Extreme events: Blending principal components analysis with independent components analysis

Malcolm Kemp

3 November 2009

Nematrian

malcolm.kemp@nematrian.com

Extreme events: blending principal components analysis with independent components analysis

- Principal Components Analysis (PCA) and Independent Components Analysis (ICA)
  - Their main characteristics and differences
- Extreme events and fat tails
- Analysing what drives markets
  - Identifying similarities between PCA and ICA
- Blending together PCA and ICA for refined risk model and portfolio construction purposes

malcolm.kemp@nematrian.com

PCA vs. ICA – main characteristics and differences

- Both are examples of 'blind source separation', aiming to identify 'signals' (i.e. sources / factors) that explain (observed) market behaviour
- Principal Components Analysis (PCA)
  - Seeks to identify the largest contributors to variance, i.e. magnitude of impact
  - 'Signals' maximise sum of variances of returns of each security within universe
- Independent Components Analysis (ICA)
  - Seeks to identify contributors to market behaviour that are meaningful
  - 'Signals' maximise independence, non-Normality and/or complexity

malcolm.kemp@nematrian.com

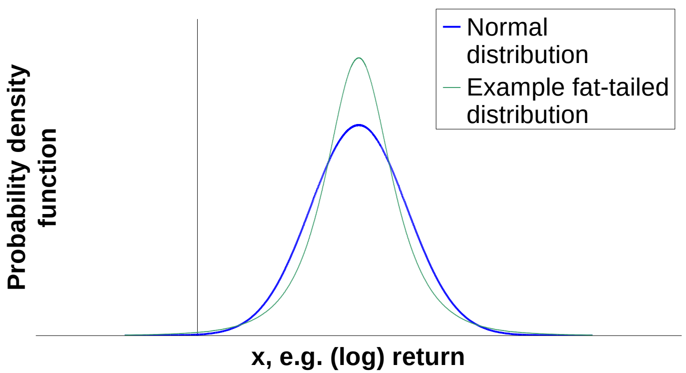
Relevance for risk and portfolio construction modelling

- Ideally risk and portfolio construction models should incorporate both magnitude and meaning
- Magnitude, because size of (adverse) event is its most important characteristic for risk management purposes, irrespective of its source
- Meaning (and hence explanatory capability), because
  - Humans are naturally curious and seek meaning (and purpose!)
  - "The only 'bad' mistakes are the ones we don't learn from"
  - Particularly for portfolio construction – as long as we are around next time!
- Extreme events probably have the most of both!

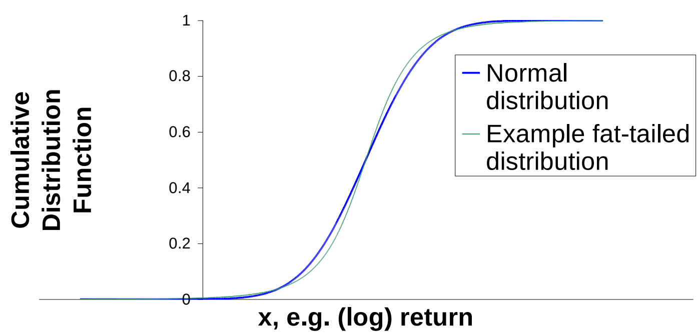
Visualising extreme events, i.e. fat tails – *single* return series

- There are various ways of visualising fat tails in a *single* return distribution. Easiest to see in format (c) below
  - By ‘fat tail’ we mean probability of extreme-sized outcomes (returns / movements / events) seems to be higher than if coming from a (log) Normal distribution

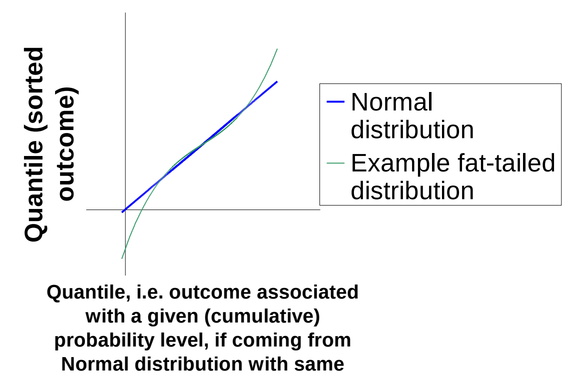
(a) Example Probability Density Function



(b) Example cumulative probability distribution plot



(c) Example quantile-quantile plot

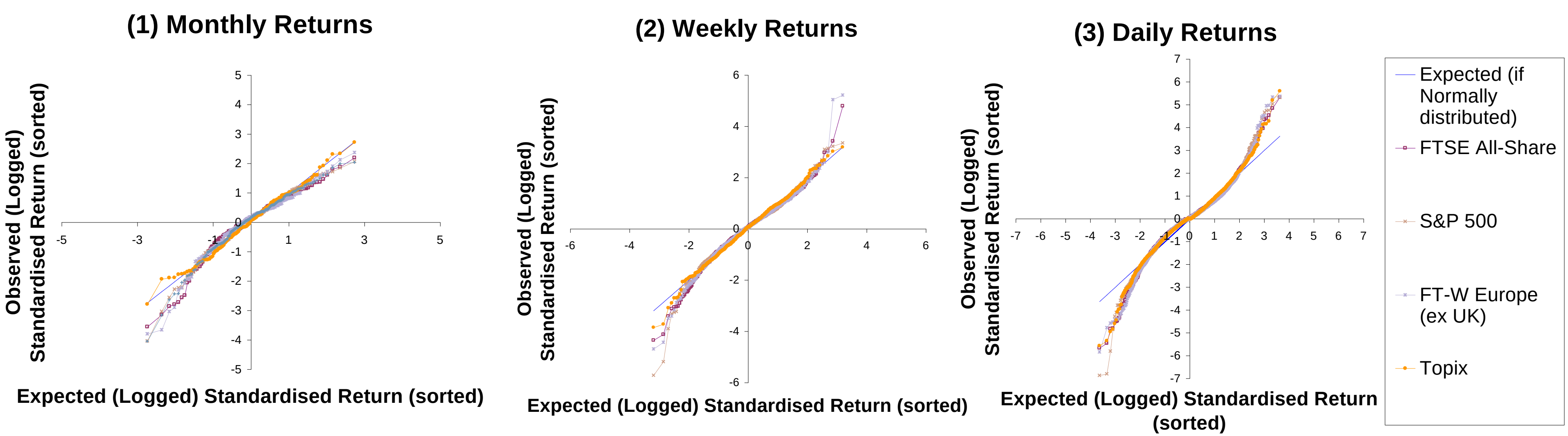


Source: illustrative

malcolm.kemp@nematrian.com

Fat-tailed behaviour depends partly on timescale

- Some instrument types intrinsically skewed (e.g. high-grade bonds, options)
- Others (e.g. equities) still exhibit fat-tails, timescale dependent
  - E.g. Monthly, weekly and daily returns for major equity market indices (end June 1994 to end Dec 2007)



Source: Threadneedle, S&P, FTSE, Thomson Datastream

malcolm.kemp@nematrian.com

Fat tails and portfolio construction

- Clients want *good performance at an acceptable level of risk*, i.e. efficient use of the available *risk budget*:
  - *Choose the right level of risk to run* (i.e. the risk budget), and
  - *Construct a portfolio* (i.e. choose between assets) to deliver versus this budget
- If *all* opportunities (and combinations) ‘*equally*’ (jointly) fat-tailed
  - Same answers as traditional mean-variance optimisation, but with risk budget adjusted accordingly
- If *different* combinations exhibit *differential* fat-tailed behaviour
  - Portfolio construction ought in principle to change, if you can reliably estimate these differentials (and if investors don’t have quadratic utility functions)

