

Investor Sentiment and Momentum Across Market States

Jiayi Zhou*

Milton International School, Thaiyur, India

**Corresponding author: Jiayi Zhou.*

Abstract

Previous studies link investor sentiment to momentum, yet the forecasting content of that link may vary with the sentiment measure and the market environment. We compare the same US momentum outcomes across two published sentiment indices, their available calendar windows, and the winner and loser legs, and then evaluate pseudo-out-of-sample forecasts. The Baker-Wurgler (BW) series covers 701 months, whereas the relative-sentiment (RIS) series covers 323 months; a common-sample exercise aligns the outcome dates. We estimate retrospective regressions with three-factor controls and HAC inference. Across the 12 primary slope and interaction tests, none remains significant after Holm adjustment at 5%. RIS has a UP-state momentum slope of 2.486 percentage points per sentiment standard deviation (adjusted $p = 0.2826$). Expanding-window forecasts use training-only scaling and five model specifications. The state-interaction model produces out-of-sample coefficients of determination of 0.745% for BW and -3.615% for RIS, lower than the best simple comparator in each full sample. Secondary associations and changes over time rule out a universal null reading. The evidence is therefore measurement-sensitive and retrospective; it does not establish broad forecast superiority, a causal investor-psychology channel, or real-time trading profitability.

Keywords

investor sentiment, momentum, market states, portfolio returns, forecast evaluation

1. Introduction

Momentum portfolios take long positions in recent winners and short positions in recent losers, so the return they deliver depends on the market conditions in which those positions are formed. Jegadeesh and Titman documented intermediate-horizon continuation [1]. Later work tied momentum to prior market performance [2] and investor sentiment [3]. These findings motivate conditional-return specifications, but an observed association alone does not show that sentiment adds forecasting power beyond simpler predictors.

Measurement complicates this comparison. A composite sentiment index and a relative-valuation index can represent different aspects of investor beliefs, while their available histories can cover different market episodes. Portfolio aggregation introduces another ambiguity: a positive winner-minus-loser association can arise from winners, losers or both. Comparing full-period spread coefficients alone therefore combines

measurement, calendar composition and portfolio-leg effects. Forecast evaluation adds a separate requirement because fitted explanatory controls may be unavailable at the prediction date.

This paper makes three contributions. First, it compares two published sentiment measures against the same US momentum outcomes, separating measurement differences from changes in portfolio targets and calendar coverage. Second, it decomposes the spread into winner and loser legs across lagged market states, which clarifies where conditional associations arise without assigning them to a single psychological mechanism. Third, it separates retrospective explanatory regressions from pseudo-out-of-sample forecasting and asks whether a sentiment-state interaction improves prediction over simpler benchmarks. To limit specification searching, the analysis uses a fixed set of primary tests, comparators and sensitivity checks. Positive and negative estimates retain the same evidential status. These design choices keep conditional-return evidence separate from claims about investor psychology or executable trading profits.

2. Related Literature and Hypotheses

Behavioral explanations connect return continuation to how information reaches investors and how they respond to it. Hong and Stein model gradual information diffusion followed by momentum trading [4]. Antoniou, Doukas and Subrahmanyam relate sentiment-dependent momentum to delayed responses to news inconsistent with prevailing optimism or pessimism [3]. These arguments motivate conditional sentiment slopes, but aggregate monthly returns cannot distinguish the underlying cognitive mechanisms. Our regressions test an observable association that such mechanisms may generate.

Limits to arbitrage provide a reason to examine the two portfolio legs separately. Shleifer and Vishny show how capital constraints can restrict arbitrage even when expected gains increase [5]. Baker and Wurgler relate sentiment exposure to stocks with subjective valuations and difficult arbitrage [6, 7]. Stambaugh, Yu and Yuan locate much of their sentiment–anomaly evidence in short legs [8]. Their later analysis addresses the concern that persistent predictors could produce spurious anomaly relations [9]. These studies make the loser leg informative, while also warning against treating any isolated significant slope as decisive evidence.

