CHARTA LEARNING · LESSON NOTES

Extreme Value Theory

FRM Part II

1 A story: the day the model called impossible

- Thornbury's desk: daily loss volatility about $1m
- Risk model: daily losses are normal
- One day: a $6.5m loss, 6.5 standard deviations
- Normal odds: about once in 100 million years
- Survival is decided by days like this one

IN THE LESSON

- Thornbury Bank runs a trading desk whose daily losses have a volatility, meaning a standard deviation, of about one million dollars.
- For years its risk model assumed that daily losses follow a normal distribution, the familiar bell curve.
- Then one morning the desk lost six and a half million dollars.
- That is six and a half standard deviations.
- Under the normal model, a loss that large should happen about once in a hundred million years.
- Yet it happened, at Thornbury and at many other banks, within a working lifetime.
- Whether a bank survives is decided by days like this one, and an ordinary model treats them as impossible.

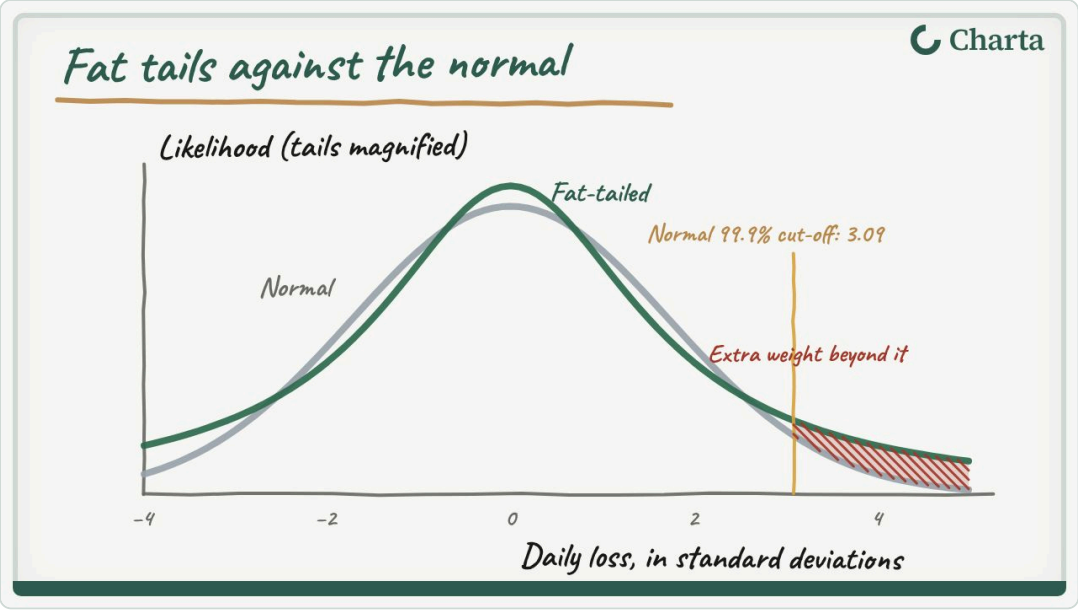
2 Why extreme values matter

- Solvency and capital are decided by rare, huge losses
- VaR at 99% or more: a far-tail quantile
- Expected shortfall: the average loss beyond VaR
- Stress tests: deliberately extreme scenarios
- All three live in the far tail

IN THE LESSON

- Why do extreme values matter so much?
- Because solvency, capital and survival are decided by rare, very large losses, and not by ordinary days.
- Value at risk at ninety nine per cent or more is a loss level that is exceeded on only one day in a hundred, or fewer.
- It is a quantile far out in the tail.
- Expected shortfall is the average loss on the days when V a R is exceeded, so it lives even further out.
- And stress tests deliberately ask about extreme scenarios.
- All three live in the far tail, which is exactly the part of the distribution that we know least about.

3 Fat tails against the normal



IN THE LESSON

- Daily loss, in standard deviations, runs along the bottom, with bigger losses to the right, and likelihood runs up the side, with the tails magnified so we can see them.
- The grey curve is the normal.
- The green curve is a fat-tailed distribution with the same average and the same volatility, which is how real losses tend to look.
- The gold line marks the normal's ninety nine point nine per cent cut-off, at three point zero nine standard deviations.
- The hatched area is the extra weight the fat tail puts beyond it.
- Out there the green curve lies well above the grey one, so a model built on the normal puts far too little weight on the very losses that matter most.

