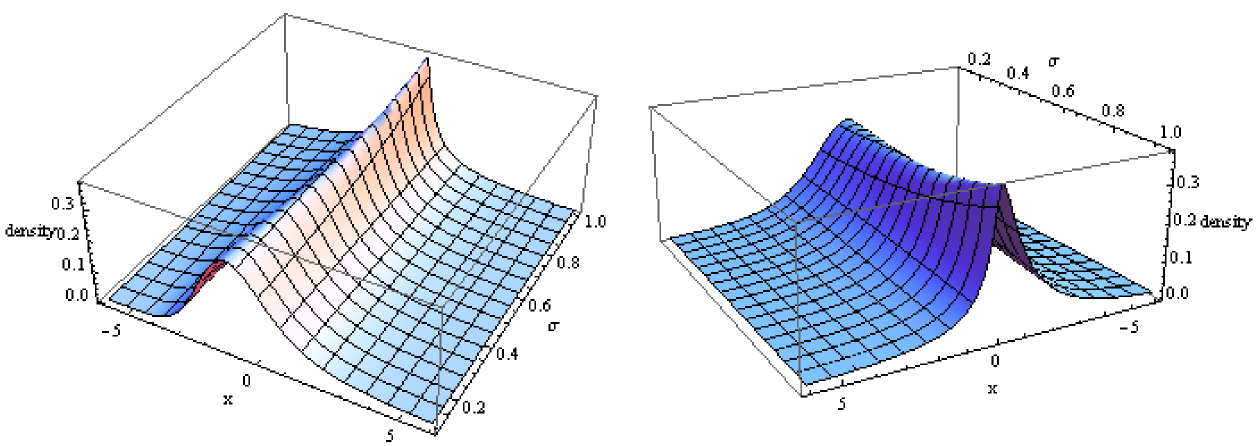

Figure 17

As σ approaches the limit zero, the distribution becomes a pure stable distribution with scale factor γ.

LNS Tail Behavior

The tail behavior of the stable and LNS distributions asymptotically is similar as x increases, but the linearity on the log-log plot arises more slowly in the LNS distribution. The right side of density plots are shown. The next three plots show the effect of increasing σ. As σ increases, it takes longer for the log-log linear phase of tail behavior to take hold.

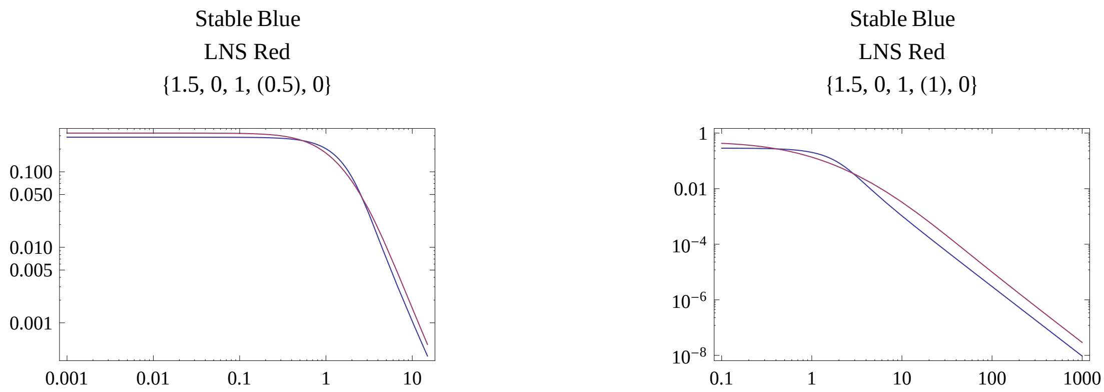

Figure 18

Figure 19 shows heavy tail of the maximally skewed density for the same α, 1.5 on the right and the light tail on the left.

