Tail fitting probability distributions for risk management purposes

Malcolm Kemp, Managing Director, Nematrian Limited

Presentation to IFoA Pensions, Risk and Investment Conference, Edinburgh

1 June 2016

Agenda

- Why is tail behaviour important?
- Traditional Extreme Value Theory (EVT) and its strengths and weaknesses
- Refinements allowing fitting of any distribution to tail data
- Other uses of such techniques

Agenda

- Why is tail behaviour important?
- Traditional Extreme Value Theory (EVT) and its strengths and weaknesses
- Refinements allowing fitting of any distribution to tail data
- Other uses of such techniques

Why is tail behaviour important? (1)

- Forecasting of any sort is challenging:
  - *"Prediction is very difficult, especially if it's about the future."* Nils Bohr
  - *"If you can look into the seeds of time, and say which grain will grow and which will not, speak then unto me."* William Shakespeare
  - *"This is the first age that's ever paid much attention to the future, which is a little ironic since we may not have one."* Arthur C. Clarke
- Extreme events, the events in the tail of the distribution, are the most difficult to forecast, but are also the ones that have the most impact
  - C.f. the impact of the 2007-09 Credit Crisis on modern financial regulation

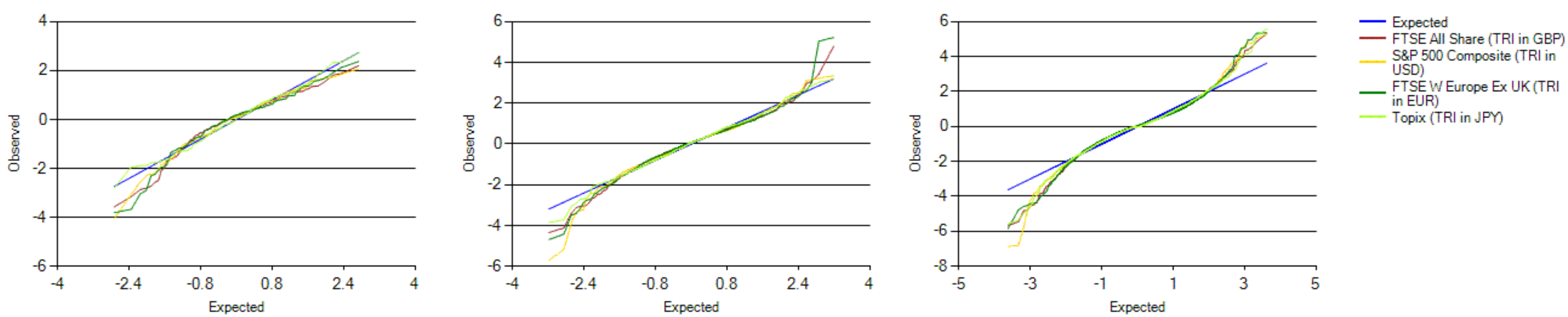
Why is tail behaviour important? (2)

- Taking due account of the possibility of extreme events occurring is important but also challenging for many market professionals
  - *Insurers:* Solvency II mandates 1 in 200 year VaR, but we do not have 200 years of relevant historical data
  - *Pension funds:* Practical likelihood of beneficiaries receiving all that they have been promised depends heavily on hopefully rare extreme credit events, e.g. the sponsor defaulting
  - Asset managers. Clients and firms themselves naturally want to understand downside risks and their potential causes
    - Even if need to balance *risk* versus *reward* means that there is a risk we can give *too much* emphasis to the downside
  - *Banks:* E.g. many recent operational risk losses have been much larger than losses previous models had considered plausible

Why is tail behaviour important? (3)

- Many return series (even well diversified ones) seem to exhibit fat-tails, often best seen using quantile-quantile plots as below, see also Appendix A.
  - Some instrument types intrinsically skewed (e.g. high-grade bonds, options)
  - Others (e.g. equities) still exhibit fat-tails, particularly higher frequency data
- Some of this is due to the time varying nature of the world, see Appendix B

(1) Monthly returns (2) Weekly returns (3) Daily returns



Nematrian

Agenda

- Why is tail behaviour important?
- Traditional Extreme Value Theory (EVT) and its strengths and weaknesses
- Refinements allowing fitting of any distribution to tail data
- Other uses of such techniques

Extreme Value Theory (EVT)

- Traditional EVT is an enticing prospect
  - Appears to offer a mathematically sound way of identifying shape of the ‘tail’ of a (univariate) distribution, and hence identifying likelihood of extreme events
  - Capital adequacy seeks to protect against (we hope) relatively rare events
  - Insurance and credit risk pricing can be dominated by potential magnitude and likelihood of large losses
- But bear in mind
  - Inherent unreliability of extrapolation, including into tail of a probability distribution
  - Possibility (indeed probability) that the world is not time stationary
  - Portfolio construction is inherently multivariate, involves choosing between alternatives

Traditional EVT results

- Suppose interested in risk measures relating to losses, x_j. EVT aims to supply two closely related results:

1. *Less relevant to risk management*: Distribution of ‘block maxima’ (or ‘block minima’), i.e. maximum value of x_j in blocks of m observations, tends to a generalised extreme value (GEV) distribution

