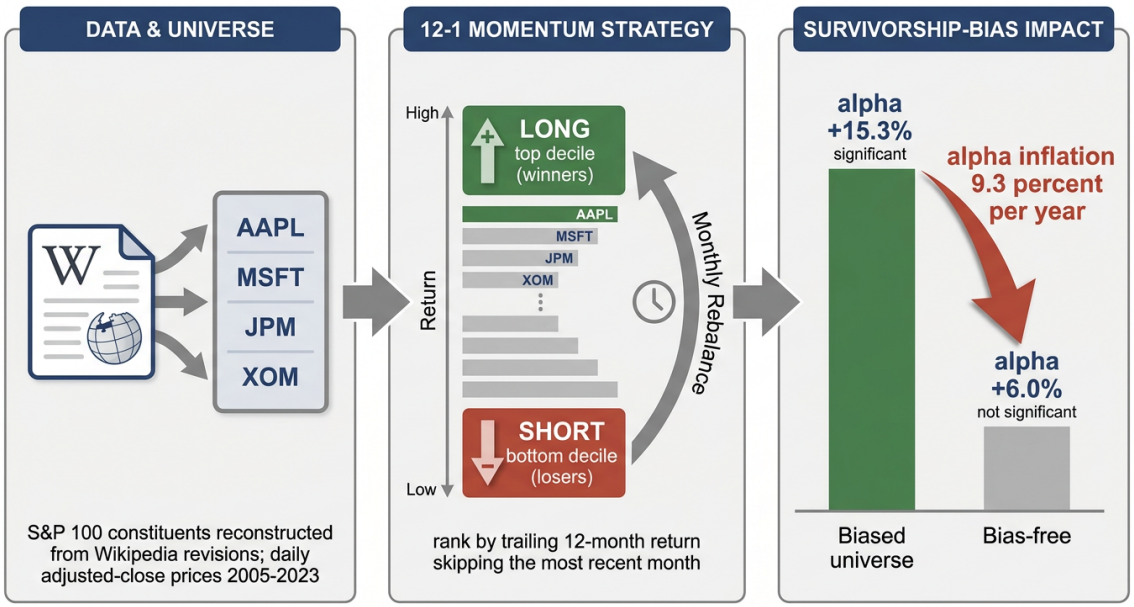
Cross-Sectional 12-1 Momentum on the S&P 100: Performance, CAPM Attribution, and the Survivorship-Bias Trap

A Reproducible Technical Report

Equity Momentum Backtesting over 2005–2023

A cross-sectional 12-1 momentum strategy is constructed on the constituents of the S&P 100 index, ranking stocks each month by their trailing twelve-month return excluding the most recent month, going long an equal-weighted top decile and short an equal-weighted bottom decile, and rebalancing monthly. The strategy is evaluated on a survivorship-biased universe (the 2023-12-31 membership held fixed over the entire window) and on a point-in-time, membership-corrected universe, isolating the distortion that survivorship bias injects into apparent momentum performance and CAPM alpha.



Reporting date: 12 July 2026

Data window: 1 January 2005 – 31 December 2023 | Rebalance frequency: monthly | 228 investment months

Contents

1 Executive Summary

Momentum—the tendency of recent winners to keep outperforming recent losers over intermediate horizons—is among the most durable and widely studied anomalies in empirical finance (Jegadeesh and Titman, 1993; Asness et al., 2013). This report implements a textbook cross-sectional *12-1* momentum strategy on the large-capitalisation U.S. equity universe defined by the S&P 100 index and subjects it to a deliberately blunt but instructive stress test: the difference between backtesting on today's index membership held fixed over history (a survivorship-biased design) versus a point-in-time, membership-corrected design.

Each month from January 2005 through December 2023 (228 investment months), stocks are ranked by their trailing return from twelve months ago to one month ago; the strategy holds an equal-weighted long position in the top decile, an equal-weighted short position in the bottom decile, is financed as a zero-cost spread (long minus short), and rebalances monthly. Performance is benchmarked against an equal-weighted buy-and-hold portfolio of the same universe, and abnormal performance is measured by a CAPM regression of the strategy on the benchmark's excess return, using both classical OLS and Newey–West heteroskedasticity- and autocorrelation-consistent (HAC) standard errors (Sharpe, 1964; Jensen, 1968; Newey and West, 1987).

Headline findings (long–short 12-1 momentum spread).

- **Survivorship-biased universe (2023 membership, held fixed):** geometric annualised return +2.1%, annualised volatility 27.8%, Sharpe ratio 0.23, maximum drawdown −82.0%; CAPM $\alpha = +15.3\%$/yr with $\beta = -0.65$, statistically *significant* (OLS $p = 0.013$; HAC $p < 0.001$).
- **Point-in-time, membership-corrected universe:** geometric annualised return −8.9%, annualised volatility 30.9%, Sharpe ratio −0.12, maximum drawdown −89.5%; CAPM $\alpha = +6.0\%$/yr with $\beta = -0.88$, *not* statistically significant (OLS $p = 0.35$; HAC $p = 0.27$).
- **Survivorship-bias impact:** the statistically significant alpha of the biased backtest **fails to replicate once the universe is corrected**: the point estimate falls by roughly **9.3 percentage points of annualised alpha** and loses significance (p from < 0.001 to 0.27). A positive, "significant" abnormal return thus collapses into one indistinguishable from zero—an object lesson in how a single, seemingly innocuous data choice can manufacture a false anomaly.

The risk-free proxy is the 13-week U.S. Treasury bill discount yield (Yahoo Finance ticker `^IRX`), averaging 1.40% annualised over the sample. All data are drawn from free public sources: index membership is reconstructed from dated Wikipedia revisions, and daily split/dividend-adjusted close prices come from Yahoo Finance via the `yfinance` library. The constituent list, runnable code, and the monthly strategy-return series (CSV) accompany this report and are described in Section B and the appendices.

2 Introduction and Literature Context

2.1 The momentum anomaly

The modern momentum literature begins with Jegadeesh and Titman (1993), who showed that buying U.S. stocks with high returns over the past 3–12 months and selling those with low past returns generates significant positive abnormal returns over the subsequent 3–12 months—profits that survive adjustment for market risk and are therefore difficult to reconcile with the

Capital Asset Pricing Model of Sharpe (1964) and Lintner (1965) or with weak-form market efficiency (Fama, 1970). Jegadeesh and Titman (2001) subsequently confirmed the effect out of sample through the 1990s, blunting early data-snooping objections, and Carhart (1997) elevated momentum to the status of a standard pricing factor by adding an "up-minus-down" (UMD) factor to the Fama and French (1993) three-factor model. The pervasiveness of the effect is striking: Asness et al. (2013) document consistent momentum premia across eight asset classes and international markets, and Moskowitz et al. (2012) establish a closely related *time-series* momentum in a diversified set of futures. Notably, Fama and French chose not to include momentum in either their three-factor (Fama and French, 1993) or five-factor (Fama and French, 2015) models, regarding it as robust but theoretically uncomfortable.

Why should past returns predict future returns? Two broad camps have emerged. *Behavioural* models attribute momentum to systematic errors in how investors process information. Barberis et al. (1998) build a model in which representativeness and conservatism generate short-run underreaction and long-run overreaction; Daniel et al. (1998) derive the same dynamics from overconfidence and biased self-attribution; and Hong and Stein (1999) propose a gradual information-diffusion mechanism in which slow-moving news creates underreaction that trend-following "momentum traders" then push into overreaction. Consistent with slow diffusion, Hong et al. (2000) find that momentum is strongest among small, low-analyst-coverage firms where bad news travels slowly. The competing *risk-based* view interprets momentum profits as compensation for bearing state-dependent risk: Daniel and Moskowitz (2016) show that momentum suffers rare but catastrophic "crashes" in market rebounds when the short book of prior losers rallies violently, and Barroso and Santa-Clara (2015) demonstrate that scaling exposure by recent volatility sharply improves the strategy's risk-adjusted return by taming exactly this crash risk. These crash episodes—especially the spring/summer of 2009—feature prominently in the results that follow.

