Non-Normality Facts and Fallacies

David N. Esch, PhD
New Frontier Advisors
Article appearing in JOIM First Quarter 2010

Outline

• Introduction
• Normality, finance, and statistics
• Normality Pros and Cons
• Parsimony
• Higher moments
• Extreme value analysis
• Gaussian copulas
• Conclusion

Financial Data – A Statistician’s View

• Multivariate
• Dependent
• Data generation mechanisms always changing; need current information
• Only one replication available
• Noisy
• Non-normal
• Want a lot of results out of severely limited information.
Non-Normality and Investing

• Non-normality of return distributions matters because we want to invest properly.
• Portfolio construction depends on accurate knowledge of return distributions.
• Markowitz mean-variance frontier: optimality conditional on
  1. Normal return distribution OR quadratic utility
  2. No uncertainty in inputs.
     (Neither true in general)

Optimizers and Non-normality

• Non-normality is often invoked in finance due to condition (1) above.
• Levy-Markowitz (1979): mean-variance frontier nearly optimal for a wide variety of utility functions given expectation and variance of return.
• 2nd order Taylor approximation of U(R) about E(R) valid locally. If distribution of R is local then quadratic utility assumption plausible.
• Larger concern: condition (2) - optimality conditional on perfectly known inputs. Must account for estimation error when sample statistics are used.

Central Limit Theorem

• The Normal Distribution is the limiting distribution of many sums and averages.
• Applies generally if each summand is asymptotically negligible (not too big) and has finite variance.
• Applies to sums of non-identically distributed quantities.
• Some datasets will be close to normal because of the CLT, others less so.
Normal Distributions are a Stable Closed System

- Adding two normals gives a normal.
- Any portfolio of normals is also normal.
- Conditional distributions (regressions) of multivariate normals are normal.
- Marginal distributions of multivariate normals are normal.
- Normals can be broken into normal components.
- The normal distribution is the only distribution with this property.

Normal Models are Reliable

- Often used for smoothing
- Well-known, tractable mathematics – calculations simple
- Stable software widely available

Mean and Variance

- A normal distribution is defined by its mean (expected return) and variance (risk).
- Mean and variance map clearly to features (location, width of bell curve) of the density.
- Mean and variance are important investment considerations – expected return and risk.
Normal Models are Flexible

- Almost any study of financial returns shows that returns do not follow the normal distribution.
- Under many circumstances the normal analysis is suitable despite non-normality of data.
- Theory may justify non-normal models if procedures are sensitive to deviations from normality.
- Which non-normal probability model to use can be a difficult question whose answer may influence the result.

Non-Normality is Hard to Detect in Samples

- Frequentist Hypothesis testing
  - Anderson-Darling test
  - Cramér-von-Mises test
  - D'Agostino's $K^2$ test
  - Jarque-Bera test
  - Pearson's $x^2$ test
  - Lilliefors (Kolmogorov-Smirnov) test
  - Shapiro-Francia
  - Shapiro-Wilk etc.
- Bayesian Approach
  - Bayes Factors
- Better: Theoretical Justification for/against

Statistical Approaches to Non-Normality

- Many possibilities – must make specific choice.
- Data-based methods (e. g. resampling, bootstrapping) – may emphasize peculiarities of data.
- Model-based methods may be subject to misspecification.
- Challenge to estimate – by nature outliers are rare and highly error-prone.
- Be sure your approach makes sense.
**Error Distributions Should be Contained**

- Using a skewed or heavy-tailed error distribution reduces the explanatory power of the statistical model.
- We want the statistical model to explain as many interesting aspects of variation as possible, not the error, especially for forecasting.
- Error distributions cannot be predicted – it’s the part we throw away for forecasting.
- If error distribution can accommodate extreme values then there is less need for the model to fit the data.

---

**Parsimony**

- Heuristic: Simplest explanation is preferred.
- Guideline for scientific research
- In statistics, parsimony means fewer model parameters.
- Although statistical models with more parameters fit data better, generally perform worse out-of-sample.
- Non-normal distributions tend to introduce extra parameters to index the degree of non-normality.

---

**Overfitting**
Overfitting – 1 year out

Overfitting – 2 years out

Overfitting
Higher Moments: Basics

- **Skewness** is related to the (standardized) expected value of $X^3$, and **Kurtosis** to the expected value of $X^4$.
- Skewness measures the asymmetry of a variable, and kurtosis measures the fatness of the tails.
- Normal distributions can have any mean or (positive) variance, but all higher moments are fixed.
- Non-normal distributions can have many different values of skewness and kurtosis.

