Geometry, Differential Invariants and Improved Tail Risk Measurement

William F. Shadwick
Joint work with Ana Cascon
RiO Dec 2019

Omega Analysis

The goal of Extreme Value Theory (EVT) is to extrapolate statistical estimates of the likelihood and severity of 'random' events beyond those regularly observed–for example:

What is the '100 year flood level' in the Nile River delta? What level of flooding should be expected in excess of the 100 year flood.

What is the probability of exceeding the highest recorded temperature in July for South East England?

Or, in a financial context:

What is the 3-day loss level on a 'Short VIX' position that should only be exceeded 1 day in 100 and what is the average of losses in excess of this level?

How likely is a one week loss of more than 20% in the price of WTI Crude Oil?

The fact that there is a process for answering such questions is a result of the remarkable limit theorems of EVT. These produce the '3 types' of Extreme Value Distributions and the corresponding 3 types of Generalised Pareto Distributions.

We have shown recently that these results can be unified, explained and extended in terms of geometric invariants which are precisely analogous to the curvature of a surface.

Just as there are only three types of constant curvature surfaces, the distributions at the heart of Extreme Value Theory are the exceptional ones with constant invariants.

Omega Analysis

Our new invariants provide simple and easy to use characterisations of domains of attraction in EVT and explain the relationship between the EVT and Generalised Pareto Distributions.

The invariants also provide an intrinsic measure of the rate of convergence to the limiting distributions.

This isn't just of mathematical interest.

It has led us to highly efficient tail models which produce excellent results in financial market data.

First Approach to Extremes: Sample Maxima

Let $X_1, \ldots, X_N$ be a sample of N independent, identically distributed random variables with distribution function F. Let X_{Max} be the sample maximum.

If $X_{Max} < r$ then all of the sample draws must be less than r and the probability of this is $F^N(r)$.

Thus the distribution of X_{Max} is F^N.

In 1928 Fisher and Tippett addressed the question:

Does there exist a sequence of ‘location-scale’ transformations $x \to a_N x + b_N$ and a distribution G such that

$$F^N(a_N x + b_N) \to G(x) \qquad (1)$$

as N tends to ∞?

Intuition: Any such G *must* be the distribution of its own extremes.

Omega Analysis

Fisher and Tippett proved that there are only three families of distributions with this ‘stability property’.

$$\Phi(x,\alpha) = e^{-(-x)^{\alpha}},\ x \in (-\infty, 0],\ \alpha > 0 \tag{2}$$

$$\Psi(x,\alpha) = e^{\frac{-1}{x^{\alpha}}},\ x \in [0, \infty),\ \alpha > 0 \tag{3}$$

$$\Lambda(x) = e^{-e^{-x}},\ x \in (-\infty, \infty). \tag{4}$$

(Weibull, Fréchet and Gumbel distributions respectively)

Omega Analysis

The 'three types' of distributions Fisher and Tippett discovered are really a one-parameter family as Richard von Mises showed in 1936.

They are more conveniently denoted on variable domains depending on α as follows:

$$E_\alpha(x) = exp(\frac{-1}{(1+\frac{x}{\alpha})^\alpha}), \quad \alpha \neq 0 \tag{5}$$

This is the Weibull type, defined on $[-\infty, -\alpha]$ when $\alpha < 0$ and is the Fréchet type defined on $[-\alpha, \infty,]$ when $\alpha > 0$.

As $|\alpha| \to \infty$, both types have the Gumbel distribution $E_\infty = e^{-e^{-x}}$ as their limit.

Fisher and Tippett showed that the sample maxima limit for the Normal distribution was Gumbel type.

They gave no method for determining if a given distribution had a limit, or if it did, what the limit was.

It took 15 years to fill this gap.

Omega Analysis

In 1943 Gnedenko provided an independent derivation of the 'three types' theorem as well as necessary and sufficient conditions for convergence.

Let the distribution F be defined on $[\alpha(F), \omega(F)]$ (where we may have $\alpha(F) = -\infty$ and/or $\omega(F) = \infty$).

Gnedenko showed that Domains of Attraction (the collection of distributions which converges to a given type) depends only on the limiting *shape* of the distribution as $x \to \omega(F)$.

