

How Deep, How Long, and Why

S&P 500 drawdowns, recovery times, and measurement conventions

Vincent Wang

July 2026

THE QUESTION

98.5 years of daily data, 73 underwater
episodes:

**what can history tell us before the next
drawdown arrives?**

DISCLAIMER

Research use only; not investment advice.

All results are descriptive statistics of historical data; past drawdowns do not predict future outcomes.

| The Long-Horizon Investor's Four Drawdown Questions

01 | DEPTH



How deep

Heavy-tailed: most episodes are 5–10% pullbacks; a few systemic bears set the risk budget.

02 | DURATION



How long

Recovery-time confidence intervals span years, and milestones arrive long before full recovery.

03 | ATTRIBUTION



Why

Recession overlap is the watershed for both depth and duration.

The fourth question runs through everything: the same history changes its 4/21

| Data and Index Lineage

Four data layers

- ▶ Daily core: Yahoo ^GSPC closes, 1927-12-30 to 2026-07-02
- ▶ Monthly: Shiller composite (from 1871, intra-month averages)
- ▶ Variants: nominal/real \times price/total-return
- ▶ Macro and cross-asset: NBER recessions, nine ETFs

24,742
trading days

98.5 yrs
daily sample span

73
5%+ episodes (ATH-anchored)

! Pre-1957 daily values are the backfilled S&P 90; pre-1926 monthly data are the Cowles reconstruction.

All inputs pass automated validation (dates, gaps, extreme moves).

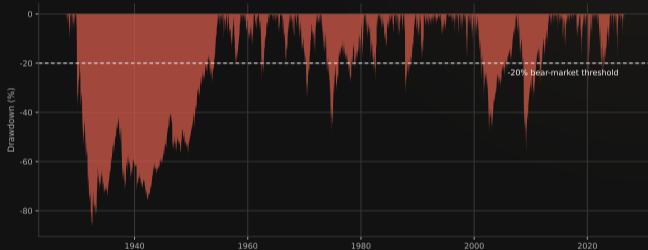
| Episode Definition: Peak to Full Recovery

Canonical episode

$$D_t = P_t / M_t - 1$$

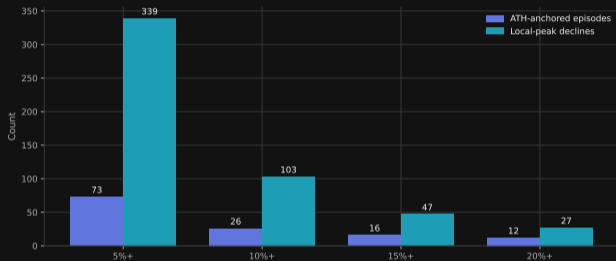
One episode = prior high → trough → recovery of that high; maximum depth assigns the 5/10/15/20% sets, one bear counted once.

Daily closes; calendar days; open episodes censored (currently 0).



The definition drives the counts: every number that follows rests on it.

| One Dataset, Two Answers



73 vs 339

5%+: ATH-anchored vs local-peak

3.4/yr

5% dips under the local-peak convention

Famous declines like 1937–38

(−54.5%) and 2011 (−19.4%) are absorbed into longer episodes.

Episode counts here are lower bounds — do not compare with “three 5% dips a year” statistics.

| Executive Summary: Four Numbers

73 / 26 / 12

episodes at 5%+ / 10%+ / 20%+
(ATH-anchored)

764 days

median 20%+ recovery; 95% CI [666,
2310] days

**-35.0% vs
-14.3%**

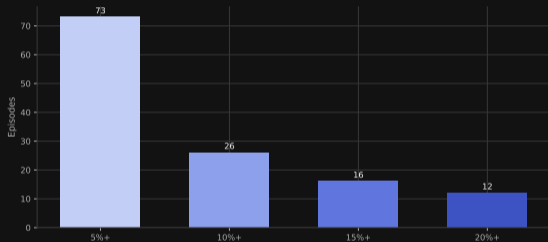
median depth: recession-overlapping vs
not

**5.4 → 2.7
yrs**

median 20%+ spell, price → total return

How deep, how long, why, and under which measurement — answered together.