Fat tails – in *joint* return series

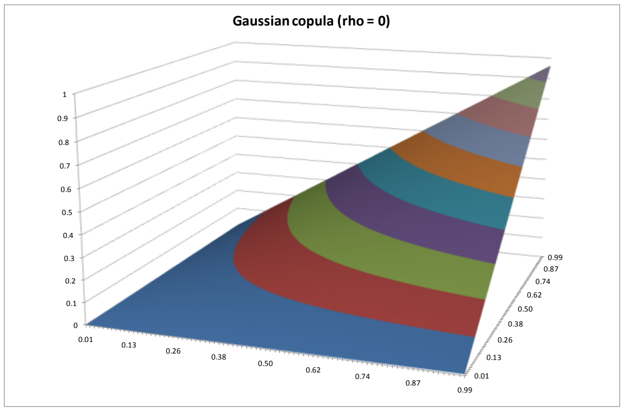
- We can subdivide joint 'fat-tailed-ness' into two parts:
  - How fat-tailed each series is in isolation, i.e. each *marginal* distribution, and
  - How fat-tailed is their co-movement, i.e. their (joint) *copula* function
- *Sklar's theorem*:
  - Suppose that X_1, X_2, ..., X_N are random variables
  - With marginal distribution functions, i.e. individual cumulative probability distribution functions, say, $F_1(x_1)$, $F_2(x_2)$, ..., $F_N(x_N)$
  - And a joint distribution function $F(x_1, x_2, ..., x_N)$
  - Then F can always be characterised by the N marginal distributions and an N-dimensional copula, C, i.e. a function that maps a vector of N numbers each between 0 and 1 onto some value in the range 0 to 1, using:

$$F(x_1, x_2, ..., x_N) = C(x_1, x_2, ..., x_N) \times F_1(x_1) \times F(x_2) \times ... \times F(x_N)$$

malcolm.kemp@nematrian.com

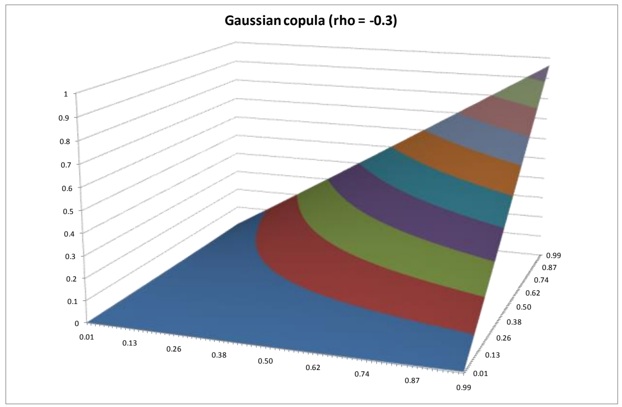
Visualisation of *joint* fat-tailed behaviour

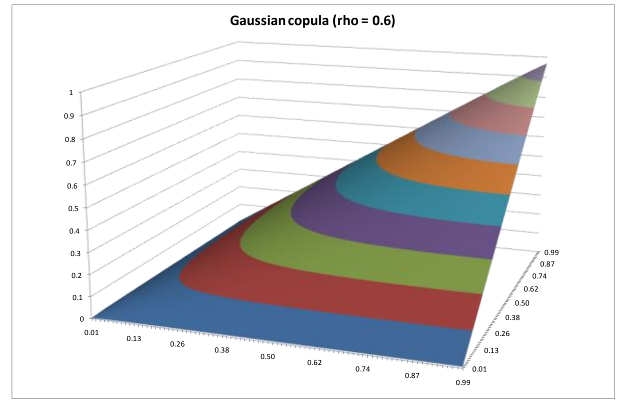
- Visualisation also tricky, easiest seems to be *differences* in copula gradients
- Effectively the same as *fractile-fractile*, i.e. *quantile-quantile box*, plots

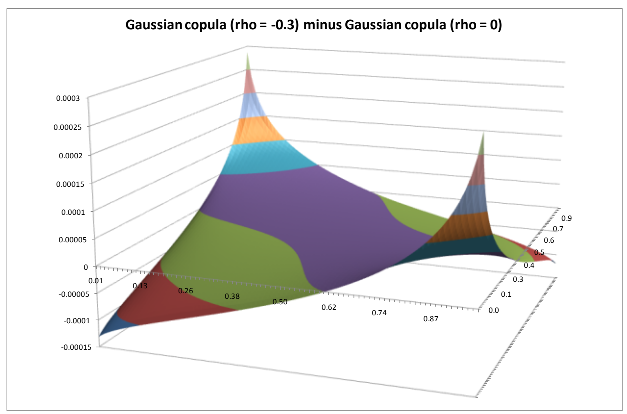


Gaussian copula with rho = 0 also called

the “product” or “independence” copula



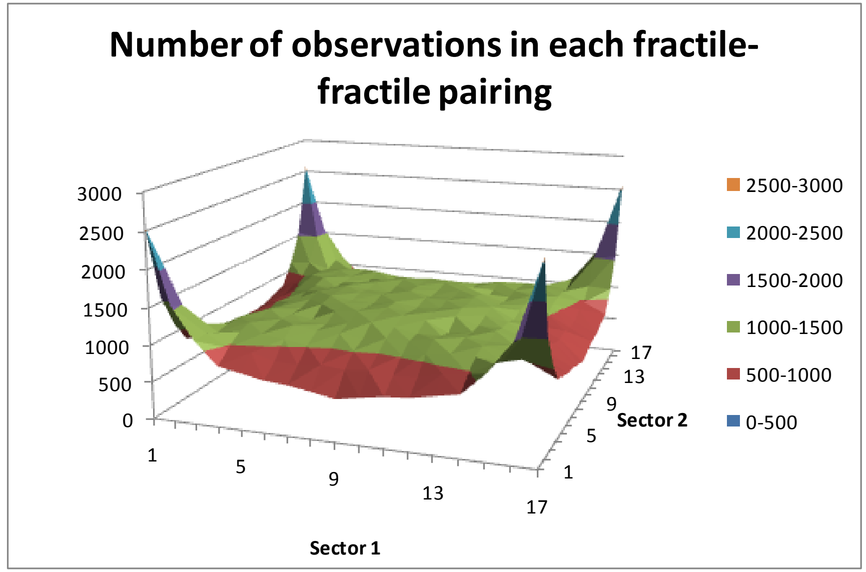


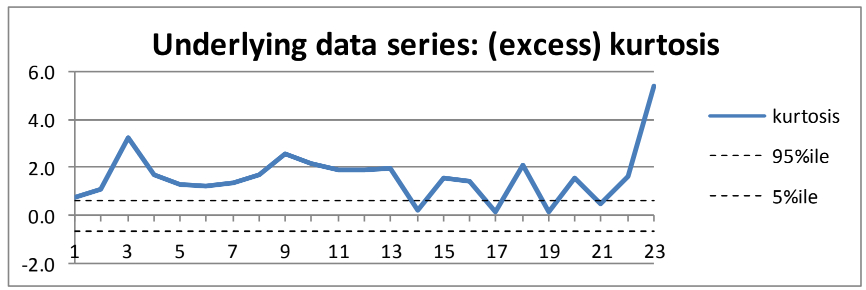


Source: Threadneedle, Nematrian

Quantile-quantile box plots

- E.g. consider two return series in tandem, bucket each into quantile boxes and plot the number of times each quantile box pairing occurs
  - Maybe include all possible unit +/- stances to make four corners of the plot symmetrical?
  - Aggregate plots for all possible sector pairs?
- E.g. chart opposite based on monthly (log) sector relative price movements for 23 MSCI AC-World sectors with complete series between (30/05/96 and 28/02/09)
- Strong evidence that correlations “tend to unity” in stressed times?
- Or merely that we are mixing different distributions together?