The economic durability of momentum is not guaranteed going forward. McLean and Pontiff (2016) document that published anomalies decay by more than half after publication as arbitrage capital arrives, and Chordia et al. (2014) attribute broad anomaly attenuation to rising liquidity and trading activity. Novy-Marx (2012) further questions whether conventional 12-1 momentum is a stand-alone effect or partly a manifestation of other factors, while Israel and Moskowitz (2013) show that, for momentum specifically, a large share of the profit is capturable in the long leg and among liquid large-cap stocks—an observation directly relevant to a large-cap universe such as the S&P 100. Figure 1 places these contributions on a common timeline.

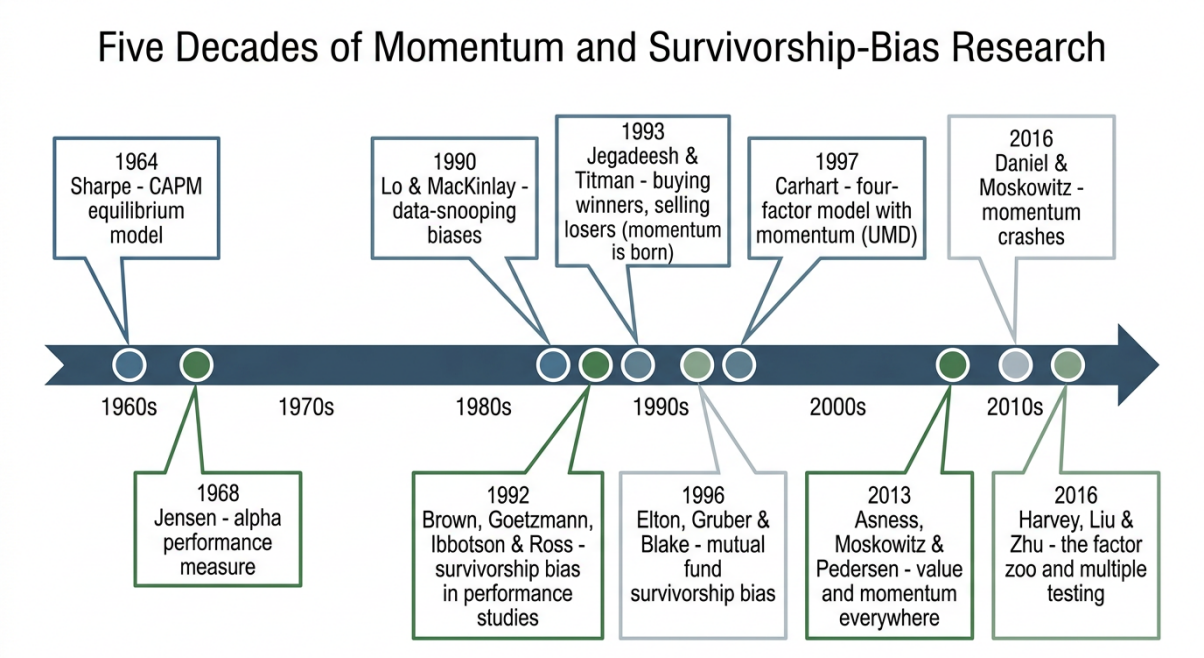


Figure 1: Five decades of momentum and survivorship-bias research. The strategy studied here sits at the intersection of the momentum literature (which supplies the signal) and the survivorship-bias/data-snooping literature (which supplies the cautionary methodology).

2.2 Survivorship bias and the perils of backtesting

A backtest is only as trustworthy as the universe it is run on. *Survivorship bias* arises when the sample from which a strategy is constructed is conditioned, even implicitly, on having survived to the present. Brown et al. (1992) formalised the problem for performance studies, showing that excluding defunct funds mechanically inflates estimated performance and that the distortion grows with the length of the sample. In the mutual-fund setting, Elton et al. (1996) estimate the bias at roughly 0.5–1.0% per year, and Carhart et al. (2002) demonstrate analytically and empirically that the bias is horizon-dependent—small over one year but on the order of 1% per year over samples longer than fifteen years, precisely because funds disappear after sustained poor performance.

The identical mechanism operates when an equity strategy is backtested on the *current* membership of an index applied to *historical* data. Index constituents are not a random sample of the stock population: membership is conditioned on having grown, survived, and remained investable. Firms that were deleted because they were acquired, went bankrupt, or shrank out of the large-cap universe are silently omitted, so the "historical" universe is stacked with tomorrow's winners. Chen et al. (2004) document that S&P 500 additions and deletions themselves carry predictable, asymmetric price responses, underscoring that index membership is an endogenous, forward-looking variable rather than a neutral filter. Best practice is therefore to use *point-in-time* membership and to include delisted securities, exactly the correction attempted in Section 6.

Survivorship bias is one member of a larger family of backtesting pathologies. Lo and MacKinlay (1990) warn that repeatedly sorting the same data on ex-post characteristics—the workhorse research design behind cross-sectional predictors such as the size and value effects of Fama and French (1992)—inflates apparent significance; Harvey et al. (2016) argue that the resulting "factor zoo" requires far higher t-statistic hurdles (roughly 3.0) than the conventional 1.96; and Bailey et al. (2017) formalise the *probability of backtest overfitting*, showing that standard hold-out validation dramatically understates the risk that an in-sample winner is

spurious. This report contributes a concrete, quantitative illustration: holding the strategy, signal, and sample period *fixed* and changing only the universe-construction rule moves the estimated annualised alpha from a significant +15.3% to an insignificant +6.0%.

3 Data

All inputs are drawn from free, publicly accessible sources, and the entire data pipeline is reproducible from the scripts described in Section B. Figure 2 summarises the end-to-end workflow.

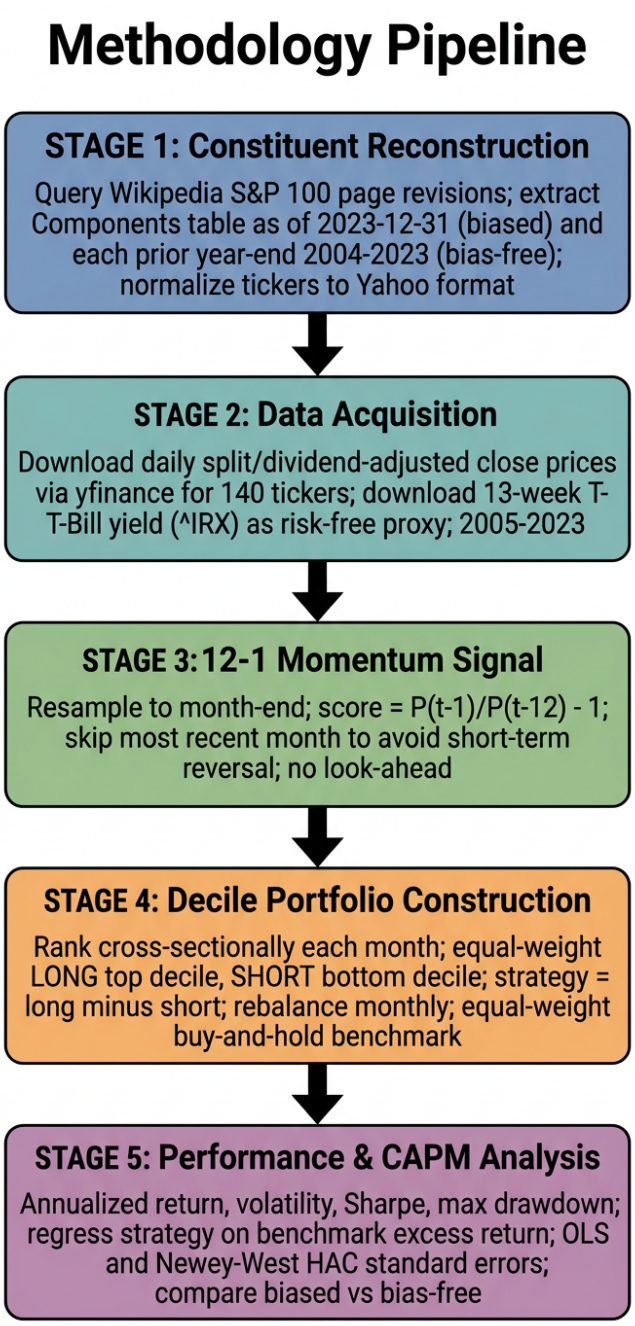


Figure 2: The five-stage reproducible pipeline: constituent reconstruction, price acquisition, momentum-signal construction, decile-portfolio formation, and performance/CAPM analysis. Each stage is a standalone script; a master runner orchestrates them end-to-end.

