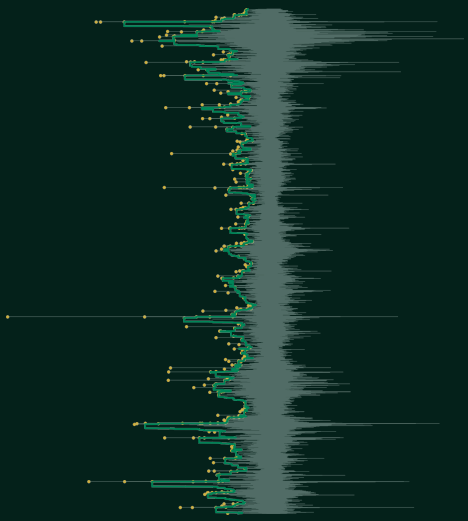
100 YEARS · 26,024 DAYS · ONE MODEL

THE RISK WITH NO HISTORY

the model every bank runs, a century of data, and what frontier AI breaks



NABIL EL MAHYAOUI

Expert IA & Digital · 2026

The Risk With No History

The Risk With No History

A century of market data, the model every bank runs, and the first systemic risk that arrives before its data

Nabil El Mahyaoui

An illustrated essay — every number computed from primary data

Copyright © 2026 Nabil El Mahyaoui.
All rights reserved.

Every figure in this book is generated by code from the Kenneth R. French Data Library (CRSP US market, daily, 1926–2026). Nothing is drawn by hand.

Contents

Opening — A model is a memory 2

ACT I — The letter 6

ACT II — The machine 11

ACT III — The days it broke 18

ACT IV — The third worst day 24

ACT V — Better mathematics 30

ACT VI — The risk with no history 36

CODA — How long, not how big **42**

Sources & reproducibility **48**

× 7.2

On 19 October 1987, the standard risk model said the worst plausible day was a loss of 2.4 %. The market lost 17.4 % before the close.

Computed from CRSP daily data, Kenneth R. French Data Library

Opening

Every risk model I have ever worked with, in a trading room or in front of a board, was a memory. It looked at what markets had done, arranged the past into a distribution, and read the future off it. This is not a flaw in the craftsmanship. It is the definition of the instrument.

Most of the time the instrument works, which is precisely what makes it dangerous. It fails on the days that matter, in the direction that hurts, by margins that decide who survives. And it fails for a reason that has nothing to do with bad engineering: the future occasionally stops resembling the sample.

On 31 August 2026, the chair of the Financial Stability Board wrote to the G20 that the most immediate concern for the financial system is frontier artificial intelligence. Not stretched valuations. Not sovereign debt. A class of risk for which there is no history at all — and therefore no sample, no distribution, and no model in the sense that finance uses the word.

This book takes that letter seriously by doing something concrete. We rebuild the risk model that nearly every bank on earth runs — historical Value at Risk — and we replay it, day after day, across a century of American market data: 26,024 trading days, from 1927 to 2026. Every chart in these pages is computed from that public record. Nothing is drawn by hand, and nothing is quoted that the code did not produce.

The thesis fits in three sentences. **A risk model is a memory, and it is excellent at measuring the losses you have already had. Better mathematics can stretch a memory, but cannot create one. A risk with no history therefore cannot be budgeted by size — only by the speed with which you recover from it — and this is exactly what the FSB has asked for.**

Seven acts. The *letter* — what the FSB actually said, and why central bankers do not use such words casually. The *machine* — Value at Risk rebuilt from scratch, honestly. The *days it broke* — a century of scorekeeping, 380 failures where the model allowed 260. The *third worst day* — what the number literally is, and the sample-size theorem that no sophistication escapes. *Better mathematics* — what extreme value theory buys you, measured on the same century, and where it stops. The *risk with no*

history — why frontier AI breaks the three assumptions that even the best tail models need. Then a coda: how to budget for a risk you cannot measure, and the five questions a board should ask.