2. *More relevant to risk management*: Distribution of ‘threshold exceedances’ (i.e. ‘peaks-over-thresholds’) tends to a generalised Pareto distribution (GPD), Here u is a predetermined high threshold and we focus on realisations of x_j that exceed u, i.e. on $y_j = x_j - u | x_j > u$, which if EVT applies means that the distribution of x_j has a cumulative distribution function $G_{\mu,\sigma,\xi}(z)$ for suitable μ, σ, ξ where:

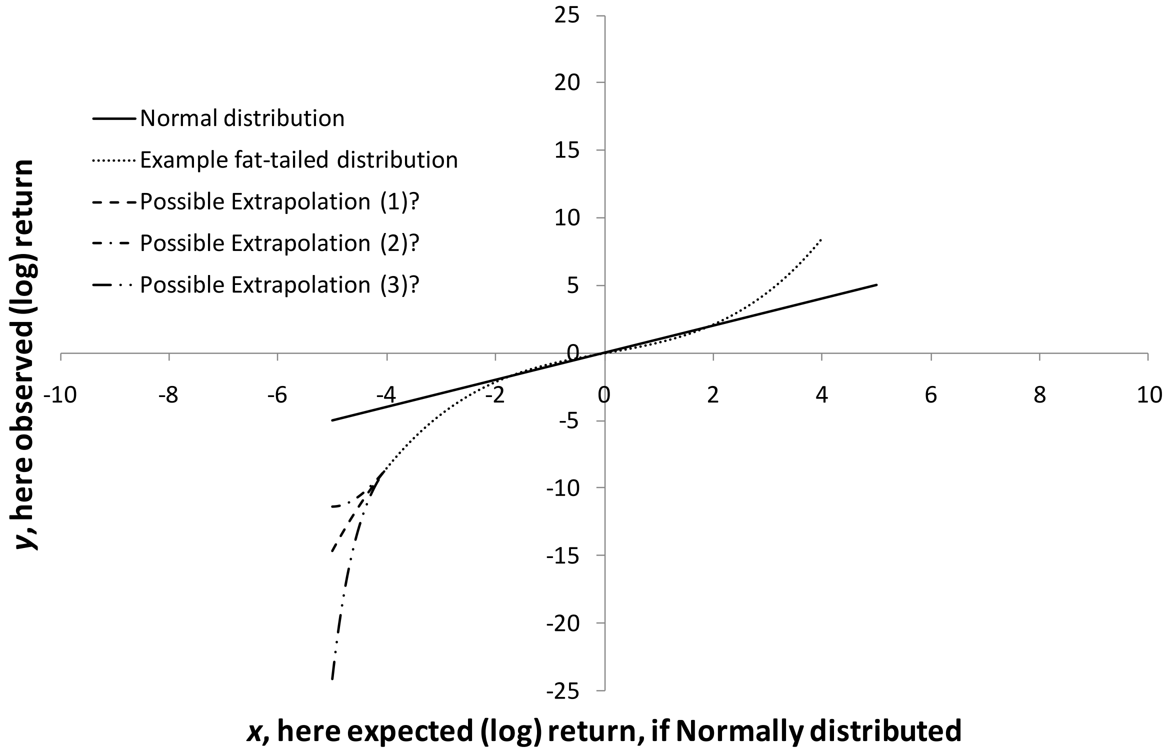
$$G_{\mu,\sigma,\xi}(x) = \begin{cases} 1 - (1 + \xi z)^{-1/\xi} & \xi \neq 0 \\ 1 - \exp(-z) & \xi = 0 \end{cases} \qquad \text{where} \qquad z = \frac{x - \mu}{\sigma}$$

But is EVT the only or best way of fitting the tail?

- In traditional EVT we assume that the limiting distribution of observations in the tail of the distribution, $F_u(y)$, is a generalised Pareto distribution (GPD)
  - Problem of estimating F and hence behaviour in the tail (e.g. tail quantiles) then in effect reduces to problem of estimating from the data the μ, σ and ξ that provide the best fit GPD to the data
  - Can be done using mean excess functions, maximum likelihood (ML) estimation, method of moments etc.
- But equally we could fit to the relevant part of the QQ-plot using any other reasonable curve fitting approach
- As long as the fit is feasible, does it have to tend to a GPD in the limit?

Potential weaknesses of EVT

- EVT seems very helpful and seems to characterise limiting distributions very succinctly
- But requires (arguably quite strong) regularity conditions that may not be satisfied
- At issue is potential unreliability of **extrapolation**
  - E.g. Press et al. (2007)



Source: Nematrian

Agenda

- Why is tail behaviour important?
- Traditional Extreme Value Theory (EVT) and its strengths and weaknesses
- Refinements allowing fitting of any distribution to tail data
- Other uses of such techniques

