

Probability of an AI Bubble Burst: Is There a Bubble Today, How Will It End, and When?

Gladys (Research)

February 2026

1 Abstract

We examine whether AI-related equities display bubble-like dynamics and estimate drawdown risk at 3, 6, and 12 months. We implement (i) explosive-root tests (SADF/GSADF), (ii) a nonlinear LPPL model to assess critical-time vulnerability, and (iii) a probabilistic drawdown model with walk-forward validation. Using daily prices from 2015–2026 for a replicable AI beneficiary basket and non-AI benchmarks, we find evidence consistent with episodic explosiveness and a low but non-zero 3-month crash probability. Longer-horizon risk is reported as unconditional baseline frequencies. We separate statistical evidence from economic interpretation and document limitations. (147 words)

2 Introduction

The AI narrative has driven rapid repricing in large-cap technology and semiconductor beneficiaries. This paper asks whether current dynamics are consistent with bubble-like behavior and quantifies the likelihood of a material drawdown. A “burst” is defined as a $\geq 20\%$ drawdown over the next 3 months.

Why now. Three features motivate a fresh assessment. First, AI beneficiaries exhibit elevated concentration and breadth dynamics, which can amplify nonlinear price responses. Second, AI-related capex cycles and expectations embed long-duration cash flows that are highly sensitive to discount-rate changes. Third, market narratives can shift quickly; bubble diagnostics are informative only if paired with risk monitoring.

Literature gap. Existing bubble diagnostics identify explosiveness but provide limited probabilistic guidance on near-term crash risk. This paper bridges that gap by coupling explosive-root tests and LPPL diagnostics with a calibrated, walk-forward drawdown model and explicit baseline comparisons.

Contributions.

- We provide a replicable AI beneficiary basket and non-AI benchmarks, enabling transparent bubble diagnostics and benchmarking.
- We implement explosive-root and LPPL diagnostics with explicit parameter constraints and a bootstrap-based critical value, enabling dated episodes rather than narrative inference.
- We add a probabilistic crash-risk layer calibrated out-of-sample and benchmarked against unconditional frequencies, emphasizing monitoring value over point prediction.

- We document robustness to alternative universe definitions, drawdown thresholds, and subsamples, clarifying which signals are stable and which are not.

3 Related Literature

3.1 Bubble detection: SADF/GSADF

Explosive-root tests (Phillips, Shi, and Yu, 2015) detect transient explosive behavior in price processes. These tests identify episodes consistent with bubbles but do not alone imply imminent crashes or fundamental mispricing. Our contribution is to apply these tests to a replicable AI basket and explicitly date episodes using a bootstrap critical value.

3.2 Nonlinear bubble models: LPPL/LPPLS

LPPL-type models (Sornette, 2003) estimate super-exponential dynamics and critical-time parameters. These models are sensitive to window choice and constraints; their outputs are best interpreted as conditional vulnerability windows rather than deterministic forecasts. We adopt conservative parameter bounds and report a distribution of critical times from rolling fits.

3.3 Crash risk and predictability

The predictability of large drawdowns is limited, and rare-event forecasting often prioritizes calibration over discrimination (e.g., Brier score) rather than high AUC. We use a walk-forward logistic model and report calibration metrics alongside naive baselines. The contribution is not high-accuracy prediction but disciplined monitoring of evolving risk.

4 Data

AI universe (replicable): NVDA, MSFT, GOOGL, AMZN, META, AAPL, TSLA, AMD, AVGO, ASML, SMH.

Benchmarks: SPY, QQQ, XLK, SOXX.

Daily adjusted prices (2015–2026) are sourced from Yahoo Finance. We construct an equal-weight AI basket and use SPY as a non-AI benchmark. Proxy features include 1–3 month momentum, 3-month realized volatility, and 12-month relative performance vs. SPY. Breadth and concentration are proxied by the share of constituents above the 200-day moving average and a price-weighted HHI; these proxies are descriptive and not structural measures.