Skewness and Kurtosis: Hard to Estimate

- Higher moments are harder to estimate from data. More data is required for reliable estimates than for mean and variance.
- Outliers, when taken to higher powers, greatly influence the estimate.
- Sample moments should not generally be used for estimating models. Maximum likelihood (ML), Maximum a posteriori (MAP) methods are preferred.

Normal Data, Non-Normal Fit, N=60
A Good Case

Sample moments match true moments fairly well.

A Bad Case

Sample skewness and true skewness match poorly.

Normal Data, Non-Normal Fit, N=24
A Good Case, N=24

Fewer good examples with N=24

A Bad Case, N=24

Sample skewness and kurtosis badly estimated.

Leptokurtotic Data, N=60
Leptokurtotic Data, N=24

Sample skewness and kurtosis badly estimated.

Variation of Moment Estimates, N=120

10,000 simulations of N(0,1) data in each histogram
**Skew: Desirable for Investors?**

- It has been said that rational investors are averse to negative skew and should prefer positive skew, all else equal.
- If all else (e.g., mean, variance) is held equal, adding a positive outlier must be balanced by shifting some of the data negatively.

**Skewness**

- The graph below shows what happens when skewness is varied.
- As skewness increases, the bulk of the data shifts to the left.

**Impact of Positive Skew**

- Investors will occasionally have large returns but will pay for it the rest of the time.
- During tranquil periods, the negatively skewed investor will have better returns.
- The existence of bubbles would seem to indicate that real investors actually prefer negatively skewed returns. (Perhaps in denial of the negative tail)
### Kurtosis

- Changes in kurtosis cause opposing changes in the variance of the bulk of the data, under constant variance.
- Changes in the tail must be balanced in the central hump for variance to remain constant.
- In exchange for outliers, distributions with excess kurtosis have closer to constant centers.
- The variance is an effective risk measure – it captures the overall scale of the data.

**The graph below shows what happens when kurtosis is varied.**

- As kurtosis increases, the bulk of the data shifts to the center.

### Extreme Value Analysis (EVA)

- Descriptive, not predictive
- Analyzes only tail data
- Discards more stable central information
- One common statistic used in EVA is CVaR.
CVaR

- Conditional value at risk, a.k.a. expected shortfall.
- Defined as $E(X|X<X_q)$, where $X_q$ is the $q$ quantile of $X$ (e.g., 0.05).
- Insensitive to density within tail as long as expectation is preserved.
- Insensitive to anything outside of tail.
- Might be a more useful summary under strict assumptions, e.g., if all return distributions were normal with fixed mean.

In general, CVaR is unsuitable as a global summary of investment value – assets with equal CVaR can be priced quite differently.

The next slide shows three probability distributions with the same CVaR.

The shaded area represents the lowest 5% of the density function.
**Dependency and Correlation**

- Correlations measure **linear** dependencies.
- Real data may have nonlinear dependencies.
- Multivariate non-normal distributions may have complex dependency structures.
- $N$ variables have $O(N^2)$ correlation parameters; more may be required to model nonlinear relationships.
- To be useful, models estimated on smaller datasets need to simplify these relationships. (parsimony)

**Copulas**

- General non-independent variables:
  - Association not always linear
  - Individual variables can be non-normal
  - Difficult mathematically since conditional densities require integration
  - Normal distributions easy since always normal, just requires regression-like calculations
- Idea: transform a well-known distribution

**Gaussian Copulas**

- Problem: How to model dependency when distributions are non-normal.
- Simplification: separate transformations for each variable. (Transformation on marginals)
- Model transformed data as multivariate normal.
Copula Misfit Causes

- Data peculiarities, sparseness
- Wrong transformation
- True dependency nonlinear after transformation, i.e., non-normality after normalizing transformation.

Copula Example

- The following slides show an example that cannot where copulas go wrong.
- No amount of data can fix the problem – the model just can’t explain the data.
- Probability model:
  - $X \sim \text{Normal}$;
  - $Y|X \sim \text{Normal}(L_1(X), L_2(X)^2)$,
  - where $L_1()$ and $L_2()$ are linear functions.

Copula Example – Data Density

[Data density plot showing skewed $Y$ and normal $X$.]
Sample from Data Density

Fitting Procedure

- Transform to normal by their empirical distributions.
- Estimate correlation via maximum likelihood.
- Transform the normal back to the original scale.

