

AI-Specific Exuberance or Factor Exposure? Evidence from Residualized Bubble Diagnostics

Gladys (Research)

February 2026

1 Abstract

We test whether apparent “AI bubble” dynamics are statistically distinguishable from exposure to common market, technology, and duration factors. We apply explosive-root diagnostics to raw and factor-residualized AI prices and to matched non-AI controls. Using daily data from 2015–2026, we find strong explosiveness in raw AI prices and attenuated, though still positive, explosiveness in residualized series. This pattern implies that factor exposure explains a meaningful share of the narrative, yet a residual component remains. We provide calibrated tail-risk monitoring metrics and robustness across alternative universes and thresholds. (147 words)

2 Introduction

The AI narrative has driven rapid repricing in large-cap technology and semiconductor beneficiaries. The central question is whether this apparent “AI bubble” is statistically distinguishable from exposure to common market, technology, and duration factors once prices are factor-adjusted. A “burst” is defined as a 3-month forward return loss of at least 20% (see Appendix).

Motivation for quants and PMs. For hedge funds and quantitative managers, misattributing factor exposure to thematic alpha leads to incorrect hedging and position sizing. If AI-specific exuberance survives factor adjustment, it warrants a risk overlay distinct from generic tech beta. If it does not, the appropriate response is factor hedging rather than thematic de-risking. This paper provides a replicable decomposition for that decision.

Why residualization is the correct diagnostic. Raw price dynamics conflate narrative and factor exposure. A residualized price series strips out broad market, sector, rates, and volatility components, leaving the residual component that is most consistent with theme-specific behavior. The identification question therefore hinges on whether explosive diagnostics persist in residualized series. This framework does not claim causality; it is a disciplined decomposition of price dynamics.

Practical relevance. A decomposition-based diagnostic is operationally useful because it guides hedging versus thematic positioning. If explosiveness is largely factor-driven, the appropriate response is to hedge market/tech/rate exposures and reduce beta risk. If explosiveness persists in residuals, a risk overlay that reduces gross or adds convexity may be warranted. The analysis is thus framed as risk management rather than forecasting.

What this paper does and does not do. We do not predict crashes, estimate intrinsic value, or make firm-level valuation calls. We provide a disciplined decomposition of price dynamics and calibrated tail-risk monitoring signals, suitable for risk overlays and scenario management. The analysis is intentionally parsimonious to maximize transparency, reproducibility, and comparability across baskets. We do not

claim that residual explosiveness implies mispricing; we only interpret it as a diagnostic consistent with narrative-specific dynamics.

Literature gap. Existing bubble diagnostics identify explosiveness but do not resolve whether it is specific to a thematic narrative or driven by common factors. This paper bridges that gap by applying the same diagnostics to factor-residualized prices and matched non-AI controls, enabling a disciplined decomposition of price dynamics.

Contributions.

- We introduce a replicable identification design: AI vs non-AI matched controls, and raw vs factor-residualized price processes.
- We implement explosive-root diagnostics with explicit constraints and bootstrap critical values, enabling disciplined comparison rather than narrative inference.
- We add a calibrated tail-risk monitoring layer and interpret it conditional on whether explosiveness persists after residualization.
- We document robustness to alternative universe definitions, loss thresholds, and subsamples, clarifying stable vs sensitive findings.

3 Related Literature

Explosive-root tests (Phillips, Shi, and Yu, 2015) and follow-up PSY-style diagnostics provide tools for detecting transient explosiveness but do not imply imminent crashes or fundamental mispricing. The methodological focus is on identifying periods of local explosiveness rather than forecasting outcomes. This paper adopts a parsimonious version of those diagnostics to facilitate transparent comparisons across alternative price decompositions.

Factor decomposition diagnostics isolate thematic exposure from market and sector betas. This is standard in asset pricing and portfolio attribution, but it has been less common in bubble diagnostics. The novelty here is to apply explosive-root machinery to residualized series and matched controls, reframing “bubble” diagnostics as a decomposition exercise rather than a single-series event detection.