4 Example 1: how little tail data there is

Thornbury: 1,000 daily losses; the worst is $6.5m

Beyond the 99% quantile	1,000 × 1%	**10 days**
Beyond the 99.9% quantile	1,000 × 0.1%	**1 day**
Beyond the 99.97% quantile	1,000 × 0.03%	**0.3 days**
Largest historical-sim loss	the worst in the sample	**$6.5m**

IN THE LESSON

- The first difficulty is scarce data.
- Thornbury has one thousand daily losses, and its worst is six and a half million dollars.
- Beyond the ninety nine per cent quantile there are one per cent of the days, so about ten observations.
- Beyond the ninety nine point nine per cent quantile there is about one observation.
- Beyond the ninety nine point nine seven per cent quantile, there are on average point three observations, so most samples contain none at all.
- Historical simulation reads V a R straight off the past losses, and it can never produce a loss larger than the worst one in its sample, here six and a half million dollars, however far out we ask it to look.

5 Example 2: the normal understates the tail

Normal fit: mean 0, volatility $1m. The 99.9% normal z-value is 3.09

Normal 99.9% cut-off	3.09 × $1m	**$3.09m**
Days beyond it, if normal	1,000 × 0.1%	**1 day**
Days beyond it in the sample	counted	**15 days**
Tail understated by	15 ÷ 1	**15 times**

IN THE LESSON

- The second difficulty is that standard models are fitted to the centre of the distribution, where most of the data sit.
- Suppose Thornbury fits a normal with a mean of zero and a volatility of one million dollars.
- Its ninety nine point nine per cent cut-off is three point zero nine standard deviations, so three point zero nine million dollars.
- If losses really were normal, one day in a thousand would go beyond it, so we would expect about one such day in the sample.
- Counting the desk's history, there are fifteen.
- The normal understates how often this tail is reached by a factor of fifteen.
- A model can match the middle well and still be badly wrong in the tail.

6 Why the tail is hard to measure

- Scarce data: few observations beyond a high quantile
- Historical simulation cannot exceed the worst loss
- Models fit the centre; the normal misses fat tails
- Tail estimates are noisy and must extrapolate
- Exam point: this is why we need extreme value methods

IN THE LESSON

- Put the two difficulties together and you get the third.
- First, data are scarce: very few observations lie beyond a high quantile.
- Historical simulation cannot go beyond the worst loss it has seen.
- Second, standard models are fitted to the centre, and a normal assumption badly understates the fat tails of real returns.
- Third, tail estimates are therefore noisy, and they must rely on extrapolation beyond the data, which means projecting the tail out past the observations we have.
- The exam point is this.
- Scarce tail data plus fat tails make high confidence V a R hard.
- That is the motivation for extreme value methods.

7 What extreme value theory is

- EVT: the theory of the largest observations
- It models the tail directly, not the whole distribution
- Averages tend to the normal (central limit theorem)
- Block maxima tend to the GEV (Fisher-Tippett theorem)
- Needs iid data and large blocks

IN THE LESSON

- Extreme value theory, or E V T, is the statistical theory of the limiting behaviour of the largest observations.
- Instead of modelling the whole distribution and hoping the tail comes out right, it models the tail directly.
- The central limit theorem says that averages of many observations tend towards the normal distribution, whatever the data look like.
- The Fisher Tippett theorem says something similar about maxima.
- If the data are independent and identically distributed, the largest value in a large block, suitably normalised, tends towards one family: the generalised extreme value distribution, or G E V.
- Those two conditions, independent and identically distributed data and large blocks, matter later.

8 The generalised extreme value family

$$H(x) = \exp\left(-[1 + \xi(x - \mu) \div \sigma]^{-1/\xi} \right)$$
$$\xi = 0: H(x) = \exp\left(-e^{-(x - \mu) \div \sigma} \right)$$

Location, μ	Where the block maxima sit
Scale, σ	How spread out they are
Shape, ξ	How fat the tail is

IN THE LESSON

- Here is the G E V distribution.
- It gives the probability that a block maximum is at most x.
- It has three parameters.
- The location, mu, says where the maxima typically sit.
- The scale, sigma, says how spread out they are.
- And the shape, written with the Greek letter ksi, which we will call the shape parameter, says how fat the tail is.
- The top line reads: H of x equals the exponential of minus the bracket, one plus the shape parameter times x minus mu over sigma, raised to the power minus one over the shape parameter.
- The second line is the special case when the shape is exactly zero: the exponential of minus e to the power minus x minus mu over sigma.
- For risk, the shape matters most.

9 Three shapes of tail

The sign of the shape parameter ξ picks the family

Shape ξ	Family	Tail
ξ = 0	Gumbel	Thin, like the normal
ξ > 0	Fréchet	Fat: power law, index 1/ξ
ξ < 0	Weibull	Bounded: a maximum loss

IN THE LESSON

- The sign of the shape parameter picks one of three families.
- When the shape is zero, the limit is the Gumbel distribution.
- Its tail is thin, and it is the limit for data such as the normal.
- When the shape is positive, the limit is the Fréchet distribution.
- Its tail is fat and falls away like a power law, and its tail index, one over the shape, measures how fat.
- When the shape is negative, the limit is the Weibull distribution, whose tail is bounded: there is a maximum possible loss.
- For financial losses, the Fréchet case, with a positive shape, is the one usually used.
- But beware: in a finite sample, the shape is estimated with real uncertainty.