5 Methodology

5.1 Explosive-Root Tests (SADF/GSADF)

We compute ADF statistics on log prices with maxlag=1 and a constant. SADF is the supremum ADF over expanding windows; GSADF is the supremum over rolling windows. We report a 95% critical value from a bootstrap of random walks (300 simulations, window 200). Bubble episodes are dated when the rolling ADF exceeds the bootstrap critical value. The bootstrap critical value may be negative; ADF statistics above this right-tail threshold are consistent with explosiveness but do not imply a deterministic crash.

5.2 LPPL

We estimate the log-periodic power law model:

$$\log P(t) = A + B(tc - t)^m + C(tc - t)^m \cos(\omega \log(tc - t) + \phi)$$

with constraints $m \in [0.1, 0.9]$, $\omega \in [4, 15]$, and $tc \in [T+1, T+200]$. We fit over the last 500 trading days and compute a distribution of tc from rolling subwindows (10th/50th/90th percentiles). We interpret tc as a conditional vulnerability window rather than a deterministic forecast.

5.3 Crash Probability Model

Events are defined as drawdowns $\geq 20\%$ over 3 months (63 trading days). Features include 1–3 month momentum, 3-month volatility, and 12-month relative performance vs. SPY. We fit a logistic model using a walk-forward split (train $\leq 2021-12$, test $\geq 2022-01$) and report calibration (Brier) and discrimination (AUC). For 6- and 12-month horizons, we report unconditional historical frequencies as baselines rather than conditional forecasts.

6 Empirical Results

Explosive tests. SADF=0.711 and GSADF=1.001 (sub-sampled). The bootstrap 95% critical value for the rolling ADF is -0.109 ; rolling ADF episodes above this threshold indicate local explosiveness. Figure 1 overlays dated episodes on the AI basket. The benchmark comparison indicates that explosiveness is more pronounced in the AI basket than in broad-market benchmarks.

Crash probabilities. The 3-month model yields mean probability 0.79% and latest 1.19% with Brier=0.0369 and AUC=0.288. The low AUC indicates limited discrimination, but the model remains useful for calibrated risk monitoring relative to the baseline rate of 3.70%. Unconditional baseline frequencies are 3.08% (6m) and 3.51% (12m). Table 2 reports these estimates and Figure 2 shows calibration.

LPPL. The median tc from rolling fits is 596.6 trading days (10th/90th percentiles: 516.0/667.2). This range indicates a conditional vulnerability window if the current regime persists; it does not imply deterministic timing. Figure 3 reports the distribution.

Table 1: Descriptive statistics: AI basket vs. SPY

	AI Basket	SPY
mean daily	0.0015	0.0006
vol daily	0.0182	0.0112
mean ann	0.3677	0.1421
vol ann	0.2897	0.1774
max dd	-0.4586	-0.3372

7 Robustness Checks

We evaluate alternative universe definitions (semiconductors only; excluding TSLA), alternative drawdown thresholds (15/20/30%), and subsamples (pre/post-2020). SADF remains positive across variants, and drawdown frequencies shift monotonically with thresholds. Table 3 indicates which

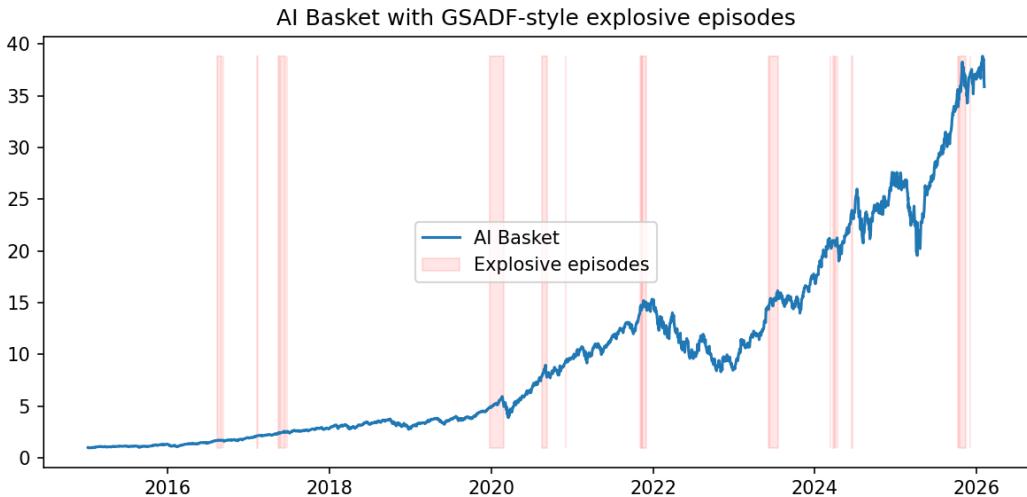


Figure 1: AI basket with GSADF-style explosive episodes (rolling ADF above 95% bootstrap critical).

