NEW FRONTIER

Non-Normality Facts and Fallacies

by

David Esch, Ph.D.

Draft
September 2009

Abstract

Recently there has been an increasing trend in the quantitative finance community to call for statistical models which explicitly model returns with non-normal probability distributions (e.g. Sheikh and Qiao, 2009; Bhansali, 2008; Harvey et al., 2004). In this paper we explain why summary rejection of normal distributions is almost always ill-advised. We first examine some of the motivations for using normal models in financial applications. These models can account for non-normal return distributions despite their normal model components. We then demonstrate some consequences of switching to more complicated and less well-known non-normal models. These models almost always have more parameters to fit from the same data. All else being equal, rational investors should prefer parsimonious models, especially when the historical signal is weak, as is often the case in finance. We survey the shortcomings of several popular non-normal financial modeling techniques, especially when implemented naıvely. Although certain problems may warrant the use of other statistical return distributions, we argue that it is still important to exhaust the possibilities of normal models before switching to them. Models with normal distributions can be extended through methods such as conditioning on other variables, inequality constraints, mixtures, integration and resampling over unknown parameter distributions, or in some cases nonlinear transformations. The mathematical properties of the normal distribution facilitate these model-building techniques and allow for thorough post-analysis and model validation to ensure the best choice for the final model. Because of the preceding arguments, we reject the popular fallacy that because return distributions have marginal non-normal distributions, normal models cannot be valid or useful.

1. Introduction

Since the global market downturn in the fall of 2008, there has been growing concern in the quantitative finance community that statistical models for return distributions are not sufficiently accounting for downside risk and negative skewness (e.g. Anson *et al.*, 2007; Bakshi *et al.*, 2003; Harvey and Siddique 2000a; Harvey and Siddique, 2000b; Martin and Spurgin, 1998; Mills, 1995; Peiro, 1999; Sears and Wei, 1985; Taylor *et al.*, 2009; Sortino and Price, 1994; McDonald and Affleck-Graves, 1989; Kat and Miffre, 2008; Fuertes *et al.*, 2009). Several recently published whitepapers claim that by using advanced statistical methods such as "Extreme Value Theory", "Higher Moments", and "Gaussian Copulas," among others, one can somehow better prepare for extremely negative returns yet invest optimally over the long term (e.g. Sheikh and Qiao, 2009; Bhansali, 2008). While it is true that advanced methods can account for these extreme events, using these methods for investment management can easily backfire for the investor. This paper will describe some of these techniques and explain why they so often fall short.

In Section 2 we briefly describe why normal distributions are often used as a framework for investment practice. In Section 3, we discuss the advantages of parsimonious models over complicated ones, especially when used in out-of-sample prediction. In section 4, we discuss some of the issues associated with higher moments, why using sample estimates for skewness and kurtosis to fit models can be dangerous, and how even relatively tame normal models can explain data with various levels of skewness and kurtosis. In section 5, we discuss Extreme Value Analysis as well as other methods of analyzing tail behavior and extreme events. In section 6, we discuss the popular notion that correlations go to 1 during turbulent markets, and why this should not cause a rational investor to drastically change strategies, especially right after catastrophic events. Although this is not strictly an issue related to normal distributions, it is relevant to our discussion of rejecting standard models in the face of one or even several unusual observations. In section 7, we discuss some advantages of using the normal distribution in statistical models, and why we should not discard one of the most useful and well-researched tools for statistical modeling. In section 8, we discuss Gaussian Copulas, a technique which has been popular in the last decade in the derivatives market, which attempts to transform joint probability distributions to multivariatenormal and back to calculate conditional or marginal properties. This calculation is often erroneous and can have unforeseen negative consequences. In section 9, we provide a brief survey of the many situations when non-normal models might in fact be useful. In the final section, we summarize and make conclusions.

2. Normal Distribution and Portfolio Optimization

For many practicing financial economists, maximum expected utility of terminal wealth is the framework of choice for rational decision making under uncertainty. Optimized portfolios, in this framework, begin with a specific utility function and an assumed return distribution. In contrast, in practical applications, Markowitz (1959) mean-variance (MV) optimization is often used to define portfolio optimality. One important reason is that Markowitz MV optimization allows linear inequality as well as equality constraints in defining an optimal portfolio. Analytical solutions are not available for most practical applications. A common critique is that Markowitz optimization requires a normal return distribution assumption. Almost any analysis of historical return data finds that the normality assumption is violated, e.g. Fama (1965), Rosenberg (1974), Rosenberg and Ohlson (1976). Consequently many analysts have proposed alternatives to the normal assumption. There are many issues that go beyond the scope of this article. However, it is worth pointing out that the proper interpretation of Markowitz MV optimization is as a very useful approximation to expected utility maximization for many utility functions in investment practice (Levy and Markowitz, 1979). There is also the additional issue of estimation error of utility function parameters that can adversely affect optimality based on complicated or hard to estimate parameters, as described by Rubinstein (1973) and Michaud and Michaud (2008a, Ch. 3). From a practical point of view, Markowitz MV optimization allows non-linear constraints where the normal return distribution framework works as a very convenient approximation to expected utility maximization. Critiques of the normal distribution assumption in portfolio optimization should properly understand its valid foundational and important practical benefits for investment practice.