Source: Nematrian, Thomson Datastream

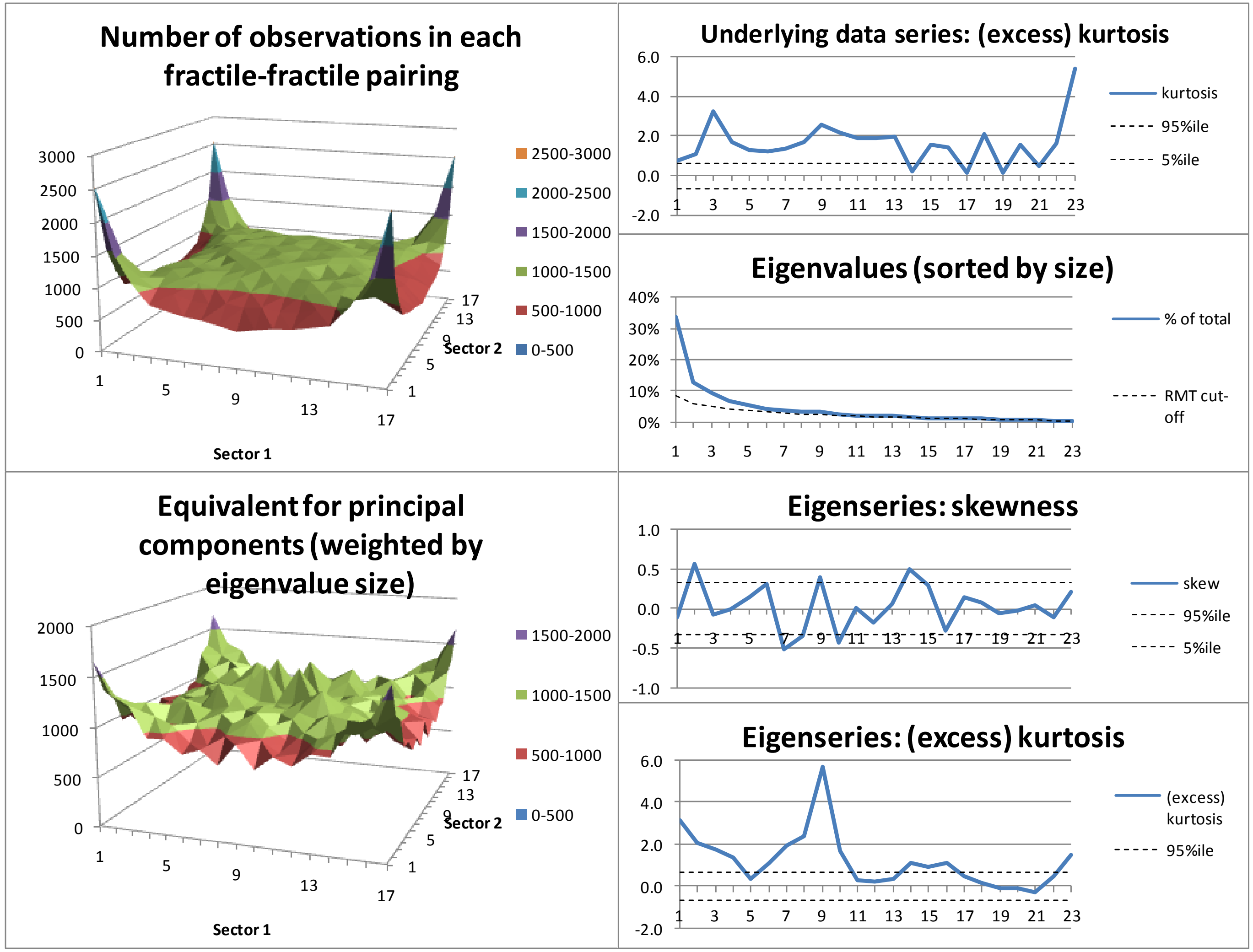
Principal components analysis (PCA)

- A common way of deriving factors that describe observed market behaviour
  - Typically introduced via *eigenvalues* and (normalised) *eigenvectors* of the return covariance matrix, V
  - i.e. solutions to $Vx = \lambda x$; the λ are the eigenvalues, the x are the eigenvectors
- Any instrument's behaviour then expressible as a linear combination of 'signals' associated with these eigenvectors
  - i.e. $r(j, t) = a(j,1) .S_1(\mathrm{t}) + \mathrm{a}(j,2).S_2(\mathrm{t}) + ... + a(j,n).S_n(t)$ for instrument *j*
- Eigenvectors are orthogonal, deemed to be 'different ' drivers of behaviour
- Usually limit merely to 'significant' factors, and add back idiosyncratic risk

malcolm.kemp@nematrian.com

Applying PCA to sector relatives

- Reduced clumping in corners of 2-dimensional principal components co-dependency
- Although not eliminated
- Individual marginal distributions for principal components still exhibit significant (excess) kurtosis



Source: Nematrian, Thomson Datastream

malcolm.kemp@nematrian.com

Principal components – explanation or noise?

- PCA focuses just on *magnitude* of contribution to variance
  - The *trace* of the covariance matrix (i.e. the sum of the variances of each security in the universe) equals the sum of its eigenvalues
- So even the most important PCA components might just be (larger magnitude) random noise
- Usually when asked to explain how something works, we expect the answers (i.e. 'drivers') to be 'causative' or 'informative', like extracting radio signals from background noise
- Is it possible instead to focus on *meaningfulness*?

malcolm.kemp@nematrian.com

Independent components analysis

- This is the basic idea behind *independent components analysis*
- Again assume output (i.e. here, observed returns) come from a linear combination of input signals
- But now focus on *meaningfulness*, e.g. ‘Independence’, ‘non-Normality’ or ‘complexity’
  - ***If source signals have some property X and signal mixtures do not (or have less of it) then given a set of signal mixtures we should attempt to extract signals with as much X as possible, since these extracted signals are then likely to correspond as closely as possible to the original source signals***

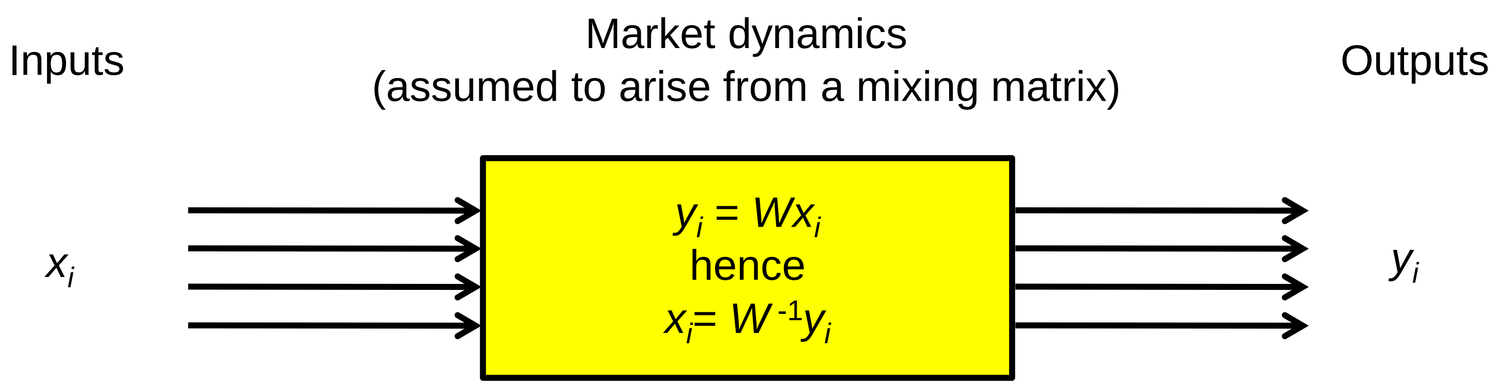
E.g. Non-Normality and Projection pursuit

- Suppose we think that ‘behaviour’ that is highly non-Normal is likely to be ‘interesting’ (i.e. worth exploring further) and probably ‘meaningful’
- Suppose we also associate non-Normality with (excess) kurtosis
- Conveniently:
  - All linear combinations of independent distributions have a kurtosis less than or equal to the largest kurtosis of any of the individual distributions
  - Kurtosis is scale independent (i.e. $k.x$ has the same kurtosis as x if k is a scalar)

malcolm.kemp@nematrian.com

Projection pursuit algorithm (1)

- If y are observed output results and x are supposed input signals then
  - We have assumed that $y = Wx$ for some W, the *mixing* matrix
  - Hence $x = Ay$, for some A, the *unmixing* matrix, where $A = W^{-1}$



Projection pursuit algorithm (2)

- Algorithm:
  - Choose an importance criterion, e.g. kurtosis
  - Choose from set of all possible unmixing coefficients the one that provides the deemed input signal (of unit strength) that maximises the importance criterion
  - Deem this to be an actual input signal, moreover the most important one
  - Take away any contribution from this signal to the output results
  - Repeat, until no further signals extracted by the algorithm appear significant / meaningful

Independent components analysis (ICA)

- Can be thought of as an ‘all-at-once version’ of projection pursuit
- Involves working out the maximum likelihood estimator of the entire unmixing matrix, assuming the signals are independent
  - Needs an *a priori* distributional form to assume for the individual signals
  - Often choose one with very high kurtosis, e.g. $p_s = \left(1 - \tanh(s)\right)^2$
- Or ‘infomax’ ICA
  - Identify how ‘surprising’ (and therefore meaningful) is the observed data given some a priori multivariate distribution in which each individual series is independent, measured using, say, *relative entropy* (aka *Kullback-Leibler divergence*)

