Financial Risk Measurement and Joint Extreme Events:
The Normal, Student-t, and Mixture of Normals

Draft, 29 January 2014

tscoleman@uchicago.edu

ABSTRACT

We all know that the normal distribution does a poor job of representing the tails of the distribution for financial returns or P&L in the univariate case – observed distributions have fat tails. What receives less attention is that for a joint normal distribution events in the tails look as if they are independent: Extreme events will not occur together, whatever the correlation (except for the boundary cases of ±1). This has important implications because it is precisely the occurrence of joint losses in multiple assets that are most important for generating large overall losses. And it is generally accepted that, empirically, financial assets do exhibit joint extreme behavior.

This paper has two objectives. First, to discuss and compare the behavior of four multivariate distributions: the normal, Student-t, two-point mixture of normals, and meta normal / Student-t (normal copula with Student-t marginals). For exposition I focus on the bivariate case. The conclusion is that the Student-t and the two-point mixture of normals both exhibit or approximate tail dependence in a way that the joint normal and meta normal / Student-t do not. The second objective is to examine, in a limited manner, the empirical behavior for some representative financial assets and compare these with the four distributions discussed.

Joint distributions of financial returns or P&L form the foundation for quantitative risk measurement, necessary for calculating portfolio measures such as the value at risk (VaR) or the contribution to risk. Joint normality has the benefit that the sum of normals remains normal (closure under convolution), providing substantive computational benefits. Assuming normality is problematic, however, among other reasons because of the tail issues discussed here. My tentative conclusion is that a simple mixture of normals may be sufficient to approximate fat tails while still retaining the analytical and computational benefits of the normal distribution.

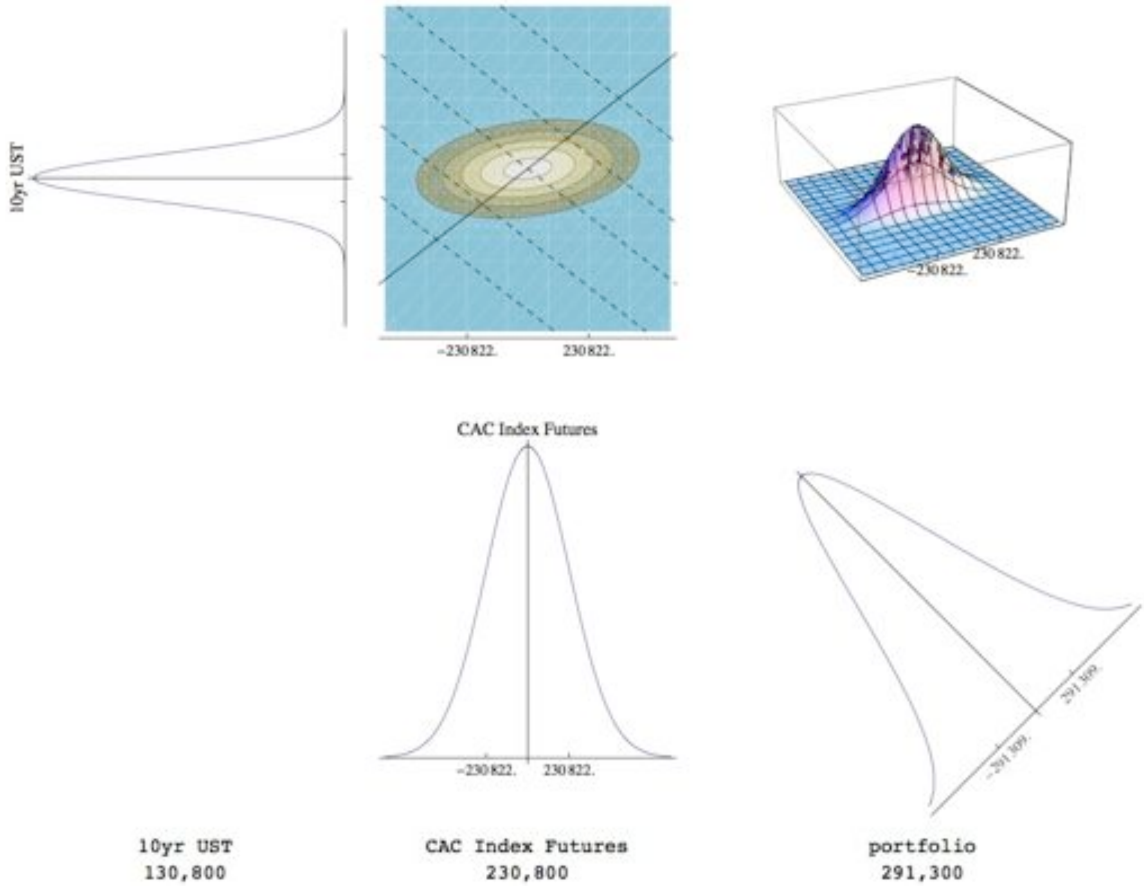
1 – **INTRODUCTION – The Role of Normality in Quantitative Risk Measurement**

Before turning to the central purpose of this paper let me take a short digression to highlight why normality is so often used in quantitative risk measurement.

The goal of risk measurement is to know the distribution of returns or profit and loss (P&L). This is a rather bold statement but it is also fundamentally true. If we know the P&L distribution we have virtually everything we need to know for risk measurement. The commonly used risk measures such as volatility, value at risk (VaR), expected shortfall, together with measures of tail events and extreme values – these are all summary measures or descriptions of one aspect or another of the distribution. Contribution to risk and other risk portfolio tools that aim to uncover the sources of risk all seek to describe how the portfolio distribution depends on the component asset distributions – again the goal is to know the P&L distribution.

Assuming normality for underlying asset distributions is popular because it provides considerable computational simplification in calculating and using the P&L distribution. A simple example will help clarify the role normality plays in risk measurement, and how and why normality provides such simplification and insight. Take a portfolio of $20 million notional of a US 30-year treasury bond and €7 million notional of CAC equity index futures. Figure 1 shows everything graphically. Our primary focus is on the portfolio P&L distribution, represented in the lower right. When we have this distribution we can calculate the volatility, the VaR, expected shortfall, probability of extreme events – indeed anything we want to know (probabilistically) about the portfolio P&L.

Figure 1 – Joint Distribution for $20mn US Treasury and €7mn CAC Index Futures



The portfolio P&L is the random variable:

$$T = X + Y$$

where

X = P&L for CAC equity index position (converted to dollars)
Y = P&L for UST bond position
T = P&L of portfolio, simply the sum of the component assets

Generally we do not know the portfolio distribution directly – we have to calculate it from the joint distribution for the component assets. If we have the joint distribution or density function

Joint density: *P[CAC P&L=x & UST P&L=y]*

then the probability density for the portfolio P&L is obtained by integrating the joint density under the condition that $x+y=t$ or $y=x-t$:

(1) Portfolio density = *P[Portoflio P&L=t] =*

$$\int_{x=-\infty}^{x=+\infty} P\left[CAC\ P\ \&\ L = x\ \&\ UST\ \ P\ \&\ L = x-t\right]dx$$

We can think of this as the projection of the joint distribution along the 45° diagonal. This is analogous to obtaining the marginal CAC density by projecting along the horizontal axis – integrating out the UST P&L:

(2) CAC density = *P[CAC P&L=x]* = $\int_{y=-\infty}^{y=+\infty} P\left[CAC\ P\ \&\ L = x\ \&\ UST\ \ P\ \&\ L = y\right]dy$

Alternatively the portfolio density is the convolution of the two densities:

Marginal CAC density:	*P[CAC P&L = x]*
Conditional UST density:	*P[UST P&L = y | CAC = x]*

Portfolio density = *P[Portoflio P&L=t] =*

$$\int_{x=-\infty}^{x=+\infty} P\left[CAC\ P\ \&\ L = x\ \right]P\left[UST\ \ P\ \&\ L = x-t \mid CAC = x\right]dx$$

However we may write the operation, deriving the portfolio density from a general multivariate density is a difficult operation.