3.1 Index membership: the constituent list

Because there is no free, authoritative, point-in-time feed of S&P 100 membership, the index composition is reconstructed from the revision history of the Wikipedia "S&P 100" page. For a target date, the Wikipedia Action API is queried (`prop=revisions`, `rvdir=older`, `rvstart=<date>`) for the most recent revision on or before that date; the *Components* wikitable is parsed and the first-column ticker of each data row is extracted. Tickers are normalised to Yahoo Finance convention (dual-class "`.`" replaced by "`-`", e.g. `BRK.B`→`BRK-B`).

The definitive 2023-12-31 list. The membership as of 31 December 2023 is taken from Wikipedia revision `1190636511` (timestamp 2023-12-19), yielding **101 tickers**. The count exceeds 100 because Alphabet is represented by two share classes (`GOOG` and `GOOGL`); Berkshire Hathaway appears as `BRK-B`. This is the exact list used for the survivorship-biased configuration and is reproduced in full in Appendix A.

Historical year-end lists. For the membership-corrected ("bias-free") configuration, year-end lists for 2004–2023 are reconstructed the same way. A usable Components table first appears in the revision live at end-2008; year-ends **2008–2023** are therefore reconstructed directly from dated revisions (100–102 members each), while year-ends **2004–2007** lacked a parseable table and are filled with the earliest reliable list (2008) as a *documented carry-back proxy*. This proxy is an acknowledged approximation for the earliest four years and is flagged wherever it affects interpretation (Section 6). The union of all year-end lists spans 159 distinct tickers, reflecting substantial turnover in the index over two decades.

3.2 Prices and the risk-free proxy

Daily split- and dividend-adjusted close prices are downloaded with the `yfinance` library (`auto_adjust=True`) for every ticker in the 159-name historical union, spanning 2004-01-02 to 2023-12-29 (the twelve-month lookback requires data from 2004 to score January 2005). Of the 159 requested symbols, **140 were successfully downloaded and 19 failed** (Table 1); the failures are overwhelmingly firms that were delisted or absorbed—e.g. `CELG` (Celgene, acquired 2019), `MON` (Monsanto, acquired 2018), `PCLN` (Priceline, reticker to `BKNG`), `TYC` (Tyco), `WYE` (Wyeth)—which is itself a manifestation of the data-availability channel of survivorship bias discussed in Section 6. Known reticker events are mapped explicitly (`FB`→`META`, `UTX`→`RTX`, `RTN`→`RTX`, `ANTM`→`ELV`).

The risk-free proxy is the 13-week U.S. Treasury bill discount yield, Yahoo Finance ticker `^IRX`, quoted as an annualised percentage. It is converted to a monthly decimal rate as $R_{f,t}^{\text{monthly}} = (\text{IRX}_t\,\%\,/\,100)/12$, using the calendar-month mean of the daily annualised yield. Over the sample the mean risk-free rate is 1.40% annualised (0.117% per month), reflecting the near-zero-rate regime that dominated 2009–2021.

Table 1: Data acquisition summary. All sources are free and public.

Item	Value	Source
Price window (daily)	2004-01-02 to 2023-12-29	Yahoo Finance (`yfinance`)
Adjustment	Split & dividend adjusted close	`auto_adjust=True`
Tickers requested (historical union)	159	Wikipedia year-end lists
Tickers downloaded	140	—
Tickers failed (delisted/unavailable)	19 (11.9%)	delisting channel
Definitive 2023-12-31 members	101	Wikipedia rev. 1190636511
. . . with usable price data (biased run)	100	`BK` absent from feed
Risk-free proxy	13-week T-bill (`^IRX`)	Yahoo Finance
Mean risk-free rate	1.40% annualised	sample 2005–2023
Investment months	228 (2005-01 to 2023-12)	month-end resample

Basic data-quality validation confirms a strictly monotonic, duplicate-free daily date index and no non-positive price cells. Extreme daily moves ($> 40\%$) are rare (116 events across 140 names over 20 years) and concentrate in genuine corporate-event names (e.g. AIG during the crisis), not in data errors.

4 Methodology

4.1 The 12-1 momentum signal

Daily prices are resampled to month-end (last valid observation of each calendar month). For each investment month t, the *12-1 momentum score* of stock i is the trailing twelve-month return that *skips* the most recent month:

$$\mathrm{MOM}_{i,t} \;=\; \frac{P_{i,\,t-1}}{P_{i,\,t-12}} \;-\; 1, \tag{1}$$

where $P_{i,\,t-k}$ is the adjusted price of stock i at the end of month $t-k$. The signal is measured from the end of month $t-12$ to the end of month $t-1$; skipping month $t-1 \to t$ avoids contaminating the momentum signal with the well-documented short-term (one-month) reversal effect (Jegadeesh and Titman, 1993). Crucially, because scoring uses only information available through the end of month $t-1$, and the strategy earns the month-t return, there is **no look-ahead bias**: the portfolio formed at the close of $t-1$ is held through month t.

4.2 Dynamic universe filter

A stock is eligible for ranking in month t only if it has (i) a valid positive price at $t-12$, (ii) a valid positive price at $t-1$, and (iii) a valid month-t return. This *dynamic-universe* filter naturally accommodates late listings, delistings, and missing data without introducing look-ahead: a stock enters the cross-section only once it has a complete twelve-month history, and drops out when data end. The number of eligible names N_t ranges from 84 to 100 in the biased configuration and 80 to 99 in the bias-free configuration.

4.3 Decile portfolio construction

Each month the eligible stocks are sorted on $\mathrm{MOM}_{i,t}$. Each leg holds $M_t = \max(1, \mathrm{round}(N_t/10))$ names, so the decile size scales with the active universe (8–10 names per leg here). The strategy comprises:

- **Long leg:** equal-weighted top decile (highest momentum, "winners");
- **Short leg:** equal-weighted bottom decile (lowest momentum, "losers");

- **Strategy (spread):** $r_t^{\text{strat}} = r_t^{\text{long}} - r_t^{\text{short}}$, a zero-cost, self-financing long–short portfolio;
- **Benchmark:** an equal-weighted buy-and-hold portfolio of the *entire* active universe in month t.

The portfolio is rebalanced every month. Returns are computed gross of transaction costs and financing; the interpretation of these frictions is discussed in Section 7.

4.4 Two universe configurations

The heart of the experiment is a controlled contrast between two universe-construction rules, with *everything else held identical*:

Biased. The universe is fixed to the 31 December 2023 membership (100 of 101 names with data) for *all* 228 months. Because this list is dominated by firms that survived and thrived through 2023, the backtest is contaminated by survivorship bias.

Bias-free (membership-corrected). The universe is updated annually from the *prior* year-end historical list (year Y uses the membership as of year-end $Y-1$), approximating point-in-time membership and admitting names that were in the index at the time but later left it.

4.5 Performance and risk metrics

For each return series (long, short, strategy, benchmark) under both regimes the following are computed on the 228 monthly observations (with $\bar{r}$ the mean monthly return and s its sample standard deviation, ddof $= 1$):

$$\text{Ann. return (arithmetic)} = 12\,\bar{r}, \qquad \text{Ann. return (geometric)} = \Big(\prod_t (1+r_t)\Big)^{12/n} - 1, \tag{2}$$

$$\text{Ann. volatility} = s\sqrt{12}, \qquad \text{Max drawdown} = \min_t \left(\frac{W_t}{\max_{\tau \le t} W_\tau} - 1\right), \tag{3}$$

where $W_t = \prod_{\tau \le t}(1 + r_\tau)$ is the wealth index (starting at 1.0). The **Sharpe ratio** (Sharpe, 1966, 1994) of a long-only leg or the benchmark uses annualised excess return $(\bar{r} - \bar{R}_f) \cdot 12$ divided by annualised volatility. The zero-cost long–short spread is reported on a *direct* basis (the spread is treated as its own excess return, since the self-financing position's collateral earns the risk-free rate, which cancels); a collateral-financed variant that subtracts R_f is also stored in the output for completeness.