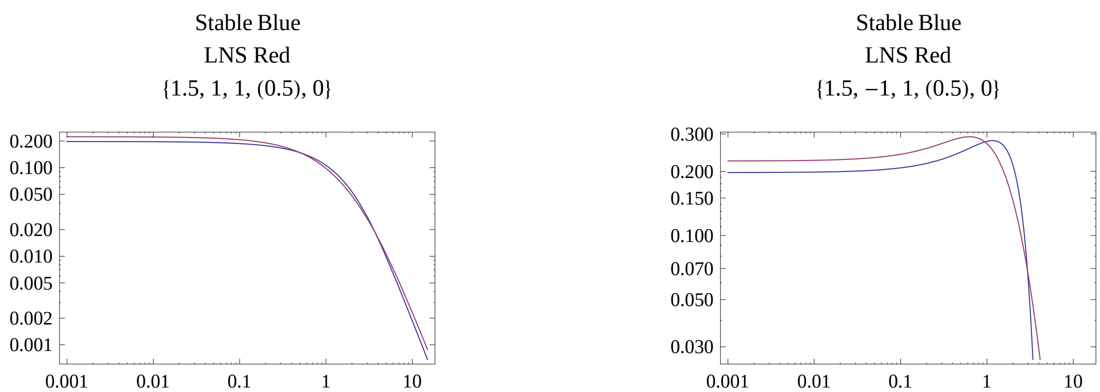

Figure 19

The tail behavior of the LNS distribution asymptotically has a Pareto tail behavior proportional to the stable distribution tail behavior from which it is derived. The proof is straight forward using the dominant term of the stable series expansion[6] to replace the stable distribution function in equation(6) and integrating the product of this term and the lognormal density over the range of the lognormal density.

Thus sums of LNS random variables have a limiting distribution that is a pure stable distribution by the generalized central limit theorem, but this might take sums of large numbers of random variables. It can be found from examining the LNS characteristic function that the properties of the moments of the LNS are similar to those of stable distributions, namely that the variance exists at $\alpha = 2$; the expectation exists when $\alpha > 1$.

In summary, the LNS distribution has another shape parameter, σ, that squeezes the middle of the distribution, causing a higher peak to the mode, and delays the stable tail behavior which occurs further from the mode of the distribution. Ultimately the distribution is heavy-tailed with the same tail exponent as the stable distribution from which it is derived.

Extreme Value Analysis

We have shown that the LNS distribution has the same tail behavior as a stable distribution but that it occurs further from the mode. This suggests that with our large data set we could calculate the tail exponent using the generalized extreme value distribution[7].

The generalized extreme value distribution in standardized form is given below.

$$e^{-(1+x\xi)^{-1/\xi}}; \text{ where } 1 + x\xi > 0. \qquad (8)$$

This function has a limit at $\xi = 0$, so it is continuous over the entire range of ξ.

$$\lim_{\xi \to 0} e^{-(x\,\xi+1)^{-1/\xi}} = e^{-e^{-x}} \tag{9}$$

A three parameter form of the distribution is needed for fitting, where μ is the location parameter and $\sigma > 0$ is the scale parameter, by substituting.

$$x \to \frac{x-\mu}{\sigma}$$

The full parameterization of the distribution function :

$$\begin{aligned} \mathrm{GEV}(x,\,\xi,\,\sigma,\,\mu) &= e^{-\left(\frac{(x-\mu)\,\xi}{\sigma}+1\right)^{-1/\xi}} \;;\; \text{where } \frac{(x-\mu)\,\xi}{\sigma}+1>0 \text{ and } \xi \neq 0 \\ \mathrm{GEV}(x,\,\xi,\,\sigma,\,\mu) &= e^{-e^{\frac{\mu-x}{\sigma}}} \;;\; \text{where } \xi = 0 \end{aligned} \tag{10}$$

The density function is.

$$\begin{aligned} \mathrm{gev}(x,\,\xi,\,\sigma,\,\mu) &= \frac{e^{-\left(\frac{(x-\mu)\,\xi}{\sigma}+1\right)^{-1/\xi}} \left(\frac{(x-\mu)\,\xi}{\sigma}+1\right)^{-1-\frac{1}{\xi}}}{\sigma} \;;\; \text{where } \frac{(x-\mu)\,\xi}{\sigma}+1>0 \text{ and } \xi \neq 0 \\ \mathrm{gev}(x,\,\xi,\,\sigma,\,\mu) &= \frac{e^{\frac{\mu-x}{\sigma}-e^{\frac{\mu-x}{\sigma}}}}{\sigma} \;;\; \text{where } \xi == 0 \end{aligned} \tag{11}$$