Gnedenko's necessary and sufficient condition for F to be in the domain of attraction of the Fréchet distribution $F(x, \alpha)$ is

$$\lim_{x \to \infty} \frac{1 - F(x)}{1 - F(tx)} = t^{\alpha} \tag{6}$$

for all $t > 0$.

He gave a similar sort of condition for the Weibull distribution.

Both conditions describe the *asymptotic scaling behaviour* of the distribution. (It's not at all clear why that should have anything to do with F^N.)

Omega Analysis

Gnedenko gave a variety of necessary and sufficient conditions for F to be in the domain of attraction of the Gumbel distribution.

He was not satisfied that any of them were either definitive or practical.

As it turns out, the key to the Fréchet and Weibul distributions is invariance under *scaling transformations* but in the Gumbel case it is *translation invariance* .

Omega Analysis

Second Approach to Extremes: ‘Peaks over threshold’

Does the distribution of a random variable X, conditional on X exceeding a threshold T, tend to a limit as T tends to $\omega(F)$ up to location scale transformations?

In this case we say F has a PoT limit.

In 1975 Picklands (and independently Balkema and de Haan) showed that there was a strong connection between PoT limits and Domains of Attraction of extreme value distributions.

F is in the domain of attraction of E_α if and only if the PoT limit of F is, up to a location-scale transformation, a Generalised Pareto distribution G_α as $x \to \omega(F)$ where

$$G_\alpha(x) = 1 - \frac{1}{(1 + \frac{x}{\alpha})^\alpha}, \quad \alpha \neq 0 \tag{7}$$

and

$$G_\infty(x) = 1 - e^{-x} \tag{8}$$

G_α is defined on $[0, -\alpha]$ when $\alpha < 0$ and on $[0, \infty)$ when $\alpha > 0$. As $|\alpha| \to \infty$, both types have the exponential distribution $G_\infty = 1 - e^{-x}$ as their limit.

But now we have a real mystery.

Everything in the Domain of Attraction of E_α is converging to everything else in that Domain of Attraction. For example, it's easy to check that for each ν, the Student t distribution $S(x,\nu)$ is in the domain of attraction of E_ν.

So what is special about Generalised Pareto distributions and what is behind the connection between Extreme Value limits and PoT limits?

Geometry answers these questions.

Omega Analysis

The Geometry of Extreme Value Distributions is the information invariant under what statisticians call the 'location scale' transformations and mathematicians call the proper affine group on the line $\mathcal{A}$.

These are the transformations of the form

$$x \to ax + b. \tag{9}$$

where $a > 0$.

The geometry is all of the information that is invariant under the group $\mathcal{A}$.

The most powerful method for discovering this geometry was produced by Elie Cartan extending the 18th Century results of Sophus Lie.

Cartan's *Method of Equivalence* allows us to construct a collection of *differential invariants* (like the curvature of a surface) which completely characterise the geometry.

The geometry always identifies *exceptional cases* such as constant curvature surfaces.

Let $I = log(F)$ and $J = I_{xx}/I_x^2$.

It turns out that all of the geometric information about F under location scale transformations is determined by the relation between I and J.

Since we only have one independent variable, J must be functionally dependent on I.

All of the geometry is encoded in the functional relation $J = H(I)$.

The Stability Property is the condition that a distribution F and F^N be in the same equivalence class with respect to $\mathcal{A}$.

But there's no need to restrict this question to integer powers.

It turns out that if we ask what distributions F are in the same $\mathcal{A}$ equivalence class as F^λ for all positive values of λ the answer is still the Extreme Value distributions.

Omega Analysis

We have a new invariant for the 1-parameter family of equivalence classes $[F^{\lambda}]$ because

$$I_{F^{\lambda}} J_{F^{\lambda}} = I_F J_F. \tag{10}$$

for all $\lambda \epsilon (0, \infty)$.

The *exceptional distributions* for which $K = IJ$ is constant are precisely the Extreme Value distributions.

Each value of the constant c determines a distinct equivalence class.

It is easy to see that the Extreme Value distributions provide normal forms for these equivalence classes.

The equivalence class of E_α is given by $c = 1 + \frac{1}{\alpha}$ for $\alpha \neq 0$ and E_∞ is given by $c = 1$.

Omega Analysis

The Geometry of Domains of Attraction

Theorem (Cascon and Shadwick)

Let F be a distribution defined on $[\alpha(F), \omega(F)]$.