But let us take another step back. We usually do not know the joint density directly but derive it by combining the marginal distributions using some observations and assumptions about the comovement between the UST and the CAC index. In practice we build the joint density from the marginals rather than calculate the marginal from the joint as in equation (2).

In the general case the process of building the joint density and then calculating the univariate portfolio P&L density is computationally difficult. As a practical matter Monte Carlo simulation is used. If we assume normality, of course, everything is simple;[1] since the sum of normal random variables is itself normal, the convolution is easy and well-known. The portfolio distribution will be normal with mean and variance:[2]

Portfolio mean: $\mu_p = \mu_x + \mu_y$
Portfolio variance: $\sigma^2_p = \sigma^2_x + 2\rho\sigma_x\sigma_y + \sigma^2_y$

Assuming normality considerably simplifies calculation of the portfolio distribution and the decomposition of risk into underlying sources. Unfortunately, financial returns diverge from normality in important respects and thus normality may not be a suitable assumption. This paper examines one of the most important divergences: the fat tails in financial returns – the observation that there are more large returns (both positive and negative) than would be predicted

[1] Note that we must assume both normality of the marginals and joint normality. Two random variables can each be normal but not jointly bivariate normal. Two normal random variables that are *independent* will be jointly normal (independent and with correlation zero).
[2] The appendix outlines some of the issues around decomposition of the portfolio distribution and the simplifications with assuming normality.

by a normal distribution. For the univariate case this is well-known. What is less well-known is that under joint normality there will be virtually no joint extreme events – an equally serious problem.

The evidence provided in this paper suggests that both the multivariate Student-t and a mixture of normals can provide substantially better approximations to financial returns than the joint normal. The Student-t requires Monte Carlo simulation while the mixture of normals retains the analytic and computational simplicity of the normal.

2 – **FAT TAILS**

It is well-known that financial returns exhibit more extreme observations than would be predicted by the normal or Gaussian distribution – financial returns or P&L distributions have fat tails. Mandelbrot highlighted the issue in the 1960s (see Mandelbrot 1962 and 1963, also Hudson and Mandelbrot 2004). A simple example is provided by returns for the Dow Jones Industrial Average over the period 1954-2004 shown in table 1 (taken from the discussion at the beginning of Beirlant, Schoutens, and Segers, 2005). This shows the 10 largest down moves of the Dow over the period. Using 25% as a reasonable (although slightly high) estimate of the volatility over this period, we can calculate the distance from zero, in standard deviations, for these 10 moves (the standard score or z-score).

All the ten returns are 3.68σ or further from zero. A normal distribution would predict that the probability of observing these events would be very low. Assuming normality a single observation 3.68σ (or worse) from the mean has probability 0.0117% (1.17×10^{-4}). Having ten such observations over 50 years (roughly 12,500 observations) would be miniscule – 0.0003%. (See Coleman 2011 section 5.5 or Coleman 2012 section 8.4 for a fuller discussion.)

Table 1 – Ten largest down moves of the Dow, 1954-2004

Date	Close	log-return	No. Sigma	NORMAL Prob % <=	NORMAL Prob n >=
19-Oct-87	3-Oct-04	-25.63%	-16.22	1.82E-59	0.00E+00
26-Oct-87	27-Nov-04	-8.38%	-5.30	5.79E-08	2.62E-07
27-Oct-97	9-Aug-19	-7.45%	-4.72	1.18E-06	5.28E-07
17-Sep-01	2-Jun-24	-7.40%	-4.68	1.43E-06	4.24E-09
13-Oct-89	12-Jan-07	-7.16%	-4.53	2.95E-06	5.50E-10
8-Jan-88	25-Mar-05	-7.10%	-4.49	3.56E-06	1.04E-11
26-Sep-55	30-Mar-01	-6.77%	-4.28	9.34E-06	5.31E-11
31-Aug-98	21-Aug-20	-6.58%	-4.16	1.59E-05	5.08E-11
28-May-62	29-Jul-01	-5.88%	-3.72	9.96E-05	6.49E-06
14-Apr-00	18-Mar-28	-5.82%	-3.68	1.17E-04	3.19E-06

The return and the number of standard deviations from zero (the z-score) assuming the standard deviation is 25% annualized. The column "Prob % <=" is the probability of observing a single daily return the relevant number standard deviations from the mean, assuming a normal distribution. The column "Prob n >=" is the probability of observing, over the 50-year period (roughly 12,500 days), the number of observed returns at that level z-score or worse given the probability of a single day that bad or worse. For the last row that is the probability of observing 10 observations -3.68σ or worse given the probability for each day is 1.17×10^{-4}, for the second-last row the probability of observing 9 observations -3.72σ or worse, etc. Source: Based on Beirlant, Schoutens, and Segers (2005, Table 1) and author's calculations.

This problem with the normal distribution – it under-predicts the probability or frequency of extreme events – appears across asset classes and across periods. Table 2 shows the problem with a different set of measures – looking at the relative frequency of 3σ-exceedance for daily log returns for individual stocks, selected stock indexes, and bond yields. This is the observed probability of being outside 3σ (either below or above). If log returns were normally distributed then the probability would be 0.270% (0.135% in the lower tail, 0.135% in the upper tail). In fact, returns for IBM's share price show a relative frequency of 1.290% (with 11,038 daily observations).[3] This is 4.8-times larger than predicted by the normal distribution. Results are similar for Proctor and Gamble, Citicorp, the S&P 500 index (capital appreciation only), capital returns for the 30-year US Treasury, and the FTSE and CAC equity indexes.

Table 2 – Relative Frequency of Observed 3σ Exceedance – Various Securities

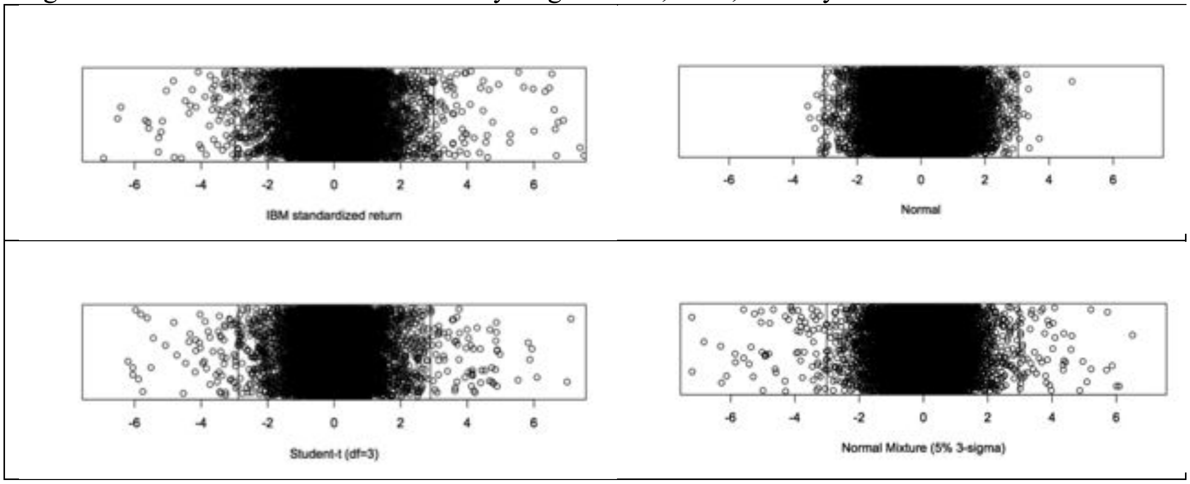
	Normal	IBM	P&G	Citi	S&P index	30yr UST	FTSE	CAC	Student-t	Mixture
Start date		Jan-70	Jan-70	Jan-77	Feb-77	Feb-77	2-Mar-90	2-Mar-90		
No. Observations		11,038	11,038	9,269	9,161	9,161	5,884	5,884		
3-sigma exceedance -	0.270	1.290	1.110	1.349	1.397	1.343	1.44	1.39	1.385	1.220
No. of times normal		4.8	4.1	5.0	5.2	5.0	5.3	5.1	5.1	4.5

Source: Yahoo Finance and author's calculations. The row labeled "3-sigma exceedance - %" shows the relative frequency of daily log-return being either below -3σ or above +3σ. The returns are daily, starting in the date shown and ending October 2013. The row labeled "No. of times normal" shows by how many times larger this observed relative frequency is than the probability implied by a normal distribution. "Student-t" is 3 degrees of freedom, "Mixture" is α=5%, β=3.