While empirically historical returns generally do not conform to normal distribution assumptions, proper interpretation of these results may be useful. Rosenberg (1974) attributes non-normality to the likely existence of mixed normal return distributions. From a practical point of view, Rosenberg's insight suggests that within a regime normality works quite well, but overlapping normal regimes result in hiding the approximate normality (within estimation error) of historical returns. The notion is that the mean and variance do approximate investors' expectations at any point in time but that the returns and variances vary over time. In this context a normal return distribution assumption is often appropriate for risk-return estimation at a given point in time.

3. Parsimony is Preferable to Complexity for Out-of-Sample Performance

Complex models have a tendency to overfit data. The more a model specifically fits the sample data, the less likely it is to fit future data not contained in the sample. There simply is not enough information in a small dataset to fit more than a few parameters. Financial data used for asset allocation or equity portfolio optimization may not contain many observations, because the markets are constantly changing and information becomes stale over time. When too many parameters are fit to a sample with not enough data, the model begins to fit the noise instead of the signal. Statisticians have devised criteria for choosing among models that often include a measure of model fit minus a penalty for model complexity, e.g. Akaike (1974) and Schwarz (1978). The optimal model for a dataset will simultaneously smooth out noise and detect information. Overly complicated models try to extract too much information from the data, resulting in a fitted model that is polluted with idiosyncratic noise and will perform poorly going forward. The argument here in favor of parsimony is a variation on the classic Occam's razor argument against unnecessary complexity in choosing among explanatory hypotheses.

Shrewd investors are always skeptical of claims that a complicated model will result in superior future performance. Although complicated models will almost certainly outperform more parsimonious models on the data from which they are fitted, their performance will suffer when subjected to new data. Even a very good model will suffer if it is reduced to a point estimate output, and put through an optimizer. It is often necessary to average over a distribution of potential outcomes to achieve optimal average out-of-sample (future) performance. Resampling is a useful statistical technique for extracting information from risk-return point estimates. Michaud Resampled Efficiency optimization takes into account many possible outcomes, and the resulting allocations, although they appear not to perform as well when measured against the sample data, are designed to be effective averaged over multiple outcomes. This means that they are designed to perform well over many return periods, because they take into account many possible outcomes (Michaud and Michaud, 2008a,b).

4. Higher Moments have Greater Estimation Error

Financial models often fit historical datasets which compromise between too little and too much data. Many applications use as little as five or ten years of historical return data. Less than a certain amount may have too little information and more may have too much stale information. Some practitioners have been recommending the use of "higher moments" for fitting statistical models. The problem with using the moments directly in analyses is that with a relatively small data size, higher moments become wildly unstable when estimated directly from sample data. Moments are calculated from modified averages of higher powers of the observations. For example, the skewness, or third moment, is calculated from an average of the third power of the data points. Similarly, the kurtosis, or fourth moment, uses an average of the fourth power of the data. Outliers in a dataset are even further removed from the rest of the data when these powers are taken, and can become highly influential, creating large sampling errors in the sample skewness and kurtosis.

Additionally, practitioners often misunderstand the implications of changing the higher moments while keeping lower moments fixed. Adding negative skewness to a distribution while keeping the mean and variance fixed will shift the bulk of the distribution positively. Similarly, adding positive skewness to a distribution while keeping the mean and variance fixed will shift the bulk of the distribution negatively. It is said (Harvey, Liechty, Liechty, and Mueller, 2004) that "most agree that *ceteris paribus* investors prefer a high probability of an extreme event in the positive direction over a high probability of an extreme event in the negative direction." The investors in this case do not realize that "*ceteris paribus*" means that most of the time they will underperform their peers who prefer zero or negative-skew return distributions.

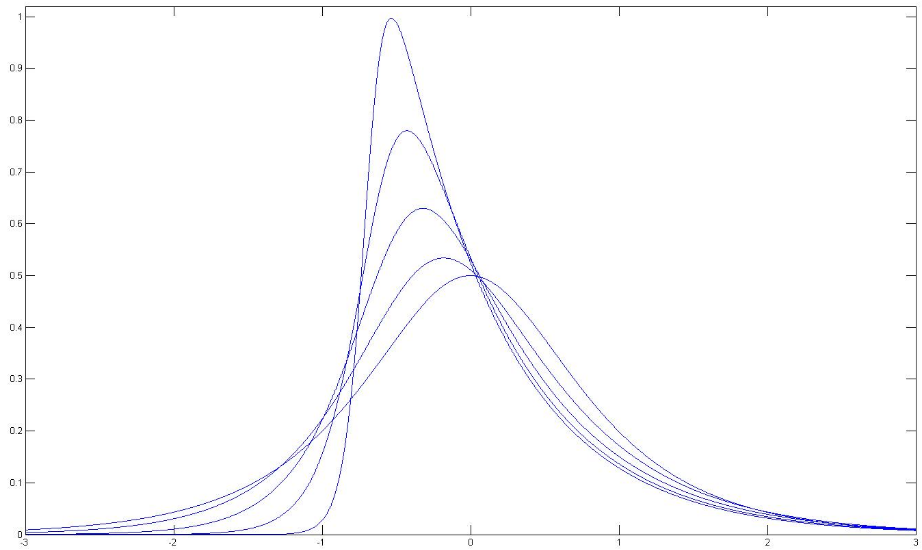

Figure 1: A family of probability density curves with fixed mean and variance, and varying skewness. As the skewness increases, the bulk of the density shifts negatively.