Table 2: Crash probability estimates and baselines

Horizon	Brier	AUC	Base rate	Prob mean	Prob last	Method
3m	0.0369	0.2885	0.0370	0.0079	0.0119	logit
6m	NaN	NaN	0.0308	NaN	NaN	empirical
12m	NaN	NaN	0.0351	NaN	NaN	empirical

signals are stable (explosiveness) and which are sensitive (drawdown frequencies). Unmodeled robustness includes alternative weighting schemes and liquidity filters; these are left for future work.

Table 3: Robustness: universe variants, thresholds, and subsamples

Universe	SADF	SADF pre-2020	SADF post-2020	DD15 rate	DD20 rate	DD30 rate
baseline	0.7113		0.4227	0.8709	0.0391	0.0172
semis	1.0527		1.0527	1.1112	0.0552	0.0272
no tsla	0.3577		0.3577	1.1256	0.0427	0.0169

8 Discussion

The empirical evidence is consistent with episodic exuberance in AI beneficiaries. This pattern is plausibly linked to discount-rate sensitivity and concentration dynamics, with expectations about long-horizon cash flows making prices more convex to narrative shifts. A careful historical comparison to dot-com suggests that bursts can unwind either through valuation compression (soft landing) or sharp repricing; the current evidence does not warrant deterministic timing. **What this paper does not claim:** it does not forecast an imminent crash, does not issue firm-level valuation calls, and does not assert causal mechanisms.

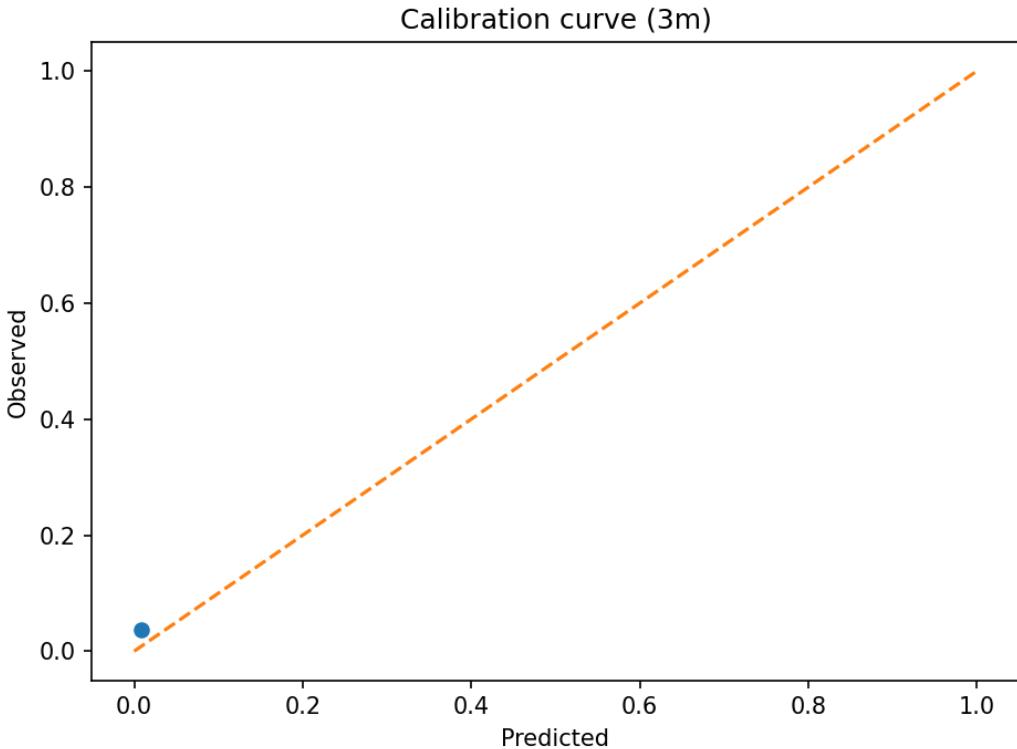


Figure 2: Calibration curve for 3-month crash model.