10 Example 3: Thornbury's tail index

Thornbury's estimated shape ξ = 0.25: the Fréchet case

Tail index	1 ÷ 0.25	**4**
Far out: double the loss level	probability ÷ 2^{4}	**÷ 16**
Thin Gumbel-type tail	probability falls much faster	**thinner**
Finite moments	order below the index, 4	**orders 1, 2, 3**

IN THE LESSON

- Thornbury's analysts estimate a shape of point two five, so its losses fall in the Fréchet case.
- The tail index is one over point two five, which is four.
- Far out in a power-law tail, doubling the loss level divides its probability by two to the power of the tail index.
- Two to the fourth is sixteen.
- Under a thin Gumbel-type tail, such as the normal's, the probability collapses much faster as the loss doubles.
- A tail index of four also tells us which moments exist: only those of order below four.
- So the mean, the variance and the skewness are finite, but the kurtosis is not.
- Keep in mind that this shape is itself an estimate.

11 How risk managers use EVT

- VaR and ES at very high confidence, such as 99.9%
- Extrapolate beyond the worst loss in the sample
- Design stress tests from the fitted tail
- Financial losses: usually $\xi > 0$, the Fréchet case
- Caveat: ξ is uncertain in finite samples

IN THE LESSON

- How do risk managers put E V T to work?
- First, to estimate V a R and expected shortfall at very high confidence levels, such as ninety nine point nine per cent, where the data alone say almost nothing.
- Second, to extrapolate beyond the sample, past the worst loss in the history, which historical simulation cannot do.
- Third, to design stress tests, by asking how large a loss the fitted tail implies at a given rarity.
- For financial losses, the fitted shape is usually positive, the Fréchet case.
- The caveat: the shape is uncertain in finite samples, so results should be reported with that uncertainty in mind.

12 Example 4: the iid trap

EVT assumes iid data; real losses cluster in volatile spells

Standardise each day's loss	loss ÷ that day's volatility	**z**
Fit EVT to z: 99% quantile	from the fitted tail	**2.80**
Calm day: volatility $0.8m	0.8 × 2.80	**$2.24m**
Stressed day: volatility $1.6m	1.6 × 2.80	**$4.48m**

IN THE LESSON

- The usual mistake with E V T is ignoring the assumption of independent, identically distributed data.
- Real losses show volatility clustering: big days come in bunches, during volatile spells.
- The fix is to standardise.
- Divide each day's loss by that day's volatility forecast, to get a standardised loss, written z.
- Fit E V T to the z values, which are much closer to independent.
- Suppose the fitted ninety nine per cent quantile of z is two point eight.
- On a calm day, with forecast volatility of point eight million dollars, V a R is point eight times two point eight, two point two four million.
- On a stressed day, with volatility of one point six million, it is four point four eight million.

13 Peaks over threshold: the idea

- Choose a high threshold u
- Keep only the losses that exceed it
- Model the excesses: loss − u
- Thornbury: u = $2m, 50 days over; $2.6m → $0.6m excess
- Every large loss is used, not one per period

IN THE LESSON

- Now to the most widely used E V T method, peaks over threshold, or P O T.
- The idea is simple.
- Choose a high threshold, written u.
- Keep only the observations that exceed it, and throw the rest away.
- Then work with the excesses, meaning each loss minus the threshold.
- Thornbury sets its threshold at two million dollars, and fifty of its one thousand days exceed it.
- A loss of two point six million is an excess of point six million.
- Every large loss in the history is used, not just one per month or per year, and the ordinary days that tell us nothing about the tail are left out.

14 The Pickands-Balkema-de Haan theorem

Excess y = Loss − u, given Loss > u: y ~ GP(ξ, β)
P(Loss > x) ≈ (Nu ÷ n) × [1 − G(x − u)]

Shape, ξ	Same tail shape as the GEV
Scale, β	Typical size of the excesses
Share beyond u	Nu ÷ n: fraction of days
Threshold, u	Chosen by the risk manager

IN THE LESSON

- The Pickands Balkema de Haan theorem says that, for a high enough threshold, the excesses follow a generalised Pareto distribution, or G P, with a shape parameter and a scale parameter.
- The shape is the same tail shape that appears in the G E V.
- The scale sets the typical size of the excesses.
- The second line turns this into a tail probability for the whole sample.
- The chance that a loss exceeds a level x above the threshold is the fraction of observations beyond the threshold, N u over n, times the G P chance that the excess goes beyond x minus u.
- That second factor is called the survival function: one minus the distribution function.
- The threshold itself is chosen by the risk manager.