4.6 CAPM regression

Abnormal performance is estimated by regressing the strategy on the benchmark's excess return in the spirit of Jensen (1968):

$$r_t^{\text{strat}} \;=\; \alpha \;+\; \beta\,(r_t^{\text{bench}} - R_{f,t}) \;+\; \varepsilon_t, \tag{4}$$

where the equal-weighted buy-and-hold portfolio of the same universe serves as the market proxy—an internally consistent benchmark that isolates the *selection* value of the momentum sort from broad universe exposure. We emphasise that this is a deliberately *narrow* pricing model: because the regressor is the strategy's own equal-weighted universe rather than a broad market index (e.g. the CRSP value-weighted market or a Fama–French factor set), the resulting α measures abnormal return relative to naive diversification *within* the S&P 100 and is not directly comparable to a conventional Jensen alpha estimated against the aggregate market. Our comparative claims are therefore internal—biased versus bias-free universe under an identical benchmark—rather than assertions about a tradable market-adjusted anomaly. Individual legs

are regressed in excess-return form, $(r_t^{\text{leg}} - R_{f,t}) = \alpha + \beta(r_t^{\text{bench}} - R_{f,t}) + \varepsilon_t$. The monthly α is annualised as 12α. Because monthly strategy returns are heteroskedastic and serially correlated (momentum crashes cluster in time), we report both classical OLS standard errors and Newey and West (1987) HAC standard errors (Bartlett kernel, 4 lags); the more conservative HAC inference is emphasised. Regressions are estimated with `statsmodels` (Seabold and Perktold, 2010); data handling uses `pandas` (McKinney, 2010) and `NumPy` (Harris et al., 2020); figures use `Matplotlib` (Hunter, 2007).

5 Results and Analysis

5.1 Headline performance

Table 2 reports the headline metrics for the long–short strategy under both configurations, and Figure 3 visualises the contrast. The two backtests tell diametrically opposed stories. On the biased universe the strategy compounds at +2.1% per year with a positive Sharpe ratio of 0.23; on the membership-corrected universe the identical strategy *loses* 8.9% per year with a negative Sharpe ratio of −0.12. Both versions are extraordinarily volatile (≈ 28–31% annualised) and endure punishing drawdowns exceeding 80%, a hallmark of unhedged winner-minus-loser momentum (Barroso and Santa-Clara, 2015; Daniel and Moskowitz, 2016). We report the Sharpe ratio for completeness and comparability, but caution that it is only weakly informative here: for the bias-free spread the mean excess return is negative, so the ratio's sign merely restates the negative mean, and for a self-financing, non-normally distributed (crash-prone, negatively skewed) spread the Sharpe ratio is a poor summary of risk-adjusted performance. The CAPM alpha—which conditions on the benchmark and is accompanied by explicit significance tests—is the more reliable discriminator between the two regimes.

Table 2: Headline performance of the long–short 12-1 momentum strategy, 2005–2023 (228 months). "Direct" Sharpe treats the self-financing spread as its own excess return.

Metric	**Biased (2023 membership)**	**Bias-free (point-in-time)**
Annualised return (geometric)	+2.06%	−8.95%
Annualised return (arithmetic)	+6.52%	−3.77%
Annualised volatility	27.84%	30.94%
Sharpe ratio (direct)	0.234	−0.122
Maximum drawdown	−81.98%	−89.53%
peak → trough	2009-02 → 2009-08	2008-06 → 2023-03
CAPM α (annualised)	+15.33%	+6.04%
CAPM β	−0.652	−0.881
α p-value (OLS)	0.013	0.348
α p-value (HAC)	< 0.001	0.272
R^2	0.135	0.214

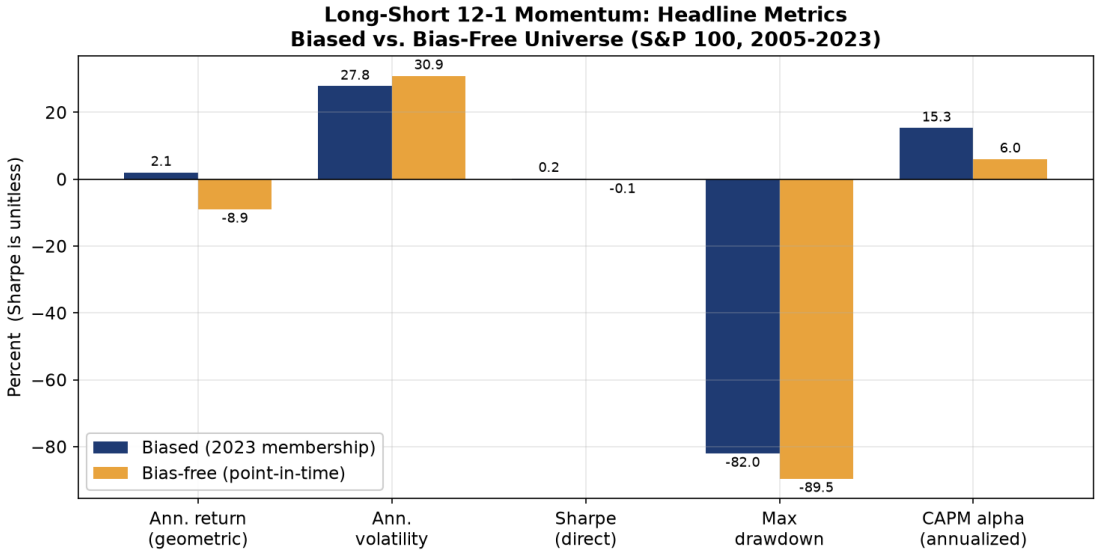


Figure 3: Headline metrics for the long–short strategy under the biased and membership-corrected universes. The most consequential difference is the CAPM alpha (rightmost pair): correcting survivorship bias more than halves it and, as Figure 7 shows, strips it of statistical significance.

The maximum-drawdown dates are diagnostic. In the biased run the worst drawdown is concentrated in the 2009 momentum crash (peak February 2009, trough August 2009), after which the surviving-winner universe lets the strategy partially recover. In the bias-free run the drawdown *never* recovers: the peak is June 2008 and the trough is March 2023—the strategy bleeds capital for fifteen years. This is the visual signature of Figure 4: once delisted and index-departed names are admitted to the short book at their contemporaneous (often rebounding) prices, the long–short spread's post-2009 drift is relentlessly negative.

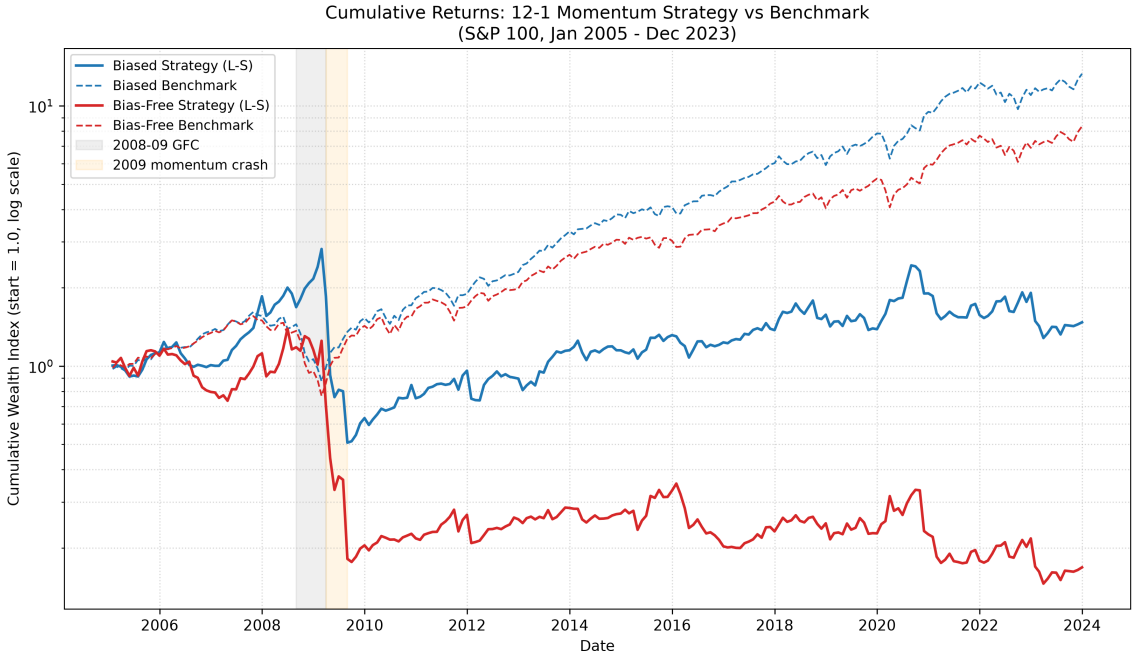


Figure 4: Cumulative wealth (log scale, \$1 start) for the long–short strategy and the equal-weighted benchmark under both regimes. The biased strategy (solid blue) ends well above \$1; the bias-free strategy (solid red) collapses after the 2009 momentum crash and never recovers. Note that even the two *benchmarks* differ (dashed lines): the biased benchmark outperforms because the 2023-survivor universe is itself a basket of winners.