Market conditions offer a competing and potentially complementary explanation. Cooper, Gutierrez and Hameed document momentum differences across prior market states [2]. Daniel and Moskowitz associate momentum crashes with panic conditions and subsequent rebounds [10]. Lagged state and volatility may therefore explain variation that would otherwise be attributed to sentiment. Gao and colleagues provide a recent, direct precedent through relative investor sentiment and its association with momentum [11]. Their valuation-based index motivates a comparison with composite sentiment on matching return months. Our estimates concern this public-data reassessment and do not reproduce the original article's complete empirical design.

Aksoy-Yurdagul, Buchner and Zareei examine the persistence of firm-level news sentiment and horizon-dependent return patterns [12]. Their news-based measure addresses a different question from aggregate sentiment levels. Our analysis relates published monthly indices to shared portfolio outcomes; it neither measures news persistence nor tests the performance of their firm-level signal.

Three primary hypotheses define the empirical tests. H1: investor sentiment is associated with momentum returns in UP market states. H2: the sentiment slope differs between UP and DOWN states. H3: a model allowing sentiment to interact with market state improves pseudo-out-of-sample forecast accuracy relative to the historical mean, sentiment-only, state-only and additive benchmarks. Winner and loser outcomes provide a mechanism decomposition rather than a separate directional hypothesis, so no coefficient sign is imposed. Risk adjustment follows the market, size and value factors of Fama and French [13], while Fama's efficiency framework cautions that return tests also depend on the expected-return model [14]. Neither an intercept nor a forecasting gain alone identifies market inefficiency.

3. Data and Methodology

We merge two published sentiment series with the same monthly US portfolio returns. Baker-Wurgler's orthogonal index (BW) removes macroeconomic components from a composite sentiment measure [6, 7], while relative investor sentiment (RIS) offers a valuation-based alternative [11]. The complete-case samples cover 701 months for BW (August 1965–December 2023) and 323 months for RIS (February 1996–December

2022). We also rerun the comparison on the 323 months shared by both indices. Thus, the exercise changes the sentiment measure while holding the underlying outcomes common; it is not a comparison of independent markets.

French's monthly prior-return deciles provide the portfolio returns [15]. The source sorts eligible NYSE, AMEX and NASDAQ stocks using prior months 2–12 and NYSE breakpoints. We use the supplied value-weighted highest and lowest deciles, without reconstructing security-level holdings. Momentum is the highest-minus-lowest return; winner and loser excess returns each subtract the risk-free return. Input percentage returns are converted to fractions, missing-value sentinels are removed, and calendar gaps are preserved before lagging. Outcomes stop at each index's last observed month. Revised histories and unavailable original release dates make all forecast exercises pseudo-out-of-sample.

Let t index the outcome month and y_t denote momentum, winner excess return or loser excess return. The lagged sentiment predictor s_{t-1} is standardized within the estimation sample. The indicator D_{t-1} equals one when the compounded total market return over the preceding 36 months is negative, and zero otherwise. These states are called DOWN and UP, respectively. The variable v_{t-1} is the standardized sample standard deviation of total market returns over the preceding 12 months. With intercept α , sentiment coefficient β , state coefficient γ , volatility coefficient δ , interaction coefficient θ and residual ε_t , the retrospective regression is

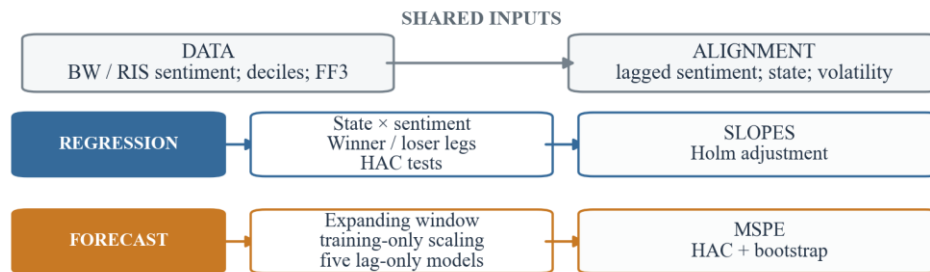
$$y_t = \alpha + \beta s_{t-1} + \gamma D_{t-1} + \delta v_{t-1} + \theta s_{t-1} D_{t-1} + \lambda^T f_t + \varepsilon_t. \quad (1)$$

Here f_t contains contemporaneous market excess, size and value factor returns, and λ contains their loadings [13]. The superscript T denotes transpose. The UP sentiment slope is β , the DOWN slope is $\beta + \theta$, and θ is their difference. Full-sample standardization is used only for retrospective associations and is not reused in forecasting.

