VaR & Expected Shortfall: The Precision Paradox

We built ever-more-precise risk models – only to discover they fail exactly when precision matters most

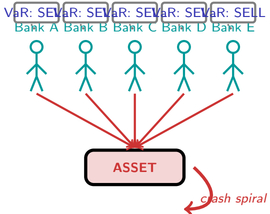
Digital Finance

Why Did Every Major Bank's Risk Model Fail on the Same Day?

The Precision Paradox

VaR models gave every bank a single, precise number: the maximum loss at a given confidence level. For decades, this number was the cornerstone of risk management and capital allocation.

- In 2008, every major bank's VaR said losses were within acceptable bounds – until they were not
- The problem was not that models were wrong individually; they were all wrong in the *same way*
- VaR promised certainty: "We will not lose more than X with 99% confidence"
- What it delivered: systematic underestimation of tail risk because Gaussian assumptions compress the very tails that matter
- The models created a shared blind spot across the entire financial system



When everyone uses the same model, the model becomes the risk.

VaR was designed to measure risk – but when every bank uses the same measure, the measurement itself creates systemic risk.

How Confident Are You in a Number That Has Been Wrong Every Time It Mattered?

Quick Exercise

Before we define VaR formally, test your own risk intuition:

1. A model tells you there is a 1% chance of losing your entire savings tomorrow. How does that number make you feel – safe, or worried?

2. Now you learn the same model failed to predict every major financial crisis in the past 30 years. Does that 1% number still mean the same thing?

3. Can you name a single crisis where VaR models gave an advance warning?

The disconnect between precision and reliability is the core of the precision paradox. VaR gives you a number with decimal places, confidence intervals, and statistical significance. It looks *exact*. But exactness is not the same as accuracy.

A thermometer that consistently reads two degrees too low is precise – it gives the same reading every time – but it is wrong in a way that matters when you are checking for a fever. VaR is that thermometer for financial risk: precise, consistent, and systematically wrong about the cases that matter most.

VaR gives the illusion of precision – a single number that summarizes infinite complexity. The precision paradox is that this precision is most misleading when it matters most.

What Are the Three Approaches to Measuring VaR – and Where Does Each Break?

Dimension	Historical	Parametric	Monte Carlo
Data needed	250+ past returns	Mean, variance, correlation	Return model + simulation
Assumption	Past repeats	Returns are normal	Model is correct
Speed	Fast	Very fast	Slow
Fat tails	Captured if in sample	Missed (Gaussian)	Depends on model
Correlation	Captured if stable	Fixed matrix	Can model regime change
Crisis behavior	Only past crises	Underestimates	Only if modeled
Regulatory use	Basel II internal	Basel II standard	Basel III preferred

Pattern to notice: Reading left to right, sophistication increases – but not necessarily accuracy. Historical VaR is limited to what has already happened. Parametric VaR is limited by the normal distribution assumption. Monte Carlo VaR is limited by whatever model you feed it.

All three share a common weakness: they estimate a *quantile* of the loss distribution but say nothing about what happens beyond it.

The shared blind spot

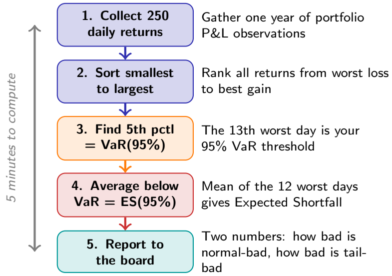
- Every VaR method answers: "What is the worst loss that will *not* be exceeded $\alpha\%$ of the time?"
- None of them answers: "If we *do* exceed that threshold, how bad could it get?"
- This is like a weather forecast that says "it probably will not flood" but refuses to say how deep the water gets if it does

Why this matters for regulation

- Basel II let banks choose their VaR method – creating inconsistency across institutions
- A bank using parametric VaR could hold less capital than one using historical VaR for the *same portfolio*
- Basel III moved toward Expected Shortfall precisely to close this gap

Three methods, one shared weakness: every VaR approach assumes something about the future that crises routinely violate.

Follow One VaR Calculation from Raw Returns to Risk Number



Worked example: CHF 100M equity portfolio

- Collect 250 daily returns (one trading year)
- Sort from worst (-4.8%) to best ($+3.9\%$)
- The 13th worst return is -2.3%

$\Rightarrow$ **VaR(95%, 1-day)** = CHF 2.3M

"On 95% of days, we lose less than CHF 2.3M."