F is in the domain of attraction of an Extreme Value distribution if and only if the limit of $K_F = I_F J_F$ as x approaches $\omega(F)$ is the constant invariant of the Extreme Value distribution.

It is easy to use our result for any of the standard probability distributions to determine which of the Extreme Value distributions E_α and E_∞ they have as their limits.

Unlike Gnedenko's theorem, there's only one test and it's just as simple for the Gumbel attractor as it is for Weibull or Fréchet cases.

It is also easy to verify that for each α the Generalised Pareto distribution G_α is in the Domain of Attraction of E_α and that G_∞ is in the Domain of Attraction of E_∞.

The Picklands Mystery Again

The utility of the Generalised Pareto distributions is that they converge rapidly to their Extreme Value limits. This means that less data is required to make a reasonable fit.

But why the Generalised Pareto distributions and not others in the same Domain of Convergence?

Omega Analysis

Another Approach to Extremes: ‘Peaks under threshold’

If F is defined on $[\alpha(F), \omega(F)]$ and $T \in (\alpha(F), \omega(F)]$ the distribution conditional on $x < T$ is $F_T = \frac{F}{F(T)}$. If $[F_T]$ tends to a limiting distribution as $T \to \alpha(F)_-$ then F is said to have a PuT limit.

Such distributions *must* be their own PuT limits so we have the question of PuT stability:

When is $[F_T] = [F]$?

It turns out that once again, geometry answers this question.

The PuT stable distributions are precisely the *exceptional ones* for which J is constant.

Each constant determines an equivalence class of distributions and all constants are possible.

It is easy to integrate $J = c$ to produce normal forms, where $\alpha = \frac{1}{c}$:

$$\widehat{G}_{\alpha}(x) = \frac{1}{(1 - \frac{x}{\alpha})^{\alpha}}, \quad \alpha \neq 0 \tag{11}$$

and

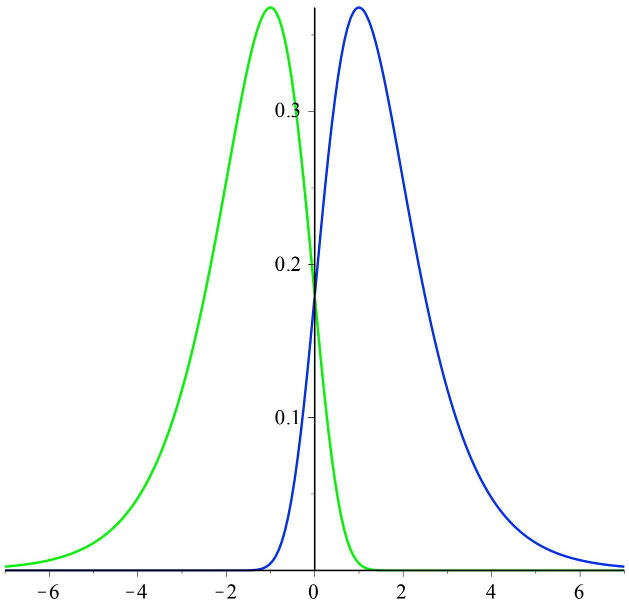
$$\widehat{G}_{\infty}(x) = e^{x} \tag{12}$$

$\widehat{G}_{\alpha}$ is defined on $[\alpha, 0]$ when $\alpha < 0$ and on $(-\infty, 0]$ when $\alpha > 0$.

As $|\alpha| \to \infty$, both types have the exponential distribution $\widehat{G}_{\infty} = e^{x}$, on $(-\infty, 0]$ as their limit.

Omega Analysis

For any probability density function f defined on $[A, B]$ there's a 'mirror image' probability density $\widehat{f}$ on $[-B, -A]$ defined by $\widehat{f}(x) = f(-x)$.



A Gumbel density (blue) and its mirror image

If F is the distribution with density f then we will refer to the distribution $\widehat{F}$ whose density is $\widehat{f}$ as the ‘mirror image’ of F.

It’s easy to check (just differentiate) that $\widehat{F}$ is given on $[-B,-A]$ by

$$\widehat{F}(x) = 1 - F(-x). \tag{13}$$

The Generalised Pareto distributions introduced by Picklands are precisely the mirror images of the exceptional distributions corresponding to constant differential invariant J.