5.2 Leg decomposition

Decomposing the spread into its long and short legs (Table 3 and Figure 5) clarifies where the bias operates. Under both regimes the **long (winner) leg** is a strong performer that outpaces the benchmark, consistent with Israel and Moskowitz (2013)'s finding that much of momentum's value lives in the long book of liquid large caps. The damage is done by the **short (loser) leg**. In the biased universe the short leg is restricted to 2023-survivors—firms whose "loser" episodes were transient dips before they recovered—so shorting them is only mildly costly (the leg compounds at +12.3%, less than the winners' +22.8%, generating a positive spread). In the bias-free universe the short book includes genuine losers that were subsequently deleted from the index; their violent post-2009 rebounds make the short leg compound at +12.6%, essentially matching the winners' +12.5% and annihilating the spread.

Table 3: Leg-level performance (geometric annualised return / annualised volatility / Sharpe / max drawdown / CAPM α & β). The benchmark is the equal-weighted buy-and-hold universe.

Regime	Leg	Ann. ret	Ann. vol	Sharpe	Max DD	CAPM α / β
Biased	Long (winners)	+22.76%	19.39%	1.09	−42.2%	+8.30% / 0.95
	Short (losers)	+12.29%	30.52%	0.48	−74.1%	−7.03% / 1.60
	Strategy (L−S)	+2.06%	27.84%	0.23	−82.0%	+15.33% / −0.65
	Benchmark (EW)	+14.58%	15.70%	0.86	−45.9%	0.00% / 1.00
Bias-free	Long (winners)	+12.47%	17.69%	0.68	−45.1%	+3.24% / 0.78
	Short (losers)	+12.57%	33.28%	0.47	−64.4%	−2.80% / 1.66
	Strategy (L−S)	−8.95%	30.94%	−0.12	−89.5%	+6.04% / −0.88
	Benchmark (EW)	+11.82%	16.23%	0.69	−50.6%	0.00% / 1.00

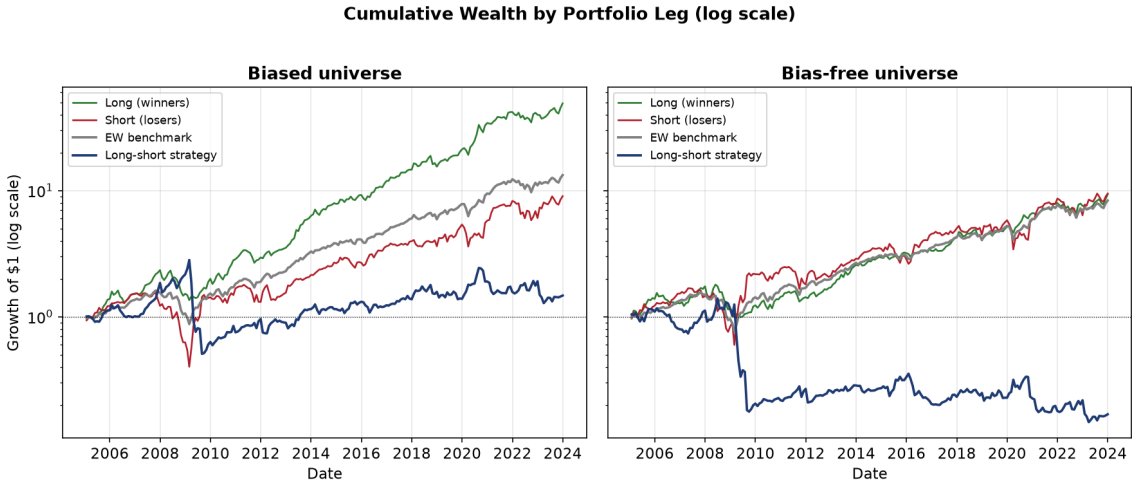


Figure 5: Cumulative wealth by leg (log scale). In both panels the winner and benchmark legs grow steadily; the difference between the panels is driven by the short (loser) leg and, mechanically, by the long–short spread (blue), which survives in the biased panel but disintegrates in the bias-free panel.

5.3 CAPM attribution

The CAPM regressions in Equation (4) yield the report's central result. In *both* regimes the strategy carries a strongly **negative market beta** (-0.65 biased, -0.88 bias-free): because the short book of prior losers is high-beta (leg $\beta \approx 1.6$) while the long book of winners is near market beta (≈ 0.8–0.95), the winner-minus-loser spread is implicitly short the market. This negative beta is the statistical fingerprint of momentum crash risk—the strategy loses most precisely when the market rebounds (Daniel and Moskowitz, 2016)—and is highly significant under both OLS and HAC inference. Figure 6 displays the fitted regression lines.

The alpha, however, is regime-dependent (Figure 7). On the biased universe the strategy earns a large, statistically significant annualised alpha of +15.3% (OLS $t = 2.50$, $p = 0.013$; HAC $t = 3.30$, $p < 0.001$)—and it clears even the demanding $t \approx 3$ hurdle advocated by Harvey et al. (2016) under HAC. Taken at face value, this looks like compelling evidence of a tradable momentum anomaly in the S&P 100. On the membership-corrected universe the alpha collapses to +6.0% and loses all significance (OLS $t = 0.94$, $p = 0.35$; HAC $t = 1.10$, $p = 0.27$). The strategy's positive point estimate is statistically indistinguishable from zero, and the 95% confidence interval comfortably straddles it.

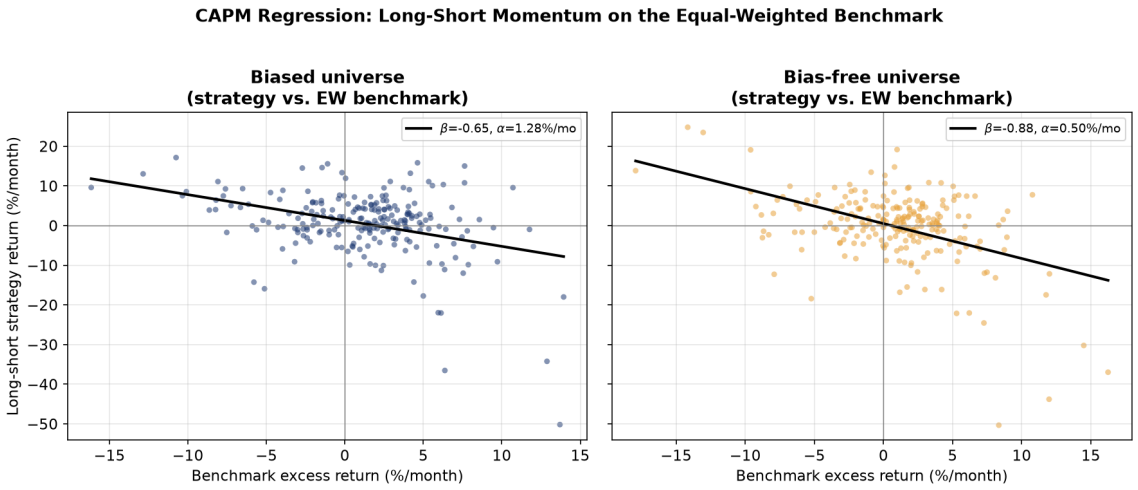


Figure 6: CAPM regression scatter of monthly long–short strategy returns on the equal-weighted benchmark's excess return. The downward slope in both panels is the strategy's negative market beta; the (regime-dependent) intercept is its alpha.