Finance is not the subject of this book. It is the microscope. What it shows applies to any organisation that runs on models calibrated to yesterday — which, in 2026, is every organisation.

ACT I

The letter

31.08.2026

the day the financial system's referee named frontier AI its most immediate concern

Act I — The letter

The Financial Stability Board is not a think tank. It was created by the G20 in the wreckage of 2009 to watch the global financial system as a whole, and its chair — Andrew Bailey, governor of the Bank of England — writes to the world's finance ministers twice a year in language that is negotiated word by word. Central bankers deal in understatement the way surgeons deal in sterile instruments. When they escalate, the escalation is the message.

The letter of 31 August 2026 escalates. It surveys the usual vulnerabilities — stretched asset valuations, fragilities in sovereign debt markets, the opacity of private credit — and then sets them all aside to name something else:

> *"The most immediate concern is the potential impact of frontier AI on cyber risk."*

Two things in the phrasing deserve a quant's attention.

First, *frontier* AI — not chatbots, not the copilots in your office suite. The letter speaks of models "showing increasingly sophisticated autonomy and problem-solving abilities", and, in the same sentence, "threat capabilities". The concern is the capability step at the edge of the field, the systems that can plan, probe, adapt and execute with less and less human help.

Second, the mechanism. The letter does not say frontier AI creates a new category of loss. It says it may "materially alter the speed, scale and economics of cyber risk". Read that as

an engineer: three parameters of an existing threat, all moving at once, in the adversary's favour. An attack that once required a rare specialist team becomes software. What was expensive becomes cheap; what was slow becomes fast; what was targeted becomes broad. When the *economics* of an attack change by orders of magnitude, its frequency distribution is not the one in your data.

For the boardroom

Nothing in your risk inventory prices this. Cyber risk, yes — as an operational-risk line calibrated on historical incident data. But the letter's point is precisely that the generating process behind that data is about to change. A model calibrated on the old process is not conservative or aggressive about the new one. It is silent.

The letter's prescription is the most revealing part, and we will return to it in the coda. It does not ask supervisors to demand better cyber models. It asks for "robust response and recovery capabilities" and resilience among critical third-party technology providers, and it asks authorities to support safe and responsible model release and deployment. Response. Recovery. Not measurement.

The referee of the global financial system has quietly told you which instruments still work. To understand why the others do not, we need to open one up. The next act rebuilds the most widely used risk model in banking — with a century of real data and full transparency about what is inside.

ACT II

The machine

26,024

trading days of the US market, 1927–2026 — the laboratory for everything that follows

Act II — The machine

Here is how a bank measures market risk, stripped of vocabulary. Take the last 250 trading days — one year, the regulatory minimum observation window under the Basel market-risk framework. Sort them from worst to best. Walk in from the bad end and stop at the loss that only 1 % of days exceed. That number is your one-day Value at Risk at 99 %: the loss you should suffer, on average, one trading day in a hundred.

On a 250-day window, one percent of the sample is two and a half days. Round up: your Value at Risk is the *third worst day of the past year*. That is the entire machine. Everything else — the reports, the limits, the capital charges hung off it — is plumbing downstream of this one sorted list.

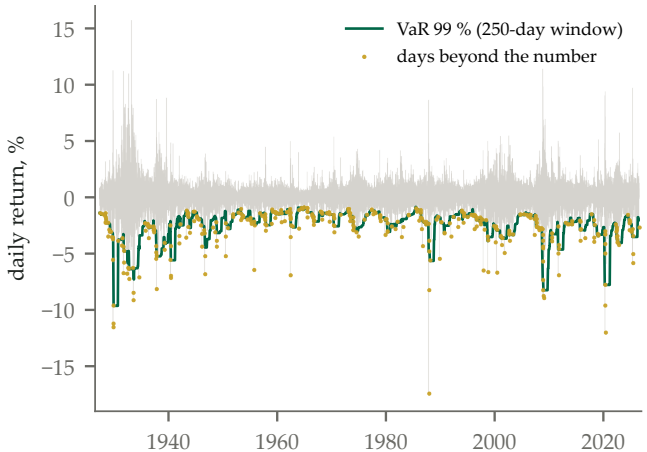


The 250 trading days before 19 October 1987, sorted. The gold bar is the third worst: 2.41%. That was the model's "worst plausible day" on the eve of the crash.