Random Sample from Copula Distribution
Copula Sample on Original Scale

Comparison of Original and Copula Samples

Comparison of Original and Copula Densities
Gaussian Copulas

- Work well when
  - Copula model is good, simple dependence structure
  - Parameters are estimated well
  - Inference is in area where model fits
- Hazards
  - Copula model inadequate, like example
  - Parameters are badly estimated, scarce/bad information in data
  - Inference is in extremities of distribution where discrepancy is maximized

Conclusions

- Nobody believes return distributions are normal.
- Probability models are often deliberately imperfect descriptions of reality.
- Normal distributions are a sensible first-pass choice for many analyses, and are flexible enough to accommodate many complicated data structures.

Conclusions

- Estimates come with sampling error and do not represent precise knowledge.
- Ignoring sampling error in optimization is a greater cause of optimal portfolio instability than choice of normal/non-normal returns in many cases.
- Non-normal analyses should be handled with care.
Sources


Sources


Black Swans and Fat Tails
New Mathematical Descriptions of Investment Risk

Johann Klaassen, Ph.D., AIF®
johann@firstaffirmative.com
and
David H. Brown, Ph.D.
dbrown@coloradocollege.edu

Introductions

Johann A. Klaassen
BA, St John’s College (Santa Fe), Liberal Arts (Great Books Program)
MA & PhD, Washington University (St Louis), Ethics & Social Philosophy
Accredited Investment Fiduciary™, Center for Fiduciary Studies
Certificate in Financial Planning, Boston University
VP, Managed Account Solutions, First Affirmative Financial Network

David H. Brown
BA, St John’s College (Santa Fe), Liberal Arts (Great Books Program)
Post-Grad study in Mathematics, University of New Mexico
PhD, University of California, Davis, Applied Mathematics
Post-Doc Research, Dept. of Agronomy and Range Science at UC Davis
Asst Professor, Dept. of Mathematics and Computer Science, Colorado College

The Upshot

This is the culmination of more than 5 years of Investment Committee research, and about 9 months of research into the mathematics of the markets.

The upshot: We can show our clients a better – if scarier – depiction of the risks of their investment portfolios.
A Modicum of Mathematics

• A simple example of mathematical modeling
• Bank account balance changes over time due to deposits, withdrawals, interest:
  \[ \Delta \text{balance} = \text{deposits} - \text{withdrawals} + \text{interest} \]
• With enough information about each of these processes, we can predict how the balance will change over time

A Modicum of Mathematics

Some uses of mathematical models:
• Predict the future
• Explain the past or present
• Analyze what if scenarios
• Verify current understanding
• Compare to data, estimate unobservable processes
• Suggest new hypotheses
• Develop coherent theory

A Modicum of Mathematics

BUT:
• All mathematical models are wrong
• They should be evaluated based on whether they are *useful*
• Building a model involves making many decisions
  • Decisions are based on scientific insight, mathematical convenience, and purpose of the model
Risk Yesterday
Most discussions of risk in the markets begin with a simple statistical model of randomness: the bell-curve of "normal distribution" (aka the "Gaussian distribution"). This approach was suggested in 1900 by Louis Bachelier, and worked into Modern Portfolio Theory by Markowitz in the 1950s.

Bell Curves Everywhere
Similar ideas influenced Fama & French, Sharpe, Samuelson, Shiller, Malkiel, and many others.
CAPM, Black-Scholes, and MPT are all built on the foundation offered by the industry-standard depictions of the normal distribution.

IPS Diagram
FAFN’s own discussions of risk have, to date, been based on the normal distribution, too:
“The graphs are based on the probabilities of total return outcomes for various periods. Normally, outcomes can be expected to occur between +1 and -1 standard deviation 68% of the time, and between +2 and -2 standard deviations 95% of the time.”
Index Returns
Normal distributions are used in MPT in part because they are easy to work with; in an era before powerful computers this was a serious issue.

These two diagrams show annualized monthly returns for two key market measures: don’t they look “normal”?

Swans & Tails
But the bear market that began in early 2000 made a lot of us think about how we present the risks inherent in any kind of investment – to ourselves, and to our clients.

This was our first try.