- Average the 12 returns worse than -2.3%
- Mean of those 12 days: -3.8%

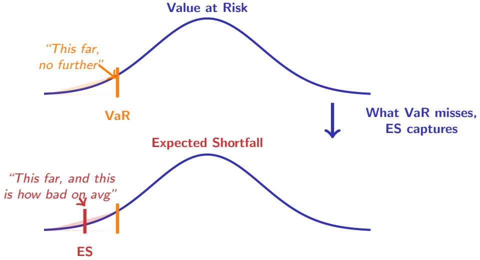
$\Rightarrow$ **ES(95%, 1-day)** = CHF 3.8M

"On the 5% of bad days, the average loss is CHF 3.8M."

Key insight: ES is $1.65\times$ VaR here. That multiplier reveals how heavy the tail is – the heavier the tail, the larger the ratio.

The historical approach is the simplest: sort your returns and count from the worst. The 13th worst day in 250 is your 95% VaR.

Why Is Expected Shortfall a Better Risk Measure Than VaR – and What Did We Sacrifice?



VaR answers: "How bad is a normal bad day?"
But it is *silent* on severity – a VaR breach of CHF 3M and CHF 300M look identical.

ES answers: "Given we are in the tail, how bad on average?"
It captures the *shape* of the tail, not just its boundary.

Mathematical properties:

- ES is **coherent** (subadditive): diversification always reduces ES
- VaR is **not** coherent – adding assets can increase VaR, penalizing diversification
- ES satisfies all four coherence axioms; VaR fails subadditivity

The Basel III shift:

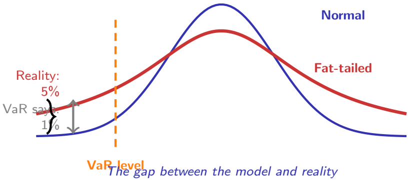
- FRTB replaced 99% VaR with 97.5% ES for market risk capital
- The sacrifice: ES is harder to backtest because tail events are rare by definition

ES answers the question VaR refuses to ask: if we are past the cliff edge, how far is the fall? That is why Basel III shifted from VaR to ES.

What Happens When Your Risk Model Assumes Normality but Reality Has Fat Tails?

Four ways VaR fails when it matters most

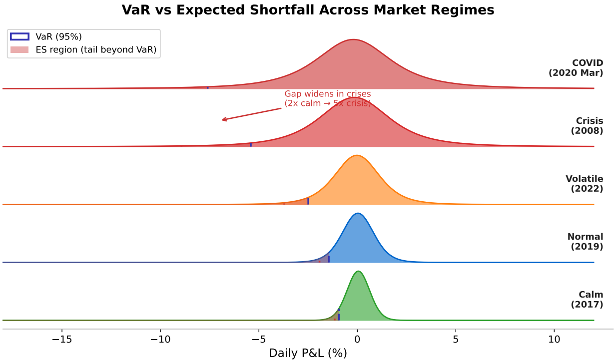
- **Fat tails:** Real financial returns have tails 10–100$\times$ heavier than the normal distribution predicts. A "4-sigma" daily loss that should occur once every 126 years happens roughly once a decade.
- **Correlation breakdown:** In calm markets, correlations are moderate. In crises, correlations spike toward 1 – all assets fall together, exactly when diversification is supposed to protect you.
- **Procyclicality:** VaR measured from recent data falls during calm periods (less capital required) and spikes in crises (more capital required, forcing sales, deepening the crisis).
- **Non-subadditivity:** VaR can report that a merged portfolio is riskier than the sum of its parts – punishing diversification, the one thing that actually reduces risk.



The gap between the model and reality is where crises live.

Fat tails mean extreme events are 10–100$\times$ more likely than normal models predict. The 2008 crisis was a 25-sigma event under normality – effectively impossible, yet it happened.

How Do VaR and ES Compare Across Different Market Regimes?