15 Example 5: how often is a loss above $5m?

u = $2m, Nu = 50, n = 1,000, ξ = 0.25, β = $0.8m. Find P(Loss > $5m)

Share of days above u	50 ÷ 1,000	**5%**
Excess over u	5 − 2	**$3m**
1 + ξ × excess ÷ β	1 + 0.25 × 3 ÷ 0.8	**1.9375**
GP survival: power −1/ξ = −4	1.9375^{−4}	**0.0710**
P(Loss > $5m)	5% × 0.0710	**0.35%**

About once every 282 days

IN THE LESSON

- Thornbury's fit above two million dollars gives a shape of point two five and a scale of point eight million.
- So how often should it lose more than five million?
- First, the share of days above the threshold is fifty over one thousand, five per cent.
- The excess we need is five minus two, three million.
- Next, one plus the shape times the excess over the scale: one plus point two five times three over point eight, which is one point nine three seven five.
- Raise that to the power minus four, which is minus one over the shape: about point zero seven one, the G P survival probability.
- Multiply: five per cent times point zero seven one is about point three five per cent, or about once every two hundred and eighty two days.

16 Choosing the threshold: bias against variance

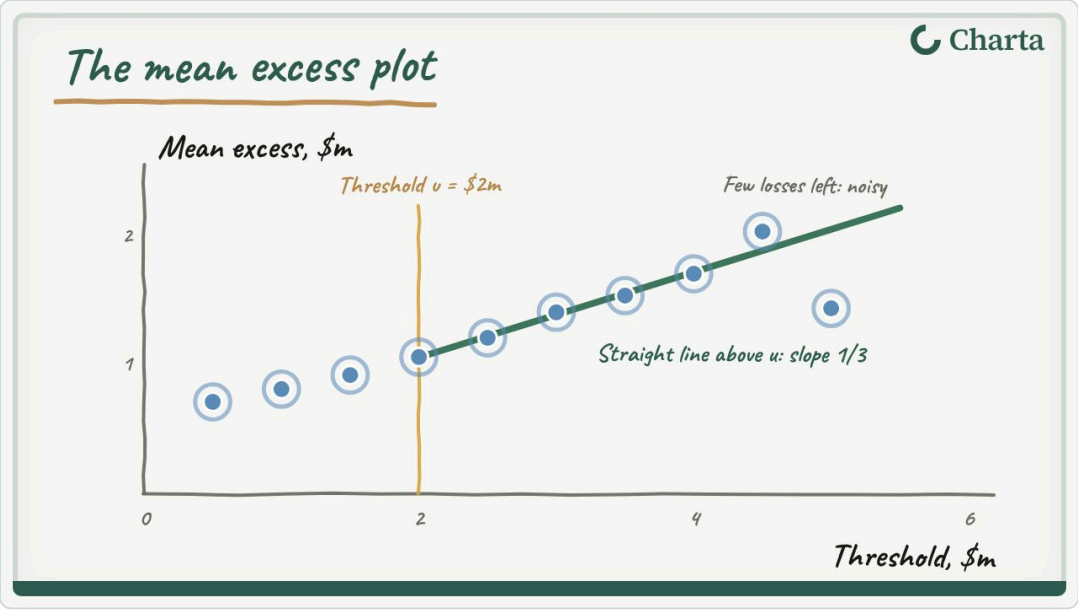
A lower threshold keeps more data but fits the GP worse

Threshold u	Exceedances	Problem
$1.0m	200	GP fit poor: bias
$2.0m	50	Balanced: chosen
$4.5m	about 5	Too few: noisy

IN THE LESSON

- A lower threshold keeps more data, but fits the generalised Pareto worse.
- So choosing it is a trade-off between bias and variance.
- At one million dollars there are two hundred exceedances, plenty of data, but many of those losses are ordinary days, so the G P approximation is poor and the estimate is biased.
- At two million there are fifty exceedances, a balance, and that is the threshold Thornbury chose.
- At four and a half million, the fitted tail implies only about five exceedances, which is too few: the estimates become noisy and swing with every new loss.
- Too low gives bias, too high gives noise.

17 The mean excess plot



IN THE LESSON

- A standard tool is the mean excess plot.
- Along the bottom is a candidate threshold, in millions of dollars.
- Up the side is the mean excess: the average amount by which losses beyond that threshold exceed it.
- Each dot is computed from Thornbury's data.
- The gold line marks the threshold of two million dollars.
- Above it, the dots lie close to a straight line, and that is the sign of a good threshold, because a G P tail has a mean excess that rises in a straight line.
- The slope is one third.
- An upward slope signals a fat tail, with a positive shape.
- At the far right only a few losses are left, so the last dots are noisy.