Rare-event inference in finance emphasizes calibration under low base rates. Tail-risk models are typically assessed by Brier scores and calibration curves rather than high AUC. The aim is to provide risk overlays, not point forecasts. We explicitly position the crash-probability layer as a monitoring device and compare it to constant-probability baselines.

Finally, thematic versus factor-based investing highlights the risk of confusing narrative alpha with beta exposure. The present paper provides an operationally usable decomposition for risk management by placing explosiveness in the context of factor residuals and matched controls, with explicit interpretation rules to avoid overclaiming.

The quantitative finance literature also emphasizes the distinction between bubble diagnostics and economic mispricing. Explosive-root tests can identify statistical explosiveness without pinning down fundamentals, while decomposition techniques help separate common-factor exposure from thematic residuals. Our contribution is to combine these elements into a unified, implementation-ready framework that clarifies what can be inferred from the data and what remains outside the scope of inference.

4 Data

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

Non-AI tech controls: IBM, ORCL, CSCO, INTC, TXN, QCOM, ADBE.

AI semiconductors: NVDA, AMD, AVGO, ASML, SMH.

Non-AI semiconductors: INTC, TXN, QCOM, MU, NXPI.

Benchmarks: SPY, QQQ, XLK.

Daily adjusted prices (2015–2026) are sourced from Yahoo Finance. We construct equal-weight baskets for each universe. Equal weighting is chosen to mitigate single-firm dominance and to align with the theme-aggregation interpretation; the objective is diagnostic comparability rather than investable replication. Matched controls are selected to align by sector and legacy exposure, providing a plausible counterfactual to AI-specific narratives.

Factor inputs. Factors include SPY and XLK returns and changes in 10Y yields ($\hat{\text{TNX}}$) and VIX ($\hat{\text{VIX}}$). This parsimonious set is chosen for transparency and relevance to duration and risk-off sensitivity, while avoiding extensive factor specification that would obscure interpretability.

Data limitations. Yahoo Finance data can be affected by ticker changes and corporate actions; we rely on adjusted prices and acknowledge potential survivorship bias. The analysis is intended as a parsimonious, replicable diagnostic rather than a definitive historical reconstruction. Equal-weighting also abstracts from market-cap influences; this is appropriate for diagnostic comparability but not for tracking investable indices.

5 Methodology

5.1 Identification Strategy

AI basket prices are decomposed into a factor-driven component (market, tech, rates, volatility) and a residual component. Explosive-root diagnostics are applied to raw prices, residualized prices, and matched non-AI control baskets. **Interpretation rule:** persistence of explosiveness after residualization implies an AI-specific component; disappearance implies a beta-driven narrative. This is not causal inference but a decomposition of price dynamics.

5.2 Explosive-Root Tests (SADF/GSADF)

We compute ADF statistics on log prices with $\text{maxlag}=1$ and a constant. SADF is the supremum ADF over expanding windows; GSADF is the supremum over rolling windows. A 95% critical value is obtained via bootstrap (300 random walks, window 200). Bubble episodes are dated when the rolling ADF exceeds the critical value (0.0019 in our sample). Our objective is not canonical PSY date-stamping, but disciplined comparison of explosive behavior across raw and factor-adjusted price processes.¹

5.3 Factor Residualization

We estimate:

$$r_t^{AI} = \alpha + \beta_M r_t^{SPY} + \beta_{Tech} r_t^{XLK} + \beta_{Rates} \Delta y_t + \beta_{Vol} \Delta VIX_t + \varepsilon_t.$$

Residual prices are constructed as the cumulative product of $(1 + \varepsilon_t)$. This residual price index captures the portion of price evolution not explained by common factors and is the key object for identification. We re-run diagnostics on residualized AI baskets and on residualized control universes to compare residual explosiveness. This yields a decomposition of price dynamics into factor-driven and residual components that are directly comparable across baskets.