Figure 1 shows the effect of increasing skewness, while holding the mean and variance fixed. The peak of the distribution shifts to the left. The effect of this shift is that random draws from the distribution will generally be of lesser value, and will be balanced only when an extreme event happens. Few rational investors would willingly opt for a strategy that underperforms their peers during times of market stability. In fact, fund managers are routinely fired for such underperformance.

There is a similar tradeoff between kurtosis and variance. Holding the variance constant while increasing the kurtosis, making the tails of the distribution fatter, will result in a portion of the probability mass being concentrated in the center of the distribution. This means that the fatter tails must be balanced by a less variable center to keep the variance constant. Figure 2 illustrates a family of probability density curves with fixed mean, variance, and skewness. The kurtosis is different for each of the five curves. The most fat-tailed distribution here is the most peaked in the center. The differences in the tails are hardly visible in this chart here, but would be clear under magnification. The most visible effect of increasing the kurtosis, all other things held constant, is the gathering of probability mass in the center of the distribution. Most investors would balk at the increase in risk inherent in fattening the tails of the return distribution without narrowing the center.

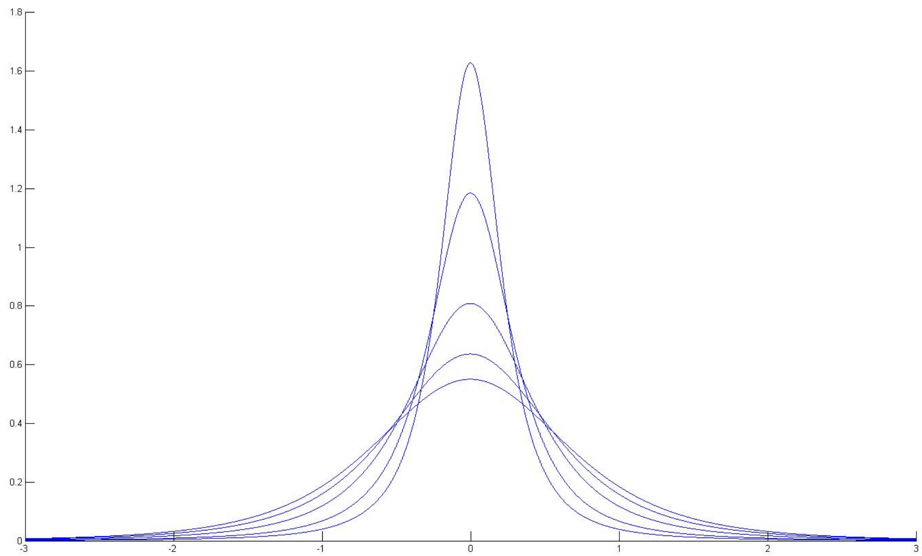

Figure 2: A family of probability densities with fixed mean, variance, and skewness. The kurtosis is different for each curve, and the most leptokurtotic distribution has the sharpest peak in the center.

The message here is twofold: higher moments are badly estimated, and there is no "free lunch." That is to say, increasing the skewness or kurtosis also has a substantial effect on the behavior of the main probability mass in the center of the distribution, not just on the tails. Furthermore, attempting to estimate higher moments from samples may substantially increase error in estimating the lower moments. Strangely enough, samples from the normal distribution may exhibit substantial skewness and/or kurtosis. Tables 1 and 2 show the ranges of skewness and kurtosis for samples of normal variates, and we can see that even samples of size 1000 produce a range of estimated skewness and kurtosis.

Skewness Variation

N:	20	40	60	90	135	200	300	450	667	1000
0.005	-1.3138	-0.9845	-0.8162	-0.6465	-0.5626	-0.4600	-0.3669	-0.2965	-0.2547	-0.2047
0.010	-1.1258	-0.8743	-0.7274	-0.5846	-0.4994	-0.4128	-0.3325	-0.2686	-0.2271	-0.1870
0.025	-0.9485	-0.7082	-0.6011	-0.4948	-0.4164	-0.3391	-0.2704	-0.2323	-0.1854	-0.1554
0.050	-0.7768	-0.5807	-0.5007	-0.4060	-0.3447	-0.2816	-0.2254	-0.1901	-0.1542	-0.1303
0.100	-0.5906	-0.4377	-0.3852	-0.3113	-0.2689	-0.2198	-0.1721	-0.1458	-0.1204	-0.1017
0.250	-0.3079	-0.2294	-0.2030	-0.1582	-0.1383	-0.1133	-0.0900	-0.0769	-0.0649	-0.0522
0.500	-0.0067	-0.0023	-0.0004	0.0000	-0.0010	0.0004	0.0032	-0.0018	0.0004	0.0001
0.750	0.3047	0.2370	0.2008	0.1662	0.1344	0.1125	0.0937	0.0749	0.0621	0.0524
0.900	0.5910	0.4485	0.3833	0.3189	0.2616	0.2171	0.1837	0.1459	0.1219	0.0994
0.950	0.7631	0.5926	0.4956	0.4059	0.3340	0.2787	0.2328	0.1833	0.1559	0.1256
0.975	0.9392	0.7098	0.6027	0.4907	0.4026	0.3338	0.2793	0.2207	0.1867	0.1492
0.990	1.1572	0.8652	0.7401	0.5827	0.4825	0.3989	0.3369	0.2670	0.2199	0.1777
0.995	1.3034	0.9722	0.8448	0.6629	0.5463	0.4494	0.3759	0.2970	0.2423	0.1970

Table 1: A table of quantiles of sample skewness of datasets drawn randomly from the standard normal distribution. The theoretical value for the skewness is zero.