Blending together PCA and ICA

- PCA offers magnitude, ICA offers meaningfulness
- Ideally we would like the best of both
  - But one focuses on variance, the other on (e.g.) kurtosis?
- Fortunately, it is possible to blend the two, e.g. by
  - Recasting PCA along the lines of projection pursuit (with an importance criterion involving maximising contribution to variance), and then
  - Choosing a different importance criterion that blends together variance and (e.g.) kurtosis

malcolm.kemp@nematrian.com

Identifying Principal Components one at a time

- Most PCA algorithms calculate all eigenvectors/eigenvalues simultaneously
- However, suppose V is an n x n covariance matrix with (sorted) eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ (largest is λ_1) and corresponding (normalised) eigenvectors $q_1, q_2, ..., q_n$.
- Suppose our importance criterion involves $f(a) = a^T V a$, and $|a|=1$
- Then a can be expressed as $a = a_1 q_1 + ... + a_n q_n$ with $a_1^2 + ... + a_n^2 = 1$
- And $f(a) = a_1^2 \lambda_1 + ... + a_n^2 \lambda_n$ so $f(a)$ is maximised when $a = q_1$
- And eigenvectors are orthogonal, so removing one from output signals leaves remainder still to be extracted

malcolm.kemp@nematrian.com

Blending PCA with ICA

- Use a blended importance criterion, e.g.
  - Maximise $f(a) = \sigma(1+cK)$, across possible *a* with $|a|=1$, where:
  - K is the kurtosis of *a*, $\sigma^2 = a^T V a$
  - *c* is some constant that represents a trade-off between concentrating on maximising variance and concentrating on maximising kurtosis (if $c = 0$ then equivalent to PCA, if *c* is large then will approximate ICA)
- Can be re-expressed to be akin to the Cornish-Fisher 4th moment asymptotic expansion for estimating quantiles of a Non-Normal distribution (with zero skew)

$$y = m + \sigma\left(x + \frac{K(x^3 - 3x)}{24}\right)$$

  - E.g. 99.5%ile, $x = N^{-1}(0.995) = -2.576$ and $c = 0.39$

Extreme events appear to be very important!

	PCA			Blended PCA/ICA			c.f. ICA	
Component	StdDev	Kurt	Criterion	StdDev	Kurt	Criterion	StdDev	Kurt
1	10.6%	3.1	10.6%	8.3%	14.9	56.6%	4.5%	24.2
2	6.5%	2.1	6.5%	4.9%	24.9	52.7%	4.2%	23.5
3	5.6%	1.7	5.6%	5.0%	22.1	48.0%	4.5%	18.1
4	4.8%	1.4	4.8%	4.5%	14.7	30.1%	6.9%	16.2
5	4.2%	0.4	4.2%	4.3%	15.0	29.7%	4.2%	15.0
6	3.7%	1.1	3.7%	4.8%	9.2	22.1%	4.2%	13.7
Av (top 6)	5.9%	1.6	**5.9%**	5.3%	16.8	**39.9%**	4.7%	18.5
Av (all 23)	3.2%	1.2	3.2%	3.6%	8.2	17.5%	3.7%	9.1

- Sizes of ‘1 in 200’ events potentially underestimated by PCA by 4- or 5-fold
  - If portfolio built on the basis of ‘meaning’ (e.g. if actively managed)

malcolm.kemp@nematrian.com

Limitations

- Both PCA and ICA assume that observations are (time stationary) linear combination mixtures
  - i.e. $a = a_1 q_1 + ... + a_n q_n$ and $y = Wx$
  - But not all mixtures are of this form
- Consider distributional mixtures, y drawn from distribution D_1 with probability p_1, from distribution D_2 with probability p_2 etc.
  - These typically result in fat-tailed behaviour
- Very important special case is modelling a time-varying world
  - c.f. GARCH, regime shifts etc.
- Also, Cornish-Fisher (and hence kurtosis) may misestimate sizes of fat tails

Summary

- PCA concentrates on magnitude (maximise aggregate contribution to variance)
- ICA concentrates on meaningfulness (and thus comes in more flavours, but often seeks to maximise kurtosis)
- In either case, most important component can be extracted by *projection pursuit* maximising a particular importance criterion
- So we can blend the two together, using a blended importance criterion
- But further refinements needed to cater for time-varying volatility and other behaviour linked to distributional mixtures
  - Such mixtures are important sources of fat-tailed behaviour in practice

malcolm.kemp@nematrian.com

Appendix A: Time-varying volatility in *single* return series

- Fat tails involve deviation from Normality

- Hence at least some of the higher *cumulants* (moments), aka *semi-invariants*, of the distribution, e.g. skew and (excess) kurtosis, must deviate from zero (Normality)

$$mean = \mu = E(x)$$

$$standard\ deviation = \sigma = E\left((x-\mu)^2\right)$$

$$skew = \gamma_1 = E\left(\left(\frac{x-\mu}{\sigma}\right)^3\right)$$

$$(excess)\,kurtosis = \gamma_2 = E\left(\left(\frac{x-\mu}{\sigma}\right)^4\right) - 3$$

- Use of these (and possibly other higher cumulants) is most common way of analysing and coping with fat tails, but it is not necessarily the best approach

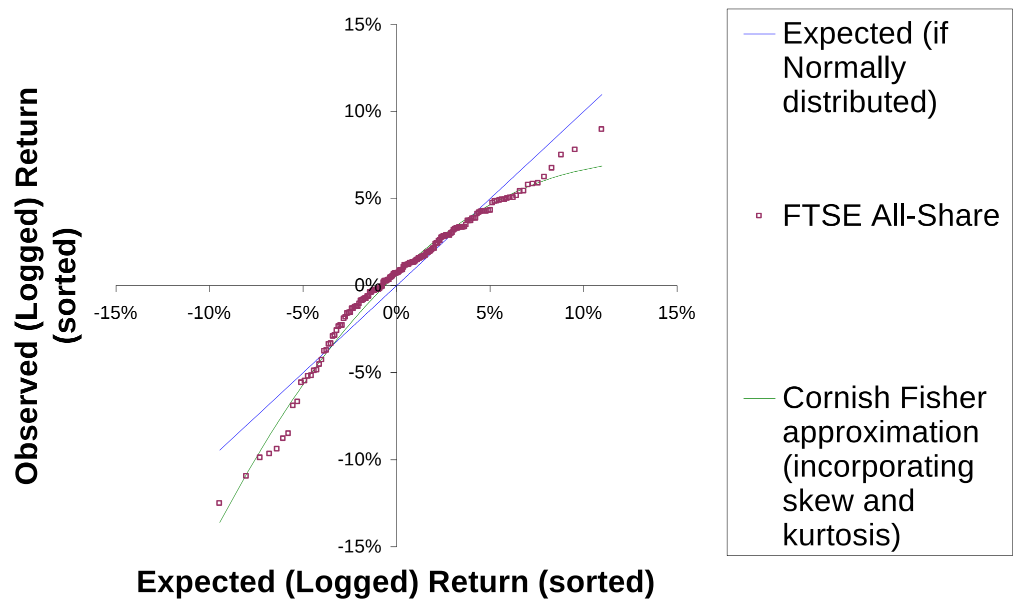
malcolm.kemp@nematrian.com

Interpretation via Cornish-Fisher asymptotic expansion

- Cornish-Fisher (4th moment version) estimates distributional form from merely the first 4 moments, i.e. mean, standard deviation, skew and (excess) kurtosis
- Regularly appears in risk management academic literature
- For standardised returns (zero mean, unit standard deviation), quantile-quantile plot estimated via a cubic equation:

$$y_{CF4}(x) = x + \frac{\gamma_1\left(x^2 - 1\right)}{6} + \frac{3\gamma_2\left(x^3 - 3x\right) - 2\gamma_1^2\left(2x^3 - 5x\right)}{72}$$

Monthly returns (end Jun 1994 to end Dec 2007)