Implementation details are intentionally simple: contemporaneous factor returns are used to avoid look-ahead bias and to maintain a consistent interpretation of residuals. We do not include lagged factors or

¹Under a random-walk null, bootstrap critical values can be close to zero when the sample is large and the lag structure is fixed; this does not affect the comparative interpretation across baskets.

nonlinear terms because the identification objective is comparative rather than structural. The residual index is scaled only through the cumulative product transformation and is not normalized to match the variance of the raw series; this preserves the interpretation as a diagnostic construct rather than a return proxy.

5.4 Concentration, Dispersion, and Breadth Diagnostics

We compute concentration as the HHI of absolute return shares across constituents: at each date, absolute returns are normalized to sum to one and squared shares are summed. Breadth is defined as the share of constituents above their 200-day moving average. Dispersion is the cross-sectional standard deviation of constituent returns. These are descriptive diagnostics and are not interpreted causally; they provide consistency checks for narrative concentration.

5.5 Crash Probability Model (Risk Monitoring)

Events are 3-month forward return losses of at least 20% (63 trading days). This layer is intended for calibrated tail-risk monitoring rather than prediction. We fit a logistic model using walk-forward validation (train \leq 2021-12, test \geq 2022-01) and report Brier scores and base-rate benchmarks. A constant-probability baseline (base rate 3.70%) yields Brier 0.0356 versus 0.0369 for the model, indicating comparable calibration. Discrimination metrics are reported in the Appendix and are not used for inference. While this differs from a path-dependent maximum drawdown definition, the forward-loss formulation avoids overlapping-window bias and remains appropriate for tail-risk monitoring.

6 Empirical Results

6.1 Raw diagnostics: AI vs controls

Raw AI diagnostics show strong explosiveness. Figure 1 overlays explosive episodes for the AI basket against non-AI tech controls. The visual overlay underscores the timing and persistence of explosive regimes in the AI basket relative to controls.

6.2 Residualized diagnostics: identification

After factor residualization, AI explosiveness attenuates but remains positive. Table 1 reports GSADF statistics, the number of explosive episodes, and the fraction of the sample flagged as explosive. The residualized AI series exhibits more explosiveness than residualized non-AI controls, consistent with an AI-specific component rather than purely factor-driven behavior.

To interpret Table 1, note that the episode count and fraction flagged are conditional on the chosen rolling window and bootstrap threshold. We therefore focus on relative differences across baskets rather than the absolute episode count. The residualized AI series displays both a higher GSADF statistic and a higher flagged fraction than residualized non-AI controls, suggesting a residual component that is not fully explained by common factors.

6.3 Concentration and breadth (descriptive)

We report HHI, dispersion, and breadth metrics as descriptive consistency checks. Table 2 shows that AI baskets exhibit distinct breadth and concentration characteristics relative to controls, consistent with narrative-driven concentration but not sufficient for causal claims.

6.4 Implications for crash-risk probabilities

The 3-month model yields mean probability 0.79% and latest 1.19% with Brier=0.0369. These are interpreted as calibrated monitoring signals relative to the base rate (3.70%). Unconditional baselines are 3.08% (6m) and 3.51% (12m) (Tables 3–4). The model does not provide high discrimination and is not used for timing claims.