An alternative, graphical, view of this is shown in Figure 2. The upper left panel shows the 11,038 log returns for IBM over the period. The light vertical bars are ±3σ. There are 142 observations outside ±3σ – the 1.290% relative frequency shown in table 2. The upper right panel shows 11,038 simulated returns from a normal distribution with the same standard deviation – there are few observations beyond 3σ and only one beyond 4σ.

[3] For IBM there are 11,038 observations and the normal probability of 0.270% outside of ±3σ (0.13499% below -3σ and 0.13499% above +3σ) means we should expect roughly 15 observations in each tail. The observed relative frequency of 1.290% for IBM corresponds to an observed frequency of 142 or roughly 71 in each tail. The probability of observing 142 extreme returns if the true probability for each observation to be outside of ±3σ were 0.270% would be miniscule.

Figure 2 – Observed and Simulated Daily Log Returns, IBM, January 1970 to October 2013



Notes: The figures show returns between ±7σ. (Returns are spread out over the vertical axis for visual clarity.) For IBM there are 11,038 observed log returns from January 1970 to October 2013; five returns are outside ±7σ and are not shown in the figure. (It turns out that October is not a good month for IBM: log returns of roughly -16.3σ for 19-oct-87 (-26.8%, the 1987 crash, but followed by +10.8% the following day), -10.3σ for 18-oct-2000 (just a bad day), and -9.9σ on 21-oct-99 (announcement of 4Q profit warning). The other two large moves were positive: +7.4σ on 25-jul-1996, +7.5σ on 22-apr-1999.) Source: Yahoo Finance and author's calculations.

There are various approaches one can take to address the issue of fat tails. Extreme value theory is one popular approach (see, for example, McNeil, Frey, Embrechts, chapter 7). I will not discuss EVT here, partly in the interests or brevity but more significantly because EVT does not apply easily to the multivariate context that is the primary focus of this paper. Instead I will focus on alternative distributional assumptions – investigating theoretical distributions that exhibit fat tails.

The Student-t distribution is well-known and often-used because it has fat tails relative to the normal distribution. In Figure 2 we can see this graphically – the lower left panel displays 11,038 simulated observations drawn from a Student-t distribution with the same standard deviation as the observed IBM returns.[4] There are many observations beyond ±3σ and, visually, the Student-t looks dramatically different from the normal distribution and closer to the empirical distribution. In table 2 we can see numbers: the empirical distribution of IBM returns has a relative frequency of 1.290% beyond ±3σ and the Student-t has a probability of 1.385%, versus the normal 0.270%.

Another simple distribution is a two-point mixture of normals: low probability of a high-volatility regime (α% chance of volatility β×σ*) and high probability of a low-volatility regime (1-α% chance of volatility σ*). The lower right panel shows 11,038 simulated observations from a distribution with 5% chance of 3× high-vol regime.[5] Once again, in Figure 2 we clearly see extreme events or fat tails and in Table 2 the probability of values exceeding ±3σ much higher than the normal (1.220% versus 0.270%).

---

[4] One technical note – the display shows a standard t-variate with 3 degrees of freedom divided by √3. A standard t-variate with df=η has standard deviation √η/(η-2) – dividing by the √3 ensures a standard deviation of one.

[5] α=5%, β=3, σ*=0.845. The mixture standard deviation σ* must be less than one so that the volatility of the mixture, $\sigma^* \sqrt{(1-\alpha)+\alpha \cdot \beta^2}$, equals 1.

The mixture of normals is only an approximation to fat tails – eventually the distribution behaves as a normal with thin tails. What the mixture does is push out the point at which the tails are thin. For many purposes such an approximation is perfectly acceptable. And if not, we could use a three-point mixture, with a very low probability of a yet higher-vol regime. The benefit of such an approximation is that virtually all the analytical and computational simplicity of the normal carry over. Furthermore, such an approach has intuitive appeal – most days the markets are well-behaved but every once-in-a-while things go wild.

3. – **JOINT TAIL EVENTS FOR NORMAL AND OTHER DISTRIBUTIONS**

3.1 – SIMULATION FOR NORMAL, STUDENT-NORMAL, STUDENT, MIXTURE OF NORMALS

The thinness of normal tails is well-known but another aspect, somewhat less well-known, is that joint normality implies extreme events will not occur together, whatever the correlation (apart from the boundary case of ± 1).

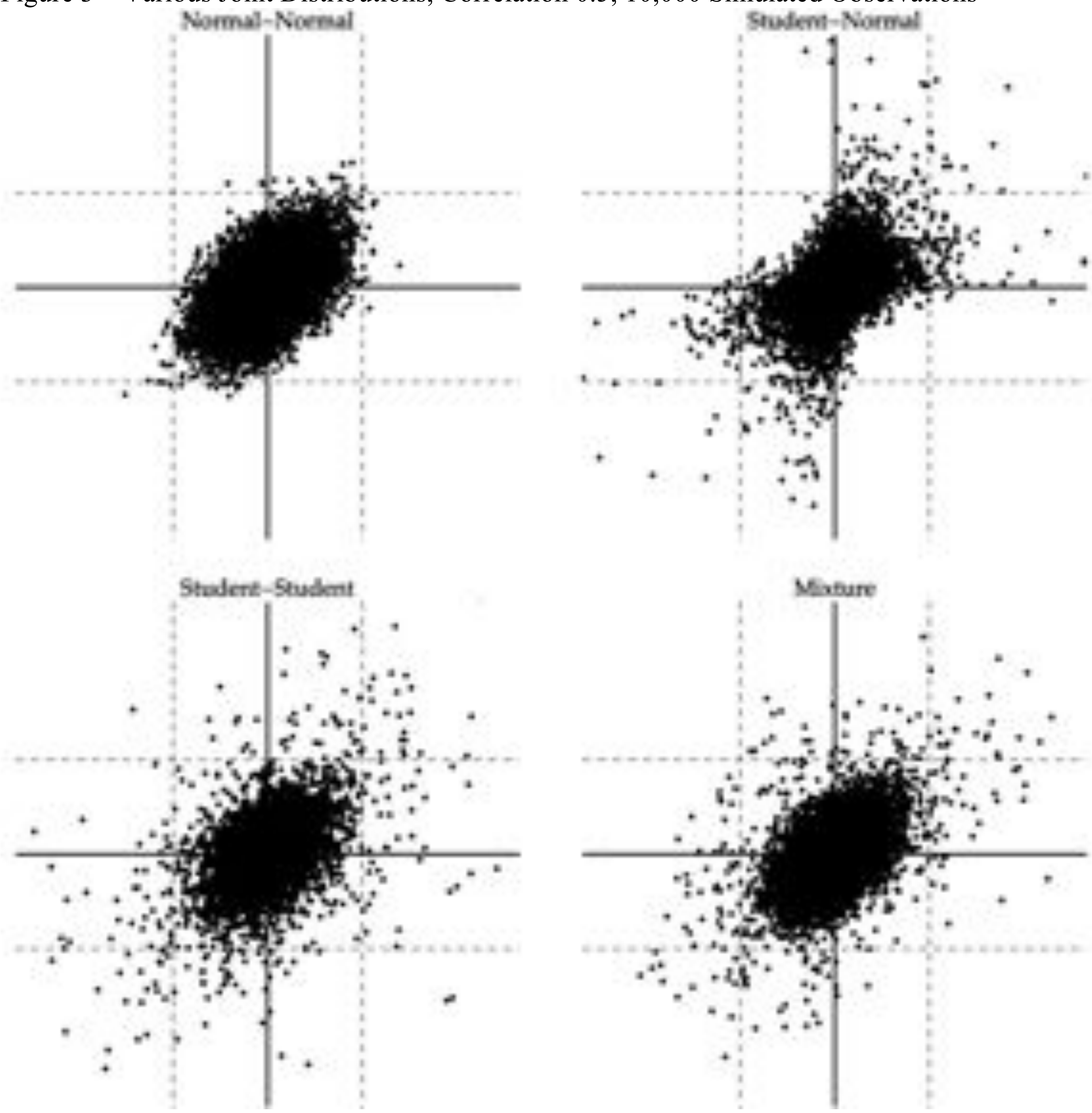
To illustrate the issues, consider Figure 3, which shows scatter plots of four simulated distributions, all with correlation 0.5. The details of the distributions will be discussed more fully below but for now focus on the top row. The top left shows a bivariate normal distribution. The dotted lines are at 3σ and there are almost no observations beyond 3σ. We should expect this since the univariate normal distribution generates few observations beyond 3σ and, similarly, the bivariate normal will produce neither marginal *nor* joint extreme observations. With the bivariate normal we see so few extreme events we cannot draw any conclusions about joint extreme events.