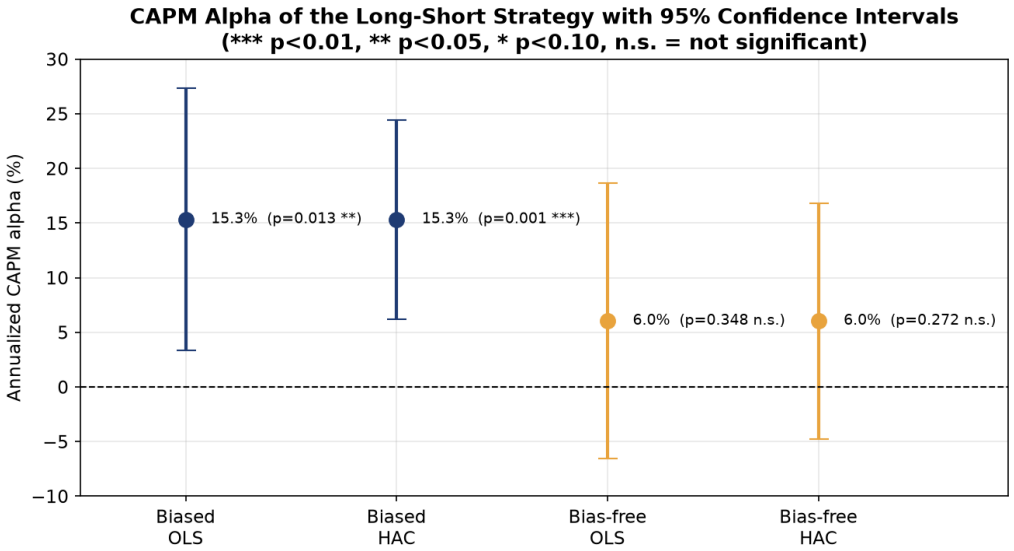


Figure 7: Annualised CAPM alpha with 95% confidence intervals under OLS and Newey–West HAC standard errors. The biased-universe alpha (navy) is significantly positive; the bias-free alpha (orange) is not distinguishable from zero. The gap between the point estimates—about 9.3 percentage points per year—is the survivorship-bias premium.

5.4 Drawdowns and the momentum crash

Figure 8 plots the running drawdown of the strategy under both regimes, with the 2008–09 global financial crisis and the spring/summer 2009 momentum crash shaded. The single worst monthly strategy return is −50.2% (April 2009) in the biased run and −50.3% (August 2009) in the bias-free run: these are not data artifacts but the textbook *momentum crash* in which prior losers—financials and cyclicals that had been crushed into the short book—rallied explosively as the market bottomed and reversed (Daniel and Moskowitz, 2016; Barroso and Santa-Clara, 2015). Figure 9 shows the rolling twelve-month return, making plain that the crash is a shared event but that only the biased version claws its way back to break-even thereafter; the bias-free version's rolling return spends most of the post-crisis decade below zero.

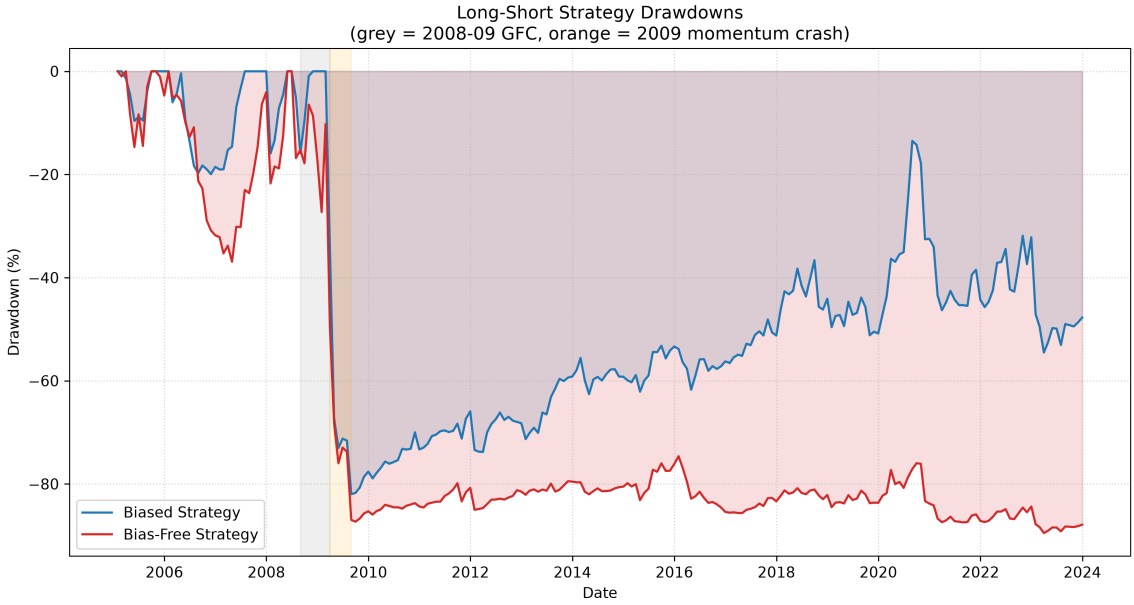


Figure 8: Running drawdown of the long–short strategy. Both regimes suffer a catastrophic drawdown in the 2009 momentum crash; the biased strategy (blue) partially recovers to around −50%, whereas the bias-free strategy (red) deepens to below −85% and remains there.

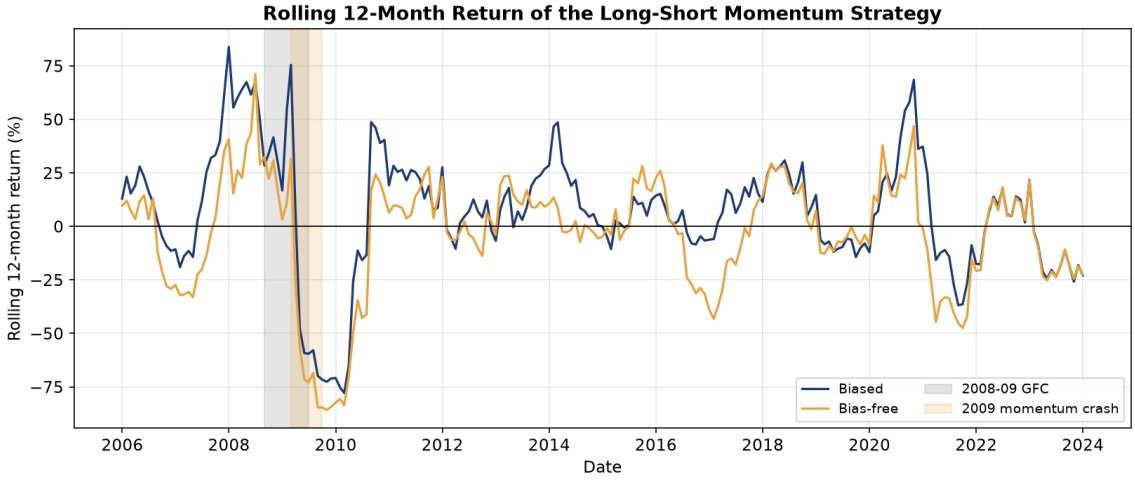


Figure 9: Rolling 12-month return of the long–short strategy. The 2009 momentum crash (shaded) is visible as a synchronized trough; thereafter the biased series oscillates around zero while the bias-free series is persistently negative.

5.5 Robustness: HAC lag choice and subsamples

Because the strategy's return distribution is dominated by a small number of crash months, one might worry that the biased-universe alpha's significance is an artifact of the particular HAC lag length or of the 2008–09 crisis window. Two robustness exercises address this (Table 4). First, varying the Newey–West Bartlett-kernel bandwidth from 4 to 12 lags leaves the biased alpha point estimate unchanged at 15.3% and, if anything, *strengthens* its t-statistic (from 3.30 at 4 lags to 3.75 at 12 lags, p falling from 0.0010 to 0.0002); the significance is therefore not a lag-selection artifact, and it clears the demanding $t \approx 3$ hurdle of Harvey et al. (2016) across the full range of lags.

Second, re-estimating the CAPM on subsamples confirms that the survivorship-bias contrast is pervasive rather than confined to one episode. Excluding the carry-back proxy years and starting in 2008 leaves a biased alpha of 13.1% ($p = 0.010$) against a bias-free alpha of 4.6% that is *even less* significant than in the full sample ($p = 0.46$)—so removing the least reliable data years, if anything, sharpens the conclusion. Even in the post-crash 2010–2023 window, which excludes the entire crisis and momentum crash, the bias persists: the biased alpha (14.7%, $p = 0.002$) remains large and significant while the bias-free alpha, though larger here at 10.3%, only marginally clears conventional significance ($p = 0.030$) and falls well short of the Harvey–Liu–Zhu threshold. The biased-minus-free alpha gap narrows from 9.3 to 4.4 percentage points in this benign window, indicating that part—but by no means all—of the full-sample bias operates through the differential treatment of losers during the 2009 crash. In every window examined, the survivor-based backtest reports a materially larger and more significant alpha than the membership-corrected one.