For maximum likelihood fitting, the corresponding log densities have an explicit form with the same corresponding parameter restrictions:

$$\begin{aligned} \mathrm{loggev}(x,\,\xi,\,\sigma,\,\mu) &= -\left(\frac{(x-\mu)\,\xi}{\sigma}+1\right)^{-1/\xi} - \log(\sigma) - \frac{(\xi+1)\log\left(\frac{(x-\mu)\,\xi}{\sigma}+1\right)}{\xi} \;;\; \text{where } \frac{(x-\mu)\,\xi}{\sigma}+1>0 \text{ and } \xi \neq 0 \\ \mathrm{loggev}(x,\,\xi,\,\sigma,\,\mu) &= -\frac{x-\mu}{\sigma} - e^{\frac{\mu-x}{\sigma}} - \log(\sigma) \;;\; \text{where } \xi == 0 \end{aligned} \tag{12}$$

For the calculations we use equations (12) above. We partitioned the data into days and selected the maximum values from each day's morning and afternoon events and fit this distribution to a generalized extreme value distribution (gev). To calculate the left tail, we take the negative value of the minima. The ξ parameter of the gev is corresponds to $1/\alpha$ of a stable distribution; thus we have another method of calculating α.

In doing the calculation, we have found that we run into the same difficulty as we experienced calculating the distribution of the scale factors-- the parameters have changed dramatically since the beginning of September of 2008; so we have divided the sample into two parts. First is the data from July 2007 through August 2008. The calculation of α is about the same for each tail and again higher than we observe when we attempt a stationary stable calculation for the data.

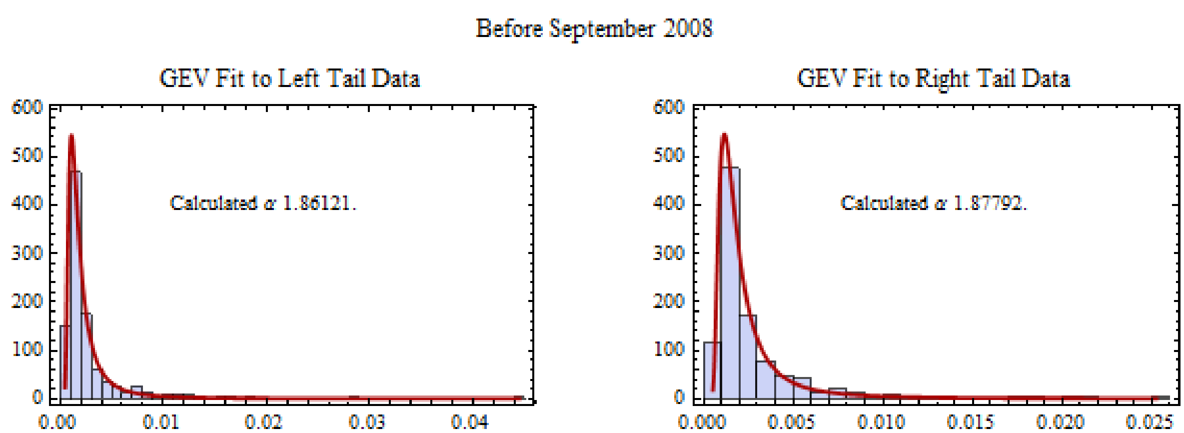

Figure 20

Figure 21 shows the calculations for the data since the start of September 2008. The sample size is smaller; on the left tail we are not yet seeing tail behavior consistent with our daily stable analysis.