18 Example 6: reading the slope

GP tail: mean excess above v is e(v) = (β + ξ(v − u)) ÷ (1 − ξ)

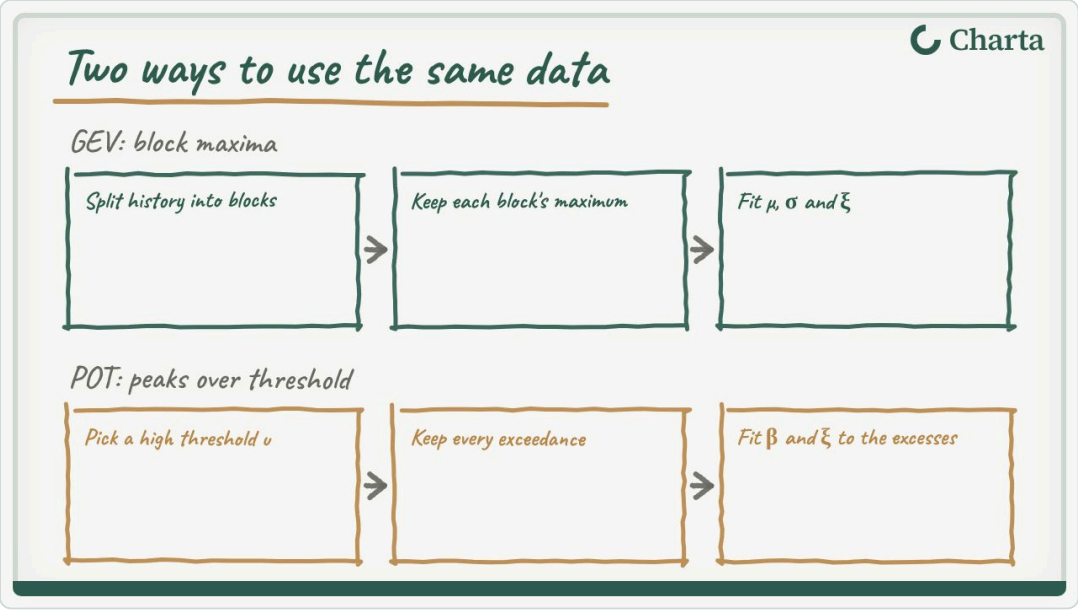
Mean excess at v = $2m	0.8 ÷ 0.75	**$1.07m**
Mean excess at v = $3m	(0.8 + 0.25) ÷ 0.75	**$1.40m**
Mean excess at v = $4m	(0.8 + 0.5) ÷ 0.75	**$1.73m**
Slope	0.25 ÷ 0.75	**1/3**
Shape from the slope	(1/3) ÷ (1 + 1/3)	**ξ = 0.25**

Trap: the threshold is not harmless; it drives the answer

IN THE LESSON

- Here is why the line is straight.
- For a G P tail, the mean excess above any level v at or above the threshold is the scale parameter plus the shape parameter times v minus u, all over one minus the shape parameter.
- At two million dollars it is point eight over point seven five, one point zero seven million.
- At three million it is point eight plus point two five, over point seven five, one point four zero million.
- At four million, point eight plus point five, over point seven five, one point seven three million.
- Each extra million adds a third, so the slope is point two five over point seven five, one third.
- Working backwards, the shape equals the slope over one plus the slope: one third over four thirds, point two five.
- And here is the trap: the threshold is not harmless.
- It drives the answer.

19 Two ways to use the same data



IN THE LESSON

- Both main E V T methods rest on the same theory, but they use the data differently.
- Look at the first row, the generalised extreme value, or block maxima, approach.
- Split the history into blocks, for example months or years.
- Keep only each block's maximum loss.
- Then fit the G E V to those maxima, with its location, scale and shape.
- Now the second row, peaks over threshold.
- Pick a high threshold.
- Keep every exceedance, however many fall in the same month.
- Then fit the generalised Pareto, with its scale and shape, to the excesses.
- The exam point: G E V is block based, and P O T is threshold based.

20 Example 7: what each method keeps

1,000 days in 50 blocks of 20 days; the 50 losses above $2m fall in 18 blocks

GEV keeps	one maximum per block	**50 maxima**
Large losses GEV keeps	one per block that has any	**18**
Large losses GEV throws away	50 − 18	**32**
GEV maxima below u	calm blocks: 50 − 18	**32**
POT keeps	every loss above u	**50**

IN THE LESSON

- Now see what each method keeps.
- For the G E V, we cut the history into fifty blocks of twenty trading days, roughly months, and keep one maximum per block: fifty maxima.
- Now, the fifty losses above two million dollars fall in only eighteen blocks, because big losses cluster in bad months.
- So the G E V keeps just eighteen of them, one per block, and throws away the other thirty two.
- Worse, the other thirty two blocks were calm, so thirty two of its maxima lie below the threshold and are not really extreme at all.
- Peaks over threshold keeps all fifty large losses.
- That is why it is called more data efficient.