Tail-weighted distribution fitting

- One possible alternative is simply to fit a curve, e.g. a polynomial, directly to the relevant tail of the observed QQ-plot, selecting its coefficients using e.g. weighted least squares, to target the best fit within the tail
  - But this does not always return a feasible probability distribution and may be difficult to interpret
- Probably better is to use ‘tail weighted’ approaches, e.g. tail weighted least squares or tail weighted maximum likelihood, see Kemp (2013). Implemented via web functions named “MnProbDistTW…” in the Nematrian function library
  - Always returns a feasible probability distribution, as the ‘best fit’ (in the tail) is automatically constrained to fall within a specified family of valid distributions
  - Maximum likelihood variant inherits the nice asymptotic properties of maximum likelihood estimation and if equally weight fit across *whole* distribution then same as traditional MLE

Tail weighted maximum likelihood (TWMLE)

- We re-express maximum likelihood to refer to the ordered observations:

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

- E.g. by writing the log-likelihood as:

$$\log L = \sum_i q_i \quad where \quad q_i = \log\left(\frac{\left(1 - F\left(x_{(i)}|\theta\right)\right)^{n-i}}{\left(1 - F\left(x_{(i-1)}|\theta\right)\right)^{n-i+1}} \frac{(n-i+1)}{i} f\left(x_{(i)}|\theta\right)\right)$$

- Instead of maximising log likelihood we maximise e.g. $\log L^* = \sum_i w_i q_i$
  - For some suitable weights, w_i, e.g. 1 if in tail, 0 otherwise
  - Allowing us to leverage intrinsic appeal of maximum likelihood estimation
  - Some subtleties if quantiles not equally spaced and complete

Tail weighted least squares (TWLS)

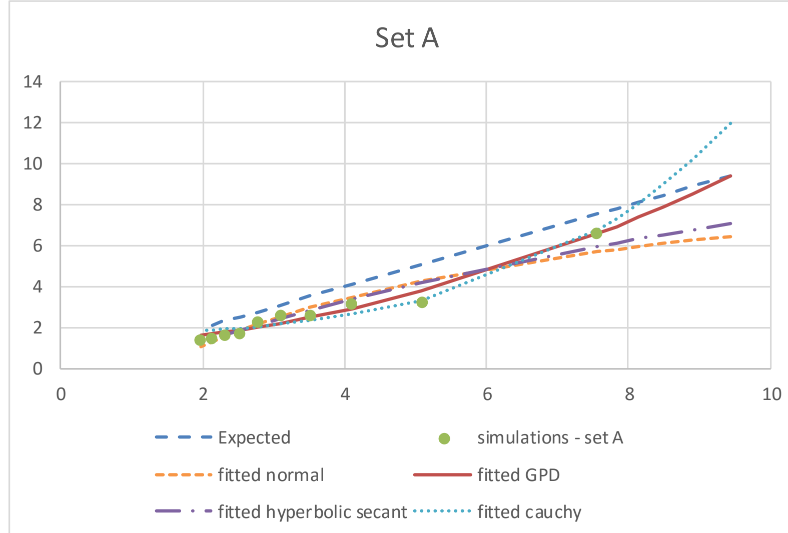
- Still use ordered observations: $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$
- But now arrange for observed and expected quantiles to align ‘as closely as possible’, with the favouring specific quantiles, e.g. ones in the tail
- I.e. minimise $\sum_i w_i C_i$ where:

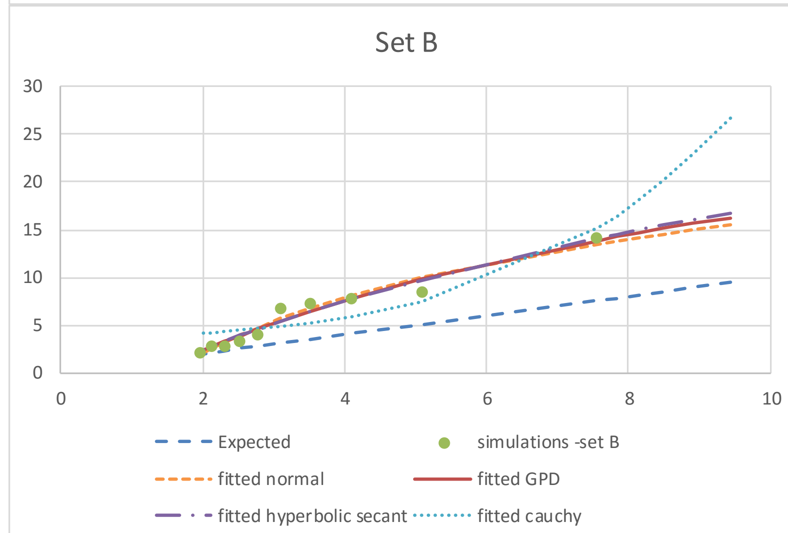
$$C_i = \left(X_{(i)} - F^{-1}\left(\frac{i - 1/2}{n} | \theta \right) \right)^2$$

