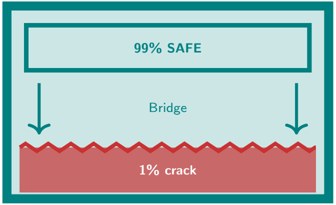
Why do risk models keep failing right when we need them most?

The paradox of precision:

- Risk models give us a single clean number: "There is a 95% chance we lose no more than this amount tomorrow"
- That precision feels reassuring – we quantified the danger
- But the number rests on assumptions: returns are normally distributed, volatility is stable, the past predicts the future
- When markets break – crashes, panics, contagion – those assumptions collapse
- The model that showed green lights yesterday shows catastrophic red today



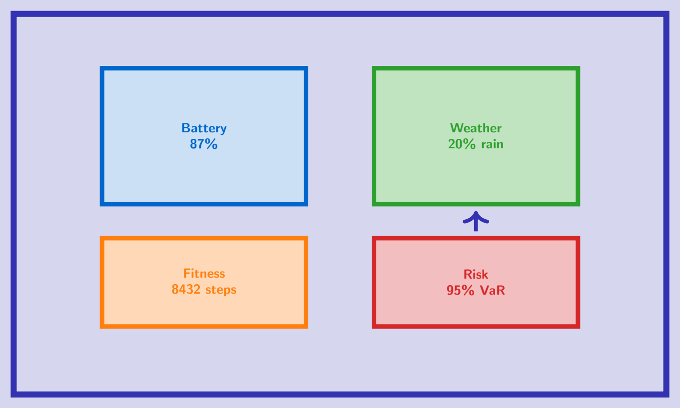
A person stands on a bridge marked "99% safe" while the 1% crack spreads underneath.

The core tension: We need numbers to manage risk, but trusting those numbers blindly creates new risk.

Risk measurement is not prediction. It is preparation for a range of outcomes – but the range itself is uncertain.

Every risk model is a map, not the territory. The map simplifies to be useful, but crises happen in the parts left off the map.

Have you ever felt safe because of a number – then realized the number was wrong?



We trust numbers daily: battery percentage, weather forecasts, step counts. Sometimes they are wrong. Risk numbers are no different – they summarize uncertainty, they do not eliminate it.

A number gives the illusion of control. Understanding what the number measures – and what it ignores – restores reality.

Risk models are tools, not oracles. The question is not "Is this number right?" but "What assumptions make this number valid?"

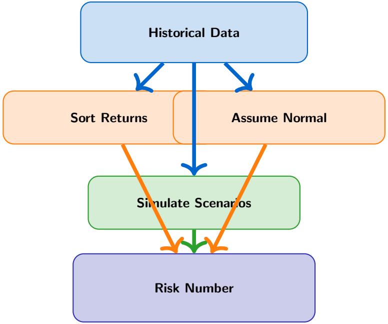
What are the main approaches to measuring market risk?

Three measurement families:

- **Value-at-Risk (VaR):** Sort historical returns and find the cutoff. "On 95% of days, losses are smaller than this."
- **Expected Shortfall (ES):** Average the losses beyond VaR. "When things go wrong, how bad is it on average?"
- **Volatility:** Measure daily swings. Standard deviation of returns becomes the input to other models.

Three estimation methods:

- **Historical simulation:** Sort actual past returns. Simple, no assumptions, but assumes the past repeats.
- **Parametric (variance-covariance):** Assume returns are normal, use formulas. Fast, elegant, but wrong when tails are fat.
- **Monte Carlo:** Simulate thousands of random scenarios. Flexible, handles complexity, but computationally expensive.



All methods start with data, make different assumptions, produce slightly different answers.

No single method is "correct". Real risk teams run all three and compare. When they disagree, that reveals model risk.

Measurement approach matters less than understanding what each method assumes and where it breaks.

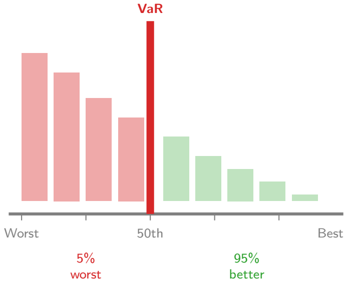
How does a Value-at-Risk calculation turn price history into a risk number?