Nature only makes so many exceptional, ‘constant curvature’ objects. The Extreme Value distributions and the (mirror image) Generalised Pareto distributions both have this property, but for different ‘curvatures’.

Here’s the relationship between them.

Omega Analysis

Duality of Domains of Attraction

Theorem (Cascon and Shadwick)

If F is a distribution on $[\alpha(F), \omega(F)]$ and $\widehat{F}$ is the mirror image distribution on $[-\omega(F), -\alpha(F)]$ then

$$\lim_{x \to \omega(F)_-} I_F J_F = 1 + c. \qquad (14)$$

if and only if

$$\lim_{x \to -\omega(F)_+} J_{\widehat{F}} = c. \qquad (15)$$

Omega Analysis

The geometry we have uncovered unifies and explains 70 years worth of discoveries in Extreme Value Theory.

But it does much more than that.

It provides an *intrinsic scale* on which we can measure the degree of convergence of a distribution to its EVT or PuT attractor.

The values of the invariants J and K at quantiles are also invariants.

The difference between one of these values and the EVT constant is an *intrinsic measure of convergence*.

The more rapidly a distribution converges to its EVT limit, the less data is necessary to discover that limit.

So being able to compare rates of convergence has a very important statistical application.

Generalised Pareto distributions converge to their EVT limits incredibly quickly (as you can see by comparing them, quantile by quantile, with other distributions in the same EVT Domain of Attraction.

For example it's easy to check that $GP(x,3)$ approaches its EVT limit much faster than $Student(x,3)$. But this doesn't necessarily translate into superior performance in modelling 99% VaR and ES from sample data.

That's because you don't model the entire distribution by the Generalised Pareto–you only use it to model the *tail*.

You may only be fitting, for example, the top 5% of the data using the Generalised Pareto distribution.

In this case you're putting the convergence at only its $q = 0.8$ level up against the $q = 0.99$ for the $Student(x, 3)$ distribution. And in that contest the $Student(x, 3)$ distribution wins, hands down. (See Appendix.)

This is *not* a recommendation to fit tails with $Student$ distributions rather than Generalised Paretos.

It's just an illustration that in spite of their remarkable convergence properties, it's easy to find examples of distributions that are closer to their EVT limits *over the quantile range that matters in practice*.

Such distributions are more efficient in use with short data sets.

This is a tremendous advantage in financial market data.

The ability to make good estimates of VaR and ES using short data windows allows you to observe and respond to changes in risk while there's still time to take advantage of the information.