I want to be fair to this instrument, because I have used it for years and it is not stupid. It is distribution-free: it assumes nothing about bell curves, only that the recent past is a fair sample of the near future. It is auditable: anyone can recompute it from the same window. And on ordinary days it is remarkably serviceable — which is why regulators standardised

it, and why it survived every crisis that embarrassed it.

Now we do what a careful engineer does with any instrument: we test it out of sample, massively. The Kenneth French Data Library publishes the daily return of the entire US stock market — every NYSE, AMEX and NASDAQ stock, value-weighted — back to July 1926. We replay the machine across the whole record: every day from May 1927 to June 2026, compute the trailing 250-day VaR, then check whether the day's actual return fell through it. 26,024 experiments, one per trading day, no hindsight in either direction.



A century in one frame: 26,024 daily returns (pale), the model's daily worst-case line (green), and every day the market fell through it (gold).

The green line is the model doing its honest work: widening after turbulent years, tightening in calm ones, always one year of memory deep. The gold dots are the days it was wrong — wrong in the only direction risk management is paid to care about.

For the boardroom

Whenever a model's output looks authoritative, ask what its memory is. This one's memory is 250 days — by regulation. Every gold dot on the chart is a day the future exceeded one year of past. The next act counts them.

260

days beyond the number the model allowed, over the century

380

days beyond the number that actually happened

Computed on 26,024 trading days, 1927–2026 — 46 % more failures than promised

ACT III

The days it broke

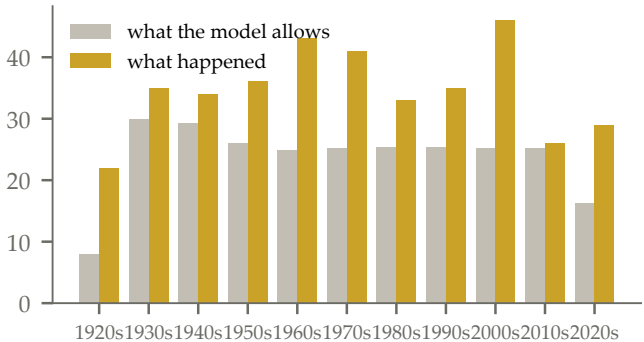
+46 %

more exceptions than a 99 % model is allowed to produce — and half of them arrive in clusters

Act III — The days it broke

A 99 % model earns its name by being wrong one day in a hundred. Over our 26,024 test days it was allowed 260 failures. It produced 380 — a rate of 1.46 %, forty-six percent above contract.

If the overshoot were spread evenly across a century, it would be a rounding story. It is not spread evenly. Fifty-four percent of all exceptions arrive inside clusters — stretches where the model fails three or more times within a few weeks. The clusters have names you know: 1929 to 1932, October 1987, autumn 2008, March 2020.

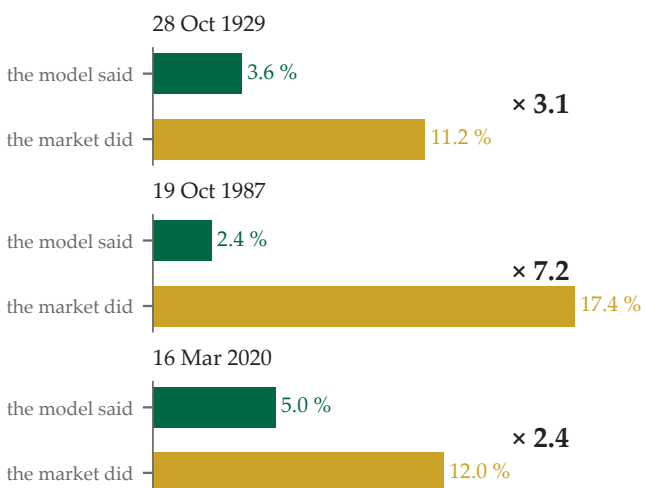


Failures by decade: what a 99 % model allows (grey) versus what happened (gold). The overshoot concentrates exactly where history concentrates its pain.