We can, however, examine the dependence behavior of the normal distribution by imposing fat-tailed marginal distributions on the normal dependence structure. This is accomplished using the idea of a copula to simulate a meta-distribution with Student-t marginals super-imposed on the bivariate normal dependence structure. (The idea behind copulas will be discussed more below.) The top right panel shows this meta-distribution or hybrid Student-t / Normal distribution. The figure show many *marginal* observations beyond 3σ but very few *joint* observations beyond 3σ.

Essentially, if we impose fat tails (extreme events) in the marginals by imposing Student-t marginals, the normal dependence structure still gives few joint extreme events. A large value of X_1 (say beyond 3σ) implies a high probability that X_2 *will not* be large. The return X_2 will be in the same direction (above or below the mean) but will not itself be large.

The normal dependence structure is in sharp contrast with the joint Student-t distribution, shown in the lower left panel. This exhibits extreme events on the marginals (as we would expect) but also many joint extreme events. This looks like we would wish – the dependence ("correlation") between the two variables extends into the tails. This is in sharp contrast to the dependence structure exhibited by the normal copula (Student-Normal distribution, upper right panel) where events in the tails look independent – there are no joint extreme events and the "correlation" between the variables completely breaks down in the tails. (The mixture of normals, shown in the lower right panel, also shows joint extreme events and will be discussed more fully below.)

Figure 3 – Various Joint Distributions, Correlation 0.5, 10,000 Simulated Observations



3.2 – CONDITIONAL TAIL PROBABILITIES FOR BIVARIATE NORMAL AND STUDENT-t

To understand better what is happening in the joint normal, consider jointly normal random variables, X_1 and X_2:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right].$$

Now consider X_2, conditional on X_1 taking a particular value x, in other words X_2 conditional on $X_1=x$. It is well known that:

$$X_2 \mid X_1 = x \sim N\left(\rho x, 1-\rho^2\right) \ .$$

Say that x is in the right tail (so that X_1 is already an extreme value) then we can ask "what is the probability that X_2 is also extreme?" The probability that X_2 will be at or above x given that X_1 is equal to x is:

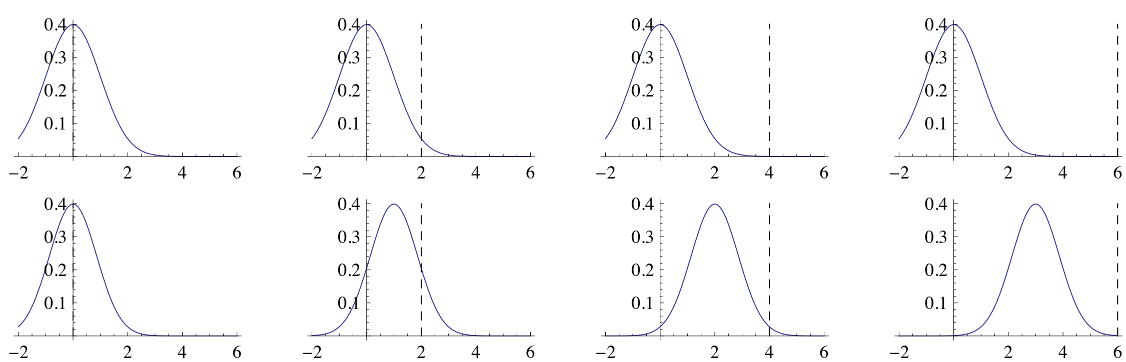
$$\mathrm{P}[X_2 \geq x \mid X_1 = x] = 1 - \Phi\left[x\sqrt{\frac{1-\rho}{1+\rho}} \right] \ .$$

As x gets more negative, $\mathrm{P}[X_2 \geq x \mid X_1 = x]$ gets smaller and, in the limit, goes to zero. McNeil, Frey, Embrechts (2005) discuss this in the context of the *coefficient of tail dependence* (section 5.2.3).

Figure 4 and Table 3 help explain what is happening, displaying the density and probabilities for the case ρ=+0.5. In Figure 4 we are looking at the upper tail with points X_1=x further out in the tail progressing from left to right. The upper row shows the unconditional probability density for X_2. As the value of X_1=x moves further out in the tail the probability $\mathrm{P}[X_2 \geq x]$ goes down. The values for $x = \{0, 2, 4, 6\}$ are shown in the first row of table 3.[6]

The second row shows the conditional density for $X_2 \mid X_1 = x$ for various values of x (0, 2, 4, 6).[7] This is still normal but it migrates to the right as x increases – the mean is ρx. This is the effect of the correlation or dependence between X_1 and X_2: the larger is X_1 the more likely is a large value of X_2. But note that the density does not migrate as fast as the value of x. The result is that conditional on an extreme value for X_1, the probability of an *equally* extreme value for X_2 becomes vanishingly small. The numbers are shown in the second row of table 3: Conditional on $X_1 = 6\sigma$ the probability that X_2 is 6σ or larger is only 0.03%.

Figure 4 – Normal Density for X_2 (Unconditional, first row) and $X_2 \mid X_1 = x$ (Conditional, second row) for $X_1 = x = \{0, 2, 4, 6\}$. For correlation ρ=+0.5.



[6] Remember that this is a standard normal with σ=1, so x=2 means 2σ.

[7] The two densities are both normal but the conditional density has a lower variance. This difference in the spread of the densities is not obvious to the eye for ρ=0.5 but becomes more marked as ρ increases towards 1.

Table 3 – Probability P[$X_2 \geq x$], both Unconditional, and Conditional on $X_1 = x$, for Various Values of $X_1 = x$. For Normal, Correlation ρ=+0.5.

x	0	2	4	6
Unconditional	0.5000	0.0228	3.17E-05	9.87E-10
Conditional	0.5000	0.1241	0.0105	2.66E-04

We can now understand the behavior seen in the upper right panel of Figure 3 (the Student-Normal meta-distribution – Student-t marginals and normal dependence structuree). The variables are *not* independent: when X_1 is large and negative then X_2 will also be negative, but it will not be *large* and negative. This is exactly what we see from the densities shown in Figure 4 – as X_1 moves out in the tails the conditional density of X_2 moves out, but not nearly as far out as the value of X_1. An extreme value of X_1 (say 4 or 6 σ) does imply an increased probability of a higher X_2, but only a remote possibility of a value as or more extreme than X_1. Far out in the tail joint extreme events are vanishingly rare and the variables do not look "correlated", no matter what the original correlation. Far out in the upper tail P[$X_2 \geq x \mid X_1 = x$] $\to 0$ (and in the lower tail P[$X_2 \leq x \mid X_1 = x$] $\to 0$).