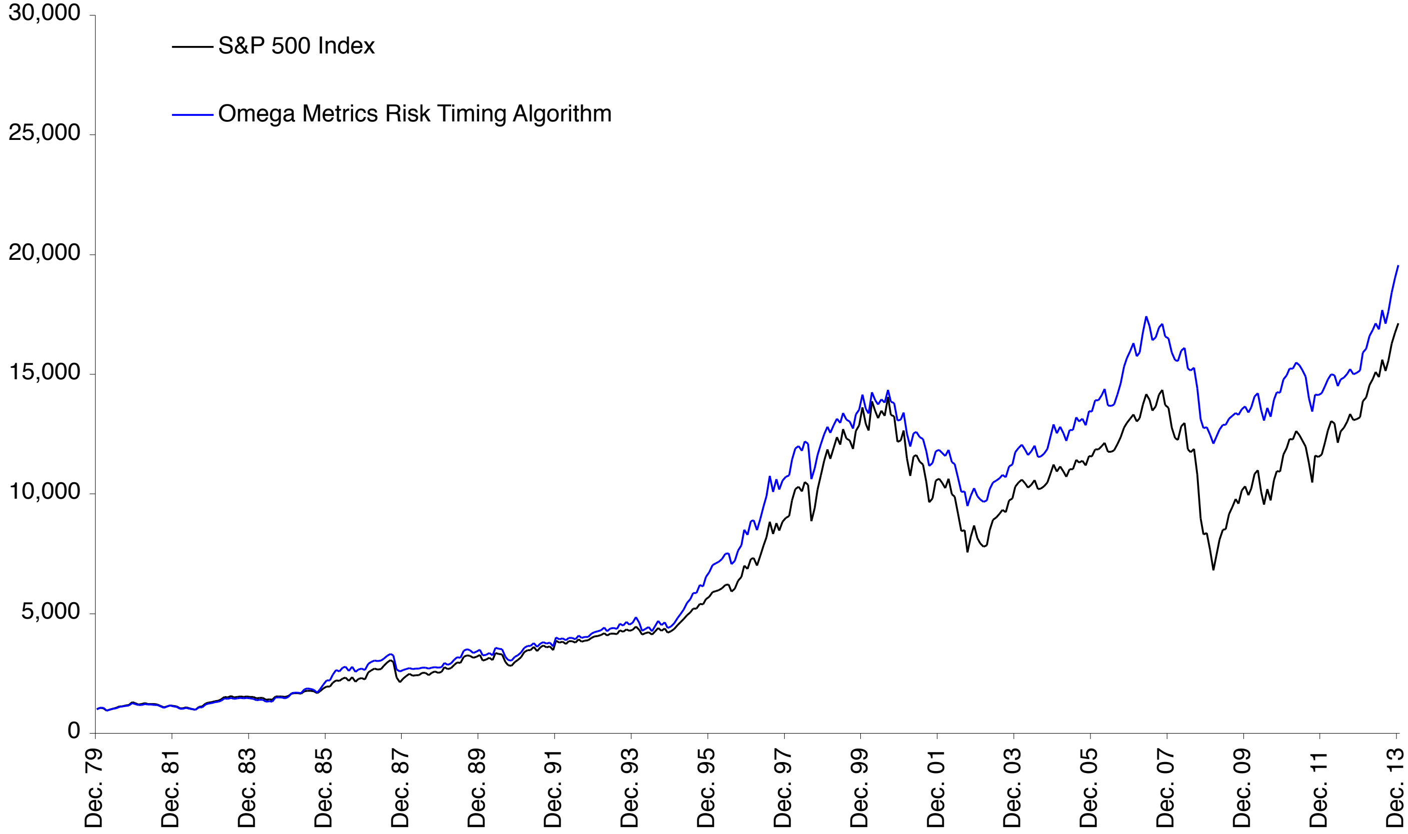
Omega Analysis has developed proprietary tail fits that converge more rapidly to their attractors than the Generalized Pareto Distributions do, over a range of quantiles of practical significance–even though the latter always converge more rapidly asymptotically.

Our distributions are very efficient models of tails in financial market returns and in other fields where decisions must be based on short data sets.

They produce consistently reliable VaR and ES measurements and have demonstrable predictive power.

Omega Analysis

Example: 'Risk Timing' the S&P 500 Index from 1980-2013



Initial value $1,000. Maintains constant 1-day 99% ES target

Omega Analysis

Example: The 2018 wipeout of the 'Short VIX' trade.

As equity market volatility reached historic lows in 2017 many investors, including large institutions and ETFs crowded into the 'Short VIX' trade.

By June 2017 JP Morgan analyst Marco Kolanovic warned that an increase in the VIX index to 20 from its value of 10.5-11 would be enough to *wipe out* the massive Short VIX positions.

But there's a big difference between knowing that, and knowing how likely a move from 10 to 20 was.

Big jumps up in the level of the VIX are in clusters– most of the damage is done in 3 days.

With an accurate estimate of the upside tail in VIX 3-day returns, for example, one could make a realistic assessment of the likelihood of the wipeout Kolanovic warned about.

In a back-test over almost 3 decades, our 3-day 99% VaR estimates showed only a handful of excess VaR breaches– so if anything we were likely to be *underestimating* the risk.

In June 2017 we published our estimate that a 3-day surge to Kolanovic's catastrophic level could be expected 1 time in 280 days, likening the Short VIX trade to Russian roulette with 1 bullet in 280 chambers.

By August 2017 the VIX reached 16 and by the end of that month our revised estimate was that there was 1 bullet and only 48 chambers.

With a new 3-day return every day, that meant that the 'game' should be over within 3 months.