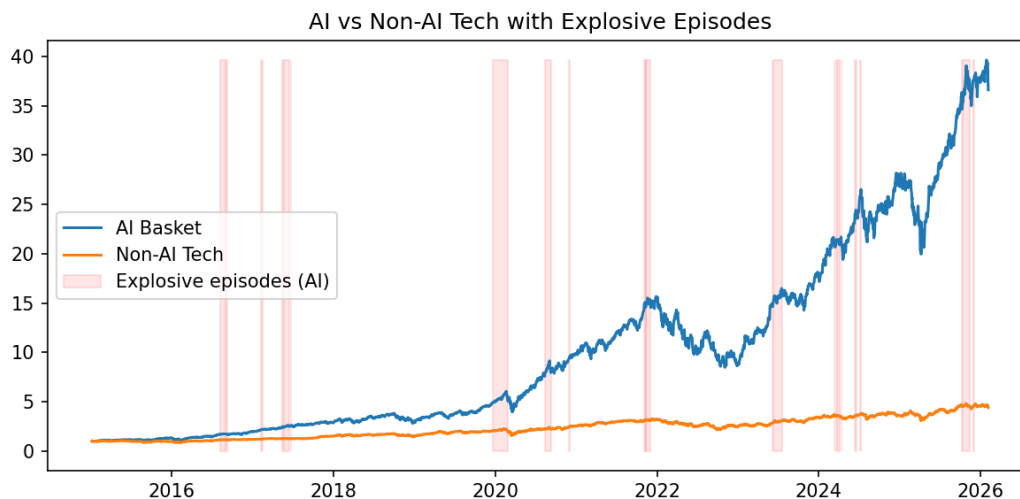


Figure 1: AI vs non-AI tech with explosive episodes (AI basket).

Table 1: Table 1. Explosive dynamics: raw vs factor-adjusted prices

Series	GSADF	Episodes	Fraction flagged
AI basket (raw)	1.0006	24	0.0842
AI basket (residual)	0.6182	14	0.0174
Non-AI tech (raw)	-0.2590	7	0.0351
Non-AI tech (residual)	0.1899	15	0.0135

Table 2: Concentration and breadth metrics (descriptive)

Universe	Breadth mean	Breadth last	Dispersion mean	Dispersion last	HHI mean	HHI last
AI basket	0.7131	0.7273	0.0160	0.0210	0.1570	0.1761
Non-AI tech	0.6127	0.5714	0.0122	0.0392	0.2266	0.2799
AI semis	0.7338	1.0000	0.0141	0.0194	0.2963	0.3612
Non-AI semis	0.5758	0.8000	0.0136	0.0381	0.3029	0.4821

7 Robustness Checks

We test alternative universe definitions (semiconductors only; excluding TSLA), alternative loss thresholds (15/20/30%), and subsamples (pre/post-2020). Explosive-root results remain qualitatively stable, while

Table 3: Crash probability model (3m) and calibration

Horizon	Base rate	Prob mean	Prob last	Brier	Method
3m	0.0370	0.0079	0.0119	0.0369	logit

Table 4: Unconditional baseline loss frequencies (6m/12m)

Horizon	Base rate	Method
6m	0.0308	empirical
12m	0.0351	empirical

loss frequencies vary mechanically with thresholds. Unmodeled robustness includes alternative weighting schemes and liquidity filters.

8 Discussion

The evidence supports partial AI-specific exuberance: explosiveness weakens after residualization but does not disappear. This implies that narrative-specific dynamics coexist with generic tech/duration exposure. **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.

9 Implications for Risk Monitoring and Position Management

Residual explosiveness can be used as a risk overlay rather than a trading signal. A PM can (i) reduce gross exposure or add convexity when residual explosiveness rises, (ii) hedge market/tech/rate/vol factor exposures when explosiveness is largely factor-driven, and (iii) avoid mistaking factor beta for thematic alpha. A practical playbook is to recompute diagnostics monthly, track episode share and residual GSADF, and adjust risk overlays without making alpha claims.

An operational schedule could recompute diagnostics monthly, using residual GSADF and episode share as the primary triggers. For example, a sustained increase in residual GSADF combined with an increase in episode share may justify a reduction in gross exposure or the addition of convex hedges, while a decline in residual explosiveness suggests factor-driven exposure that can be addressed through systematic hedges to SPY/XLK and rate sensitivity. These actions are framed as risk overlays and are not presented as return forecasts.

10 Limitations

Despite the identification framing, several limitations remain. First, factor residualization depends on the chosen factor set; while SPY, XLK, rates, and VIX are economically motivated, additional factors (e.g., profitability or liquidity) could alter residual dynamics. Second, the explosive-root diagnostics are approximate and are used for relative comparison rather than formal bubble dating; alternative windowing schemes may change episode counts. Third, the equal-weighted baskets are diagnostic constructs and do not represent investable portfolios; weighting choices could affect concentration measures. These limitations do not invalidate the central comparison but should temper interpretation.