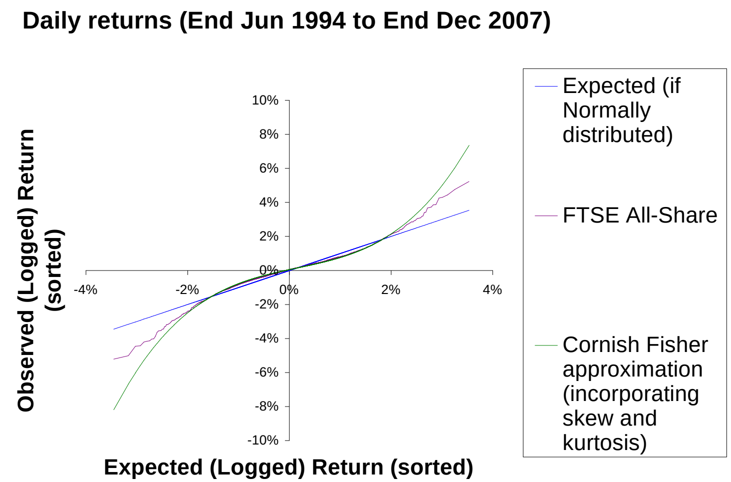


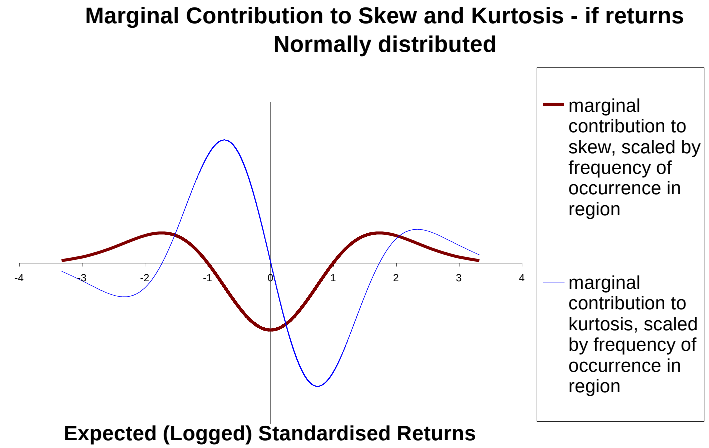
Source: Threadneedle, FTSE, Thomson Datastream

malcolm.kemp@nematrian.com

Flaws in Cornish Fisher (and hence in skew/kurtosis)

- Doesn't model index return distributions particularly well
  - Particularly parts risk managers might be most interested in, i.e. downside tails
- Computation gives less weight to tail observations (most observations are in middle of the distribution)
- Lacks a desirable stability criterion
  - Applying CF twice can lead to a more extreme distribution



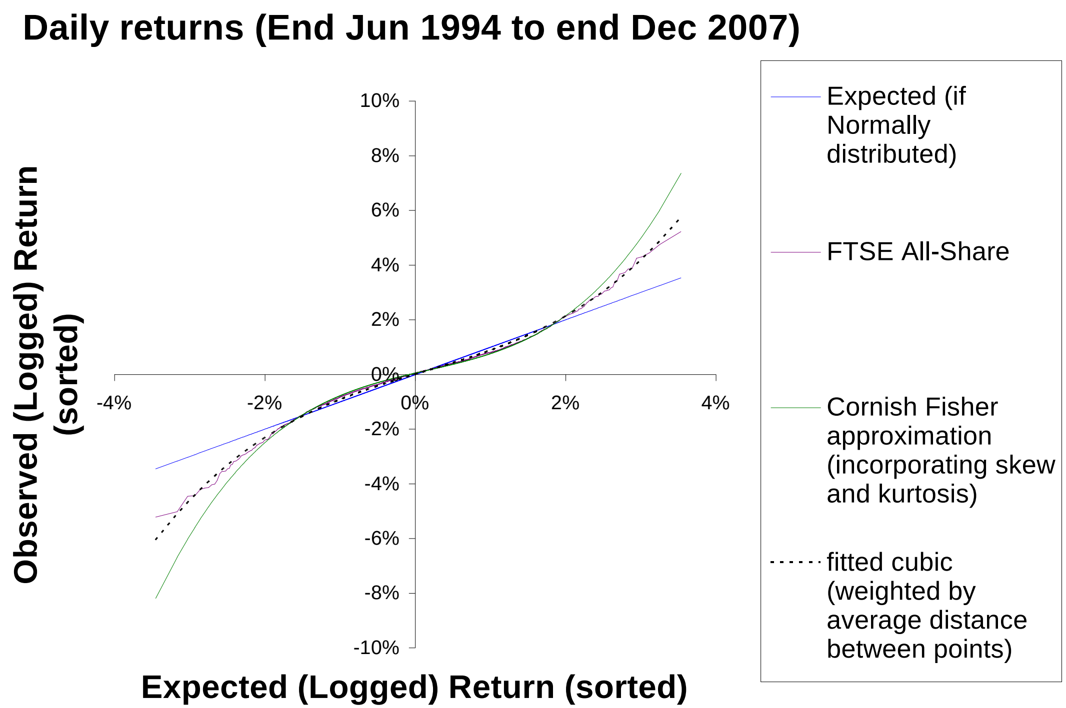


Source: Threadneedle, FTSE, Thomson Datastream

malcolm.kemp@nematrian.com

A better approach?

- Fit quantile-quantile plot directly?
  - E.g. with a cubic curve
- Calculation is more complex
- Skew and kurtosis:
  - Do not need data to be ordered
  - Come pre-canned in Microsoft Excel, SKEW() and KURT()



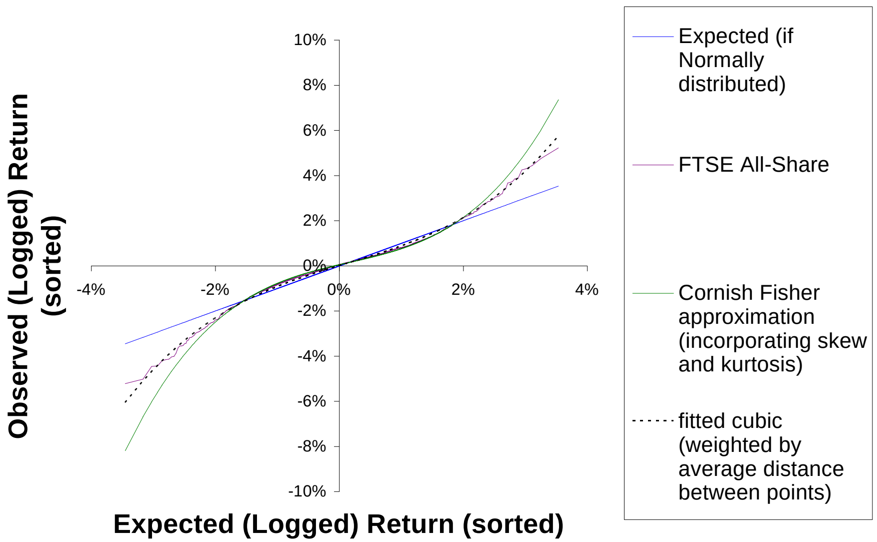
Source: Threadneedle, Thomson Datastream

malcolm.kemp@nematrian.com

Time varying volatility explains some market index fat tails, particularly on the upside

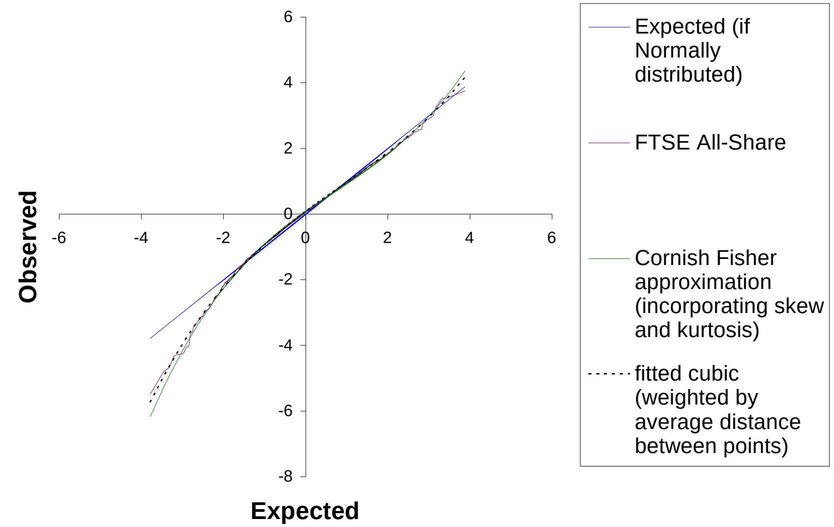
Raw Data

Daily returns (End Jun 1994 to end Dec 2007)



With Short-term Volatility Adjustment

Daily returns (end Jun 1994 to end Dec 2007, scaled by 50 business day trailing daily volatility)