The clustering is not a detail; it is the mechanism. The model's only defence is its memory, and its memory is one year deep. When a regime breaks — a leverage unwind, a pandemic, a margin spiral — the first blow arrives while the window is still full of calm days, so the number is small precisely when the danger is large. Then the window fills with turbulence, the number balloons, and the model

becomes cautious exactly when the storm is ending. It is a rear-view mirror on a mountain road: accurate, polished, and pointed the wrong way.

Now look at the three days that define the problem.



The eve-of-day model estimate (green) against the day's actual loss (gold), for the three worst regime breaks of the century.

October 19, 1987 is the purest case in the record. The preceding year had been a roaring bull market; the sorted window said the worst plausible day was 2.41 %. The market fell 17.41 % — *seven point two times* the model's number, the largest single-day failure in a hundred years of data. March 2020 tells the same story in modern dress: the model said 5 %, the pandemic said 12 %. And 2008 is the exception that proves the rule — by 15 October the window was already full of September's carnage, so the model was only wrong by half. It had finally learned, one month too late to matter.

The model was not wrong about the past. It only knew the past.

For the boardroom

Three questions travel well beyond finance. Which of your models is calibrated on a window of calm? What in your business corresponds to the cluster — the failure that arrives *because* another failure just arrived? And who is responsible for the day after the model's worst case?

ACT IV

The third worst day

2.5%

the chance that a one-in-ten-thousand-day event even appears in one year of history

Act IV — The third worst day

Strip the vocabulary again and say what the number is. Your 99 % Value at Risk, computed on 250 days, *is* the third worst day of the past year. Not an estimate refined by theory. Not a synthesis of scenarios. An order statistic: you sorted a list and picked the third item.

State it that plainly and the structural limit becomes visible. A sorted list of 250 days contains no information whatsoever about events rarer than one-in-250. The quantile you can honestly estimate from a sample ends at the edge of the sample. Beyond that edge the empirical method does not degrade gracefully — it simply has nothing to say, while continuing to print a number every morning with the same typographic confidence.

How likely is it that the event you actually fear is even *in* your window? Elementary probability answers in one line. An event with a one-in-N daily probability appears in a window of n days with probability $1 - (1 - 1/N)^n$.

The sample-size theorem

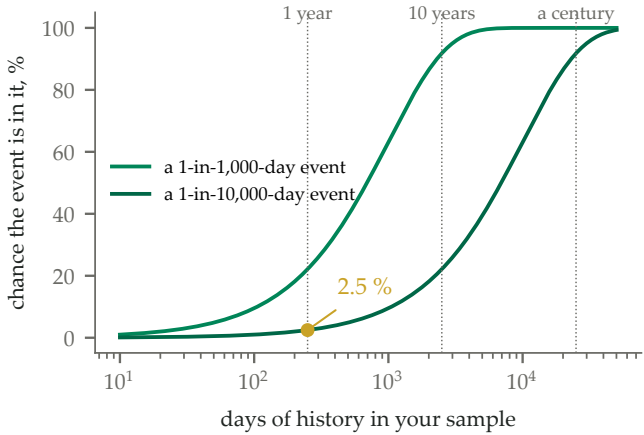
For a 1-in-10,000-day event (once in forty years):

in 250 days of history: **2.5 %**

in 10 years of history: **22.1 %**

in a full century: **92.6 %**

The model prints a number about the tail every day. The tail has a 97.5 % chance of never having been in its sample.



The chance that a rare event appears in your calibration window, as the window grows. One year of history is almost blind to the tail; even ten years sees a 1-in-10,000-day event only once in five tries.