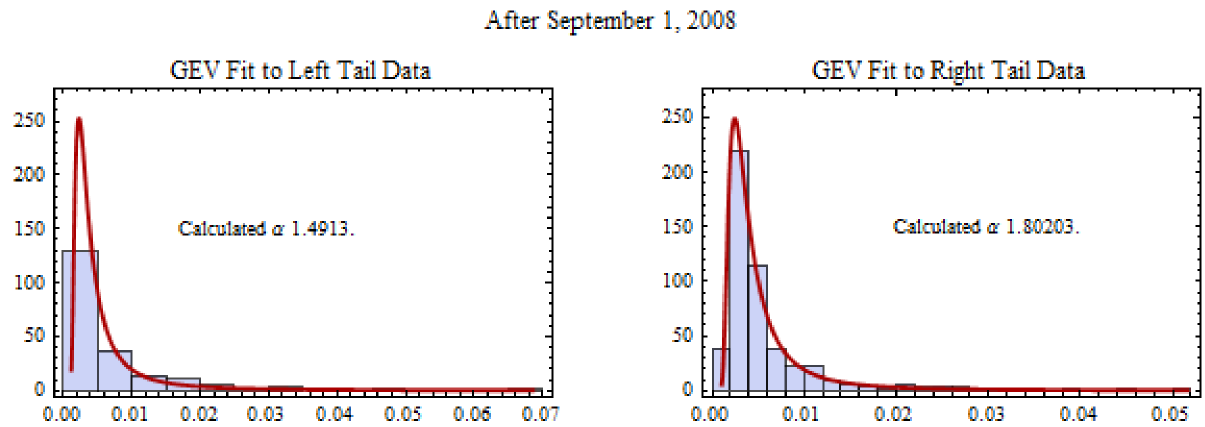

Figure 21

Fitting the LNS to a Daily Time Series

Here is the daily time series of prices for the SPY ETF since it began trading in 1993.

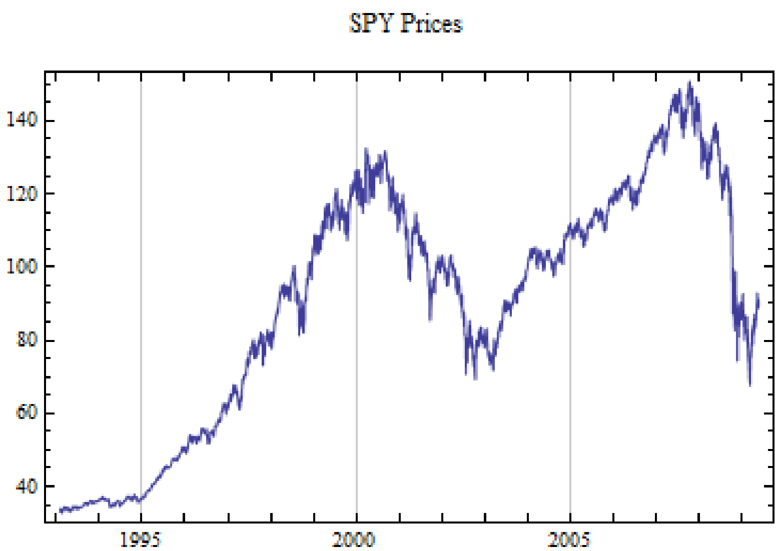

Figure 22

Data from Fri 29 Jan 1993 through Thu 21 May 2009.

So far we do not have a good method to directly fit data to a LNS distribution to data, so we use a few tricks, taking advantage of the known serial dependent structure of the scale factor. First we partition the data into sections of 30 trading days and calculate the scale factor for each of the partitions. Below we show the structure of the resulting scale factors and the fit to a lognormal distribution. The fit is taken using the median and a quantile method for the shape parameter.

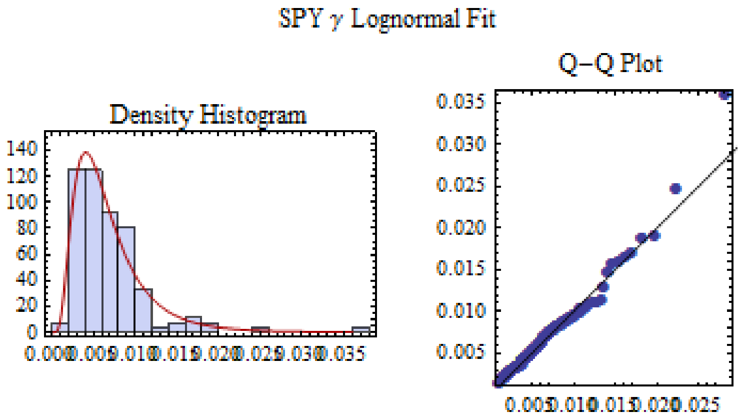

Figure 23

We then take the scale factors from each partition and rescale the data; stable parameters are fit to the rescaled data, resulting in a higher α, than we would get using a stationary stable fit. The parameters we use in the final fit are α and β from the rescaled stable calculation, γ and σ are taken from the above fit to the scale factors at 30 day intervals, and δ is the mean of all the returns. A Fast Fourier Transform method is used to calculate the LNS density as we will reuse this later to quickly calculate the log likelihood of the fit parameters, $\{\alpha, \beta, \gamma, \sigma, \delta\}$.