This behavior gives the distribution seen in the simulations of the Student-t – Normal meta-distribution in the upper right panel of Figure 3: in the tails the two variables do not move together and the "correlation" breaks down. Unfortunately this is not what we would intuitively want as a description of financial returns. And, as we shall see, it does not appear to be how financial returns actually behave.

The case for the Student-t distribution is quite different. Figure 5 shows the unconditional and conditional densities.[8] For the Student-t the conditional density not only migrates to the right (the mean shifts) but it also spreads out. As a result the conditional probability P[$X_2 \geq x \mid X_1 = x$] does not tend towards zero but rather to a finite limit:

$$t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right).$$

The probability that there will be joint extreme events does not tend to zero. And this gives the behavior we see in the bottom left panel of Figure 3: the "correlation" or dependence between the two variables extends out into the tails.

[8] According to McNeil, Frey, Embrechts (2005 section 5.2), conditional on $X_1 = x$, $\sqrt{\frac{\nu+1}{\nu+x^2}}\frac{X_2-\rho x}{\sqrt{1-\rho^2}} \sim t_{\nu+1}$

Figure 5 – Student-t Density for X_2 (Unconditional, first row) and $X_2 \mid X_1 = x$ (Conditional, second row) for $X_1 = x = \{0, 2, 4, 6\}$. For correlation ρ=+0.5.

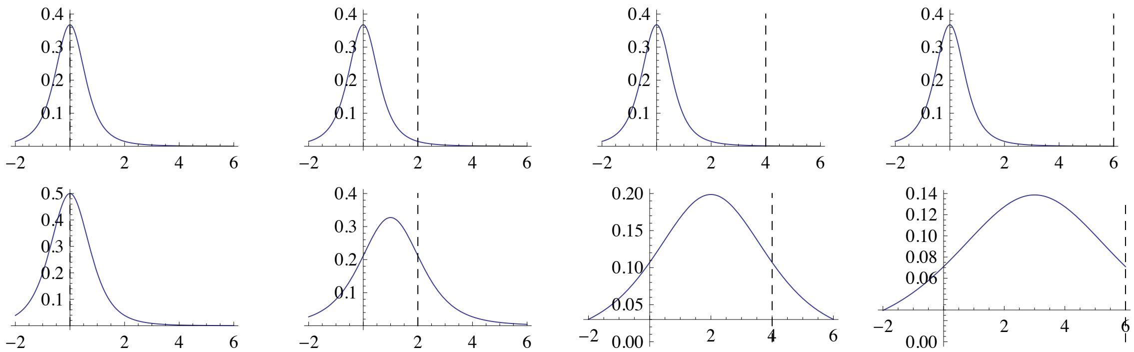


Table 4 – Probability P[$X_2 \geq x$], both Unconditional, and Conditional on $X_1 = x$, for Various Values of $X_1 = x$. For Student-t, Correlation ρ=+0.5.

x	0	2	4	6
Unconditional	0.5000	0.0203	0.0031	0.0010
Conditional	0.5000	0.2160	0.1745	0.1647

3.3 – BIVARIATE STUDENT-t AS MIXTURE OF NORMALS

To better understand *why* the bivariate Student-t shows the tail behavior it does, consider the definition of the Student-t as a mixture of normals (see McNeil, Frey, Embrechts 2005 chapter 3 or Shaw and Lee 2007):

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} / \sqrt{\chi_\nu^2 / \nu}$$

where the X_i are bivariate standard normal with correlation ρ and χ^2_ν is chi-squared with ν degrees of freedom. Now it is clear both why the Student-t (univariate or multivariate) has fat tails and why the multivariate Student-t has joint extreme events. Extreme events occur when the χ^2 is small and thus inflates the normal variates. Because it is the same χ^2 that inflates both X_1 and X_2, extreme events will tend to occur together. In a sense, the χ^2–variate acts as a common factor that inflates both dimensions.

3.4 – COPULAS

I will take a short digression into copulas to more fully explain Figure 3. One way to think of copulas is as an alternative method for writing a multivariate distribution. Consider a multivariate random variable of d dimensions, with distribution function $F(x_1, \ldots, x_d)$ and marginals $\{F_1(x_1), \ldots, F_d(x_d)\}$. It turns out that this multivariate distribution can be written in either of two forms:

- Usual multivariate distribution: $F(x_1, \ldots, x_d)$
- In terms of marginals and copula: $C(F_1(x_1), \ldots, F_d(x_d))$

There will always exist a function $C(F_1(x_1), \ldots, F_d(x_d))$, called a copula, which is itself a d-dimensional distribution function on $[0, 1]^d$ with standard uniform marginals.[9] The power of the copula approach is that it isolates the dependence across variables in the copula, with the marginals separate. (This is somewhat analogous to the linear (normal) case, where in which the dependence structure can be isolated in the correlation matrix, with the variances separate.) Using copulas we can specify the marginals and the dependence structure separately; can focus on the dependence structure separately from the marginals. A hybrid or meta-distribution can be created by mixing and matching marginals and copulas.

I briefly review the computational approach for simulating the meta distributions, although for a more detailed discussion see McNeil, Frey, and Embrechts (2005, p. 193, also pp. 66, and 76). For the simulation of Student-t marginal combined with normal copula, the process for each draw of our bivariate random variable is:

Step1. Generate a normal copula:

- For each draw, generate a standardized bivariate normal with mean zero, unit variances, and the desired correlation matrix:

$$X = (X_1, X_2)' \quad \sim \quad N_2(0, R) .$$

- Calculate
$$U = (U_1, U_2)' \quad = (\Phi(X_1),\ \Phi(X_2)).$$
where $\Phi(\cdot)$ is the normal CDF.

- This will be a normal copula. More precisely, the random vector U will have a normal copula distribution with correlation matrix R.

- The numbers U_i will be in the interval [0,1] and will look like probabilities.

Step 2. Generate the joint distribution with the desired marginals:

- Calculate

$$Y = (t_v^{-1}(U_1),\ t_v^{-1}(U_2))' \quad ..$$

where $t_v^{-1}(\cdot)$ is the univariate Student- t distribution inverse CDF or quantile function.

- This will now have a marginal Student- t distribution with normal dependence structure.

The hybrid or meta-distribution of Student-t marginals and normal copula shown in the upper right panel of Figure 3 has Student-t marginals to give fat tails, but a normal copula to impose the normal dependence structure. The simulations in Figure 3 clearly show that the normal dependence structure does not produce tail behavior that we would want for financial returns. Even though the correlation is 0.5, and the variables move together as we would expect in the central part of the distribution, when we move far out in the tails there is not the dependence

[9] See McNeil, Frey, and Embrechts (2005, section 5.1) secton 5.1. Copulas are most appropriate for continuous distributions.

structure we would want. A large value of X_1 (say beyond 3σ) implies a high probability that X_2 *will not* be large – X_2 will be in the same direction (above or below the mean) but will not itself be large.

3.5 – MIXTURE OF NORMALS

The normal distribution is popular in risk measurement circles for one very good reason: analytic and computational tractability. Assuming normality for individual assets implies normality of any linear portfolio of those assets. The portfolio distribution's parameters can be calculated from the constituent assets by simple matrix multiplication. Portfolio volatility and VaR are simple to calculate, as are portfolio analytics such as contribution to risk or best hedges. Alternative distributions such as the multivariate Student-t invariably require simulation with all the concomitant computational and convergence issues.

A distribution that exhibits (or approximates) fat tails and joint tail dependence, in addition to the Student-t, is the two-point (or three-point) mixture of normals. The two-point mixture of normals has a distribution function that is the sum of two normal distribution functions: the first with probability $(1 - \alpha)$) and standard deviation $1/\gamma$ and the second with probability α and standard deviation β/γ (both with mean 0, and $\gamma = \sqrt{1-\alpha+\alpha\beta^2}$). The variance for such a random variable will be 1. Parameters on the order of $\alpha = 0.02$ and $\beta = 4$ give a distribution roughly similar to a Student-t (3 degrees of freedom) out to maybe 5 or 6σ.