Sit with the regulatory irony for a moment. The Basel framework requires *at least* one year of observation — a floor that became an industry default. So the world's standard market-risk measure is an instrument whose defining property is that the events it exists to warn about are almost never in its field of view.

Nassim Taleb made this argument famous as rhetoric. The point of this act is that it is not rhetoric; it is arithmetic you can verify on one line of a napkin, and the 380 gold dots of Act III are its empirical signature.

One more honest step. This limitation does not depend on the market being irrational, on traders panicking, or on any story at all. It would hold for a perfectly stationary casino. It is a property of *samples*, not of markets — which is why no amount of data engineering, machine learning, or governance can vote it away. The only escape anyone has found is to stop treating the sample as the answer and start treating it as the input to a theory. That escape exists, it is beautiful, and it is the subject of the next act — including what it cannot do.

For the boardroom

When a model reports a "once in a decade" or "worst case" figure, ask one question: was an event of that size in the calibration data? If the answer is no, the number is an extrapolation wearing the costume of a measurement. Extrapolations can be excellent — but they are theory, and theory must be interrogated, not audited.

ACT V

Better mathematics

$$\xi = 0.20$$

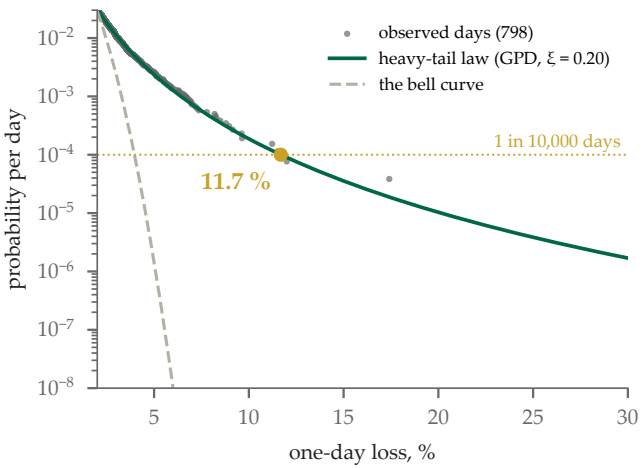
the tail index of a century of US market losses — the signature of a heavy tail, measured, not assumed

Act V — Better mathematics

There is a better instrument, and it deserves an honest exhibition rather than a caricature. Extreme value theory does for tails what the central limit theorem does for averages. A remarkable result — Pickands, Balkema and de Haan, 1974–75 — says that for essentially any reasonable distribution, the excesses over a high threshold converge to a single two-parameter family: the generalised Pareto distribution. You do not have to know the true law of markets. You only have to fit its tail.

So we fit it — on the same century, with everything shown. Threshold: a 2 % one-day loss. The record contains 798 such days. Maximum-likelihood fit of the generalised Pareto to those 798 excesses gives a tail index $\xi = 0.20$. Posi-

tive ξ is the mathematical signature of a heavy tail: a law whose extremes have no characteristic scale, only a rate of decay — and 0.20 is comfortably, measurably positive.



The tail of a century, on log scale: 798 observed loss days (dots), the fitted heavy-tail law (green), and the bell curve that classical finance would fit to the same data (dashed). The gold marker is the fitted 1-in-10,000-day loss: 11.7 %.

Look first at what the theory buys you. The dashed line is the bell curve fitted to the same century — mean 0.04 %, standard deviation 1.08 %. Under that curve, the crash of 1987 is a loss of sixteen standard deviations: an event expected once in roughly 10^{56} years, which is 10^{46} times the age of the universe. The bell curve does not underestimate the tail; it *denies* it. Under the fitted Pareto law, the same crash is an event expected about once every 213 years. Rare, violent, and — crucially — *on the map*. The same fit prices the 1-in-10,000-day loss at 11.7 %, with a bootstrap confidence band of 10.0 to 13.8 %. This is what real mathematical progress looks like: the model stops calling history impossible.

Now look at what the theory costs, because every extrapolation is a loan against assumptions.