| Threshold Frequency and Sampling Uncertainty



12

observations in the whole 20%+ bucket

[666, 2310] d

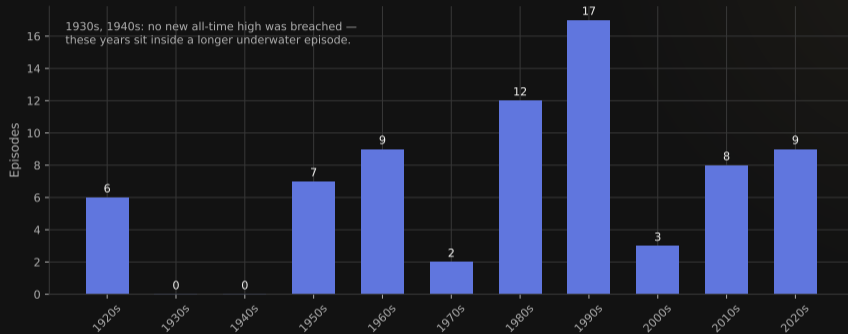
bootstrap 95% CI of the median recovery

10,000 resamples, fixed seed; the

interval spans nearly 3 \times .

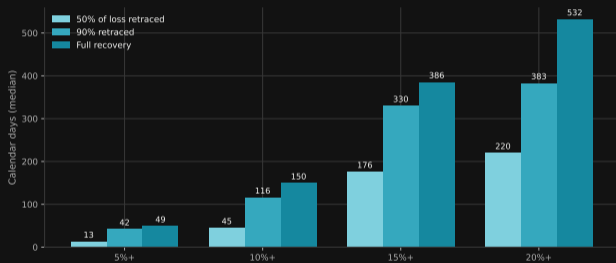
Plan on the interval, not the point estimate.

| Clustered Across Decades, Not Uniform



Zero-episode decades (1930s, 1940s) are a counting artifact: those years sit inside longer underwater spells, not calm markets.

Recovery Milestones: Most of the Loss Is Regained Early



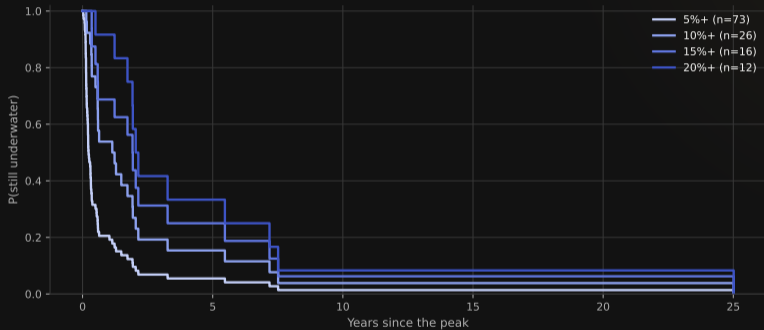
220 d
half the loss retraced (median)

383 d
90% retraced

532 d
full recovery

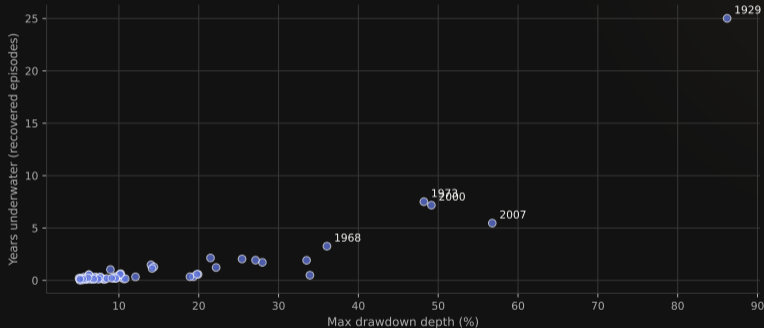
Schedule withdrawals, refinancing, and client communication around the 50%/90% milestones instead of waiting for the full-recovery date.

| The Survival View: How Much Longer Underwater?



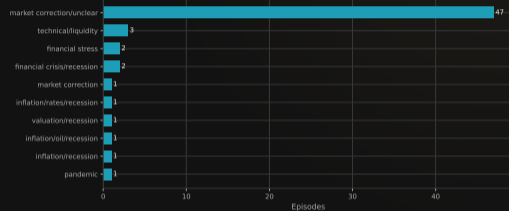
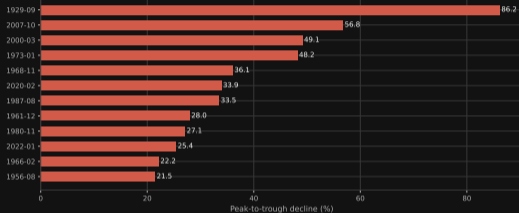
Kaplan–Meier estimates; censoring handled by construction. The 20%+ tail is driven by 1929, 1973, 2000, and 2007.