Research Begins

Introductions Risk Yesterday Swans & Tails Risk Today Implications Conclusions

Introductions Risk Yesterday Swans & Tails Risk Today Implications Conclusions

Introductions Risk Yesterday Swans & Tails Risk Today Implications Conclusions

Introductions Risk Yesterday Swans & Tails Risk Today Implications Conclusions
Interesting, but Less Useful

Bienhocker suggests that the markets move in “Lévy flights”: random small fluctuations, punctuated with large jumps.
Tick by Tick

On very short time-scales, the Lévy flight approach appears to work—the distribution is a good fit for a power law distribution, as Mandelbrot suggested.

Introductions Risk Yesterday Swans & Tails Risk Today Implications Conclusions

Swans & Tails

Four Bad Bear Markets: Adjusted for Inflation

12 Months – S&P 500

24 Months – S&P 50

Serious Error

But: We’ve been making a fundamental error. We’ve used historical data to tell us what the mean return and standard deviations are, then tried to apply those to the historical data. Of course these diagrams don’t look wild: we’re evaluating their past behavior using the parameters we derived from their past behavior.
Paint the Target, Then Shoot

What if, we thought, we were investment advisors in the early 1960s? How might we have explained risk to our clients then? We began taking ten-year backward-looking analyses, and projecting them forward for 6-, 12-, 24-, and 48-month periods. Then we compared those projections to the actual data, and looked to see how often and by how much we were wrong.

Another Look at Index Returns

These two diagrams show annualized monthly returns for two key market measures: don’t they look “normal”?

They’re not. They’re skewed and leptokurtic.

Not Skin Diseases

Upper Left: Leptokurtosis
Upper Right: Fat Tail
Lower Right: Skew
Risk Today

We evaluated 8 different indices, over as wide a variety of time periods as possible, and averaged the results together:

<table>
<thead>
<tr>
<th>“Normal” Actual Multiple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Better than +3 SD: 0.14% 0.5% ~3.7x</td>
</tr>
<tr>
<td>Better than +2 SD: 2.3% 3.0% ~1.3x</td>
</tr>
<tr>
<td>Better than +1 SD: 16% 15.1% ~0.9x</td>
</tr>
<tr>
<td>Within 1 SD of Mean: 68% 58.7% ~0.9x</td>
</tr>
<tr>
<td>Worse than -1 SD: 16% 26.2% ~1.6x</td>
</tr>
<tr>
<td>Worse than -2 SD: 2.3% 11.3% ~4.9x</td>
</tr>
<tr>
<td>Worse than -3 SD: 0.14% 3.3% ~23.7x</td>
</tr>
</tbody>
</table>

Weaknesses of This Approach

• Data-hungry, so we don’t really have as much data as we’d like from newer indices
• Based on very simplistic portfolio projections (no correction for CPI, etc.)
• In principle, the probabilities of extreme events are hard to estimate, since they are by definition rare
• A significant part of the probability of large negative returns comes from what happened in 2008; if we had conducted this study a couple of years ago, we would have gotten somewhat different results

Strengths of This Approach

• Based on actual data, not simply the convenience of the normal distribution
• Different data is used for fitting model parameters and estimating probabilities
• Fairly robust w.r.t. window size
• The pattern in the results agrees with obvious facts about the skew and kurtosis of a wide range of asset classes
Implications

- Mathematical theory (the Central Limit Theorem) predicts that a large enough sum of independent, identically distributed random variable will always converge to the normal distribution.
- But the history of market returns has ended up pretty far from the normal distribution.
- So we can infer that either (a) we don't have enough observations, or (b) the series of market returns does not represent an "independent, identically distributed random variable".

More Implications

- Our figures should not be used blindly; they are themselves based only on a simple statistical analysis of historical data, and are somewhat sensitive to the details of our approach.
- The key is for advisors to know where these risk estimates come from; just as we can't predict the expected rate of return with certainty, we also can't predict the risk of large deviations with certainty just from the mathematical models.
- This is by no means a definitive study, but it should help motivate the modelers who generate the portfolio projections to think harder about how to truth-test their predictions, particularly the downside risks.

A Familiar Graph, Changed
Conclusions

Remember:
- All mathematical models are wrong
- They should be evaluated based on whether they are useful

No simple rule of thumb emerges that is valid across all asset types and time periods, but it appears that the risk of large downward movements (3 standard deviations) over short time periods (up to 4 years) is nearly 20 times as high as what is implied by the normal distribution.

Financial advisors need to adjust upwards the risk of large negative returns that they report to clients, and simultaneously to emphasize the difficulty of assigning precise probabilities to such risks.