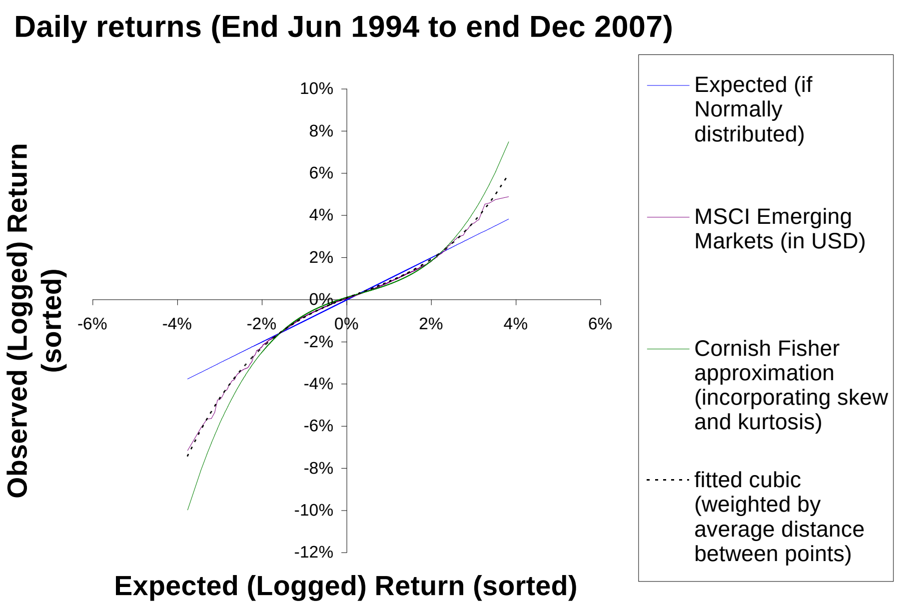
Average extent to which tail exceeds expected level (average of 6 most extreme outcomes)				
	Downside (%)		Upside (%)	
	Unadj	Adj for vol	Unadj	Adj for vol
FTSE All-Share (in GBP)	54	41	42	3
S&P 500 (in USD)	68	70	50	7
FTSE Eur ex UK (in EUR)	48	53	54	-3
Topix (in JPY)	54	72	42	39