- Meaning to assign to weights and asymptotic properties no longer so obvious

Example analysis

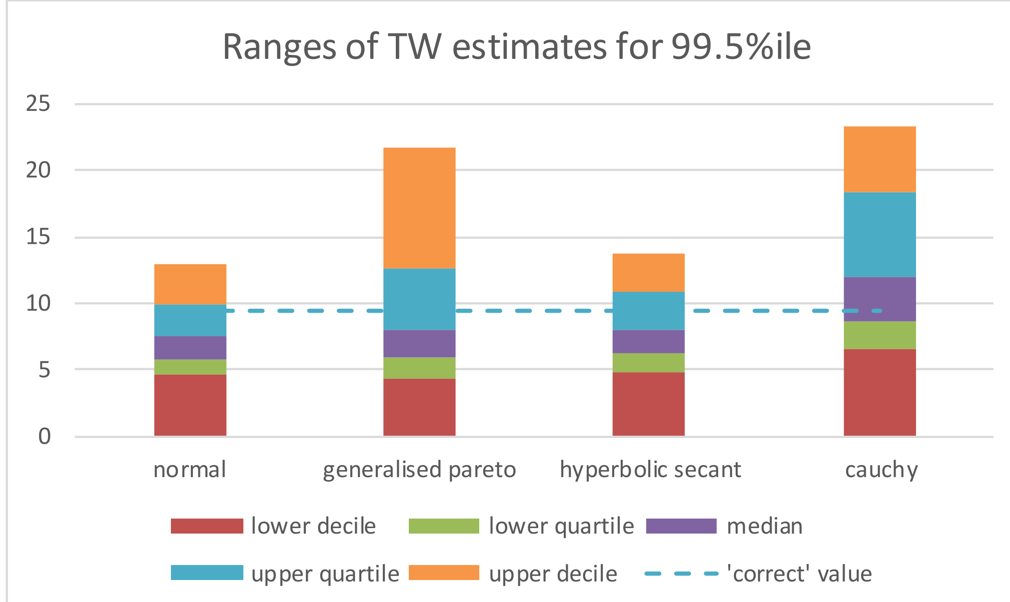
- Suppose want to estimate 99.5%ile, but only have 50 observations (so can't avoid extrapolation)
  - Say observations come from a GPD with $\mu = 0, \sigma = 1, \xi = 0.2$. Expected quantiles shown by blue dashed line
  - Use TWLS applied to selected distributional families (including GPD) to extrapolate 99.5%ile from 10 highest observations (i.e. top 20 percentiles)
- Two different random draws (Set A and Set B) each of 50 observations, 99.5%ile is right hand end of chart
  - GPD good fit for Set A, less good for set B





Key takeaways

- Nice mathematical idea
- Unfortunately, extrapolation is inherently problematic however sophisticated the mathematics we throw at the problem
  - Randomly simulate 100 such draws of 50 observations and re-estimate. Range of extrapolated answers is wide
  - Even for GPD, the distribution the observations are assumed to come from! Indeed, other distributions such as hyperbolic secant perhaps a better fit.



	Normal	Generalised Pareto	Hyperbolic secant	Cauchy
‘Correct’ value	9.4	9.4	9.4	9.4
Lower decile	4.6	4.4	4.8	6.5
Lower quartile	5.8	5.8	6.2	8.6
Median	7.5	8.0	8.0	12.1
Upper quartile	9.9	12.6	10.8	18.3
Upper decile	12.9	21.6	13.8	23.4

Agenda

- Why is tail behaviour important?
- Traditional Extreme Value Theory (EVT) and its strengths and weaknesses
- Refinements allowing fitting of any distribution to tail data
- Other uses of such techniques

Fitting distributions around specific quantiles

- Maybe we 'know' specific quantile values
  - E.g. because we trust expert judgement and these experts have for example identified the upper decile, median and lower decile of the distribution
- If we have the same number of quantiles as we have parameters to fit then can use e.g. TWLS to fit quantiles exactly (if quantiles are feasible)
  - E.g. lower quartile = -6, upper decile = +5 is fitted by $N(-2.21, 5.62)$
  - Likewise if fewer quantiles and we fix sufficient numbers of distributional parameters
- If we have more quantiles than we have available parameters then unlikely to get exact fit to all quantiles, but can select between possible 'good' alternatives by giving suitable weights to fit at different quantile points

Quantile interpolation (1)

- Can also use technique for interpolation rather than extrapolation
  - I.e. fit to a quantile within range of (simulated) observations, e.g. as part of an internal model, asset-liability modelling or other simulation exercise
  - Time to carry out a single simulation may be material, so any improvement in accuracy for the same number of simulations may be appealing
- Test idea using a very simple simulation exercise
  - Target 99.5%ile ("1 in 200")
  - Exposure assumed to be driven by 5 independent normal factors, i.e. involve multivariate normal distribution $(X_1, \ldots, X_5)^T \sim N(0, \mathbf{I})$ and overall exposure deemed to be $5X_1 + 4X_2 + 3X_3 + 2X_4 + X_5$
  - So can solve analytically, but still try using quantile interpolation (assuming distribution is normal)

Quantile interpolation (2)