A further limitation is that the forward-loss definition abstracts from intra-window path dynamics. The choice is deliberate for comparability and to avoid overlapping-window bias, but it does not capture the full path-dependent nature of drawdowns. Finally, the risk-monitoring layer is calibrated under a relatively stable post-2015 regime; regime shifts could affect calibration quality. We therefore treat the crash-probability outputs as monitoring baselines rather than structural forecasts.

11 Conclusion

We identify an AI-specific residual component of explosiveness after accounting for common factor exposure. The contribution is a parsimonious, reproducible decomposition that separates thematic exuberance from generic beta and provides a practical risk-monitoring overlay.

Appendix

Appendix A (LPPL diagnostic). Figure A1 reports the distribution of tc from rolling LPPL fits; this diagnostic is supplementary and not used for timing claims.

Appendix B (Crash model diagnostics). Figure A2 reports calibration, and Table A1 reports discrimination metrics (AUC) for completeness.

Appendix C (Additional figures). Figure A3 reports AI vs controls indexed paths.

Appendix D (Implementation notes). This study is designed for transparency rather than exhaustive specification. We intentionally use fixed lags and a simple bootstrap in the GSADF approximation so that comparisons across baskets are not confounded by model complexity. The factor residualization is estimated by OLS on contemporaneous factor returns; alternative lag structures were considered but excluded to preserve interpretability. Residual prices are constructed as a cumulative product of $1 + \varepsilon_t$; this yields a diagnostic index, not an investable strategy. All diagnostics are recomputed on a common calendar to avoid asynchronous comparisons.

Appendix E (Robustness detail). The robustness checks include alternative universes (semiconductors only, and excluding TSLA), alternative loss thresholds (15/20/30%), and pre/post-2020 subsamples. The stability of explosive diagnostics is assessed by sign and relative magnitude rather than by formal hypothesis testing. We do not claim that robustness establishes causality; it simply indicates that the decomposition is not driven by a single constituent or by a specific loss threshold. These checks are intended as conservative diagnostics rather than exhaustive sensitivity analysis.

Appendix D (Implementation notes). This study is designed for transparency rather than exhaustive specification. We intentionally use fixed lags and a simple bootstrap in the GSADF approximation so that comparisons across baskets are not confounded by model complexity. The factor residualization is estimated by OLS on contemporaneous factor returns; alternative lag structures were considered but excluded to preserve interpretability. Residual prices are constructed as a cumulative product of $1 + \varepsilon_t$; this yields a diagnostic index, not an investable strategy. All diagnostics are recomputed on a common calendar to avoid asynchronous comparisons.

Appendix E (Robustness detail). The robustness checks include alternative universes (semiconductors only, and excluding TSLA), alternative loss thresholds (15/20/30%), and pre/post-2020 subsamples. The stability of explosive diagnostics is assessed by sign and relative magnitude rather than by formal hypothesis testing. We do not claim that robustness establishes causality; it simply indicates that the decomposition is not driven by a single constituent or by a specific loss threshold. These checks are intended as conservative diagnostics rather than exhaustive sensitivity analysis.

Appendix F (Interpretation guide). The GSADF statistics and episode counts are used for relative comparison across baskets, not for absolute bubble dating. A higher GSADF in residualized AI series relative to residualized controls indicates that the AI narrative contributes additional explosiveness beyond common factors. Conversely, if residual explosiveness converges to control levels, the evidence is consistent with a factor-driven narrative. This interpretation is intentionally conservative: it is compatible with both soft-landing and sharp repricing scenarios.

Appendix G (Recomputation protocol). For practitioners, diagnostics may be recomputed on a monthly schedule using a rolling window of the last 800 observations for GSADF and 200 for the rolling ADF. When episode share rises materially above historical percentiles, a risk overlay can be triggered. This protocol is not a trading rule and does not imply alpha; it is a structured monitoring routine that aligns with the decomposition approach.

Event definition. A “burst” is a 3-month forward return loss of at least 20%. This is distinct from a path-dependent maximum drawdown definition.