The mixture of normals is appealing because it combines simplicity with a straightforward motivation for fat tails. Say the P&L distribution for each day is normal but periodically there is a day which that is still normal but with higher volatility. This is an attractive assumption since it is what trading in the markets actually feels like – periods of relative quiescence interspersed with days of mayhem. The conditional distribution, conditional on knowing whether the day is low or high volatility, will be normal. In contrast, the unconditional distribution, not knowing whether a day is low or high volatility, will be non-normal and will exhibit fat tails. A two-point mixture of normals provides a rough approximation to this.

In symbols, the two-point mixture of normals with mean zero, variance 1, and correlation ρ, will be:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim (1-\alpha)N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} / \gamma\right] + \alpha N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \cdot \beta / \gamma\right] .$$

with $\alpha \in [0,1]$, $\beta > 1$, $\gamma = \sqrt{1-\alpha+\alpha\cdot\beta^2}$.

Figure 6 and Table 5 show the conditional density for X_2, conditional on $X_1 = x$, for various values of x in the upper tail (0, 2, 4, 6). Like the Student-t distribution the conditional mixture of normals both migrates to the right (the mean goes up as ρx) but also spreads out. The reason it spreads is simple: if we observe an extreme value for X_1 we are much more likely to be in the high volatility regime. Conditional on observing $X_1 = x$, the probability of the high-vol regime will be (using Bayes' rule):

$$\text{P[high vol regime} \mid X_1 = x] = \frac{\alpha \cdot \varphi(x\gamma / \beta)}{\alpha \cdot \varphi(x\gamma / \beta) + (1-\alpha) \cdot \varphi(x\gamma)} .$$

Figure 6 – Normal Mixture (5% 3x) Density for X_2 (Unconditional, first row) and $X_2 \mid X_1 = x$ (Conditional, second row) for $X_1 = x = \{0, 2, 4, 6\}$. For correlation ρ=+0.5.

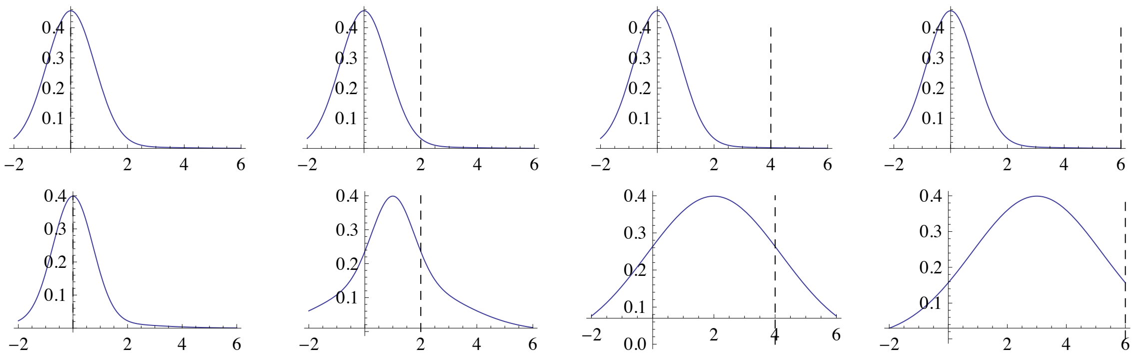


Table 5 – Probability P[$X_2 \geq x$], both Unconditional, and Conditional on $X_1 = x$, for Various Values of $X_1 = x$. For Normal Mixture (5% 3x), Correlation ρ=+0.5.

x	0	2	4	6
Unconditional	0.5000	0.0193	0.0029	0.0004
Conditional	0.5000	0.1785	0.1810	0.0859

For extreme values in the range of 3σ-6σ, this two-point mixture of normals behaves similarly to the Student-t. The simulations in Figure 3 show this. Clearly the mixture of normals only approximates fat tails – go out far enough and the mixture will behave just as the standard normal with miniscule probability of joint extreme events. But in the range considered here this two-point mixture does a good job of approximating both fat tails in the marginals and joint extreme events.[10]

4. – SOME EMPIRICAL OBSERVATIONS FOR JOINT TAIL EVENTS

This is a preliminary analysis of joint returns for a small number of assets:

- Individual stocks: IBM, Citicorp, and Proctor & Gamble
- Equity indexes: FTSE and CAC
- Equity and fixed income: S&P 500 index and US Treasury 30-year bond

Table 2 (reproduced here) shows summary statistics for the assets viewed on their own. As pointed out in section 1 all these series show more extreme events (defined here as exceeding 3σ or three standard deviations) than predicted by the normal distribution.

[10] Using a three-point mixture, with a yet smaller probability of a yet higher vol regime, can of course push the region of approximation further out in the tails.

Table 2 – Relative Frequency of Observed 3σ Exceedance – Various Securities

	Normal	IBM	P&G	Citi	S&P index	30yr UST	FTSE	CAC	Student-t	Mixture
Start date		Jan-70	Jan-70	Jan-77	Feb-77	Feb-77	2-Mar-90	2-Mar-90		
No. Observations		11,038	11,038	9,269	9,161	9,161	5,884	5,884		
3-sigma exceedance -	0.270	1.290	1.110	1.349	1.397	1.343	1.44	1.39	1.385	1.220
No. of times normal		4.8	4.1	5.0	5.2	5.0	5.3	5.1	5.1	4.5

Source: Yahoo Finance and author's calculations. The row labeled "3-sigma exceedance - %" shows the relative frequency of daily log-return being either below -3σ or above +3σ. The returns are daily, starting in the date shown and ending October 2013. The row labeled "No. of times normal" shows by how many times larger this observed relative frequency is than the probability implied by a normal distribution. "Student-t" is 3 degrees of freedom, "Mixture" is α=5%, β=3.

4.1 – JOINT BEHAVIOR FOR INDIVIDUAL STOCKS – IBM, Citicorp, Proctor & Gamble

Figure 7 shows the joint returns (normalized) for IBM and Citigroup. There are clearly many returns exceeding 3σ, both marginal and joint. Over the period 1970-2013 the correlation between the daily returns was 0.32.

There are two issues in examining Figure 7. First, there is no direct way, from Figure 7 alone, to assess how much the returns deviate from normality and whether the observations match any of the three distributions examined here – we need to examine some more formal statistics. Before doing so, however, we need to address the second issue – whether it is appropriate to consider the raw returns shown in Figure 7 or whether we should consider some factor structure. IBM and Citigroup move together and exhibit at least some joint extreme events, but how much of that is simply due to a common market factor.

The simplest factor structure is the CAPM, which posits that individual stock returns are composed of a systemic component and an idiosyncratic component, say:

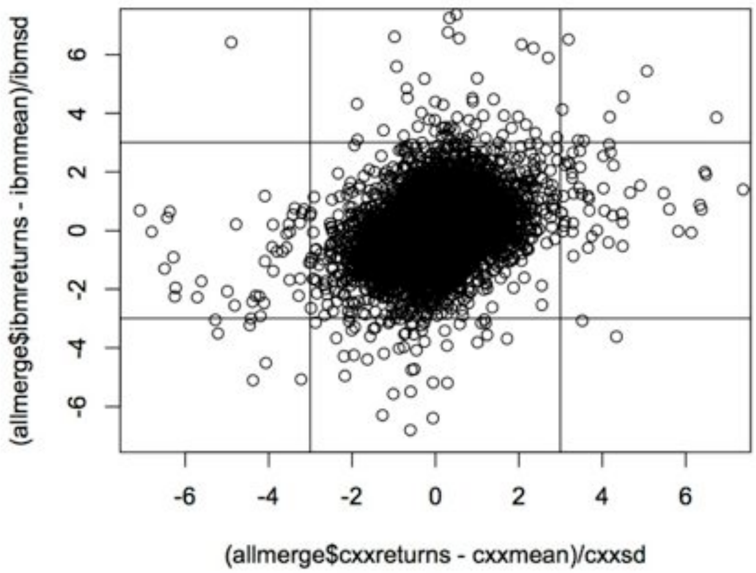
$$R_{\mathrm{IBM}} = \beta_{\mathrm{IBM}} * R_{\mathrm{S\&P}} + \text{idiosyncratic} \quad .$$

From a risk measurement perspective this is a useful way to decompose the market return for IBM: into a common component (say the S&P 500 index) and an idiosyncratic component. In this case Citigroup will similarly be:

$$R_{\mathrm{Citi}} = \beta_{\mathrm{Citi}} * R_{\mathrm{S\&P}} + \text{idiosyncratic} \quad .$$

This provides a useful decomposition, with individual stock returns being dependent on a market-wide factor (the return on the S&P index) and idiosyncratic components. The common component can then be managed or hedged, and the idiosyncratic components will (presumably) average out to zero.