Kurtosis Variation

N:	20	40	60	90	135	200	300	450	667	1000
0.005	1.5953	1.8343	1.9707	2.1085	2.2410	2.3305	2.4268	2.5067	2.5888	2.6473
0.010	1.6408	1.8989	2.0293	2.1600	2.2822	2.3752	2.4657	2.5492	2.6231	2.6842
0.025	1.7359	1.9836	2.1154	2.2403	2.3561	2.4407	2.5321	2.6008	2.6695	2.7228
0.050	1.8296	2.0706	2.2072	2.3098	2.4249	2.5096	2.5884	2.6536	2.7133	2.7623
0.100	1.9565	2.1858	2.3128	2.4140	2.5107	2.5879	2.6611	2.7135	2.7644	2.8072
0.250	2.1973	2.4018	2.5238	2.6098	2.6784	2.7397	2.7891	2.8253	2.8569	2.8881
0.500	2.5494	2.7141	2.8024	2.8599	2.8945	2.9327	2.9514	2.9622	2.9706	2.9850
0.750	3.0482	3.1364	3.1693	3.1700	3.1657	3.1614	3.1335	3.1215	3.1060	3.0946
0.900	3.6612	3.6510	3.6040	3.5304	3.4782	3.4095	3.3384	3.2859	3.2408	3.1973
0.950	4.1651	4.0322	3.9399	3.8019	3.6861	3.5808	3.4763	3.3876	3.3237	3.2649
0.975	4.6284	4.4450	4.2829	4.0649	3.9275	3.7481	3.6067	3.4852	3.4053	3.3282
0.990	5.3260	5.1515	4.7243	4.4479	4.2794	4.0119	3.7878	3.6203	3.5199	3.4075
0.995	5.9194	5.6762	5.1227	4.6899	4.4963	4.2067	3.9446	3.7212	3.6117	3.4799

Table 2: A table of quantiles of sample kurtosis of datasets drawn randomly from the standard normal distribution. The theoretical value for the kurtosis is 3.0.

Modeling skewness and kurtosis is not intrinsically wrong. There are models which, when carefully applied, will yield beneficial results when there is enough data to support them. However, it is important to get the first two moments, namely mean and variance, right before moving on to consider higher moments. Usually in investment applications information content in the sample data is simply not enough to consider much more than the first two moments.

5. Extreme Value Analysis and Conditional Value at Risk

Conditional Value at Risk, or CVaR, is defined as the mean of the upper (or lower) portion of a sample. Since this statistic is a descriptive measure of tail behavior its use in finance has been gaining popularity (e.g. Embrechts *et al.*, 1997; Embrechts, 2000; Cakici and Foster, 2003; Einmahl *et al.*, 2005; Jorion, 2007; Dowd *et al.*, 2008). CVaR is also known as "expected shortfall" (Acerbi *et al.*, 2001). For example, in a sample of 100 points, the $CVaR_{95}$ is defined as the average of the five points with the largest values. This statistic can be a useful measure of tail behavior. However, there are problems with using the $CVaR_{95}$ to fit models and optimize portfolios. First, it may not directly measure the quantity of interest. Consider two datasets of 100 points:

1. Lower 95 points have -10% return and the upper 5 points have +15% return.
2. Lower 95 points have +10% return and the upper 5 points have +15% return.

Clearly (2) would be preferable for any rational investor even though both datasets have the same CVaR. Additionally, even though CVaR is being used only to fit the tail of the distribution, it may not adequately summarize utility to the investor. Five spread out points or five equal points as the top five percent in a dataset of 100 points would have equal $CVaR_{95}$, and it is not clear that they should all be considered equally in terms of risk. Again, CVaR can be a meaningful descriptive summary, but it collapses too much of the information in the data to be useful for fitting many models or as a surrogate for approximating many expected utility functions.

Extreme events occur from time to time even under relatively conservative risk models. By nature, extreme events are unusual events. Eliminating risk altogether is impossible when investing; attempts to remove risk due to extreme downside events will generally be paid for in lower expected returns. Studying these events and modeling them statistically is helpful when the knowledge can be combined effectively into the entire system without distorting the other parts of the analysis. Separate analyses on the outliers and the central values of a distribution are likely to get both parts wrong if they are not done well. Further information on the history and practice of modeling extreme values in finance can be found in Dehann (2006) and Nawrocki (1999).

6. "Correlations Go to One"

It has been repeated often recently in the world of finance that assets become highly correlated during extreme events. All assets move up and down together. This phenomenon can be observed during turbulent times in the market, and it leads investors to fear that diversification benefits are disappearing. However, a thoughtful analysis may shed some light on this phenomenon and reduce the fear associated with this statement. Returns can be thought of as the net effect of many component forces. Sometimes certain system-wide effects may dominate all the returns of a set of assets. These systemic shocks to the returns, because of their large magnitudes, will indeed produce a large measured correlation close to one when the assets' returns are considered together. However, net of this dominating effect, there are idiosyncratic risk factors within assets which will be once again observable after the storm. When datasets are corrected for systematic effects, the "correlations go to one" effect will largely disappear. In other words, residual returns from suitable risk models do not show correlations going to 1 during extreme market volatility. It should also be mentioned that this effect applies equally to upmarkets as it does to downmarkets. Investors are probably less likely to complain about the systematic effects when they benefit from them.