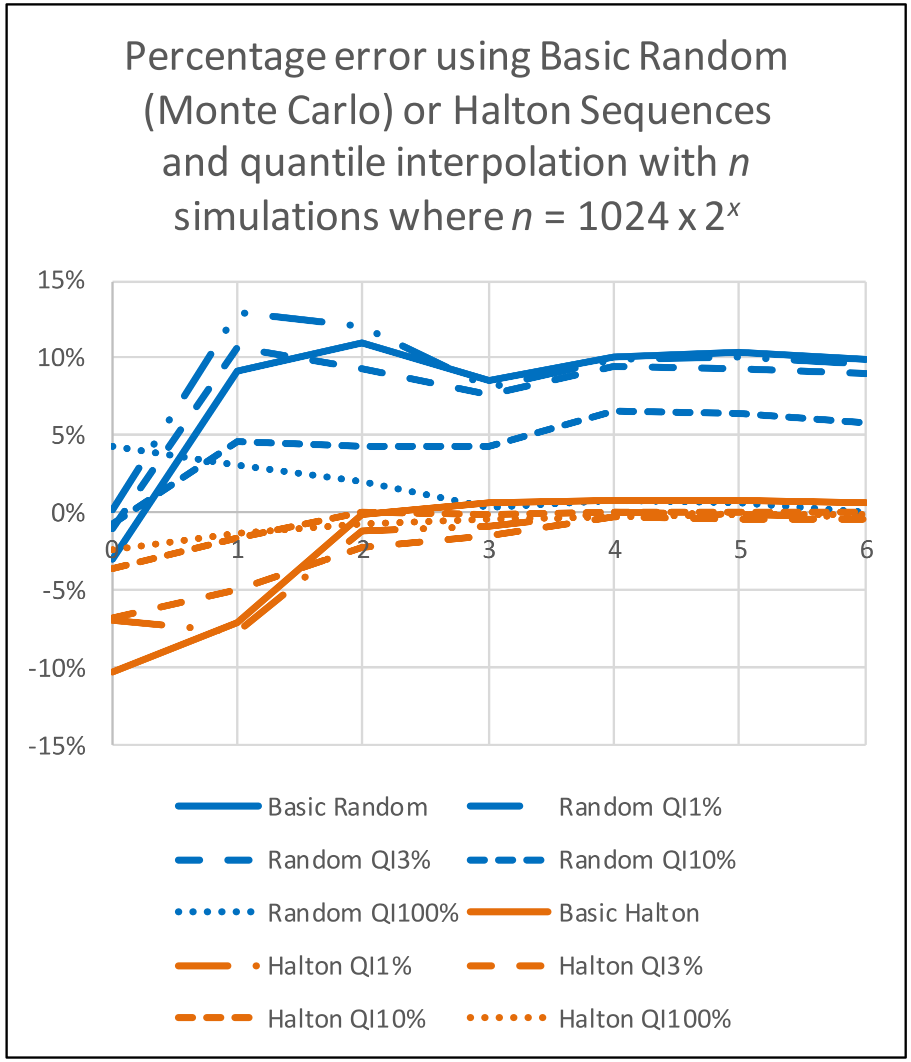
- Interpolate over what quantile range?
  - If fit to 100% of observations then akin to MLE, but the wider the range the more we have to assume that we understand the underlying distributional form
  - See impact of fitting to, say, worst 1%, 3%, 10% or 100% of simulations (using TWMLE, since clearer convergence to MLE as %age → 100%)
  - Using:
    a) Basic Monte Carlo (simulations chosen ‘at random’)
    b) (Basic) low discrepancy (Halton) sequences
    c) As a) or b) but replacing original draw sequences with their principal components (which are orthogonal by construction) and with the principal components adjusted to match assumed means and standard deviations of factors
  - Approach c) forces distribution to have overall observed moments and correlations very closely aligned to underlying distribution, so if interpolating over 100% of observations should then get almost exact answer

Quantile interpolation: Results (1)

- If using *basic* Monte Carlo or low discrepancy (Halton) then benefits look mixed for narrow quantile window but better for wider quantile window

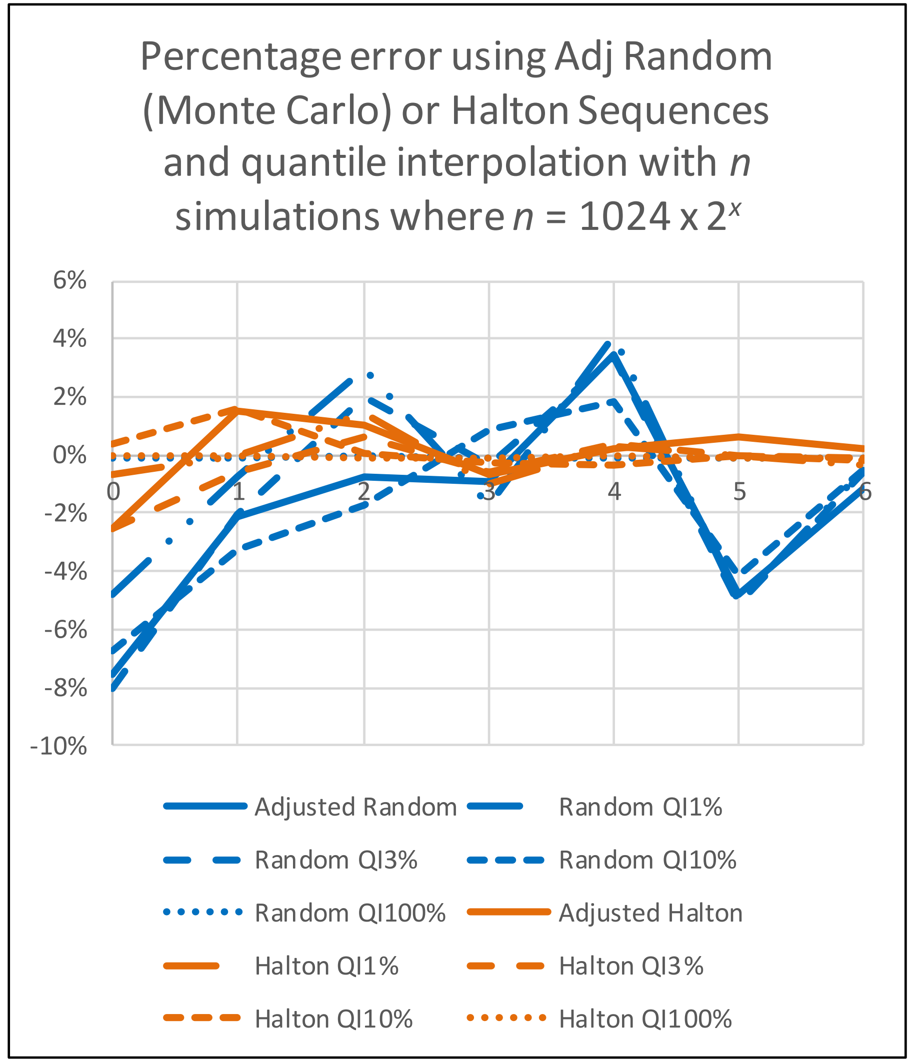
- Basic Monte Carlo
  - Errors seem very sensitive to random seeds. Possible benefit from forcing equal numbers of observations to be in each ‘quadrant’