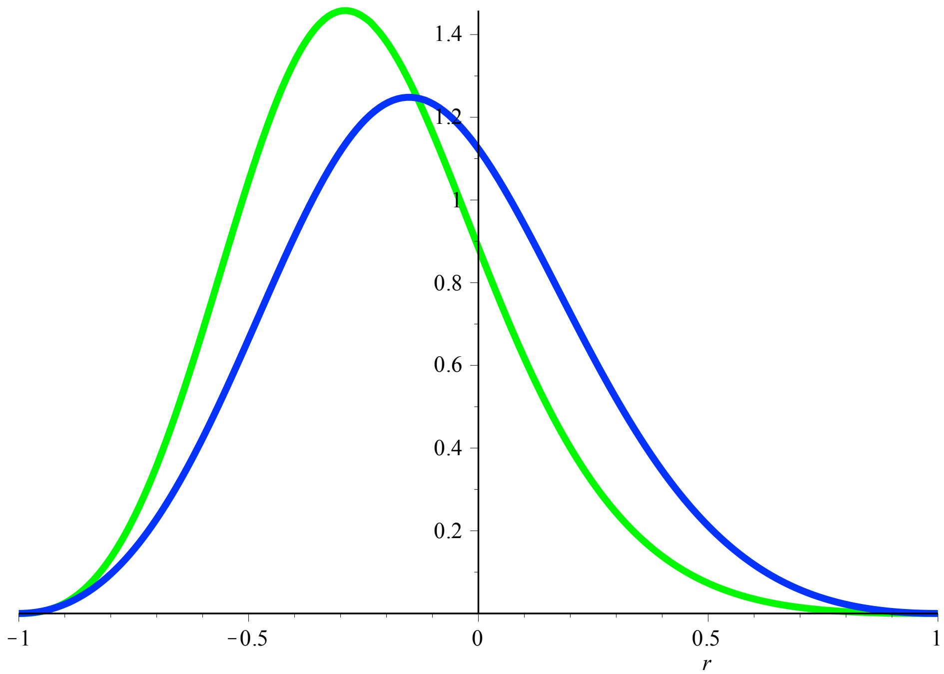
It lasted 5 months.

In February 2018 the VIX surge came, wiping out the Short VIX traders and the ETFs that brought the trade to retail investors.

But some people never learn. Short positions in the VIX reached a new record recently, as the S&P 500 Index has soared to new heights.

As of last week,the 3-day 99% VaR and ES for the Short VIX are significantly higher than they were in January 2018. A surge in the VIX of over 50% is now a 1 time in 9 month event. One bullet in 180 chambers?

Example: VIX (Green) and S&P 500 (blue) pdfs for returns data transformed to Omega Functions on $[-1, 1]$ as in Ana Cascon's talk. Shape Scores VIX 3.2, S&P 500 0.9.



Data: Daily Returns, VIX, last 6 months, S&P500, 2008

Example: Flash crashes and Jamie Dimon's grasp of statistics

Bloomberg the Company & Its Products | Bloomberg Anywhere Remote Login | Bloomberg Terminal Demo Request

Menu Search **Bloomberg** Sign In **Subscribe**

Business

'Once-in-3-Billion-Year' Jump in Bonds Was a Warning Shot, Dimon Says

Hugh Son, Sonali Basak and Doni Bloomfield

April 8, 2015, 10:36 PM GMT+1 *Updated on April 9, 2015, 5:13 PM GMT+1*

Dimon Says JPMorgan Stress Tests for Grexit

Omega Analysis

Example: Flash crashes and Jamie Dimon's grasp of statistics

In April 2015 Jamie Dimon's shareholder letter was headline news but not for the right reason. He was trying to make important points about liquidity in the U.S. Treasury market and the Swiss National Bank's impact on the Euro Swiss Franc exchange rate.

But his message was swamped by the reaction to his ridiculous observation that the October 2014 'flash crash' in U.S. Treasuries was "...supposed to happen once in every 3 billion years or so..."

Omega Analysis

Of course you can only generate a claim like that by using a Normal distribution to turn a number of standard deviations into a probability estimate. Not a very smart thing to do.

Our tail model showed that the 40 basis point move was, in fact, a daily high-low that should be expected every two to three years.

The really important point was that *the sort of move which an entire day's trading should produce only a few times per decade occurred in less than 15 minutes*

I sent our analysis of this to Jamie Dimon but he hasn't gotten back to me yet.

The other important point he hinted at without being explicit was pretty obviously aimed at what some market participants would call the Swiss National Bank's market vandalism.