| Depth and Time Under Water: Related, Far From Identical



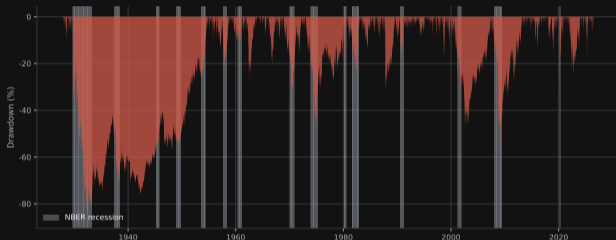
Similar depths differ in recovery by multiples (contrast 2020 with 1973) — report both.

| The Deepest Bears and Their Causes



Labels cover all 26 episodes of 10%+ (curated, confidence-rated, not causal). Crises and recessions dominate depth; technical/liquidity events (1987, Feb-2018) are deep but fast.

Macro Regimes: Recession Is the Watershed



RECESSION-OVERLAPPING (10)

-35.0% | 988 d

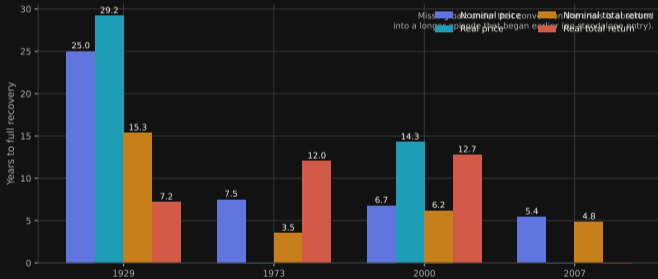
NO OVERLAP (16)

-14.3% | 216 d

Association, not causal identification.

Ask "recession or not" before "drawdown or not".

Measurement Rewrites the Recovery Narrative



25.0y →
7.2y

1929: nominal price → real total return

7.5y →
12.0y

1973: inflation makes real terms worse

In real terms 2000–2013 is one continuous spell that swallowed both crashes — match the reporting convention to the liability.

Cross-Asset Transmission in S&P Stress Windows

S&P 500 (price index)	-12	-49	-57	-14	-10	-20	-34	-25	-19
SPY (S&P 500)	-12	-48	-55	-13	-10	-19	-34	-24	-19
QQQ (Nasdaq-100)	-12	-83	-53	-16	-11	-23	-29	-35	-23
IWM (Russell 2000)	n/a	n/a	-58	-26	-9	-26	-41	-27	-23
XLK (Technology)	-12	-82	-53	-14	-10	-24	-31	-33	-26
XLF (Financials)	-21	-37	-82	-22	-12	-22	-43	-26	-16
XLE (Energy)	-15	-42	-57	-34	-14	-30	-56	-26	-19
XLU (Utilities)	-10	-52	-46	-11	-7	-9	-36	-25	-8
TLT (20y+ Treasuries)	n/a	n/a	-27	-6	-5	-4	-16	-40	-9
GLD (Gold)	n/a	n/a	-29	-13	-13	-5	-13	-21	-7
	1999-07	2000-03	2007-10	2015-05	2018-01	2018-09	2020-02	2022-01	2025-02

! The Treasury hedge is regime-dependent: TLT fell 40% alongside equities in the 2022 inflation bear. Dividend-adjusted closes; n/a = before inception. 17/21

| Tiered Risk Rules: Depth Tiers Are Different Problems

5% dips = noise

About 3.4 per year under local peaks.

No action.

10%+ = discipline triggers

Pre-agreed communication, rebalancing, leverage limits.

The recession call decides whether it deepens.

20%+ = capital preservation

Median recovery 764 days; CI spans years.

Plan on intervals and milestones, not points.

Write the three tiers into the risk playbook: triggers, actions, communication.

| As a Strategy-Evaluation Baseline

Not a timing signal

This study is the control group for any defensive overlay: trend filters, volatility control, and options programs should state what they improve — depth, time under water, or milestones — and by how much.

Next steps

- ▶ Formal conditional statistics on rate/inflation regimes
- ▶ Extend the ETF panel with recovery times

A defense without a baseline cannot be evaluated.

| Limitations

! The daily core is a price index; total-return/real robustness uses monthly data (intra-month averaging smooths extremes).

! Cause labels are narrative attributions; recession overlap is association, not causal identification.

! The ETF panel uses adjusted closes with differing inceptions; pre-1957 daily = S&P 90 backfill, pre-1926 monthly = Cowles reconstruction.

Full limitations and reproducibility notes: paper Section 12.

Long-run returns are earned by bearing and surviving drawdowns.

History offers expectation-setting, not prediction:
how deep · how long · why · under which measurement

Vincent Wang | vincentwang.research@outlook.com