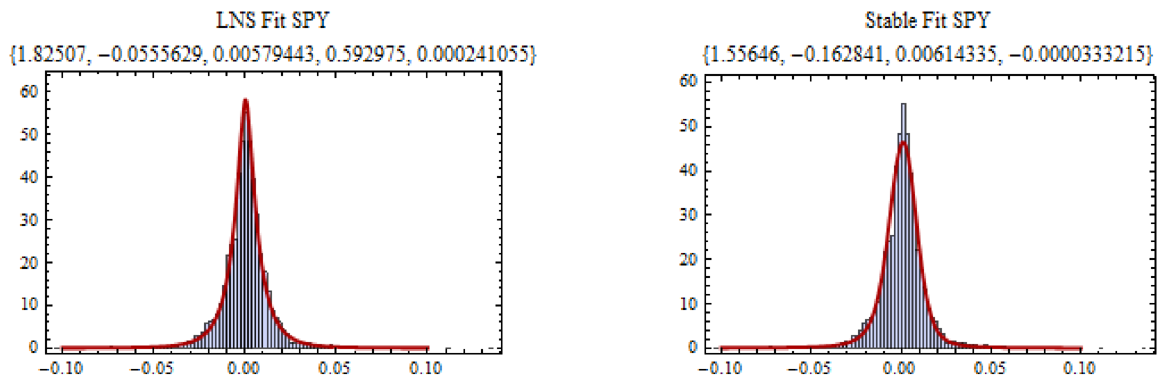

Figure 24

Comparing the two fits, we see that the LNS fit handles the peak of the distribution better and the tail exponent α is higher so that the parameters would be less likely to over-estimate extreme events if they are used for simulation. Below, the log likelihood is calculated for the stable fit and the LNS fit which is higher.

Stable log likelihood: 12682.9

LogNormalStable log likelihood: 12700.2

Below is a simulation with LNS random variables, using the parameters we calculated, cumulatively summed and converted back to prices. The simulation assumes both the lognormal and the stable random variables are independent, thus the simulated returns do not show clustering seen in true financial returns.

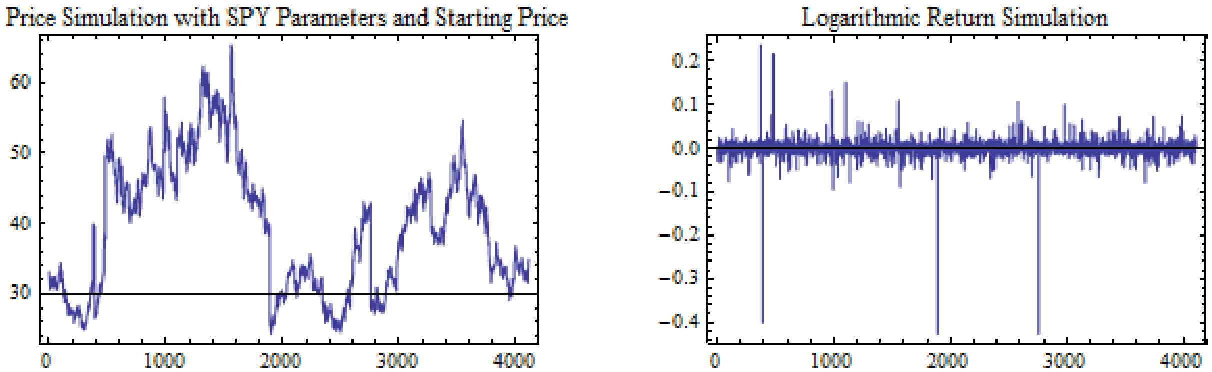

Figure 25

LNS Random Variables

Although the LNS distribution and density functions are difficult to calculate, it is relatively easy to generate random variables. In the demonstration we use *Mathematica*'s built in method for the lognormal random variable, and John Nolan's STABLE *MathLink* interface, although the algorithm to generate stable random variables directly is straight forward[9]. The algorithm is shown below.

```
LNStableRV[n_, α_, β_: 0, γ_: 1, σ_, δ_: 0, iparam_Integer: 1] := RandomReal[LogNormalDistribution[Log[γ], σ], n] StableRandom[n, α, β, 1, 0, iparam] + δ;
```

We can generate a million LNS random variables in about a second, with parameters $\{\alpha, \beta, \gamma, \sigma, \delta\} = \{1.8, 0.1, 0.006, 0.5, 0.0002\}$. The histogram and density with a LNS fit and a stable fit. As we have seen with financial data the stable fit has less central density than the LNS sample.