When they pulled the plug in January 2015, there was, according to ECB data, a 14.4% "38 standard deviation" move in the Euro Swiss Franc exchange rate.

Central bankers I have talked to seem to think this all worked out just fine.

But there were a lot of losers. If you weren't big enough for threats of legal action to be effective you probably had your guaranteed stops blown out.

There were fund managers who were just on the right side of that line while other market participants took major losses.

To see just how outrageous that 14.4% move was, we can ask a really ridiculous question.

What was the 1 day in 10 year VaR and ES before the Swiss National Bank's action?

Our model provides an answer that's perfectly reasonable. The VaR was 4.6% and the ES conditional on a VaR breach was 7.9%. In the 15 year history of the Euro there was in fact one prior move of about 8% (in the opposite direction) in September 2011.

This shows why some people think the term market vandalism is appropriate.

Even if you had the best current risk technology and even if you were willing to believe and act on a prediction at the 1 day in 10 year level, *you could not possibly have been prepared for the* 14.4% *move the SNB precipitated.*

Imagine what the FCA's reaction would have been if a non-Central Bank market participant in London had pulled this off!

Example: The dramatic slump in U.S. and European bank shares following the 2016 U.K. 'Brexit' referendum.

Our tail models allow us to make very good VaR and ES estimates for 5-day returns at the 99% level. (As judged by comparing the number of VaR breaches over long histories with the number that should have been observed.)

Here's what we predicted the prospects for drawdowns were *prior to the UK referendum* compared with what happened by 6 July 2016.

Omega Analysis

Instrument	Value at Risk (VaR) 99% 5-day	Expected Shortfall (ES) 99% 5-day	Worst 5-day Loss (since 23 June 2016)
KBW Nasdaq Bank Index	-10.9%	-16.8%	-9.3%
Stoxx® Europe 600 Banks	-14.3%	-21.8%	-16.8%
Banca Monte dei Paschi	-31.8%	-47.4%	-32.5%
Barclays	-14.2%	-21.3%	-27.1%
Deutsche Bank	-19.4%	-28.4%	-21.5%
HSBC	-10.8%	-15.7%	N/A
JPMorgan	-10.9%	-17.8%	-7.6%
UniCredit	-21.5%	-31.3%	-27.7%

VaR and ES predictions and realised losses as of 6 July 2016

Omega Analysis

With the right technology, the losses were entirely predictable as our measured risk levels for banks had *doubled* in the previous year.

This also showed the information gap between market prices and CDS spreads. Our risk measure showed that the tails of the distribution of JP Morgan returns were significantly fatter than those of HSBC's returns.

But the CDS rate for JP Morgan was only about two thirds that of HSBC at the time.

Omega Analysis

There are many more applications.

Asset Bubble early warning indicators (e.g. the U.S. equity market and most sub-sectors)

Asset Anti-Bubble early warning indicators (e.g. Lehman Brothers starting 2007, Deutsche Bank from 2018)

Risk analysis for commodities, FX etc.

‘Big Data’ . . .

Omega Analysis

Important Notice

LEGAL NOTICE
Please read this notice carefully:

The contents of this document are for illustrative and informational purposes only. No information in this document should be considered a solicitation or offer to buy or sell any financial instrument or to offer any investment advice or opinion as to the suitability of any security in any jurisdiction. All information is subject to change and correction due to market conditions and other factors. This document has been created without any regard to the specific investment needs and objectives of any party in any jurisdiction. Specific instruments are mentioned in this document but this should not be construed in any way as a recommendation to invest in them or in funds or other instruments based on them. They are used for informational purposes only.

Omega Analysis Limited provides statistical analysis services. Omega Analysis Limited does not provide investment advice. Investors need to seek advice regarding suitability of investing in any securities or investment strategies. Any decisions made on the basis of information contained herein are at your sole discretion and should be made with your independent investment advisor.