Source: Threadneedle, FTSE, Thomson Datastream

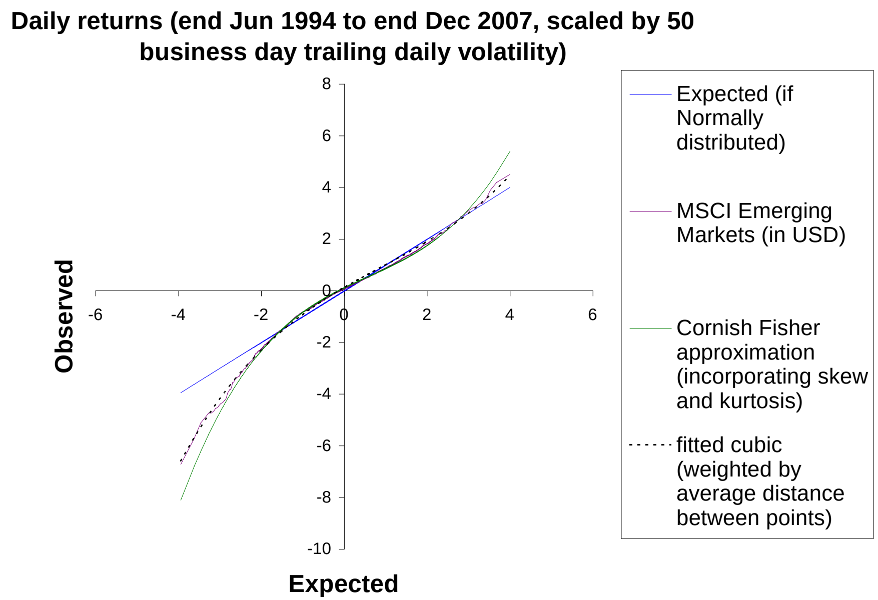
malcolm.kemp@nematrian.com

Not just a developed market phenomenon

Raw Data



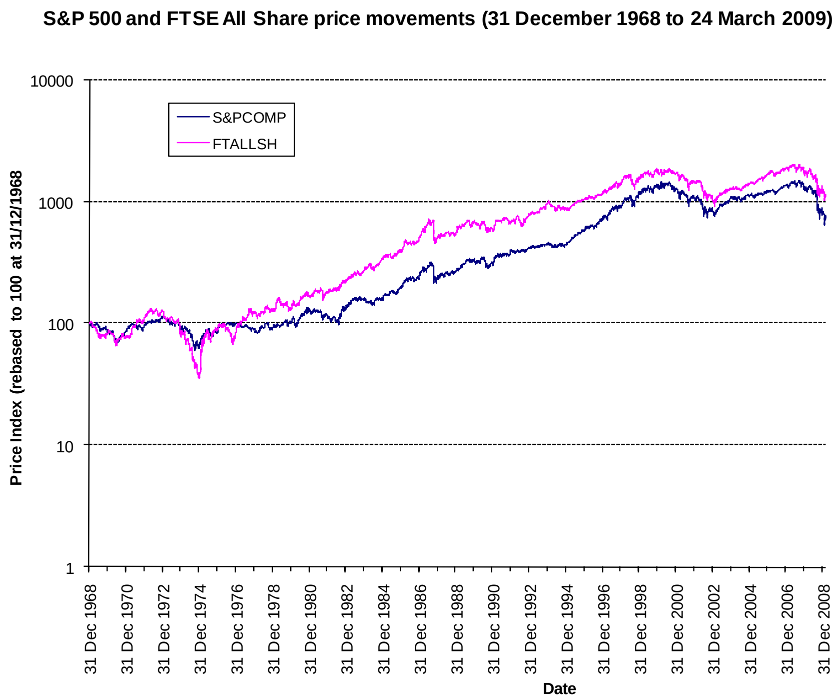
With Short-term Volatility Adjustment

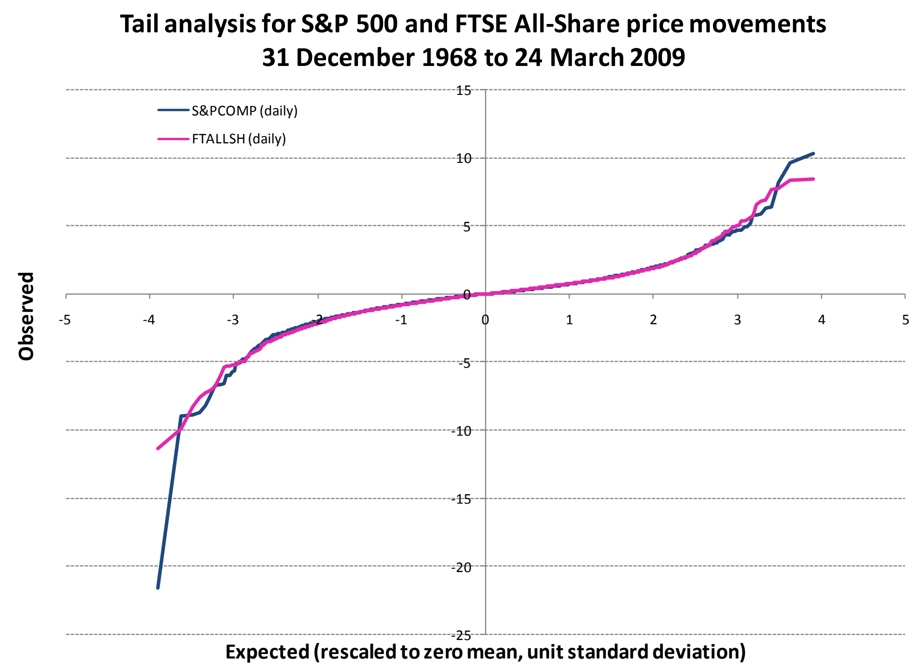


Source: Threadneedle, FTSE, Thomson Datastream

malcolm.kemp@nematrian.com

More periods give more scope for extreme events



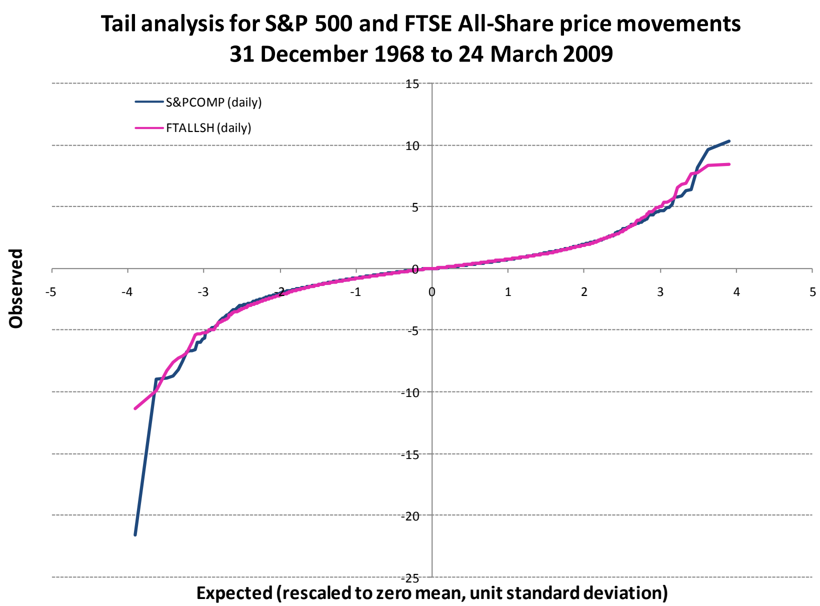


Source: Threadneedle, S&P, FTSE, Thomson Datastream

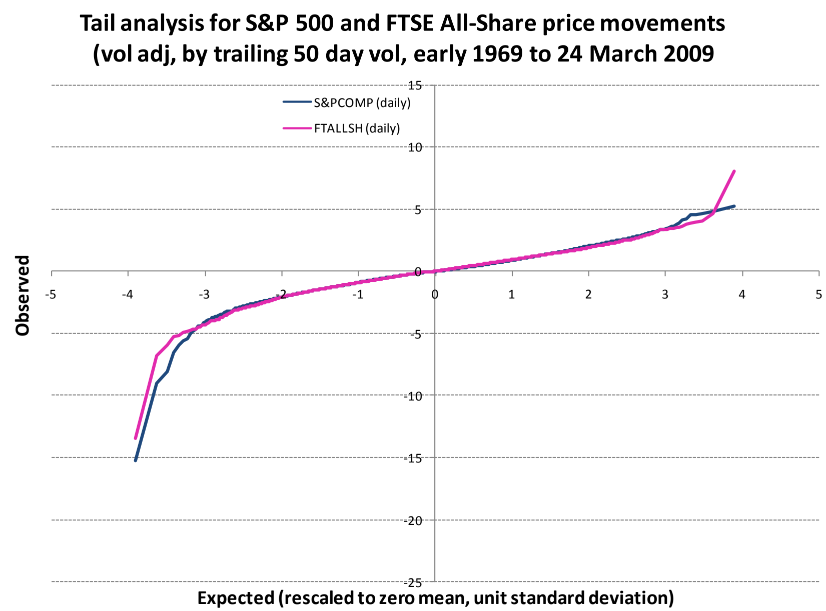
malcolm.kemp@nematrian.com

Time-varying volatility remains an important contributor

Raw Data



With Short-term Volatility Adjustment



Source: Threadneedle, S&P, FTSE, Thomson Datastream

malcolm.kemp@nematrian.com

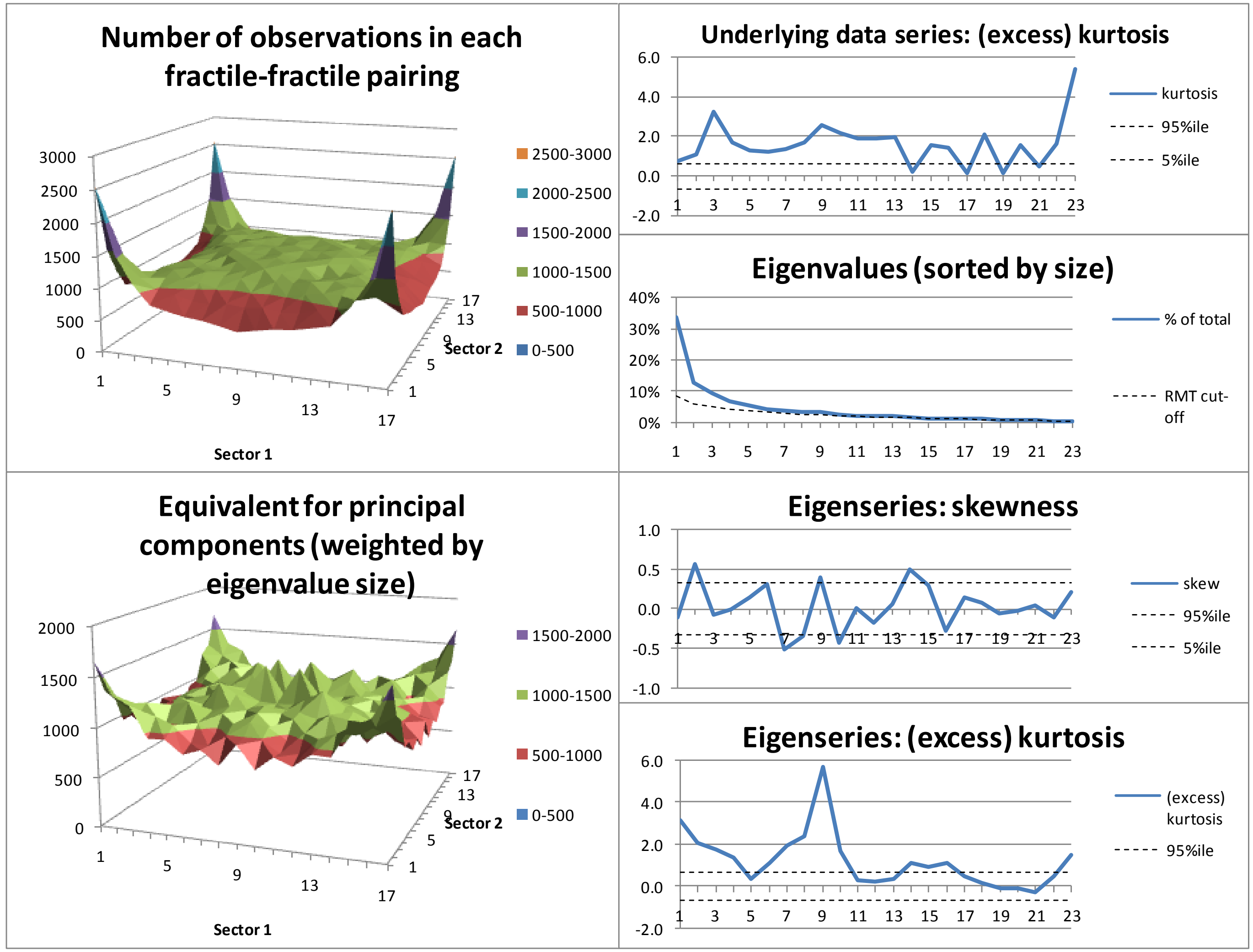
Appendix B: Time-varying volatility in *joint* return series

- Quantile-quantile box plot had peaks in four corners
- One reason is that chart includes a mixture of distributions
  - Different pairs of sectors have different correlations hence different distributions
- We can eliminate this effect by focusing on principal components
  - Orthogonal by construction
  - Hence all disjoint pairs of principal components have the same (i.e. zero) correlation

malcolm.kemp@nematrian.com

Applying PCA to sector relatives

- Reduced clumping in corners of 2-dimensional principal components co-dependency
- Although not eliminated
- Individual marginal distributions for principal components still exhibit significant (excess) kurtosis