The sorted-returns intuition:

1. Collect 1000 days of past returns
2. Sort them from worst to best
3. Count from the bottom: the 50th worst return is your 95% VaR
4. Interpretation: on 95% of days you did better than this; on 5%, you did worse

Example: Your portfolio is worth 100,000. The 50th worst return out of 1000 days is negative 2.1%. Your 95% daily VaR is 2,100. On 95% of days, you expect to lose no more than 2,100. On 5% of days, you could lose more.

What VaR tells you: The boundary of the worst 5%.
What VaR does NOT tell you: How bad it gets beyond that boundary.



The red bars are the 5% worst days. VaR is the line separating them from the rest.

VaR summarizes thousands of scenarios into one threshold. It is simple and widely used, but blind to what happens beyond the threshold.

VaR is a percentile, not an average. It marks the edge of the tail, not the severity inside the tail.

How do parametric, historical, and simulation risk models compare in architecture?

Historical simulation:

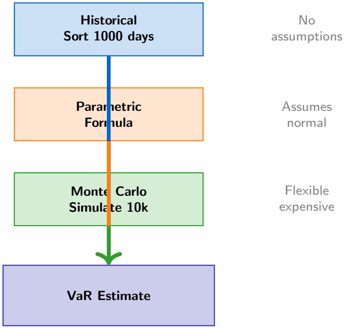
- Data in: actual past returns
- Process: sort them
- Output: the 5th percentile
- Pro: no assumptions
- Con: assumes the future looks like the past

Parametric (variance-covariance):

- Data in: mean and standard deviation
- Process: assume normal distribution
- Output: 1.65 times the standard deviation
- Pro: fast, elegant
- Con: underestimates fat tails

Monte Carlo:

- Data in: volatility, correlations
- Process: simulate 10,000 random paths
- Output: sort simulations, find 5th percentile
- Pro: handles complex portfolios
- Con: computationally expensive



Three paths to the same goal. Each makes different trade-offs between simplicity and realism.

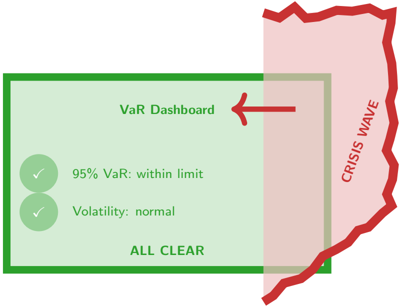
What happens when a risk model misses a fat-tailed event?

The fat-tail problem:

- Parametric models assume returns follow a bell curve
- Real returns have "fat tails": extreme events happen far more often than the bell curve predicts
- A 5-sigma event should happen once every $\sim$3.5 million trading days ($\sim$14,000 trading years) under a normal distribution (two-sided $P \approx 5.7 \times 10^{-7}$)
- In reality, 5-sigma moves happen roughly once every 10–20 years on major equity indices (1987, 2008, 2020, 2022, 2023)

Consequence: The model shows green (all clear) because 95% VaR looks fine. But the worst 1% of outcomes – the catastrophic tail – is much worse than the model predicted. When crisis hits, the model fails precisely when it is needed most.

Examples: 2008 financial crisis, 2020 pandemic crash, Long-Term Capital Management collapse. All were "impossible" under normal distribution assumptions.



The dashboard shows green while the crisis wave approaches from outside the model's range.

Fat tails mean the "impossible" happens regularly. Models built on normal distributions are dangerously overconfident.

Kurtosis measures tail fatness. Financial returns typically have excess kurtosis between 3 and 10. Normal distribution has zero.

Where do different risk measures disagree most about the same portfolio?

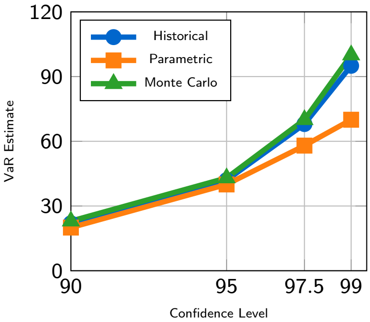
What you see: Three risk estimates for the same portfolio at different confidence levels.