References

1) *Limiting forms of the frequency distribution of the largest or smallest member of a sample*, R.A. Fisher and L.C.H. Tippett, Proceedings of the Cambridge Philosophical Society **XXIV**, part II, pp 180-190 (1928)

2) *Sur La Distribution Limite Du Terme Maximum D'Une Série Aléatoire*, B. Gnedenko, Annals of Mathematics **44**, No. 3, pp 423-453 (1943)

3) *Statistical Inference Using Extreme Order Statistics*, J. Picklands III, Annals of Statistics, **3**, No.1, pp. 119-131 (1975)

References Cont'd

4) *Residual Life Time at Great Age*, A.A. Balkema and L. de Haan, Annals of Probability, **2**, No.5, pp. 792-804 (1974)

5) *What Just Happened?*Bank risk is rocketing up on both sides of the Atlantic. 6 July 2016. OmegaAnalysis.com.

6) *JP Morgan Needs Better Statistics* W.F. Shadwick. 10 April 2015. OmegaAnalysis.com

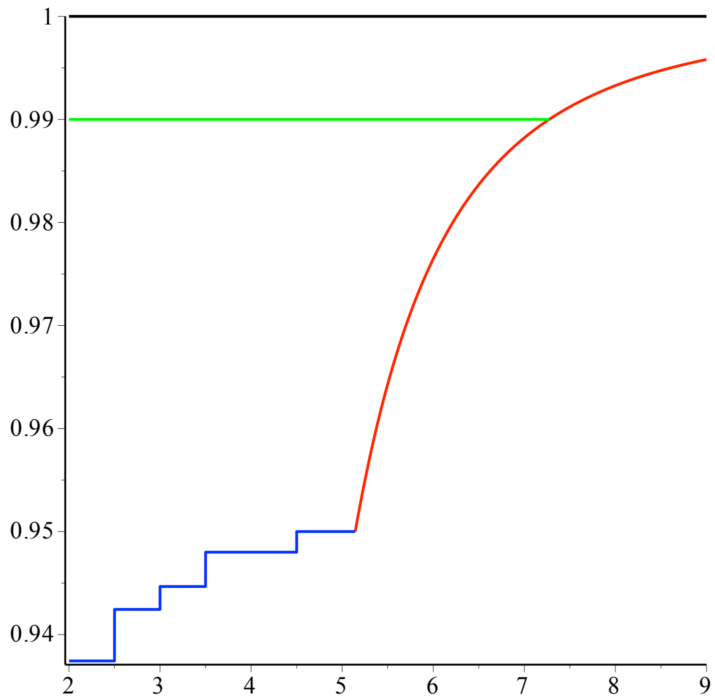
Appendix 1

Suppose we want to find the VaR at the 0.99% quantile by fitting a Generalised Pareto tail.

The model for the distribution is

$$D = \frac{19}{20} Empirical + \frac{1}{20} G(x - u, 3) \tag{16}$$

where u is the point at which the Empirical distribution reaches $q = 0.95$.



$\frac{1}{20}GP(x-u,3)$ attached at $q = .95$.

The $q = 0.99$ level for the tail plus empirical distribution is the solution to $\frac{19}{20} + \frac{1}{20}G(x-u,3) = \frac{99}{100}$.

For $x > u$ the distribution is $D = \frac{19}{20} + \frac{1}{20}G(x-u, 3)$ so the answer to the to the question "When is $D(x) = 0.99$?" is the answer to the question

"When is $\frac{19}{20} + \frac{1}{20}G(x-u, 3) = \frac{99}{100}$?" .

The answer to that is the value at which $\frac{1}{20}G(x-u, 3) = \frac{4}{100}$ or $G(x-u, 3) = 0.8$.

If instead, the model for the distribution is

$$D = Student(x, 3), \tag{17}$$

the question is "When is $Student(x, 3) = 0.99$?".

So we're comparing the efficiency of $G(x, 3)$ at $q = 0.8$ with $Student(x, 3)$ at $q = 0.99$.

Note that the value of the invariant K at the $q = 0.8$ level is the same for every one of the family of Generalised Pareto distributions $GP(\frac{x-b}{a}, 3)$. (In the example, I've just set the scale parameter a equal to 1 for convenience.)

So we can make the comparison with the value of the invariant for the distribution $Student(x, 3)$ at $q = 0.99$ using the standard Generalised Pareto distribution $GP(x, 3)$.

At those values the Generalised Pareto distribution's K is twice as far from its EVT limit of $4/3$ as the Student distribution's K is.

This is a handicap that not even the Generalised Pareto distribution can overcome.

(But there's no way to know that without our invariant measure of convergence.)