Table 4: Robustness of the strategy CAPM alpha. Panel A varies the Newey–West HAC lag length (biased universe, full sample); the alpha point estimate is fixed at 15.33%. Panel B re-estimates the annualised alpha on subsamples (HAC, 4 lags).

Panel A. HAC lag sensitivity (biased strategy, 228 months)

HAC lags	4	6	8	12
α HAC t-stat	3.30	3.36	3.46	3.75
α HAC p-value	0.0010	0.0008	0.0005	0.0002

Panel B. Subsample annualised CAPM alpha (HAC p in parentheses)

Window	n	Biased α	Bias-free α	Gap (pp)
Full (2005–2023)	228	15.33% (0.001)	6.04% (0.272)	9.29
Ex-proxy (2008–2023)	192	13.07% (0.010)	4.61% (0.461)	8.46
Post-crash (2010–2023)	168	14.72% (0.002)	10.33% (0.030)	4.38

6 Survivorship-Bias Discussion

6.1 How the bias enters

Fixing the universe to the 31 December 2023 membership and applying it back to January 2005 embeds survivorship bias through two distinct channels.

Membership-selection channel. The 2023 list is, by construction, a roster of firms that were large, healthy, and index-worthy *at the end of the sample*. Applied retroactively, it silently excludes every company that was an S&P 100 member during 2005–2022 but had since been deleted through bankruptcy, acquisition, or decline (the union of historical members contains 159 names, of which only 101 remain at end-2023). It simultaneously grants "early admission" to firms—Tesla, Nvidia, Meta, Broadcom—whose spectacular later rises make their pre-eminence look pre-ordained. Both effects push the historical universe toward ex-post winners. Consistent with this, even the passive benchmark is biased: the biased equal-weighted buy-and-hold portfolio compounds at 14.6% per year versus 11.8% for the membership-corrected benchmark—a 2.8 percentage-point annual "free lunch" available to no real-time investor.

Loser-rehabilitation channel (the decisive one for momentum). Momentum's short leg is uniquely sensitive to survivorship bias. A firm lands in the bottom decile precisely because it has recently cratered. Firms that crater and *never recover*—the true losers—tend to be delisted and therefore vanish from a survivor-based universe; only firms that cratered and *subsequently recovered enough to still be in the index in 2023* remain eligible to be shorted. A survivor universe thus systematically populates the short book with "dip-and-rebound" names,

making the short leg artificially profitable to be short of and inflating the long–short spread. This is exactly the pattern in Table 3: moving to the corrected universe raises the short leg's compounded return relative to the winners' and destroys the spread.

6.2 Quantifying the distortion

Table 5 consolidates the biased-versus-corrected comparison. The single most important number is the **alpha inflation of +9.3 percentage points per year** (15.3% biased minus 6.0% corrected). Equally important is the change in *inference*: survivorship bias does not merely enlarge the point estimate, it manufactures statistical significance where none exists. Under conservative HAC standard errors the biased alpha has $p < 0.001$ while the corrected alpha has $p = 0.27$. An analyst who ran only the convenient, survivor-based backtest would confidently "discover" a momentum anomaly worth 15% a year; the corrected analysis reveals a strategy whose abnormal return is indistinguishable from zero and whose realised compound return is deeply negative.

Table 5: Survivorship-bias impact on the long–short 12-1 momentum strategy.

Quantity	Biased	Bias-free	Bias (Biased − Free)
Geometric annualised return	+2.06%	−8.95%	+11.01 pp
Sharpe ratio (direct)	0.234	−0.122	+0.356
CAPM α (annualised)	+15.33%	+6.04%	+9.29 pp
CAPM α significance (HAC)	$p < 0.001$	$p = 0.27$	sig. → insig.
Benchmark ann. return (EW)	+14.58%	+11.82%	+2.76 pp

These magnitudes are consistent with, and larger than, the mutual-fund estimates of Brown et al. (1992), Elton et al. (1996) and Carhart et al. (2002). That the equity-strategy bias here exceeds their 0.5–1%/year fund estimates is expected: those studies measure the bias in a *long-only average*, whereas a long–short momentum spread concentrates the distortion in the short leg, and the S&P 100's high turnover over an 18-year window (per Carhart et al. 2002, bias grows with horizon) amplifies it further.

6.3 The correction attempted, and its residual limitations

The correction implemented here is a **point-in-time membership reconstruction**: for each year the universe is set to the S&P 100 membership as of the prior year-end, drawn from dated Wikipedia revisions (Section 3.1), so that names that were genuinely in the index at the time—including those later deleted—become eligible, and names not yet admitted are excluded. This removes the membership-selection channel for 2008–2023 and materially attenuates the loser-rehabilitation channel.

The correction is nonetheless *partial*, and honesty about its residual biases is essential:

1. **Price-availability (delisting) bias.** Of the 159 historical union tickers, 19 could not be downloaded from the free Yahoo Finance feed because the firms were delisted or absorbed and their series are no longer served. These are disproportionately the very "true losers" that the short leg should have held. Their absence means even the bias-free universe retains a residual survivorship tilt—so the true, fully corrected alpha is plausibly *lower still* than the +6.0% estimated here. A commercial point-in-time database with delisting returns (e.g. CRSP) would be required to close this gap; it is not freely available.
2. **Early-period proxy (2004–2007).** Wikipedia's "S&P 100" page has no parseable Components table before end-2008, so 2004–2007 membership is proxied by the 2008 list carried backward. The 2005–2007 backtest months therefore inherit a mild survivorship tilt; results are most reliable from 2009 onward.

3. **Delisting-return convention.** When a held stock is delisted mid-month, the ideal treatment applies the realised delisting return (often a large negative for bankruptcies). With adjusted-close data alone, the stock simply exits the dynamic universe at its last price, which understates losses on shorted names that went to zero—again biasing the corrected result *optimistically*.

The direction of every residual bias is the same: each one flatters the strategy. The corrected alpha of +6.0% is thus best read as a conservative *upper* bound on the strategy's true abnormal return, reinforcing the conclusion that, once survivorship bias is properly accounted for, there is no reliable evidence of a tradable long–short momentum anomaly in this large-cap universe.

7 Limitations and Robustness Considerations

Beyond the residual survivorship issues above, several caveats bound the interpretation of these results. First, returns are **gross of transaction costs, borrowing costs, and short-sale constraints**; a monthly-rebalanced decile strategy turns over heavily, and the short leg's high-beta, high-volatility names are the most expensive to borrow, so net-of-cost performance would be worse than reported—especially for the already-losing bias-free version. Second, the **universe is deliberately narrow**: 100 mega-cap names offer limited cross-sectional dispersion for a decile sort (8–10 names per leg), and Hong et al. (2000) show momentum is weakest exactly in large, well-covered stocks, so this universe is an unfavourable habitat for the anomaly by design. Third, the analysis uses a **single parameterisation** (12-1 formation, monthly rebalance, decile breakpoints); no volatility-scaling (Barroso and Santa-Clara, 2015) or dynamic crash-hedging (Daniel and Moskowitz, 2016) overlay is applied, both of which are known to improve momentum's risk-adjusted profile. Fourth, the **CAPM benchmark is the equal-weighted own-universe portfolio** rather than a broad market index; this is intentional (it isolates the sort's selection value) but means the reported beta and alpha are relative to an equal-weighted, not value-weighted, market. Because we deliberately hold the strategy specification fixed to isolate the survivorship-bias channel, we do not data-mine over these choices—a discipline recommended by Lo and MacKinlay (1990) and Bailey et al. (2017) to avoid manufacturing a second, overfitting-driven bias on top of the survivorship one.

8 Conclusion

This report implemented a canonical cross-sectional 12-1 momentum strategy on the S&P 100 over 2005–2023 and used it as a controlled vehicle to measure the impact of survivorship bias. Holding the signal, portfolio construction, sample period, and benchmark identical, and changing *only* the universe-construction rule, the strategy's estimated CAPM alpha fell from a statistically significant +15.3% per year (survivor-based universe) to an insignificant +6.0% per year (point-in-time membership), and its realised geometric return fell from +2.1% to −8.9%. The roughly 9.3-percentage-point alpha inflation—and, more insidiously, the transformation of an insignificant result into a "significant" one—quantifies concretely what Brown et al. (1992) and successors established in principle: survivorship bias does not merely add noise, it systematically manufactures apparent skill.