First: **a stationary generating process**. The fit pools 1929, 1987 and 2020 into one law — the assumption that the machine generating extremes has not fundamentally changed in a century. For markets, a defensible approximation. It is the assumption to watch.

Second: **data in the tail's direction**. The theorem needs excesses to fit — here, 798 of them. Fewer observations widen that confidence band brutally; with none, there is no fit at all.

Third: **an exogenous adversary**. Nothing in the theory requires the extremes to refrain from adapting to your defences. Market crashes do not study your risk committee. This assumption has always been so obviously true in finance that nobody bothered writing it down.

Better mathematics extends a memory. It does not create one.

For the boardroom

Extreme value theory belongs in your toolkit: it is the difference between a model that calls 1987 impossible and one that expects it every other century. But notice what the best tail science on earth still needed here: seven hundred and ninety-eight observed catastrophes and a hundred stationary years. Hold that requirement in mind as you turn the page.

ACT VI

The risk with no history

0 days

observations of a frontier-model cyber campaign at systemic scale — the sample the FSB is worried about

Act VI — The risk with no history

Now place the FSB's warning against the three assumptions that even our best instrument required, and watch them fail in order.

No data in the tail's direction. A cyber attack conducted end-to-end by a frontier model at systemic scale has never been observed. Not rarely observed — never. The empirical method of Act II has no window; the theory of Act V has no excesses to fit. Ordinary cyber crime has a history, and insurers price it. But the letter is explicit that the concern is the *frontier* — the capability step, not the running average. For the step, the sample size is zero.

No stationarity — by construction. The letter's operative phrase is that frontier AI may alter "the speed, scale and economics" of cy-

ber risk. Economics is the word a quant should underline. When the cost of mounting a sophisticated campaign collapses — when what needed a state-sponsored team becomes an API call and a subscription — the frequency of attempts is not drawn from last year's distribution. Model capability improves in steps, several times a year. A process whose parameters jump every few months is the textbook definition of non-stationary, and each capability release makes yesterday's incident data a little more like archaeology.

An adversary that adapts. Market crashes, for all their violence, do not probe your defences and reroute. An attacker equipped with models showing — the letter's words — "increasingly sophisticated autonomy and problem-solving abilities" does exactly that, at machine speed, in parallel, against thousands of insti-

tutions at once. The extremes are no longer weather. They are an opponent, and an opponent's tail is whatever your defences fail to cover.

There is a fourth feature, and for the financial system it may be the sharpest. The frontier is *concentrated*. A handful of model providers, a handful of cloud platforms, one dominant chip designer. The same letter that names frontier AI asks for resilience among "critical third-party technology providers", because supervisors can read a dependency graph: when a thousand institutions stand on three platforms, exposures that used to be independent become one correlated position. Diversification — the oldest risk instrument of all, older than any formula — quietly stops working at the infrastructure layer.

Be precise about what this act does *not* claim. It does not claim catastrophe is imminent, and it does not claim AI in finance is net-negative — I spend my working days building the opposite case. The claim is narrower and harder: **this risk cannot be measured into a budget, because measurement is memory and the memory does not exist yet.** A VaR for frontier-AI cyber risk would be a number wearing the costume of knowledge. The honest statement is the one the gold dashed box makes in the figure of Act IV: no sample.

Which forces the real question. If you cannot budget the size of the loss, what *can* you budget? The FSB's letter already answered it, in the part everyone skimmed.

For the boardroom

Inventory the models your institution trusts, and sort them by the question of Act IV: is the feared event in the calibration data? Column A has a history — manage it with Acts II through V. Column B does not. For column B, every euro spent refining a number is a euro taken from the only budget that works — the one in the final act.