21 Example 8: a GEV estimate of 99.9% daily VaR

Block maxima fit: μ = $1.9m, σ = $0.7m, ξ = 0.25; blocks of 20 days

P(block max ≤ VaR)	0.999^{20}, iid days	**0.9802**
Minus its natural log	20 × −ln 0.999	**0.02001**
Raise to the power −ξ	0.02001^{−0.25}	**2.6588**
Scale over shape, σ ÷ ξ	0.7 ÷ 0.25	**2.8**
GEV VaR: μ + (σ ÷ ξ)(… − 1)	1.9 + 2.8 × (2.6588 − 1)	**$6.54m**

IN THE LESSON

- The G E V can still give a daily V a R.
- Thornbury's block maxima give a location of one point nine million dollars, a scale of point seven million and a shape of point two five.
- If days are independent, the block maximum stays below a level only when all twenty days do.
- So a daily ninety nine point nine per cent level matches a block probability of point nine nine nine to the power twenty, about point nine eight zero two.
- Minus its natural log, twenty times minus the log of point nine nine nine, is point zero two zero zero one.
- Raise that to the power minus the shape: two point six five eight eight.
- The scale over the shape is point seven over point two five, two point eight.
- Subtract one from the power, to get one point six five eight eight, so the quantile is one point nine plus two point eight times that, about six point five four million dollars.

22 GEV against POT

Same theory and the same shape ξ; different use of the data

	GEV	POT
Data used	One maximum per block	All losses above u
Parameters	μ, σ, ξ	β, ξ, given u
Key choice	Block size	Threshold u
Weakness	Wastes data; long history	Exceedances cluster
Usual in finance	Less often	Yes: data-efficient

IN THE LESSON

- Here is the side-by-side comparison.
- Both share the same shape parameter.
- Data used: the G E V uses one maximum per block, while P O T uses every loss above the threshold.
- Parameters: the G E V fits location, scale and shape; the generalised Pareto fits scale and shape, given the threshold.
- Key choice: for the G E V it is the block size, itself a bias and variance trade-off, since small blocks break the theory and large blocks leave few maxima.
- For P O T it is the threshold.
- Weakness: the G E V wastes data and needs a long history, while P O T exceedances cluster in time.
- Usual choice in finance: P O T, because it is more data efficient.

23 The generalised Pareto distribution

$G(y) = 1 − (1 + ξy ÷ β)^{−1/ξ}$, $y ≥ 0$
$ξ = 0$: $G(y) = 1 − e^{−y ÷ β}$ (exponential)

Excess over u, y	Loss minus u, given Loss > u
Shape, ξ	Positive: a fat tail
Scale, β	Typical size of an excess

IN THE LESSON

- Now the central tool of peaks over threshold: the generalised Pareto distribution.
- It models the excess over the threshold, written y, which is zero or more.
- Given that a loss exceeds the threshold, the chance that the excess is at most y is G of y, equal to one minus the bracket, one plus the shape parameter times y over the scale parameter, raised to the power minus one over the shape parameter.
- Two parameters: the shape, the same as before, and the scale, the typical size of an excess.
- When the shape is zero, as the second line shows, the excess is exponential: G of y is one minus e to the power minus y over beta.
- A positive shape means a fat tail.

24 VaR and expected shortfall from the GP

$$q = (n \div Nu) \times (1 - c)$$
$$VaR = u + (\beta \div \xi) \times (q^{-\xi} - 1)$$
$$ES = (VaR + \beta - \xi u) \div (1 - \xi)$$

Rescaled tail probability, q	Turns 1 − c into exceedance terms
Threshold, u	Add it back to the excess
Expected shortfall	Exists only if ξ < 1

IN THE LESSON

- From the generalised Pareto we get V a R and expected shortfall.
- Start with q.
- With n observations, N u exceedances and confidence c, q equals n over N u, times one minus c.
- It rescales the tail probability, one minus c, into the sample of exceedances.
- Then V a R equals the threshold, u, plus the scale parameter over the shape parameter, times q to the power minus the shape parameter, minus one.
- So V a R is the threshold plus a scaled quantile of the excess.
- Expected shortfall equals V a R plus the scale parameter minus the shape parameter times u, all divided by one minus the shape parameter.
- It only exists when the shape is below one, because otherwise the mean loss in the tail is infinite.