- Low discrepancy (Halton)
  - Further smooths spread of data points. Relative appeal of quantile interpolation perhaps improves as simulation numbers rise



Quantile interpolation: Results (2)

- Typically smaller errors if we adjust simulations to match 1st and 2nd moments of distribution
  - E.g. by using principal components to arrange for simulations to have the same means, standard deviations and correlations as the assumed underlying distribution
- Low discrepancy (Halton)
  - Again relative appeal of quantile interpolation perhaps improves as simulation numbers rise



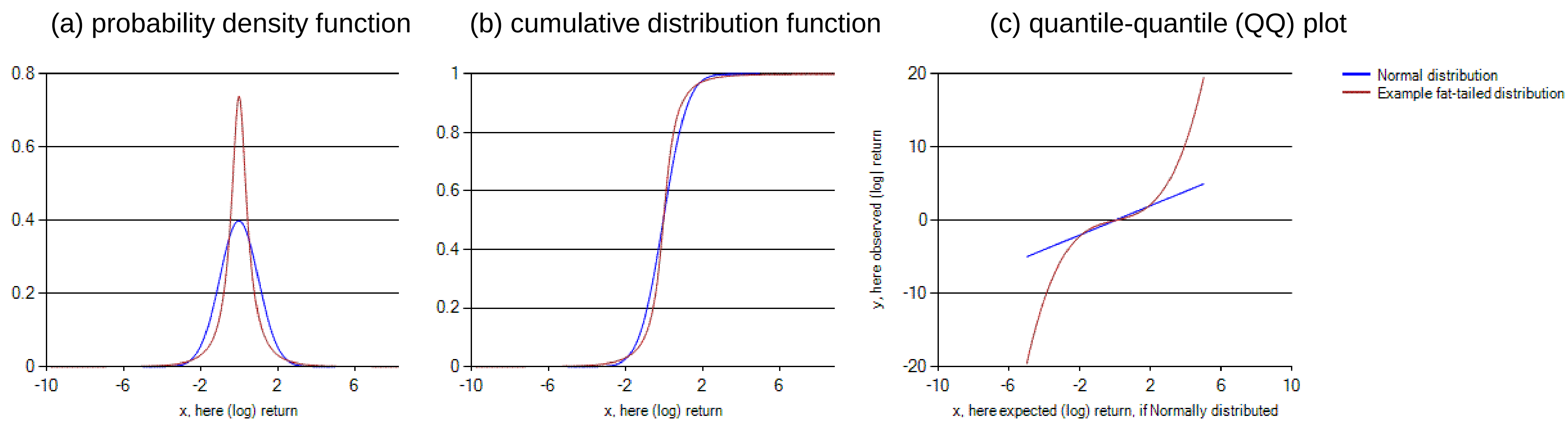
Summary

- Why is tail behaviour important?
  - Drives capital, perceptions and regulation, and is typically non-normal
- Traditional Extreme Value Theory (EVT) and its strengths and weaknesses
  - Conceptually appealing, but overemphasises robustness of extrapolation into the tail of a distribution (relies on applicability of generalised Pareto distribution)
- Refinements allowing fitting of any distribution to tail data
  - No need to use generalised Pareto, if we think another distribution might be better, but this doesn’t solve inherently problematic challenge of extrapolation
- Other uses of such techniques
  - Refinements can also be used to process expert judgement or for interpolation purposes in simulation exercises

Appendix A: Visualising fat-tailed behaviour

- Fat-tailed' means probability of extreme-sized outcomes seems to be higher than if coming from (usually) a (log) normal distribution