Figure 7 – Normalized Returns for IBM and Citigroup, 1977-2013

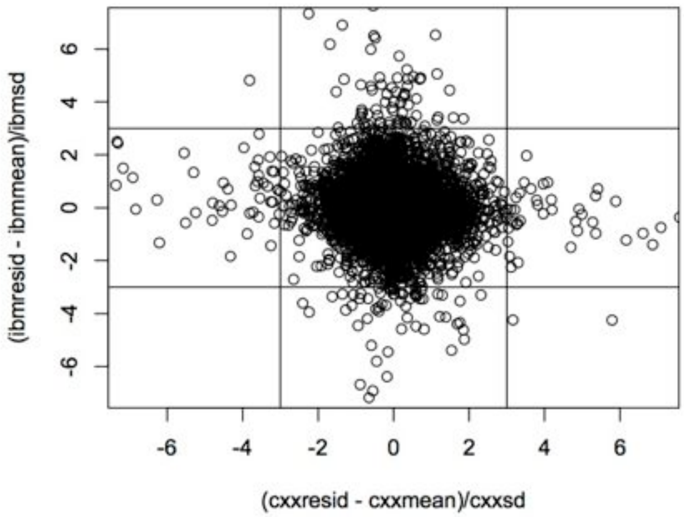


Source: Yahoo Finance and author's Calculations. There are 22 observations for Citigroup and 5 for IBM beyond ±7σ.

Using this approach it is easy to understand why IBM and Citigroup have a positive correlation and show joint extreme events: they are both driven by the underlying common factor, the S&P index.

We may, then, wish to consider the idiosyncratic or residual returns, after removing the component due to the S&P index for each stock. When we do this we find: zero correlation between the residuals for the two companies; many marginal extreme events; few joint extreme events.

Figure 8 – Normalized Returns for IBM and Citigroup, Residual after Removing S&P Common Factor, 1977-2013



Source: Yahoo Finance and author's Calculations. There are 24 observations for Citigroup and 10 for IBM beyond $\pm 7\sigma$.

The pictures are useful but to compare with normal or other distributions we need to examine some statistics. Table 6 shows statistics for IBM and Citigroup in the first two rows. The first row shows the raw returns. Under the normal distribution there would be 0.27% probability of 3σ exceedance (0.135% in each tail). For IBM raw returns for 1977-2013 there are 4.96-times the number of exceedances predicted by normality (observed frequency 1.338%). For reference a Student-t (3 degrees of freedom) predicts 5.13-times the normal, which the mixture of normals (α=3%, β=5) predicts 4.52-times.

A bivariate normal distribution with correlation 0.32 predicts a joint probability of 3σ-exceedance of about 0.005% – less than one joint 3σ-exceedance for 9,000 days. In fact we observe roughly 45-times more, a relative frequency of 0.248% (23 observations). The Student-t predicts 73-times the normal, the mixture of normals 62-times, and the Student-normal meta-distribution 12.6-times.

Table 6 – Statistics for Joint 3σ Exceedances

	Marginal exceedance					Joint 3-sigma exceedance				
	Probability	x Normal		No. observ.	Correlation		x Normal			
	Normal	Asset 1	Asset 2			Norm Prob	Observed	Student	Mixture	Copula
IBM & Citi - raw returns	0.270	4.96	4.99	9,269	0.32	0.005	45	73	62	12.6
IBM & Citi - residuals	0.270	5.93	4.19	9,269	0.00	0.001	89	473	384	26.1
IBM & P&G - residuals	0.270	5.40	4.15	11,038	0.00	0.001	99	473	384	26.1
US 30yr & S&P500 index	0.270	4.97	5.17	9,161	0.00	0.001	299	473	384	26.1
CAC index & FTSE index	0.270	5.15	5.33	5,884	0.83	0.086	9.7	9.5	8.5	6.4
Alternatives - Marginal excee	Probability	x Normal								
Student-t	1.385	5.13								
Mixture (5%, 3x)	1.220	4.52								

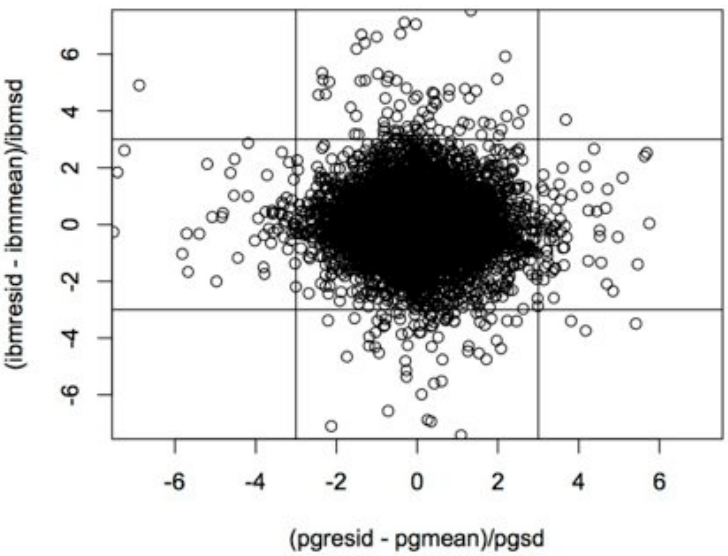
Turning to the residuals for IBM and Citigroup, removing the common factor of the S&P return produces residuals that have zero correlation. A normal distribution would predict less than 0.001% probability of joint 3σ-exceedances, whereas the observed frequency is 89-times higher

(0.065% or 6 observations). The Student-t and mixture of normals both predict yet higher frequency of 3σ exceedances – roughly 473x and 384x.[11]

The daily returns for the S&P500 and the US Treasury 30-year bond are roughly zero correlation. In contrast to the individual stock returns, however, there are a large number of joint 3σ-exceedances: roughly 300-times that predicted by the normal distribution (0.218% or 20 observations).

The FTSE and CAC equity indexes have a correlation of 0.83 over the period 1990-2013. With a correlation that high the difference between the predicted 3σ-exceedances for the normal and the alternate distributions are not as dramatic – the observed is 9.7-times the normal predicted, the Student-t is 9.5-time the normal.

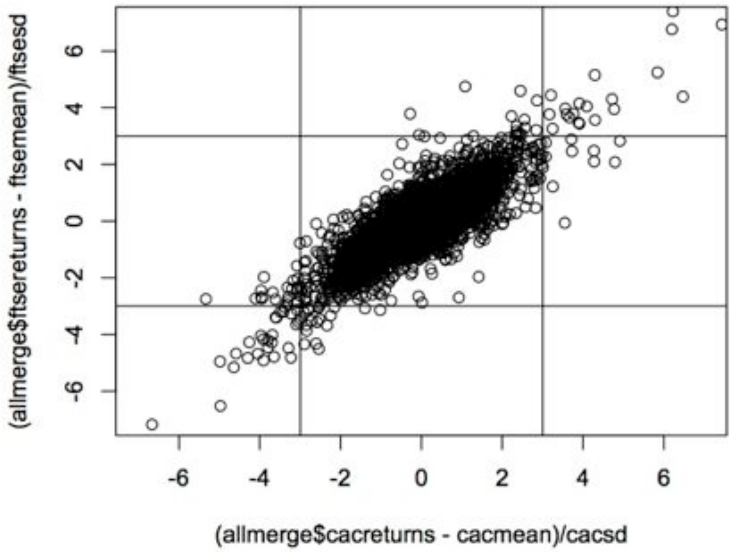
Figure 9 – Normalized Returns for IBM and Proctor & Gamble, Residual after Removing S&P Common Factor, 1970-2013



Source: Yahoo Finance and author's Calculations. Note, there are 7 observations for P&G and 13 for IBM beyond ±7σ.