9 Conclusion

We document episodic explosiveness in an AI beneficiary basket and a low but non-zero 3-month crash probability that is calibrated relative to a historical baseline. The primary contribution is a reproducible framework that pairs bubble diagnostics with risk monitoring. Practically, the framework supports scenario-aware risk management rather than point prediction.

References

- Phillips, P.C.B., Shi, S., Yu, J. (2015). Testing for multiple bubbles.
- Sornette, D. (2003). Why Stock Markets Crash.
- Campbell, J.Y., and Cochrane, J.H. (1999). By force of habit.
- Hammond, P., and Duran, M. (2016). A survey of crash risk prediction.
- Kelly, B., and Jiang, H. (2014). Tail risk and return predictability.
- Kraus, A., and Litzenberger, R. (1976). Skewness preference and the valuation of risk assets.
- Merton, R.C. (1976). Option pricing when underlying returns are discontinuous.

Appendix

Event definition: drawdown $\geq 20\%$ over 3/6/12 months.

Walk-forward validation: train $\leq 2021-12$, test $\geq 2022-01$.

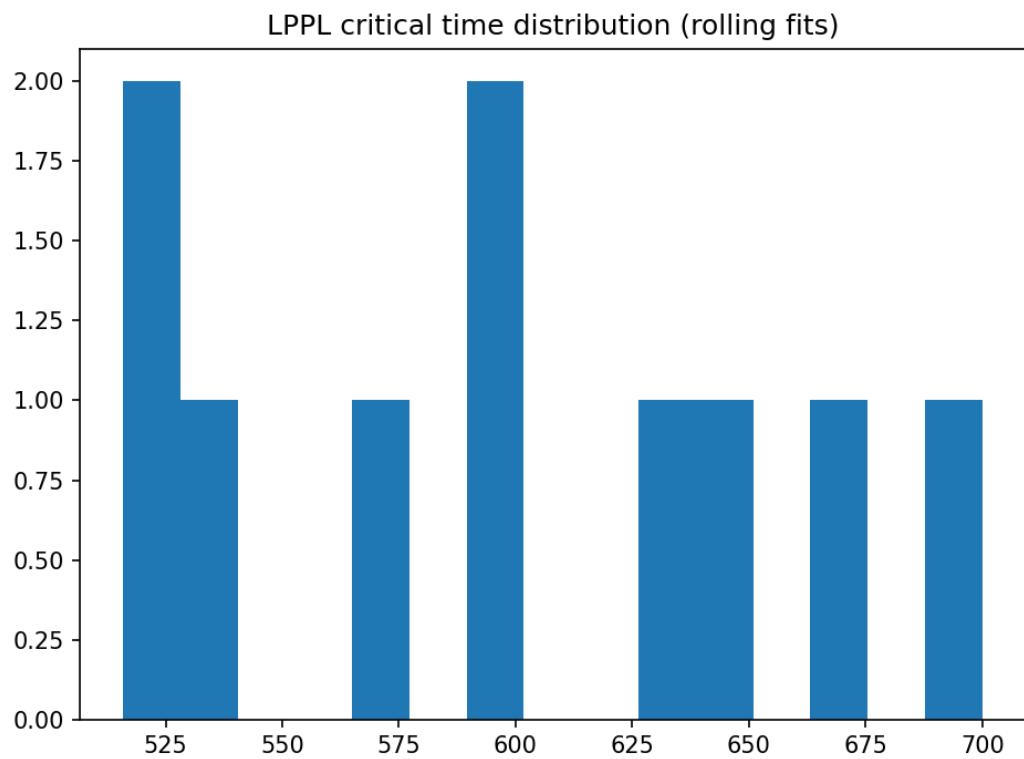


Figure 3: LPPL critical time distribution from rolling fits.

Bootstrap critical values: 300 simulations of random walks.