CODA

How long, not how big

how long ?

the only risk question that still has an answer when the sample size is zero

Coda — How long, not how big

Reread the FSB's prescription with everything the century taught us. The letter does not ask banks for a better cyber model. It asks for "robust response and recovery capabilities", resilience among critical third-party providers, and safe, responsible release and deployment of the models themselves. This is not regulatory hand-waving. It is precise mathematics for the situation, translated into supervisory prose: *when a distribution cannot be estimated, stop optimising over it, and shrink the cost of being wrong instead.*

Size and duration are different variables, and only one of them requires a history. How big the next loss will be is a question about a distribution you do not have. How long you would take to detect, isolate, restore and communi-

cate is a property of *your own system* — measurable this quarter, on your premises, by drill. You cannot rehearse the size of a flood. You can absolutely rehearse the evacuation.

The shift sounds philosophical and is entirely practical. A recovery-time objective is a number you can test on a Tuesday. Backup isolation either survives a red-team exercise or it does not. Failover to a second model provider either works under load or it does not. None of these numbers wear a costume: each one is a measurement of the present, not an extrapolation of an absent past. And every hour you shave off recovery truncates the loss of *every* tail scenario at once — the ones you modelled and the ones you cannot. On a fat tail it is the duration lever, not the prediction lever, that carries the expected value.

So: five questions for a board, in place of the model this risk will not grant you.

Five questions, quarterly

1. Which of our critical models are calibrated on events that have never happened? (The column-B inventory.)

2. What is our measured — not aspirational — time to detect, isolate and restore our three most critical processes? When was it last rehearsed against an adversary, not a checklist?

3. If our primary AI or cloud provider failed or was compromised for a week, what runs? What did the failover drill find?

4. What share of our risk budget buys sharper estimates of measurable risks, versus faster recovery from unmeasurable ones — and who decided that split?

5. Which decisions are we taking today that only make sense if the future resembles the last 250 days?

A closing word from the trading room. Nothing in this book argues against models; I have built them, run desks on them, and would again. The argument is about knowing which instrument you are holding. Value at Risk, backtested across a hundred years, is a fine speedometer and a false crystal ball — 380 gold dots say so. Extreme value theory is a genuine telescope, and it still needs a sky full of stars to point at. For the risk the FSB has now put at the top of the world's agenda, the sky is empty, by definition, until the first catastrophe fills it.

Measure what you can. Prepare for what you cannot. And never confuse the two budgets.

Value at Risk is an excellent measure of the losses you have already had. The institutions that thrive in what comes next will be the ones that stopped asking their models to be prophets — and started timing their own recovery with a stopwatch.

Sources & reproducibility

Primary data

Kenneth R. French Data Library, *Fama/French 3 Factors* [*Daily*], CRSP value-weighted US market return (Mkt–RF + RF), July 1926 – June 2026. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Financial Stability Board, *FSB Chair's letter to G20 Finance Ministers and Central Bank Governors*, 31 August 2026; and the accompanying

press release, *FSB Chair warns of risks arising from frontier AI models*. https://www.fsb.org

Methods

Historical Value at Risk: 99 % one-day, empirical quantile on a rolling 250-trading-day window (the Basel minimum observation period for internal market-risk models); the estimator on 250 days is the third order statistic of losses. Basel Committee on Banking Supervision, *Minimum capital requirements for market risk*.

Extreme value theory: peaks-over-threshold with a generalised Pareto fit (maximum likelihood) to the 798 one-day losses exceeding 2 %; confidence band by bootstrap (500 resamples). J. Pickands (1975); A. Balkema & L. de Haan (1974); A. McNeil, R. Frey & P. Embrechts,

Quantitative Risk Management, Princeton University Press.

Reproducibility

Every number and every chart in this book is produced by a single script, `analysis/apparatus.py`, from the raw daily file. No figure is drawn by hand; no quoted statistic bypasses the code. The market series is the CRSP value-weighted index of all NYSE, AMEX and NASDAQ stocks — on 19 October 1987 it lost 17.41 %, where the narrower S&P 500 lost 20.5 %; this book always quotes the broad market.

Nabil El Mahyaoui is an AI & Digital expert — applied mathematics (École Polytechnique, Ponts), capital markets (Société Générale CIB, CDG Capital), and enterprise AI leadership. He writes at nabil-el-mahyaoui.com.