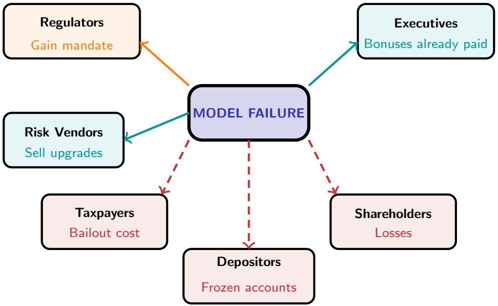
Reading the chart

- Each distribution represents a different market regime – from calm to crisis
- In **calm regimes**, distributions are narrow and VaR and ES are close together – both give similar risk estimates
- In **volatile regimes**, distributions widen and ES begins to exceed VaR noticeably as tails grow heavier
- In **crisis regimes**, the left tail extends dramatically – ES can be 2–3× VaR because the average tail loss far exceeds the threshold
- The gap between VaR and ES *widens precisely when it matters most* – during the regimes where risk is highest

Key takeaway: A single VaR number hides regime dependence. ES reveals it.

Illustrative distributions based on historical regime analysis. The chart reveals what summary statistics hide: the shape of the tail changes dramatically across regimes.

Who Benefits and Who Suffers When Risk Models Underestimate Tail Risk?



Harmed:

- **Taxpayers** bear the bailout cost when models underestimate risk and banks become insolvent
- **Shareholders** suffer losses that VaR said were virtually impossible
- **Depositors** face frozen accounts and potential losses above insurance limits

Benefit:

- **Executives** collected bonuses during the boom – bonuses tied to VaR-adjusted returns that understated true risk
- **Risk vendors** sell model upgrades after each failure

Complex:

- **Regulators** gain mandate for stricter rules but face questions about why models were approved

Moral hazard: Those who benefit from precision during calm times do not bear the cost when it fails in crises.

The precision paradox creates moral hazard: those who benefit from VaR precision during calm times are not the ones who pay when it fails in crises.

Three Questions That Reveal Whether a Risk Model Is Trustworthy

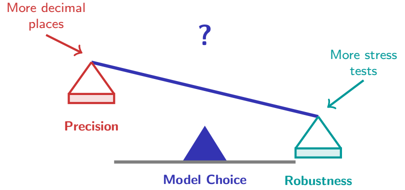
When evaluating any risk model – as a risk manager, regulator, or investor – ask these three diagnostic questions:

1. Does it account for fat tails?
If the model assumes normally distributed returns, it will underestimate extreme losses by an order of magnitude. Ask: what distribution does it use, and how were the tails calibrated?

2. Does it capture time-varying correlations?
Correlations spike in crises. A model using a fixed correlation matrix from calm periods will massively underestimate joint losses when markets crash together.

3. Has it passed out-of-sample backtesting?
A model that fits history perfectly but has never been tested on unseen data is an exercise in curve-fitting, not risk management. Ask for the backtest results – especially during stress periods.

The framework: If a model fails any one of these three, treat its output with extreme caution regardless of how precise the numbers appear.



More precision often means less robustness. Choose wisely.

A model that is precisely wrong is more dangerous than a model that is approximately right. The three questions test for robustness, not precision.

Your Challenge: Diagnose a Risk Model and Recommend Improvements

Your Challenge

The scenario: A Swiss bank reports a 99% 1-day VaR of CHF 15M using a Gaussian parametric model. Over the past 250 trading days, the bank experienced 8 VaR breaches (days where actual losses exceeded VaR).

1. Is 8 breaches consistent with a 99% VaR model?
   Hint: Expected breaches $= 250 \times 0.01 = 2.5$. Eight is the red zone under Basel's traffic light system.
2. What does this tell you about the assumed distribution?
   Hint: Too many breaches means tails are heavier than the model assumes.
3. The breach losses were CHF [18, 22, 25, 19, 31, 27, 20, 24] million. Calculate the Expected Shortfall.
   Answer: $\text{ES} = \frac{18+22+25+19+31+27+20+24}{8} = \text{CHF } 23.25\text{M}$ (that is $1.55\times$ the VaR of 15M)
4. Recommend one improvement to the bank's risk model.

Discuss: Where do you and your neighbor disagree on the recommendation? That disagreement reveals the precision-robustness tradeoff in practice.

The traffic light system says 8 breaches in 250 days is a red zone – the model is clearly wrong. The question is not whether to fix it, but how.