Standard financial models such as the Capital Asset Pricing Model (CAPM)(Treynor, 1961; Treynor, 1962; Lintner, 1965; Sharpe, 1964; Black, Jensen, and Scholes, 1972; French, 2003) and Arbitrage Pricing Theory (APT) (Ross, 1976) do an excellent job of explaining the effects of systematic risk. Both predict that, even in the presence of such risk, the best strategy remains a diversified portfolio. The fact that certain systematic market risks dominate the observed returns has little impact on the optimal strategy. Assuming that it is impossible to predict the market, everyone suffers (or benefits) when there is a shock to the market. It remains advisable to be diversified simply because a non-diversified portfolio is making active bets, leaving the investor exposed to diversifiable risk. In the absence of good information motivating these bets, it is not sensible to make them.

7. Normal Models and the Importance of Mean and Variance

Quantitative managers must care about mean and variance of their portfolios. This is because the mean and variance are excellent summaries of the center and precision of a probability distribution. They use information from all the data. Using only part of the data is only advisable when that portion of the data contains the only relevant information for the analysis. When analyzing noisy financial data it is wise to use as much of the data as possible. When the mean and variance are of interest, as is usually the case with financial data, it makes sense to use the normal distribution in statistical models for the data, since the mean and variance are the natural parameters of the normal distribution.

Normal models are useful even if they are wrong. Normal models are commonly used to test for linear relations among observed variables. Although the normal distribution itself is not necessary to extract correlations or regression coefficients themselves, the traditional hypothesis tests for these calculations rely on the normal model for residuals. It should be noted that with a good regression model, a non-normal or simply sparsely or unevenly observed input distribution can produce a highly non-normal response or output variable in spite of the normal model framework. A vast literature exists on techniques to improve regression models to be more resistant to outliers; moreover, software tools are extremely well-researched and provide excellent feedback so that the results from these models can be considered carefully. Thoughtful analysis with good software tools can produce useful results. Software for non-normal models is also available, but the tools for analyzing output and diagnosing potential problems are much more experimental and may not produce stable or valid results. Many convenient properties possessed by the normal distribution do not carry over to more complicated distributions, so much more thorough model validation and diagnosis is required for non-normal models.

Non-normal models may be the right solution for certain problems and certain cases, but they should not be haphazardly or needlessly employed. They require more care on the part of the analyst, and it usually makes sense to use them only after applying and rejecting the normal models. Many times the normal models can be adapted to unusual circumstances, thus reducing the need to resort to more experimental non-normal models.

8. Gaussian Copulas: Error-prone Calculations on Non-Normal Data Using Normal Distribution Functions

A class of non-normal models which has become fashionable lately is based on Gaussian copulas (e.g. Sklar, 1959; Ruschendorf, 1985; Nelsen, 2006; Friend and Rogge, 2005). These copulas are essentially correlation matrices. Analysis can be based on multivariate normal distributions, and the results can be run through transformations to convert the normal marginal and/or conditional distributions of the individual variables to non-normal distributions. A problem with this type of analysis is that the normal distribution is the only distribution where the marginal and conditional distributions of the individual variables are also normal. This means that the transformation to non-normal negates the validity of the analysis. The errors created by doing this transformation are compounded when some variables are far away from center, and the more variables jointly analyzed using the copulas, the greater the chance of this type of error occurring. Unfortunately, copulas are often used as a technique for including correlations in extreme value analysis, where the errors created by the nonlinearity of the transformation to non-normality are at their greatest. Copulas are probably most effectively used as a tool for simulating jointly distributed non-normal variables, but the simulations should not be expected to obey the properties of a normal distribution, and the correlations among sets of variables will also be distorted by the transformations, especially in the tail areas. These copula analyses may be partly responsible for the popularity of the “correlations go to 1” mantra, since the incorrect analysis of the nonlinear relationships among variables could lead an analyst to underestimate the joint probability of failure in two separate situations, and thus believe that the correlations must be high in order to observe the systematic failure of so many situations at once. Copulas have been widely used to price derivatives (Burtschell *et al.*, 2009) and can be particularly wrong in estimating joint tail probabilities, especially when estimating joint default probabilities on collateralized debt-based securities (Li, 2000; Jones, 2009; Whitehouse, 2005; Andersen, 2007; Embrechts *et al.*, 2002; Laurent and Gregory, 2005).