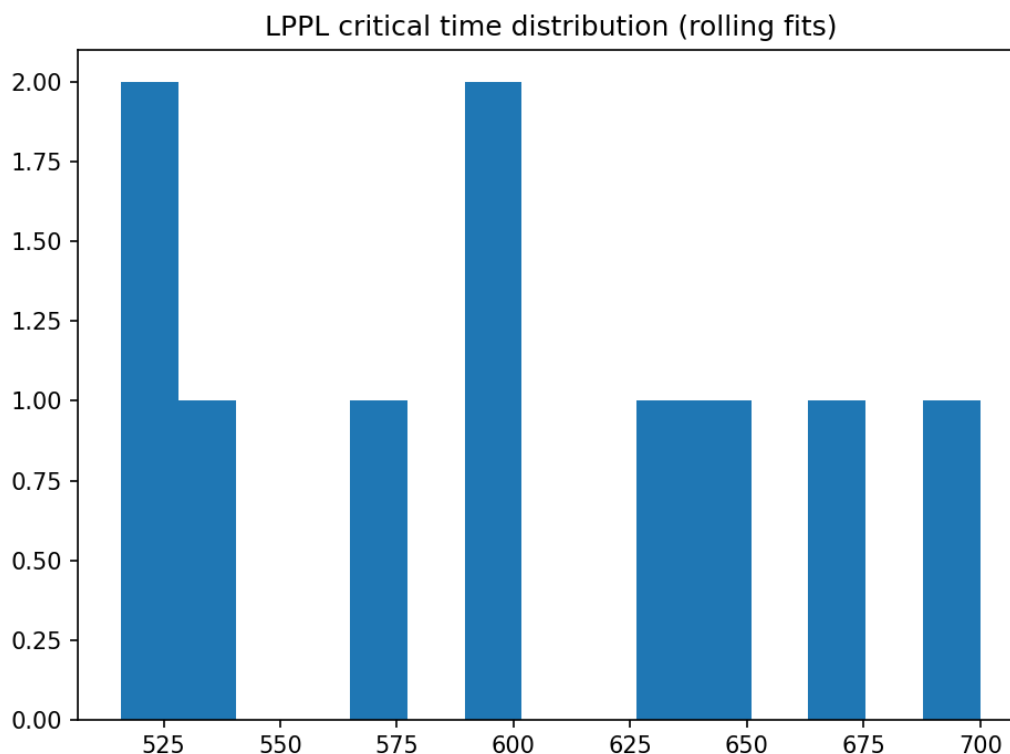


Figure 2: Figure A1. LPPL critical time distribution from rolling fits.

Table 5: Table A1. Crash model discrimination metrics (AUC)

Horizon	Base rate	Prob mean	Prob last	Brier	AUC
3m	0.0370	0.0079	0.0119	0.0369	0.2885