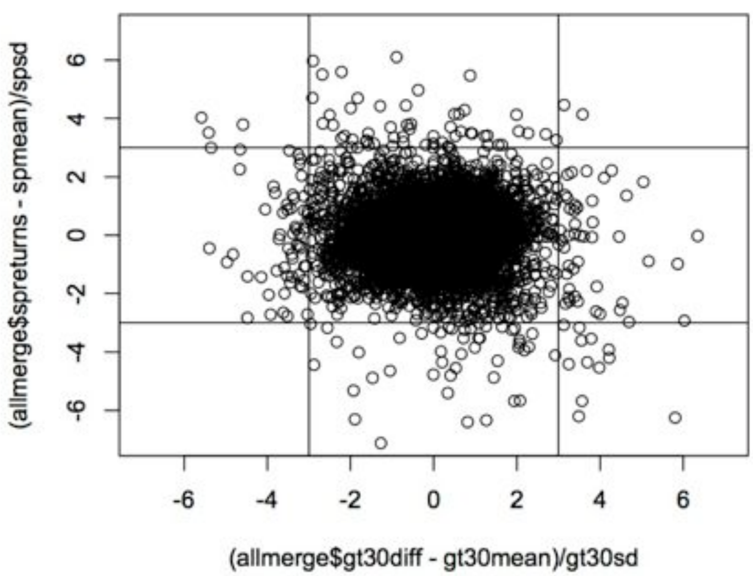
[11] Note, however, that the degrees of freedom for the Student-t and the mixture parameters for the normal mixture can be varied. For df=6 the predicted Student-t joint exceedance is 113x the normal and for α=10% and β=2 the mixture joint exceedance is 76x the normal.

Figure 10 – Normalized Returns for FTSE and CAC Equity Indexes, 1990-2013



Source: Yahoo Finance and author's Calculations. Note, there are 4 observations for the FTSE index and 1 for the CAC index beyond ±7σ.

Figure 11 – Normalized Returns for S&P 500 Equity Index and Price Returns for US Treasury 30-year bond, 1977-2013



Source: Yahoo Finance and author's Calculations. Note, there are 1 observations for the 30-year UST and 9 for the S&P500 index beyond ±7σ.

5. – **CONCLUSION**

Normality fails to capture the frequency of large returns. For assets considered on their own, the normal distribution predicts too few large returns or extreme events. This appears to hold across a variety of types of assets. In this paper I focused on the frequency of returns exceeding 3σ but the conclusion holds for various measures of extreme events.

Normality also fails to adequately represent the frequency of joint extreme events. Empirically single extreme events are rare and joint extreme events doubly-rare – but far less rare than predicted by the normal distribution. Joint normality, in fact, predicts that if one goes far enough out in the tails joint extreme events will never occur. This simply does not appear to be the case. The lack of joint extreme events appears to be inherent in the dependence structure of the joint normal distribution. Even when we impose fat tails in the marginals (by creating a synthetic distribution with Student-t marginals and a normal copula) there are few joint extreme events.

The Student-t and mixture of normals have both fat tails and joint extreme events relative to the normal distribution and thus have the potential to approximate empirical distributions more realistically than the normal. They predict both more extreme events and more joint extreme events than the joint normal distribution.

This paper has examined a small number of assets and found that there are more single and joint extreme observations than predicted by the normal distribution. The paper focused on only a single set of parameters for the Student-t and mixture of normals. Clearly 3 degrees-of-freedom and (α=5%, β=3) are not the appropriate parameters for many assets – further work needs to be done to determine realistic parameters. Nonetheless, the ability of the Student-t and the mixture of normals to more accurately approximate empirical observations is valuable. The mixture of normal is particularly useful because it retains the analytical and computational simplicity of the normal while capturing an important aspect of the financial markets (fat tails) that the simple normal does not.

APPENDIX
Normality and Contribution to Risk

Now let us turn to portfolio decomposition and understanding the sources of risk in the portfolio distribution. Once again normality provides substantial computational simplification. For any reasonable risk measure $R(\omega)$ we can decompose the measure and changes into constituent parts[12]:

$$R(\omega) \equiv \sum_{i=1}^{n} \frac{\partial R(\omega)}{\partial \omega_i} \omega_i$$

$$dR(\omega) \equiv \sum_{i=1}^{n} \frac{\partial R(\omega)}{\partial \omega_i} \omega_i d \ln \omega_i$$

[12] More specifically this applies to any linearly homogeneous risk measure and so applies to volatility, VaR, expected shortfall and indeed any coherent risk measure. This property is simply Euler's law (Euler's homogeneous function theorem).

$$\text{Marginal Contribution for asset } i = \frac{\partial R(\omega)}{\partial \omega_i}\omega_i$$

where ω_i is some appropriate scale measure for the amount held of asset i. (For our example the notional amounts, \$20mn and €7mn, are fine. Note that ω appears in both numerator and denominator so the exact units do not matter.)

Say that we want to estimate the marginal contribution for VaR. If the individual asset P&Ls are the random variables X_i the portfolio P&L is $\sum_i \omega_i X_i$, the Z% VaR is $\text{VaR}_z = \{Y \text{ s.t. } P[\sum_i \omega_i X_i \leq Y] = Z\}$. Then the marginal contribution to volatility and VaR will be[13]

$$\text{MC to Volatility for asset } i = \omega_i \left[\text{cov}(X_i, \textstyle\sum_k \omega_k X_k) / \sqrt{\text{variance}(\sum_k \omega_k X_k)}\right]$$

$$\text{MC to VaR for asset } i = \omega_i \, E[X_i \mid \textstyle\sum_i \omega_i X_i = \text{VaR}_z] \, .$$

When we use a general multivariate distribution (not assuming normality) we need to resort to Monte Carlo simulation. In that case the calculation of marginal contribution for volatility is straightforward but for VaR is not. The contribution to VaR will depend on the single simulated value for which $\sum_i \omega_i X_i = \text{VaR}_z$ and this will have large sampling variability which will not decrease as we increase the number of draws.

When we assume normality, however, the contribution to VaR has a simple form:

MC to VaR for asset i assuming normality =

$$\omega_i \left[\text{cov}(X_i, \textstyle\sum_k \omega_k X_k) / \sqrt{\text{variance}(\sum_k \omega_k X_k)}\right] \cdot \Phi^{-1}(z)$$

REFERENCES

Beirlant, Jan, Wim Schoutens, and Johan Segers. 2005. "Mandelbrot's Extremism." *Wilmott Magazine* (March).

Coleman, Thomas S. 2011. *A Practical Guide to Risk Management*. Research Foundation of the CFA Institute.

Coleman, Thomas S. 2012. *Quantitative Risk Management*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Hudson, R.L. and Mandelbrot, Benoit B. 2004. *The (Mis)Behavior of Markets – A Fractal View of Risk, Ruin and Reward*. Perseus Books Group.

Mandelbrot, Benoit B. 1962. "The variation of certain speculative prices". *IBM Research Report NC-87*.

[13] See McNeil, Frey, Embrechts equation 6.24.

Mandelbrot, Benoit B. 1963. "The variation of certain speculative prices". *Journal of Business* 36, 394–419.

McNeil, Alexander, Rudiger Frey, and Paul Embrechts. 2005. *Quantitative Risk Management*. Princeton, NJ: Princeton University Press.

Shaw, W.T., and K.T.A Lee. 2007. "Copula Methods vs Canonical Multivariate Distributions: the multivariate Student T distribution with general degrees of freedom. Kings College London, April 24.