9. When Non-Normal Models are Appropriate

Certain situations absolutely necessitate the use of non-normal distributions. The study of quantities that naturally follow certain distributions absolutely warrants the use of those natural distributions in statistical models. In fact, whenever anything is known to follow a certain distribution, it makes sense to use that distribution in a model. Although financial returns are known not to follow normal distributions, (Fama, 1965; Rosenberg, 1974; Rosenberg and Ohlson, 1976; Fama and French, 1992; Figelman, 2009; Richardson and Smith, 1993, and many others) they do not follow any particular other family of distributions either. Although many practitioners have argued for modeling returns plus one as lognormal (e.g. Aitchison and Brown, 1957; Elton and Gruber, 1974), there are considerable reasons why this model cannot be exactly true. Not the least of these reasons is that there must be a probability mass at minus infinity to capture the nonzero population of companies or funds which go out of business, thus having a -100% return and a negatively infinite logarithm of one plus return. Others have modeled returns using non-normal stable Paretian distributions, e. g., Mandelbrot (1963), but nobody has presented conclusive evidence that returns actually do follow these distributions, which were chosen more for their convenient properties of closure under convolution (Levy, 1937) than for accurately describing returns. Rather than directly changing the return distribution, a sensible way to better results is to model the known sources of variation rather than to substitute another incorrect distribution to directly model the data. If data exhibits skewness, then what is causing the data to skew? Investigating the subtleties of the data-generation mechanism is generally a much more successful way to model data than simply applying a different model directly to the observations.

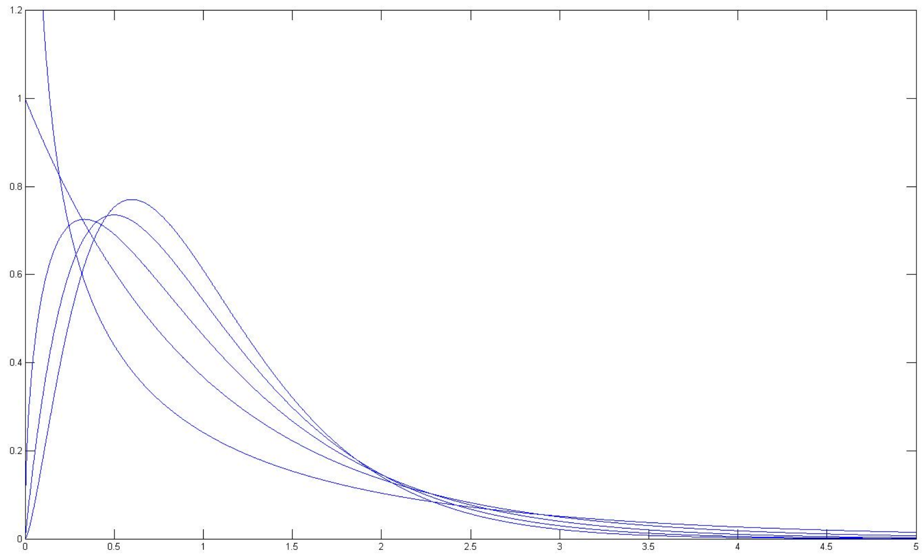

Figure 3: A family of scaled chi-squared distributions. This family arises as the distribution of variances in normal models. It features a highly skewed non-normal distribution with a bell-shaped normal curve as its limit.

Several non-normal probability laws arise directly from modeling data with normal distributions. For example, sample variances of normally distributed random variables have scaled chi-square distributions, which are a type of gamma distribution. Chi-square distributions and inverse chi-square distributions are used in Bayesian analyses for variance parameters since they arise naturally for normal models (Gelman, Carlin, Stern, and Rubin, 1995). The results of such analyses will also have inverse chi-square distributions. Both chi-square and inverse chi-square distributions are bounded below, skewed positive, and have greater excess kurtosis than a normal distribution, yet they arise naturally as a result of a normal model for the data. Similarly, T distributions, F distributions, and Beta distributions arise from simple functions of normal random variables. "Non-central" versions of these distributions also occur naturally as the distributions of test statistics when the null hypothesis is false. All of these distributions are familiar and rather well known to the student of basic statistical models. They all also include normal distributions as limiting cases. Models can add complexity to all of these distributions though the processes of nonlinear transformation, convolution (adding together random variables), mixture (using random parameters), or constraint (limiting the allowable values). The combination of these processes can yield almost any shape for a distribution. A good example of a complex model generated from simple components is the Michaud Resampled Efficient Frontier produced with New Frontier Advisors' Optimizer. By constraining, mixing over different parameter values, and transforming over the nonlinear function of producing Markowitz Efficient Frontiers, a complicated result which has quite non-normal features is generated from simple normally-distributed components. Although non-normal distributions can also be used to model the data, they are not necessary in order to account for non-normal features in the output.

More complicated distributions have been proposed to directly model data as skewed and/or heavy tailed. Some examples of these distributions are the Skew-normal, Skew-T, Pearson type IV, Inverse-Gaussian. The complete list is infinite since there are infinitely many possible non-normal distributions, of which only a few have been named or studied. Models using these probability laws may have desirable properties in certain situations but they are often little-known and not well-researched. They should be considered when there is a good reason why they should be better than the standard set of models.

10. Conclusions

It is easy to be overwhelmed by everything encompassed by "non-normality," since the term is only specific about what it is not. It almost seems deliberately chosen to describe what is not understood and make excuses for poor past performance. It is hoped that the reader of this article will gain some perspective and look on claims of model improvement simply through rejecting normal distributions with some skepticism.
Now as much as ever, it is important to stay on course with a balanced investment strategy and not react irrationally to recent market turmoil. Changing mathematical models to arbitrarily more complex ones could be a big mistake for an investor. Careful, sensible analyses which balance many outcome scenarios will win on the average when played out over future, out-of-sample returns. When it comes to marketing materials designed to impress and intimidate in an attempt to explain or conceal poor performance, *caveat emptor.*

References

Acerbi, C., Nordio, C., and Sirtori, C. (2001), "Expected Shortfall as a Tool for Financial Risk Management," Working paper, http://www.gloriamundi.org/var/wps.html.