Key pattern:

- At 95%, all three methods agree closely
- At 99%, parametric underestimates risk by ignoring fat tails
- Historical simulation and Monte Carlo capture tail events better

Why this matters: Regulators care about 99% VaR or Expected Shortfall. If your model underestimates tail risk, you hold too little capital and face insolvency when crisis hits.

Takeaway: Model choice is not academic. It determines how much capital you hold and whether you survive stress.



Disagreement widens in the tail. That is where crises live.

Method choice is invisible in calm markets but critical in crises. Tail risk is where models diverge and banks fail.

Expected Shortfall addresses this by measuring average loss beyond VaR, not just the threshold.

Who relies on risk numbers and who suffers when they are wrong?

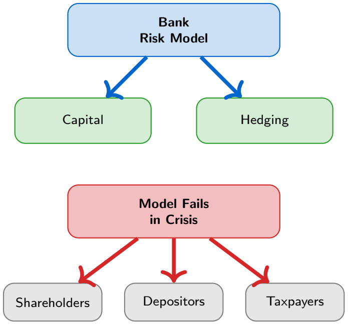
Who relies on risk models:

- **Banks:** Calculate capital requirements (Basel rules)
- **Regulators:** Set minimum capital, approve models
- **Fund managers:** Allocate across assets, set position limits
- **Corporates:** Decide how much FX or commodity risk to hedge
- **Exchanges:** Set margin requirements for futures

Who suffers when models fail:

- **Shareholders:** Lose value when undercapitalized bank fails
- **Depositors:** Face losses if deposit insurance is insufficient
- **Employees:** Lose jobs in crisis-driven layoffs
- **Taxpayers:** Fund bailouts when systemic banks collapse
- **Economy:** Credit freezes, recession follows bank failures

Asymmetry: The people who build models are not always the people who bear the consequences of model failure.



Risk models drive decisions that protect some and expose others.

Risk models are not neutral tools. They allocate capital, shape incentives, and distribute losses when they fail.

Regulatory capital based on flawed models creates moral hazard: banks appear safe while systemic risk grows.

Three questions to stress-test any risk measurement before trusting it

The Risk Model Stress Test:

(a) What distribution assumptions is the model making?

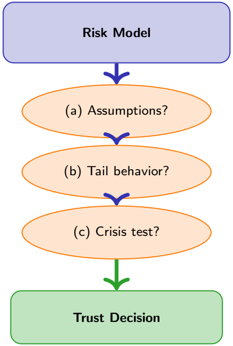
- Is it assuming returns are normal?
- Does it allow for fat tails, skewness, volatility clustering?
- Are correlations fixed or time-varying?

(b) What happens in the tail beyond the confidence level?

- Does the model report Expected Shortfall or only VaR?
- What is the worst-case scenario in the model's 1% tail?
- Is there a hard floor on maximum loss?

(c) Has the model been tested against actual historical crises?

- Run the model on 2008 financial crisis data. Does it predict the magnitude of losses?
- Test on 2020 pandemic crash, 2022 inflation shock
- If the model missed past crises, why trust it for the next one?



Three filters between model output and decision.

1. **Identify distribution assumptions and fat-tail treatment**
2. **Examine tail risk beyond the reported VaR threshold**
3. **Backtest against historical crises to validate predictions**

A risk model is only as good as its assumptions. Stress-testing reveals where assumptions break.

Your Challenge

Scenario: You manage a portfolio worth 500,000. You have collected 1,000 days of historical returns.

Task:

1. Calculate the 95% daily VaR using **historical simulation**: sort the returns, find the 50th worst, multiply by portfolio value.
2. Calculate the 95% daily VaR using **parametric method**: estimate the daily standard deviation, multiply by 1.65, multiply by portfolio value.
3. Compare the two results. Which is larger? Why do they differ?
4. Which estimate would you trust more for setting a risk limit? Explain your reasoning in two sentences.

Extension: If you were reporting this to a regulator, would you report the higher or lower estimate? What does that choice reveal about risk appetite versus regulatory compliance?

Reflection: When the two methods disagree, that gap is *model risk*. The gap widens in the tail. That is where crises happen.

Risk measurement is not about finding the "true" number. It is about understanding the range of plausible answers and the assumptions that produce each.

Real risk teams report multiple estimates and investigate disagreements. Convergence is comforting; divergence is informative.