Source: Nematrian, Thomson Datastream

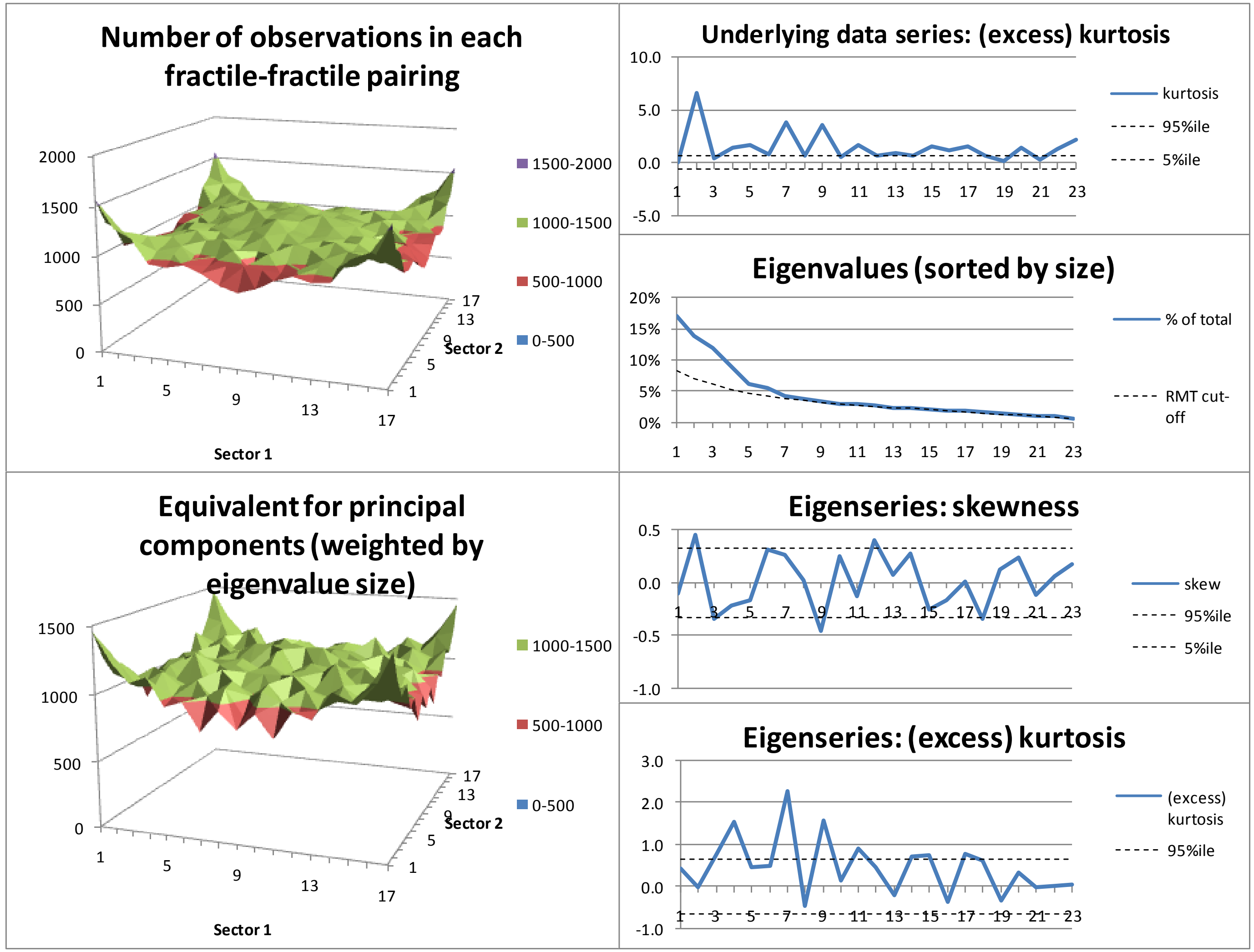
malcolm.kemp@nematrian.com

Adjusting for time-varying volatility in *joint* return series

- Possible ways of adjusting for recent past time-varying volatility include
  - *Longitudinal*: adjust each series in isolation by a *different* (time-varying) factor dependent its recent past volatility, or
  - *Cross-sectional*: adjust every series by the *same* (time-varying) factor dependent on the average spread of returns across the sectors in the recent past
  - Using *contemporaneous* data, such as implied volatilities and correlations (not analysed further here, discussed in more detail in “Market Consistency”)
- E.g. use rolling 12 month window for both longitudinal approach and cross-sectional approach
  - Choice of window a trade-off between “immediacy” and sample error

Longitudinal time-varying volatility adjustment

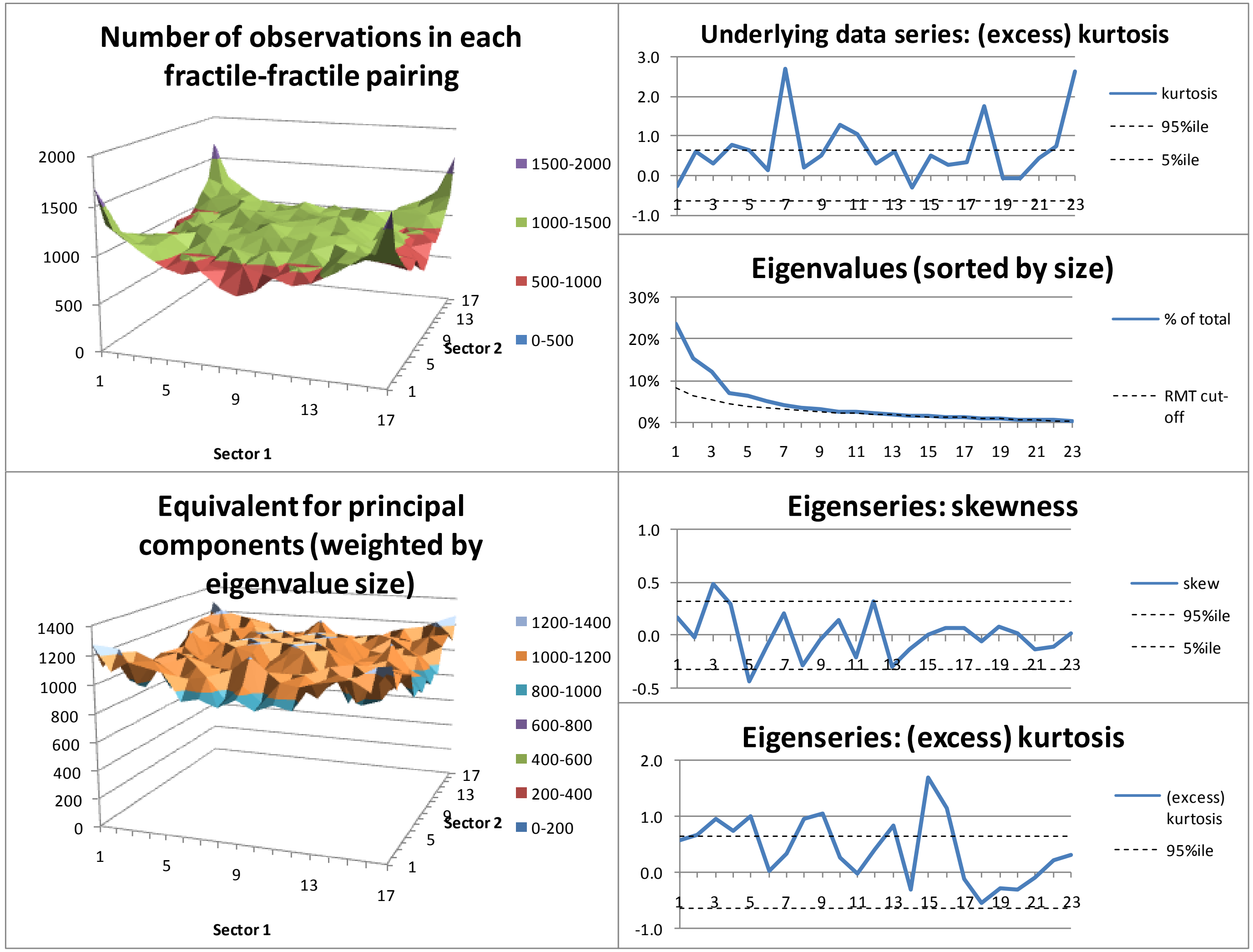
- Flatter 2-dimensional co-dependency
- Less (excess) kurtosis in marginals, particularly for principal components



Source: Nematrian, Thomson Datastream

Cross-sectional time-varying volatility adjustment

- Even flatter 2-dimensional co-dependency for principal components
- Even less (excess) kurtosis in marginals
- Although average (excess) kurtosis still noticeably positive
- Particularly for “significant” principal components



Source: Nematrian, Thomson Datastream

Back-testing time-varying volatility adjustments

- Calculate through time observed return divided by estimated tracking error
  - Each month, estimate out-of-sample covariance matrix and hence tracking error using prior monthly relative returns. Start 36 months into dataset. Apply to 100 x 23 random portfolios (100 with 1 sector position, 100 with 2 sector positions etc.)
  - Calculate percentiles and moments for observed spread of this statistic
- Cross-sectional adjustment not quite as effective as we might have hoped
  - Refine with "contemporaneous" estimates of volatility and average correlation?

	kurtosis	90%ile	99%ile	99.9%ile
Unadjusted data	2.3	1.2	2.7	4.3
Longitudinal adjustment	1.2	1.2	2.5	3.8
Cross-sectional adjustment	0.8	1.3	2.6	3.8
c.f. expected if Gaussian	0.0	1.3	2.3	3.1

Source: Nematrian, Thomson Datastream

malcolm.kemp@nematrian.com

Other sources of fat tails?

- Some fat tails still seem to come "out of the blue"
  - E.g. Quant funds in August 2007
  - Too many investors in the same crowded trades? Behavioural finance implies potentially unstable
  - For less liquid investments , impact may be via an apparent shift in price basis
  - Should only affect specific investors?
- System-wide equivalents via leverage?
  - Leverage introduces/magnifies *liquidity* risk, *forced unwind* risk and *variable borrow cost* risk

malcolm.kemp@nematrian.com

Important Information

malcolm.kemp@nematrian.com