Aitchison, John and Brown, James A. C. (1957), *The Lognormal Distribution with Special Reference to its Use in Economics*, Cambridge University Press.

Akaike, Hirotugu (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control* 19 (6): 716–723.

Andersen, L. (2007), "Main Extensions of the One Factor Gaussian Copula," *The Ultimate Guide to CDOs–Market, Application, Valuation, and Hedging*, RISK Books.

Anson, Mark, Ho, Ho, and Kurt Silberstein (2007), "Building a Hedge Fund Portfolio with Skewness and Kurtosis," *Journal of Alternative Investment,* 10(1): 25-34.

Bakshi, Gurdip, Kapadia, Nikunj, and Madan, Dilip (2003), "Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options," *Review of Financial Studies*, 16: 101–143.

Bhansali, Vineer (2008), “Tail Risk Management,” *Journal of Portfolio Management,* 34(4): 68-75.

Black, Fischer., Jensen, Michael C., and Scholes, Myron (1972), “The Capital Asset Pricing Model: Some Empirical Tests,” pp. 79-121 in M. Jensen ed., *Studies in the Theory of Capital Markets*. New York: Praeger Publishers.

Burtschell, X., J. Gregory, and Laurent, J.-P. (2009), “A Comparative Analysis of CDO Pricing Models under the Factor Copula Framework,” *Journal of Derivatives,* 16(4): 9-37.

Cakici, Nusret and Kevin R. Foster (2003), “Value at Risk for Rate-Dependent Securities,” *Journal of Fixed Income,* 12(4): 81-95.

Dowd, Kevin, Cotter, J., Humphrey, C., and Woods, M. (2008), “How Unlucky Is 25-Sigma?” *Journal of Portfolio Management,* 34(4): 76-80.

Einmahl, John H.J., Foppen, Walter N., Laseroms, Olivier W., and de Vries, Casper G., (2005), “VaR Stress Tests for Highly Non-Linear Portfolios,” *The Journal of Risk Finance*, 6(5): 382-387.

Elton, e. j., and Gruber, M. J. (1974), “Portfolio Theory when Investment Relatives are Lognormally Distributed,” *Journal of Finance*, 29(4): 1265-1273.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modeling Extremal Events for Insurance and Finance*. Springer, Berlin.

Embrechts, P. (2000), “Extreme Value Theory: Potential and Limitations as an Integrated Risk Management Tool.” Working paper, ETH Zürich.

Embrechts, P., McNeil, A.J. and D. Straumann (2002), “Correlation and Dependency in Risk Management: Properties and Pitfalls,” in *Risk Management: Value at Risk and Beyond* (ed. M. Dempster), 176–223, Cambridge University Press, Cambridge.

Fama, Eugene F. (1965), "The Behavior of Stock-Market Prices," *Journal of Business,* 38(1): 34-105.

Fama, Eugene F. and Kenneth French (1992), “The Cross-Section of Expected Stock Returns,” *Journal of Finance*, 47(2): 427-466.

Figelman, Ilya (2009), “Effect of Non-Normality Dynamics on the Expected Return of Options,” *Journal of Portfolio Management,* 35(2): 110-117.

French, Craig W. (2003), “The Treynor Capital Asset Pricing Model,” *Journal of Investment Management,* 1(2): 60-72.

Friend, A., and Rogge.E. (2005), “Correlation at First Sight,” *Economic Notes,* 34(2): 155–183.

Fuertes, Ana-Maria, Miffre, Joëlle and Wooi-Hou Tan (2009), "Momentum Profits, Nonnormality Risks and the Business Cycle," *Applied Financial Economics*, 19(12): 935-953.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (1995), *Bayesian Data Analysis.* Chapman and Hall, (2nd ed. 2007).

Harvey, Campbell R., Liechty, J., Liechty, M. W., and Mueller, P. (2004), "Portfolio Selection With Higher Moments," (Duke University, Working Paper).

Harvey, Campbell R., and Siddique, A. (2000a), "Conditional Skewness in Asset Pricing Tests," *Journal of Finance*, 55: 1263-1295.

_______ (2000b), "Time-Varying Conditional Skewness and the Market Risk Premium," *Research in Banking and Finance,* 1: 27-60.

de Haan, L. and Ferreira, A. (2006), *Extreme Value Theory: An Introduction.* Springer, Berlin.

Jondeau, Éric, Poon, Ser-Huang, and Rockinger, Michael (2007), *Financial Modeling under Non-Gaussian Distributions,* Springer.

Jones, Sam (2009), "The Formula that Felled Wall Street," *The Financial Times,* April 24 2009, 22:04, http://www.ft.com/cms/s/2/912d85e8-2d75-11de-9eba-00144feabdc0.html.

Jorion, P. (2007), *Value at Risk.* 3rd Ed. McGraw Hill, New York.

Kat, Harry M., and Joëlle Miffre (2008), "The Impact of Non-normality Risks and Tactical Trading on Hedge Fund Alphas," *Journal of Alternative Investments,* 10(4): 8-21.