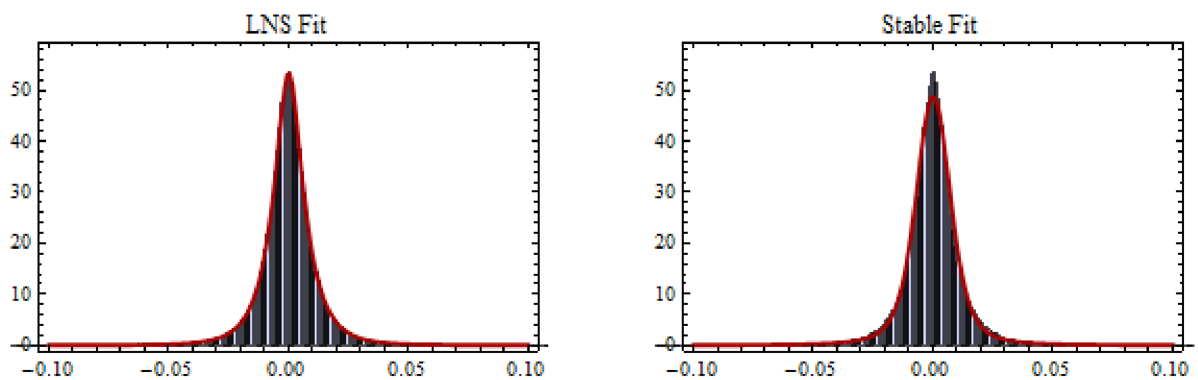

Figure 26

When the sample is fit to a pure stable distribution, we get parameters shown above the graphs, Figure 27. α as our previous experience has indicated is considerably smaller than the tail exponent in the original random variables. We cannot rescale the data, in the same manner as we did for the long series of SPY, because there is no serial dependence in our random sample. With this very large sample, we can show that the stable tail behavior is present upon summation of the random variables. The sample was partitioned into 1000 segments each of which is summed. The stable fit shows that sums of LNS random variables do indeed converge to a stable distribution with the proper tail exponent.

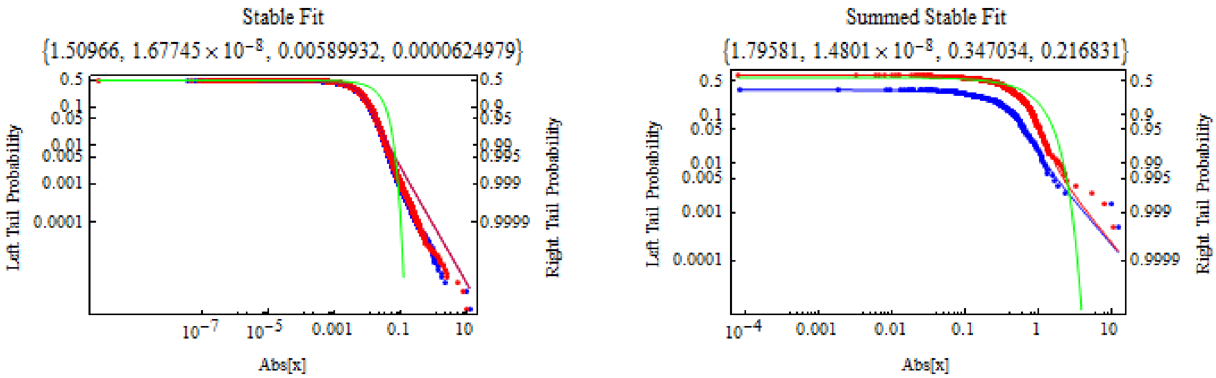

Figure 27

Discussion

We have examined logarithmic return data at one-minute intervals. We have found that over time frames as long as a day, these returns are adequately fit by a stationary stable distribution, although it is possible to show that the scale factor is continuously varying, this variation over the course of a day is smaller than the variation found between days[8]. Between days the magnitude of the scale factor variation becomes quite significant; it shows a high degree of autocorrelation. It is not random. It appears that financial market behavior for the SPY dataset can be modeled by a component of random alpha stable noise, $\alpha \cong 1.8$, *multiplied* by a scale factor, γ, which is constantly varying with a pattern of strong serial dependence. Over the time interval we have studied, the scale factor, γ, has varied 10 fold.