25 Example 9: Thornbury's 99% VaR and ES

n = 1,000, Nu = 50, u = $2m, ξ = 0.25, β = $0.8m, c = 99%

Rescaled tail probability q	(1,000 ÷ 50) × 1%	**0.2**
q to the power −ξ	0.2^{−0.25}	**1.4953**
VaR	2 + (0.8 ÷ 0.25) × (1.4953 − 1)	**$3.585m**
ES	(3.585 + 0.8 − 0.5) ÷ 0.75	**$5.18m**

Normal 99% VaR: 2.326 × $1m = $2.33m

IN THE LESSON

- Now Thornbury's ninety nine per cent figures: one thousand observations, fifty above the two million dollar threshold, a shape of point two five and a scale of point eight million.
- First, q is one thousand over fifty, which is twenty, times one per cent: point two.
- Next, point two to the power minus point two five is one point four nine five three.
- Subtract one, to get point four nine five three, and multiply by the scale parameter over the shape parameter, point eight over point two five, which is three point two.
- So V a R is two plus three point two times point four nine five three, about three point five eight five million dollars.
- Expected shortfall is three point five eight five plus point eight minus point five, over point seven five: about five point one eight million.
- The normal model would have said only two point three three million.

26 Example 10: going out to 99.9%

Same fit; confidence c = 99.9%

Rescaled tail probability q	20 × 0.1%	**0.02**
q to the power −ξ	0.02^{−0.25}	**2.6591**
VaR	2 + 3.2 × (2.6591 − 1)	**$7.31m**
ES	(7.31 + 0.8 − 0.5) ÷ 0.75	**$10.15m**

Beyond the worst loss, $6.5m. GEV gave $6.54m; the normal $3.09m

IN THE LESSON

- Now push out to ninety nine point nine per cent, where the data alone gave us about one observation.
- q is twenty times point one per cent, point zero two.
- Point zero two to the power minus point two five is two point six five nine one.
- Subtract one, to get one point six five nine one.
- V a R is two plus three point two times one point six five nine one, about seven point three one million dollars.
- Expected shortfall is seven point three one plus point eight minus point five, over point seven five, about ten point one five million.
- Notice that the V a R lies beyond the worst loss in the sample, six and a half million.
- Historical simulation could never reach it.
- The block maxima estimate was six point five four million, and the normal said three point zero nine.

27 Example 11: three classic slips

Same fit, 99% confidence: what the usual errors give

Correct VaR	u + scaled GP quantile	**$3.585m**
Slip 1: drop n ÷ Nu, q = 1%	2 + 3.2 × (0.01^{−0.25} − 1)	**$8.92m**
Slip 2: forget to add u back	3.2 × 0.4953	**$1.585m**
Slip 3: VaR formula for ES	should be 5.18	**$3.585m**

ES exists only if ξ < 1; otherwise the tail mean is infinite

IN THE LESSON

- Here are three classic slips, using the same fit at ninety nine per cent.
- The correct V a R, as we found, is three point five eight five million dollars.
- Slip one: leave out n over N u, and use one per cent itself as q.
- Then q to the power minus point two five is about three point one six, and V a R becomes about eight point nine two million, far too high.
- Slip two: forget to add the threshold back.
- That gives only one point five eight five million, the scaled excess alone.
- Slip three: use the V a R formula for expected shortfall, and report three point five eight five million, when the right answer is five point one eight.
- And remember that expected shortfall exists only when the shape is below one; otherwise the tail mean is infinite.

28 Extremes in several risk factors at once

- Step 1: each factor's own tail, with GEV or GP
- Step 2: how the extremes move together
- Tools: an extreme value copula or tail dependence
- Ordinary-day correlation says little about crises
- Marginals alone are not enough

IN THE LESSON

- A real bank has many desks and risk factors, and the danger is that they crash together.
- Multivariate E V T extends tail modelling to several variables at once, in two steps.
- Step one: model each variable's own tail with univariate E V T, using the G E V or the generalised Pareto.
- These are the marginals.
- Step two: model, separately, how the extremes move together.
- The tools are an extreme value copula, which is a function that joins the marginals into a joint distribution, or a tail dependence coefficient.
- Why bother?
- Because correlation estimated from ordinary days says little about how assets behave in a crisis.
- That is the exam point: the marginals alone are not enough.

29 Tail dependence

λ = lim (α → 1) P(Y beyond its α-quantile, given X is)
Gaussian copula: λ = 0 for any ρ < 1

Tail dependence, λ	How often extremes coincide
Correlation, ρ	Ordinary-day correlation
Copula	Gaussian: no tail dependence

IN THE LESSON

- The tail dependence coefficient, lambda, measures how often extremes happen together.
- Take two variables, X and Y.
- Lambda is the limit, as the tail level, alpha, goes to one, of the probability that Y is beyond its alpha quantile, given that X is beyond its alpha quantile.
- If lambda is zero, joint extremes become vanishingly rare relative to single ones as we go further out, however correlated the two are on ordinary days.
- Here is the trap.
- The Gaussian copula, the dependence structure of the multivariate normal, has a lambda of zero for any correlation, rho, below one.
- So it understates the chance of joint crashes and overstates diversification.