Laurent, J.-P., and Gregory, J. (2005), "Basket Default Swaps, CDOs and Factor Copulas," *Journal of Risk*, 7(4): 103–122.

Levy, H. and Markowitz, H. M. (1979), "Approximating Expected Utility by a Function of Mean and Variance," *The American Economic Review*, 69(3): 308-317.

Levy, Paul (1937), *Theorie de L'addition des Variables Aleatoires,* Paris: Gauthier-Villars.

Li, David X. (2000), "On Default Correlation: A Copula Function Approach," *Journal of Fixed Income* 9: 43-54.

Lintner, John (1965), "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets," *Review of Economics and Statistics*, 47(1): 13-37.

Mandelbrot, Benoit B. (1963), “The Variation of Certain Speculative Prices,” *Journal of Business,* 36: 394-419. Reprinted in Cootner, Paul H. (1965), *The Random Character of Stock Market Prices,* Cambridge: MIT Press.

Markowitz, H. M. (1959), *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley and Sons. Reprinted by Yale University Press, 1970, 2nd ed. Basil Blackwell, 1991.

Martin, George A., and Spurgin, Richard (1998), “Skewness in Asset Returns: Does It Matter?” *Journal of Alternative Investments,* 1(2): 66-75.

McDonald, B. D. and Affleck-Graves, J. (1989), “Nonnormalities and Tests of Asset Pricing Theories,” *Journal of Finance,* 44(4): 889-908.

Michaud, Richard O. (1989), “The Markowitz Optimization Enigma: Is 'Optimized' Optimal?” *Financial Analysts Journal,* 45(1): 31-42.

________ (2003), “A Practical Framework for Portfolio Choice,” *Journal of Investment Management,* 1(2): 14-29.

Michaud, Richard O. and Robert O. Michaud (2008a), *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation.* 2nd ed., Oxford University Press, USA.

________ (2008b), “Estimation Error and Portfolio Optimization: A Resampling Solution,” *Journal of Investment Management,* 6(1): 8-28.

Mills, T. C. (1995), “Modeling Skewness and Kurtosis in the London Stock Exchange FT-SE Index Return Distributions,” *The Statistician,* 44(3): 323-332.

Nawrocki, David N. (1999), “A Brief History of Downside Risk Measures,” *Journal of Investing,* 8(3): 9-25.

Nelsen, R.B. (2006), *An Introduction to Copulas,* 2nd Edition. Springer, New York.

Peiro, A. (1999), “Skewness in Financial Returns,” *Journal of Banking and Finance*, 54: 67-121.

Richardson, Matthew and Smith, Tom (1993), "A Test for Multivariate Normality in Stock Returns," *Journal of Business,* University of Chicago Press, 66(2), 295-321.

Rosenberg, Barr (1974), “Extra-Market Components of Covariance in Security Returns,” *Journal of Financial and Quantitative Analysis*, 9(2), 263–273.

Rosenberg Barr, and Ohlson, James (1976), “The Stationary Distribution of Returns and Portfolio Separation in Capital Markets: A Fundamental Contradiction,” *Journal of Financial and Quantitative Analysis,* 11(3): 393-402.

Ross, Stephen (1976), "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory* 13(3): 341–360.

Rubinstein, Mark E. (1973), "A Comparative Statics Analysis of Risk Premiums," *The Journal of Business,* 46(4), 605-615.

Rüschendorf, L. (1985), "Construction of Multivariate Distributions with Given Marginals," *Annals of the Institute of Statistical Mathematics* 37: 225–233.

Schwarz, Gideon E. (1978), "Estimating the Dimension of a Model," *Annals of Statistics,* 6(2), 461–464.

Sears, R. S., and Wei, K. (1985), "Asset Pricing, Higher Moments, and the Market Risk Premium: A Note," *Journal of Finance* 40: 1251-1253.

Sharpe, William F. (1964), "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19 (3), 425-442.

Sheikh, A. Z., Qiao, H., et al. (2009), "Non-normality of Market Returns: A Framework for Asset Allocation Decision-Making," White paper, J. P .Morgan Asset Management, JPMorgan Chase & Co.

Sklar, A. (1959), "Fonctions de Répartition à n Dimensions et Leurs Marges," *Publications de l'Institut de Statistique de L'Université de Paris* 8: 229-231.

Sortino, Frank A., and Price, Lee N. (1994), "Performance Measurement in a Downside Risk Framework," *The Journal of Investing* 3(3): 50-58.

Taylor, Stephen J., Yadav, Pradeep K., and Zhang, Yuanyuan (2009), "Cross-Sectional Analysis of Risk-Neutral Skewness," *Journal of Derivatives* 16(4): 38-52.

Treynor, Jack L. (1961), "Market Value, Time, and Risk," Unpublished manuscript.

Treynor, Jack L. (1962), "Toward a Theory of Market Value of Risky Assets," Unpublished manuscript. A final version was published in 1999, in *Asset Pricing and Portfolio Performance: Models, Strategy and Performance Metrics*. Robert A. Korajczyk (editor) London: Risk Books, pp. 15-22.

Whitehouse, M. (2005), "How a Formula Ignited Market that Burned Some Big Investors," *The Wall Street Journal*, September 12, 2005.