In time series notation we suggest a model of financial returns, where X_t is the return over interval t.

$$X_t = \sum_{i=1}^{t} X_i; \quad \{i \in \{1, 2, \ldots, t\}\}, \quad X_0 = 0 \tag{13}$$

$$X_i = c_i\, Z_i \tag{14}$$

Z_i is a standardized stable random variable with parameters (α, β). c_i effectively is the stable scale parameter for interval, i. If the distribution of c_i is constrained sufficiently then the tail behavior of the mixture distribution should converge to that of the stable distribution at the extremes as was shown for the lognormal distribution. c_i could be studied with time series' methods but we think it unlikely that any causal method would predict the dramatic increase in magnitude that was seen in October 2008. We have shown that c_i has a histogram and fits over selected time frames to a lognormal distribution, but that it has strong serial dependence; consequently the product with a random variable leaves a situation where X_i is not random. Although we created a mixture distribution, the LNS, to explore a random stable mixture, the strong serial dependence in c_i invalidates its use except possibly over long time frames. The lognormal relationship may be useful in studying the serial decay of volatility from a peak which may be a function of $\text{Log}(c_i)$. Figure 28 shows a power tail model for decay of volatility since October 2008, based on daily calculations of γ, a day trader would also have to be concerned about the intraday cycle in volatility.

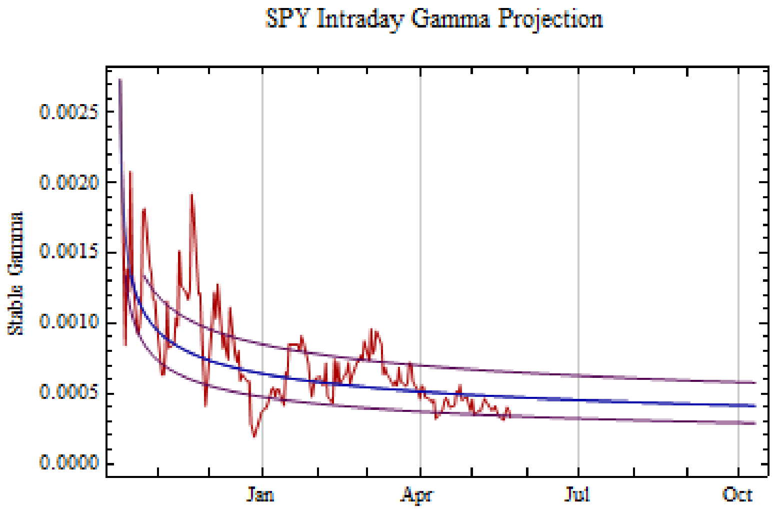

Figure 28

In this model, we are assuming that the α parameter of the stable distribution is relatively constant. We have shown that the α component, of the market model can be measured over longer time frames; it can be analyzed using the generalized extreme value distribution, but with less success than measuring α by rescaling the data. The results of such analysis confirm the heavy-tailed nature of the continuous double auction model, and the magnitude of the shape parameter obtained by this method is close to that found by fitting intraday data to a stable distribution. It seems reasonable to measure α over a long time frame and simplify the analysis with the assumption that it is constant over the interval.

We have presented a mixture distribution, the lognormally scaled stable distribution (LNS), mainly as a theoretical construction with the knowledge that the assumption of independence for the scaling parameter is false. The LNS is heavy-tailed, but the tail fit to the data is significantly lighter than found by fitting market data to a stationary stable distribution. This finding is consistent with the behavior of daily financial market data over many months to years. The LNS may be useful in assessing market risk over long time frames without overestimating extreme events. Although it is difficult to fit data to the LNS, because it is difficult to compute, it is easy to generate random variables for simulation with the LNS model. We have presented the LNS as a five parameter distribution, but for risk analysis over months, it may be adequate to simplify the model, setting the β and δ parameters to zero. Although sums of random variables from the LNS will by the generalized central limit theorem converge to a stable distribution, we do not expect that such behavior will be true of market returns over intervals longer than a few minutes. We emphasize that the serial scaling parameters are not independent. The the LNS distribution also includes the special case of a lognormally scaled Normal distribution by setting $\alpha = 2$.