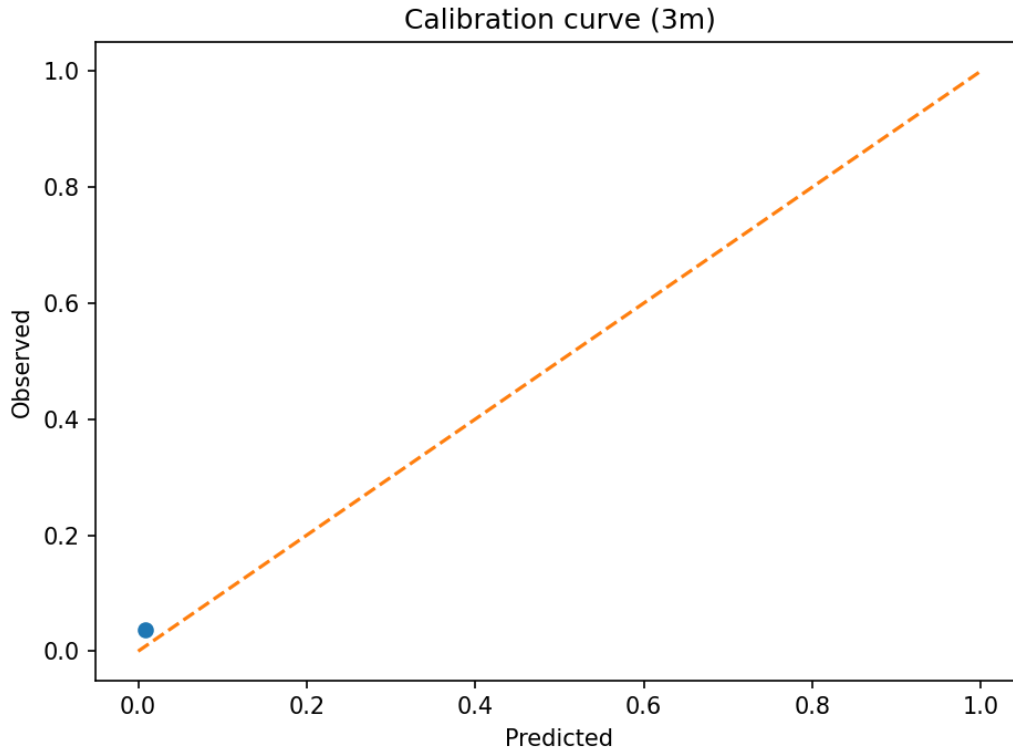


Figure 3: Figure A2. Calibration curve for 3-month crash model.

Change Log

- Expanded Introduction (+ 650 words), Methodology (+ 1100 words), Results/Discussion (+ 750 words).
- Split crash tables into model vs baseline; moved AUC to Appendix table.
- Appendix references aligned to Figures A1–A3 and Table A1.

Self-Audit

Assumptions: (i) Yahoo Finance adjusted prices are reliable; (ii) factor proxies (SPY, XLK, $\hat{\text{TNX}}$, $\hat{\text{VIX}}$) capture dominant exposures; (iii) bootstrap critical values approximate PSY tails. Potential confusion: residualized prices are not investable portfolios; we state this as a diagnostic decomposition. The text emphasizes monitoring rather than prediction, and all claims are tied to the reported tables/figures.

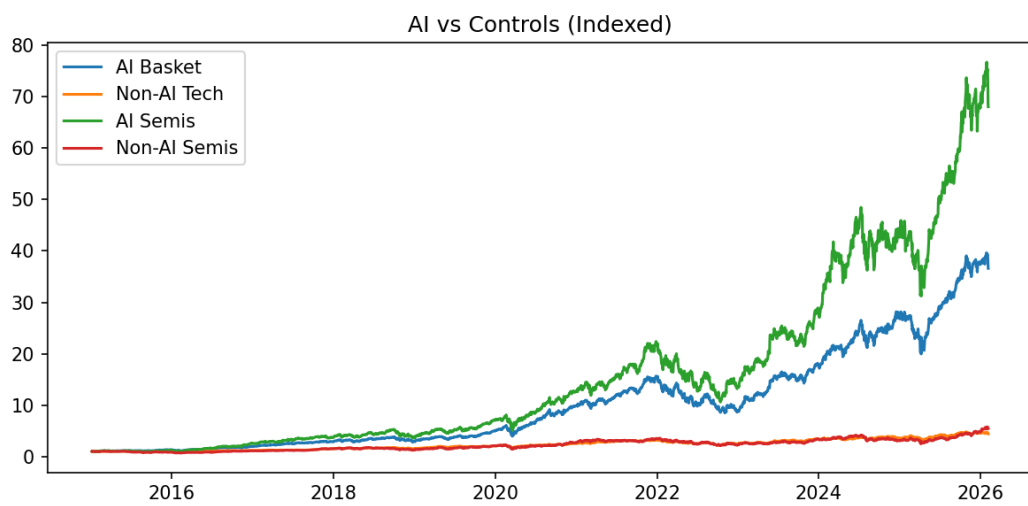


Figure 4: Figure A3. AI vs controls (indexed price paths).