- There are various ways of visualising fat tails in a single return distribution. They are easiest to see in format (c) below, i.e. using QQ-plots



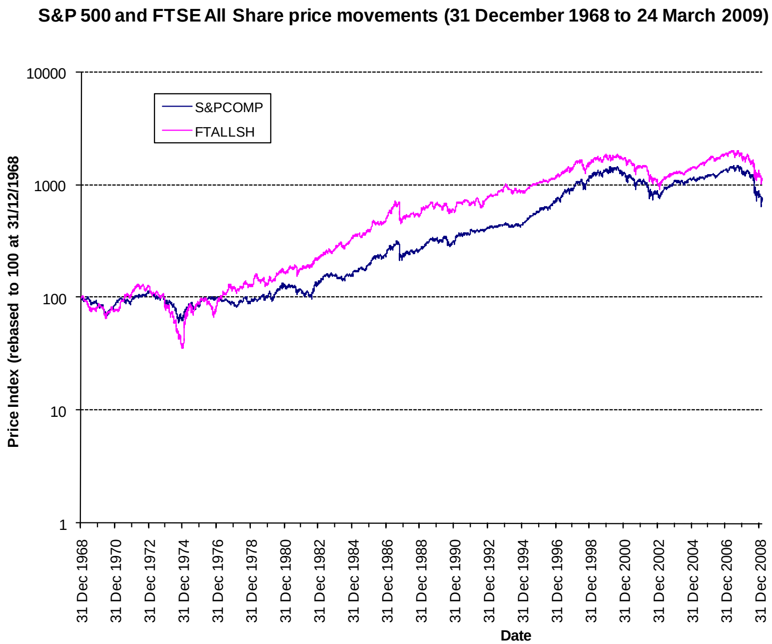
Quantile-quantile plots: other comments

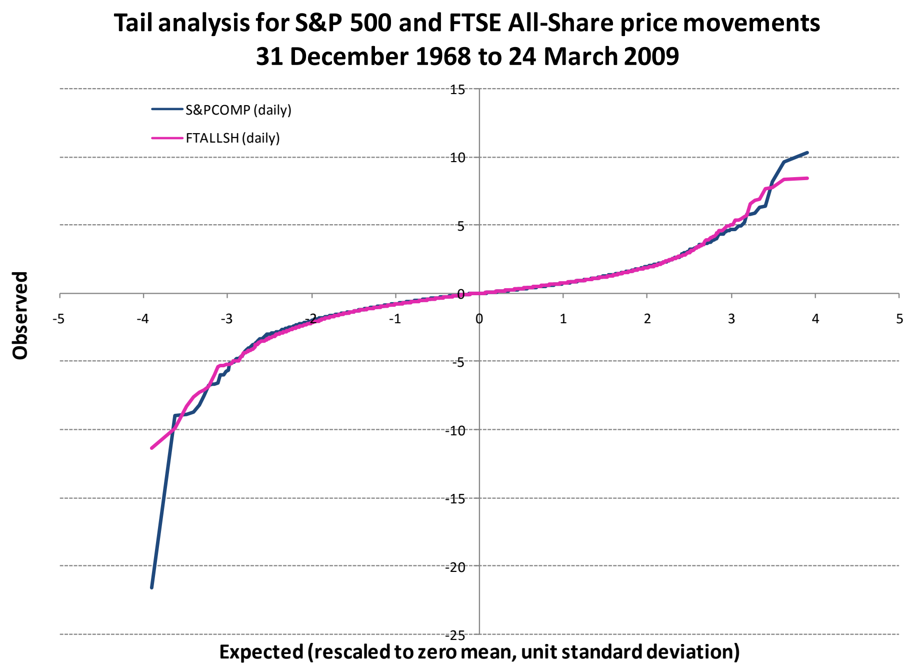
- Used for analysing whether distribution of outcomes is ‘as expected’
- Asserting that something exhibits fat-tailed behaviour requires us to have some prior view about what it might otherwise ‘reasonably’ be expected to do
- E.g. is a 2 year old an ‘outlier’ because he/she is much shorter than the average of the general population?
  - Not really, growing taller as you grow up a feature of the natural order
- With time series analysis such views are heavily influenced by time period for which data is available
  - And therefore on our perception about whether secular trends apply

Quantile-quantile plots

- In principle do not need to use normal distribution as the ‘expected’ distribution
  - C.f. definition of extreme event necessarily has in mind some prior view about what the distribution would be if it were not ‘fat-tailed’
- In practice, normal distribution is the most common reference distribution
- Need quite a few points to go ‘into the tail’