Our most important finding may be the concept that the market returns can be modeled as the product of a market volatility signal and heavy-tailed noise. The dependent structure of the volatility signal component explains the clustering of volatility and can be studied separately from the noise. The idea that the noise, which is the random variation of logarithmic returns, is multiplied by volatility gives a compound interest effect to prices. We have not considered it here, but the varying speed of transactions is a likely component of the volatility parameter and should be a subject for future research.

Appendix

Outline a rapid method for estimating the stable scale factor, γ. Equations (15) show the characteristic function written in the 1 -parameterization rewritten in polar form, r exp($i\ \theta$).

$$\begin{aligned} \phi(t) &= e^{-\gamma^\alpha |t|^\alpha}\, e^{i\,\beta\,\mathrm{sgn}(t)\,\gamma^\alpha \tan\left(\frac{\pi\alpha}{2}\right)|t|^\alpha + i\,t\,\delta}; \quad \alpha \neq 1 \\ \phi(t) &= e^{-\gamma |t|}\, e^{i\,t\,\delta - \frac{2\,i\,\beta\,\gamma\,|t|\,\mathrm{sgn}(t)\log(|t|)}{\pi}}; \qquad \alpha = 1 \end{aligned} \tag{15}$$

The absolute value of the characteristic function for all α.

$$|\phi(t)| = e^{-\gamma^\alpha |t|^\alpha} \tag{16}$$

The empirical characteristic function derived from a data sample

$$\mathrm{ecf}(t) = \frac{1}{n}\sum_{k=1}^{n} e^{i\,t\,X_k} \tag{17}$$

When t and γ are both equal one you would have the characteristic function equal e^{-1}. This can be accomplished by dividing the sample by the γ for the sample. Equation (18) may be solved numerically.

$$\frac{1}{n}\left|\sum_{k=1}^{n} e^{i\, X_k/\gamma}\right| = e^{-1} \tag{18}$$

Another useful result follows from equation (19) given by Nolan for fractional absolute moments[3]. The fractional absolute moment exists for -1 < p < α.

$$E\,|X|^p = \frac{\sec\left(\frac{\pi\, p}{2}\right)\Gamma\left(1-\frac{p}{\alpha}\right)\cos(p\,\theta_0)\left(\gamma\cos^{-\frac{1}{\alpha}}(\alpha\,\theta_0)\right)^p}{\Gamma(1-p)};\ \text{where}\ \theta_0 = \frac{\tan^{-1}\left(\beta\tan\left(\frac{\pi\,\alpha}{2}\right)\right)}{\alpha} \tag{19}$$

and $\delta = 0$ and $p \in (-1,\ \alpha)$

In the case where $\alpha > 1$, equation (20) exists as a limit as p → 1. Since δ is the mean of the distribution, the left hand side of the equation can be estimated by the absolute mean deviation. The right hand side of the equation is directly proportional to γ; if α and β are known for the distribution, γ may be estimated.

$$E\,|X-\delta| = \frac{2\,\gamma\,\Gamma\left(\frac{\alpha-1}{\alpha}\right)\cos(\theta_0)\cos^{-\frac{1}{\alpha}}(\alpha\ \theta_0)}{\pi};\ \text{where}\ \alpha > 1 \tag{20}$$

References

[1]McNeil, A.J., Frey, R., Embrechts, P. *Quantitative Risk Management, Concepts, Techniques, Tools,* Princeton University Press 2005.

[2] Smith, E., Farmer, J.D., Gillemot, L., Krishnamurthy, S., Statistical theory of the continuous double auction, Quantitative Finance, 3, 481-514.

[3] Nolan, J. (2009) Stable Distributions - Models for Heavy Tailed Data, Birkhauser. Note: in progress, Chapter 1, available at: http://academic2.american.edu/~jpnolan/stable/stable.html

[4] Feller, W., *An Introduction to Probability Theory Vol. II,* John Wiley & Sons. 1971. p 504.

[5] *Ibid*. p 53.

[6] Bergström, H. (1952). On some expansions of stable distributions. Arkiv für Matematik 2, 375-378.

[7] Embrechts, P., Klüppelberg, C., Mikosch, T. (2008) Modelling Extremal Events for Insurance and Finance, Springer.

[8] For a closer look at the intraday cycle, see the demonstration on the Market Data page of mathestate. The site has downloadable *Mathematica* software and notebooks as well as a link to the data so that others may reexamine the results.

[9] Nolan, J.P., STABLE interfaces for *Mathematica*, matlab, S-Plus. www.robustanalysis.com.