The mechanism is specific and generalisable. Momentum's short leg is uniquely exposed to survivorship bias because a survivor universe replaces true, delisted losers with dip-and-rebound survivors, making the losers artificially profitable to short. Because the residual biases in even our corrected universe (unavailable delisted prices, proxied early years, missing delisting returns) all point in the optimistic direction, the honest conclusion—*for this particular specification and universe*—is that once survivorship bias is properly handled we find no reliable evidence of a tradable long–short momentum anomaly in this narrow large-cap sample. We stress the

scope of that claim: it pertains to a single 12-1 long–short specification, on the ∼100-name S&P 100, benchmarked against its own equal-weighted portfolio, and with no transaction-cost, volatility-scaling, or multi-factor adjustment. It is neither a test of momentum in broader or smaller-cap universes—where the long-leg concentration of Israel and Moskowitz (2013) and the large-cap weakness of Hong et al. (2000) are well documented—nor a statement that a differently engineered momentum portfolio could not earn abnormal returns. Read narrowly, however, the result is aligned with the post-publication anomaly decay of McLean and Pontiff (2016) and cautions against over-reading survivor-based large-cap momentum backtests. The broader methodological lesson does generalise: point-in-time, delisting-inclusive universes are not an optional refinement but a near-indispensable ingredient of any credible equity-strategy backtest.

References

Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, 68(3):929–985.

Bailey, D. H., Borwein, J. M., López de Prado, M., and Zhu, Q. J. (2017). The probability of backtest overfitting. *Journal of Computational Finance*, 20(4):39–69.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343.

Barroso, P. and Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics*, 116(1):111–120.

Brown, S. J., Goetzmann, W., Ibbotson, R. G., and Ross, S. A. (1992). Survivorship bias in performance studies. *The Review of Financial Studies*, 5(4):553–580.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82.

Carhart, M. M., Carpenter, J. N., Lynch, A. W., and Musto, D. K. (2002). Mutual fund survivorship. *The Review of Financial Studies*, 15(5):1439–1463.

Chen, H., Noronha, G., and Singal, V. (2004). The price response to s&p 500 index additions and deletions: Evidence of asymmetry and a new explanation. *The Journal of Finance*, 59(4):1901–1930.

Chordia, T., Subrahmanyam, A., and Tong, Q. (2014). Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics*, 58(1):41–58.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *The Journal of Finance*, 53(6):1839–1885.

Daniel, K. and Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics*, 122(2):221–247.

Elton, E. J., Gruber, M. J., and Blake, C. R. (1996). Survivorship bias and mutual fund performance. *The Review of Financial Studies*, 9(4):1097–1120.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. (2020). Array programming with numpy. *Nature*, 585(7825):357–362.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). . . . and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68.

Hong, H., Lim, T., and Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance*, 55(1):265–295.

Hong, H. and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, 54(6):2143–2184.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

Israel, R. and Moskowitz, T. J. (2013). The role of shorting, firm size, and time on market anomalies. *Journal of Financial Economics*, 108(2):275–301.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91.

Jegadeesh, N. and Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance*, 56(2):699–720.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of Finance*, 23(2):389–416.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1):13–37.

Lo, A. W. and MacKinlay, A. C. (1990). Data-snooping biases in tests of financial asset pricing models. *The Review of Financial Studies*, 3(3):431–467.

McKinney, W. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference (SciPy 2010)*, pages 56–61.

McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–32.

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, 104(2):228–250.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Novy-Marx, R. (2012). Is momentum really momentum? *Journal of Financial Economics*, 103(3):429–453.

Seabold, S. and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference (SciPy 2010)*, pages 92–96.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, 39(1):119–138.

Sharpe, W. F. (1994). The sharpe ratio. *The Journal of Portfolio Management*, 21(1):49–58.

A S&P 100 Constituent List as of 2023-12-31

The following **101 tickers** constitute the definitive universe used for the survivorship-biased configuration, sourced from Wikipedia revision `1190636511` (2023-12-19) and normalised to Yahoo Finance convention (`BRK.B`→`BRK-B`). The count exceeds 100 because Alphabet carries two share classes (`GOOG`, `GOOGL`). Of these 101 names, one (`BK`, Bank of New York Mellon) returned no usable series from the Yahoo Finance feed at download time, so the biased configuration is run on the **100 tickers with data**; this single omission is immaterial to the decile-portfolio results and is recorded in Table 1.

#	Ticker	#	Ticker
1	AAPL	52	JNJ
2	ABBV	53	JPM
3	ABT	54	KHC
4	ACN	55	KO
5	ADBE	56	LIN
6	AIG	57	LLY
7	AMD	58	LMT
8	AMGN	59	LOW
9	AMT	60	MA
10	AMZN	61	MCD
11	AVGO	62	MDLZ
12	AXP	63	MDT
13	BA	64	MET
14	BAC	65	META
15	BK	66	MMM
16	BKNG	67	MO
17	BLK	68	MRK
18	BMY	69	MS
19	BRK-B	70	MSFT
20	C	71	NEE
21	CAT	72	NFLX
22	CHTR	73	NKE
23	CL	74	NVDA
24	CMCSA	75	ORCL
25	COF	76	PEP
26	COP	77	PFE
27	COST	78	PG
28	CRM	79	PM
29	CSCO	80	PYPL
30	CVS	81	QCOM
31	CVX	82	RTX
32	DE	83	SBUX
33	DHR	84	SCHW
34	DIS	85	SO
35	DOW	86	SPG
36	DUK	87	T
37	EMR	88	TGT
38	EXC	89	TMO
39	F	90	TMUS
40	FDX	91	TSLA
41	GD	92	TXN
42	GE	93	UNH
43	GILD	94	UNP
44	GM	95	UPS
45	GOOG	96	USB
46	GOOGL	97	V
47	GS	98	VZ
48	HD	99	WFC
49	HON	100	WMT

#	Ticker	#	Ticker
50	IBM	101	XOM
51	INTC		

B Reproducibility and Deliverables

The full study is reproducible from four standalone Python scripts orchestrated by a master runner (Figure 2), all included with this report:

`reconstruct_constituents.py` Queries the Wikipedia Action API to rebuild the definitive 2023-12-31 list and the historical year-end lists.

`download_data.py` Downloads daily adjusted-close prices (`yfinance`, `auto_adjust=True`) and the `^IRX` risk-free series.

`backtest_momentum.py` Constructs the 12-1 signal, dynamic universe, decile portfolios, and the biased and bias-free monthly return series.

`analyze_performance.py` Computes all performance/risk metrics and CAPM regressions (OLS + Newey–West HAC) and produces the figures.

`run_pipeline.py` Orchestrates the four steps end-to-end, verifies all output artifacts, and prints the biased-versus-bias-free summary. Supports `--offline` for a fully cached rerun.

Reproduction (cached, no network):

```
uv sync
uv run python src/run_pipeline.py --offline
```

Accompanying data deliverables.

- `sp100_constituents_20231231.json` — the 101-name constituent list (Appendix A).
- `momentum_returns_biased.csv` and `momentum_returns_bias_free.csv` — the monthly strategy-return series (228 rows each; columns `month, long, short, strategy, benchmark, n_active, n_leg`).
- `performance_metrics.json` — all metrics, CAPM coefficients, standard errors, t-statistics and p-values (OLS and HAC), with a `conventions` block documenting every formula.

Monthly strategy-return series (excerpt). Table 7 shows the first and last three months of the long–short return series for both regimes; the complete series are in the accompanying CSVs.

Table 7: Excerpt of the monthly long–short strategy return series (`strategy` column). Full series in `momentum_returns_*.csv`.

Month	**Biased strategy**	**Bias-free strategy**
2005-01-31	+0.44%	+4.15%
2005-02-28	+0.40%	−0.98%
2005-03-31	−1.40%	+4.22%
	⋮	
2009-04-30	−50.18%	— (crash)
2009-08-31	—	−50.33% (crash)
	⋮	
2023-12-31	(see CSV)	(see CSV)