30 Example 12: joint crash days

Rates and equity desks: each has a 0.1% chance of a 99.9% loss day

Joint chance if independent	0.1% × 0.1%	**0.0001%**
Wait, at 250 days a year	1 ÷ 0.000001 ÷ 250	**4,000 years**
Joint chance, 40% tail link	0.1% × 40%	**0.04%**
Wait, with the tail link	1 ÷ 0.0004 ÷ 250	**10 years**
Joint crashes more likely by	0.04% ÷ 0.0001%	**400 times**

IN THE LESSON

- Thornbury has a rates desk and an equity desk.
- Each has a point one per cent chance, on any day, of a ninety nine point nine per cent loss day.
- If the two were independent, the chance of both on the same day would be point one per cent times point one per cent, which is point zero zero zero one per cent.
- At two hundred and fifty trading days a year, that is once in four thousand years.
- Now suppose that, given one desk has such a day, the other has one forty per cent of the time.
- The joint chance becomes point one per cent times forty per cent, point zero four per cent, or once in ten years.
- Joint crashes are four hundred times more likely.

31 Example 13: the joint tail runs out of data

1,000 days; each desk exceeds its own 99% level on 10 days

Two desks, if independent	1,000 × 1% × 1%	**0.1 days**
Two desks, seen together	counted	**4 days**
Three desks, if independent	1,000 × 1%^{3}	**0.001 days**
Three desks, seen together	counted	**2 days**

Joint extremes are rare, and grow rarer with each variable

IN THE LESSON

- The limitation of multivariate E V T is data.
- Thornbury's one thousand days give each desk ten days beyond its own ninety nine per cent level.
- For two desks, if they were independent, we would expect one thousand times one per cent times one per cent, point one joint days.
- The history shows four, which is strong evidence of tail dependence, but four observations are very little to fit a model to.
- Add a third desk.
- Independence would give one thousand times one per cent cubed, point zero zero one days, and the history shows just two.
- Joint extremes are far rarer than single-asset extremes, and the count in the joint tail shrinks quickly as the number of variables grows.

32 What multivariate EVT is for

- Portfolio tail VaR across desks
- Contagion: does a crash in one market spread?
- Stress scenarios with joint extreme moves
- Gaussian copula understates joint crashes
- Limit: joint extremes are far rarer than single ones

IN THE LESSON

- Multivariate E V T supports portfolio tail V a R across desks, because the bank's loss in a crisis depends on how many desks crash at once.
- It supports contagion analysis, asking whether a crash in one market spreads to another.
- And it supports stress scenarios built on joint extreme moves, rather than on ordinary-day correlations.
- Remember that a Gaussian copula understates joint crashes, however high its correlation, short of one.
- And remember the limitation: joint extremes are far rarer than single ones, so the estimates are uncertain.

33 Mistakes to avoid

- Clustered data: standardise by volatility first
- The threshold or block size is not harmless
- GP: keep n ÷ Nu, and add u back
- ES has its own formula, and needs ξ < 1
- Ordinary-day correlation fails in a crisis

IN THE LESSON

- Before the quiz, here are the mistakes to avoid.
- One: applying E V T to clustered data as if it were independent.
- Standardise by a volatility forecast first.
- Two: treating the threshold, or the block size, as harmless.
- It drives the answer, so check the mean excess plot.
- Three: in the generalised Pareto formulas, dropping the factor n over N u, or forgetting to add the threshold back.
- Four: using the V a R formula for expected shortfall, and forgetting that expected shortfall needs a shape below one.
- Five: trusting ordinary-day correlation, or a Gaussian copula, to describe what happens in a crisis.

34 Try three quick questions

1. n = 2,000, Nu = 40. For 99.5% VaR, what is q?
   Answer: q = (2,000 ÷ 40) × 0.5% = 0.25.
2. A GEV fit gives ξ = 0.5. Family and tail index?
   Answer: Fréchet, fat tail; index 1 ÷ 0.5 = 2.
3. GP: VaR $4m, β = $1m, ξ = 0.2, u = $2.5m. ES?
   Answer: (4 + 1 − 0.5) ÷ 0.8 = $5.625m.

IN THE LESSON

- Now try three quick questions.
- Pause, think, then listen.
- Question one.
- A sample has two thousand observations and forty exceed the threshold.
- For ninety nine point five per cent V a R, what is q?
- Two thousand over forty is fifty, times point five per cent, so q is point two five.
- Question two.
- A G E V fit gives a shape of point five.
- Which family, and what is the tail index?
- A positive shape means Fréchet, a fat tail, with a tail index of one over point five, which is two.
- Question three.
- A generalised Pareto fit gives a V a R of four million dollars, a scale of one million, a shape of point two and a threshold of two and a half million.
- What is expected shortfall?
- Four plus one minus point five, over point eight, is five point six two five million dollars.