More periods give more scope for extreme events

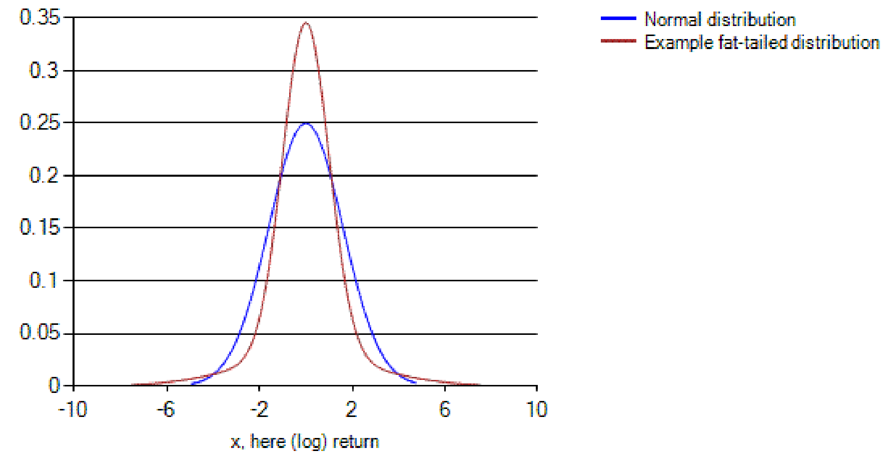


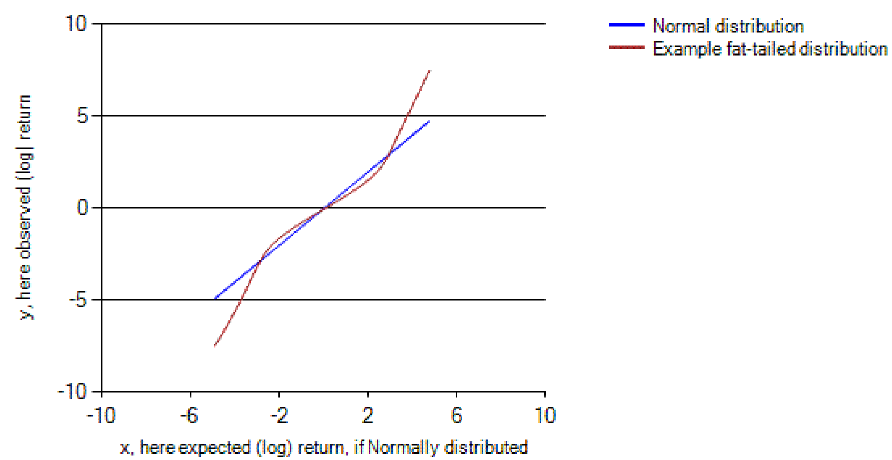


- N.B. There are also more daily observations than there are weekly (or monthly ones in the same overall time period

Appendix B: Time-varying volatility

- Very widely observed phenomenon
  - E.g. draw X with prob p from N_1 and prob (1-p) from N_2
  - Quite different behaviour to *linear combination mixtures*, i.e. $a.X_1 + b.X_2$
- If N_1 and N_2 have same mean but different s.d.'s then distributional mixture is fat-tailed (if $p \neq 0$ or 1), c.f. charts on the right of this page
  - Time-varying volatility is similar, involves draws from different distributions at different times





Important Information

All contents of this presentation are based on the opinions of the relevant Nematrian employee or agent and should not be relied upon to represent factually accurate statements without further verification by third parties. Any opinions expressed are made as at the date of publication but are subject to change without notice.

Any investment material contained in this presentation is for Investment Professionals use only, not to be relied upon by private investors. Past performance is not a guide to future returns. The value of investments is not guaranteed and may fall as well as rise, and may be affected by exchange rate fluctuations. Performance figures relating to a fund or representative account may differ from that of other separately managed accounts due to differences such as cash flows, charges, applicable taxes and differences in investment strategy and restrictions. Investment research and analysis included in this document has been produced by Nematrian for its own purposes and any investment ideas or opinions it contains may have been acted upon prior to publication and is made available here incidentally. The mention of any fund (or investment) does not constitute an offer or invitation to subscribe to shares in that fund (or to increase or reduce exposure to that investment). References to target or expected returns are not guaranteed in any way and may be affected by client constraints as well as external factors and management.

The information contained in this document is confidential and copyrighted and should not be disclosed to third parties. It is provided on the basis that the recipient will maintain its confidence, unless it is required to disclose it by applicable law or regulations. Certain information contained in this document may amount to a trade secret, and could, if disclosed, prejudice the commercial interests of Nematrian or its employees or agents. If you intend to disclose any of the information contained in this document for any reason, including, but not limited to, in response to a request under the Freedom of Information Act or similar legislation, you agree to notify and consult with Nematrian prior to making any such disclosure, so that Nematrian can ensure that its rights and the rights of its employees or agents are protected. Any entity or person with access to this information shall be subject to this confidentiality statement.

Information obtained from external sources is believed to be reliable but its accuracy or completeness cannot be guaranteed.

Any Nematrian software referred to in this presentation is copyrighted and confidential and is provided “as is”, with all faults and without any warranty of any kind, and Nematrian hereby disclaims all warranties with respect to such software, either express, implied or statutory, including, but not limited to, the implied warranties and/or conditions of merchantability, of satisfactory quality, or fitness for a particular purpose, of accuracy, of quiet enjoyment, and non-infringement of third party rights. Nematrian does not warrant against interference with your enjoyment of the software, that the functions contained in the software will meet your requirements, that the operation of the software will be uninterrupted or error-free, or that defects in the software will be corrected. For fuller details, see license terms on www.nematrian.com. Title to the software and all associated intellectual property rights is retained by Nematrian and/or its licensors.