Ordinary least squares estimates use Newey–West heteroskedasticity and autocorrelation consistent (HAC) covariance with six lags, Bartlett weights and a finite-sample correction [16]. Intervals use the residual degrees of freedom of the corresponding Student t distribution. The primary family contains 12 two-sided tests: the UP slope and state difference for three outcomes and two indices. Holm adjustment controls familywise error at 5%. DOWN slopes, common-sample estimates and specification checks are secondary; their pointwise intervals do not replace the primary tests.

Expanding-window forecasts start with 120 valid training months and predict the next month's momentum return. Each origin refits coefficients and estimates sentiment and volatility means and standard deviations using training observations only. The historical mean uses an intercept; sentiment only adds lagged sentiment; market state only adds state and volatility. The additive model combines these predictors, and the state interaction model adds their sentiment–state product. Forecasts exclude contemporaneous factor returns. All five models use identical test months within each index. Figure 1 separates the retrospective and forecasting paths and their information sets.

Figure 1: Study design. Shared monthly inputs feed retrospective regressions and expanding-window forecasts



Mean squared prediction error (MSPE) is the average squared difference between observed and predicted momentum. It is the primary forecast metric; mean absolute error and out-of-sample R^2 are secondary. With $MSPE_0$ denoting the historical-mean model's error over identical forecast months, the latter is

$$R_{OOS}^2 = 1 - \frac{MSPE}{MSPE_0}. \quad (2)$$

An MSPE below the historical mean gives a negative loss difference. For each month, we subtract the comparator's squared error from the interaction model's squared error and use these paired differences for inference. The intervals are HAC intervals supplemented by a circular moving-block percentile bootstrap with 12-month blocks and 2,000 draws (seed 20260904). The bootstrap resamples the saved losses and does not refit the forecasts. Comparisons among nested models are reported descriptively.

Sensitivity checks vary the sentiment lag (1, 3, or 6 months), the state lookback (12, 24, or 36 months), and the weighting scheme (value or equal). They also use raw BW sentiment, alternative factor controls, HAC bandwidths of 3 or 12, and a 180-month initial forecast window. Calendar partitions are 1966-1989, 1990-2006 and 2007-2023, subject to data availability. For each index, the three best and three worst momentum months are selected symmetrically from realized returns. The checks are intended to show stability, rather than to choose a favorable significant specification. Separate computational checks verified feature timing, return identities and forecast calculations.

4. Results

4.1 Sentiment Slopes and Portfolio Legs

After Holm adjustment at the 5% familywise level, none of the 12 primary tests is significant. Table 1 gives the full-sample UP slopes and DOWN-minus-UP differences with pointwise intervals and adjusted p-values. BW's momentum slope is essentially zero. RIS's UP-state slope is 2.486 percentage points per sentiment standard deviation; its pointwise interval runs from 0.336 to 4.635, the nominal p-value is 0.0235, and the adjusted p-value is 0.2826. RIS therefore supplies the clearest directional indication of a sentiment-momentum link, but the indication does not withstand familywise multiplicity correction. Neither index yields a significant primary difference between states.

The RIS spread estimate pairs a positive winner coefficient with a negative loser coefficient. This pattern is compatible with a larger contribution from the loser leg, although the adjusted leg-specific tests do not confirm that interpretation. The regressions describe return associations; they do not measure short-sale constraints or identify the investors trading the portfolios. In the UP state, the winner and loser estimates are 0.680 and -1.805 percentage points per sentiment standard deviation, respectively. Their nominal p-values are about 0.077, while both adjusted values equal 0.8431. BW shows similar winner and loser slopes, so its difference is close to zero.

Table 1: Full-sample primary tests. MOM is the winner-minus-loser momentum return. Entries report the UP slope and DOWN - UP difference with pointwise 95% HAC intervals; adjusted p-values are reported separately for the two tests across the 12-test family

Index	Outcome	UP slope [95% CI]	DOWN - UP [95% CI]	Adj. p (UP)	Adj. p (Δ)
BW	MOM	-0.001 [-0.498, 0.496]	-0.016 [-1.850, 1.818]	1.0000	1.0000
BW	Winner	-0.130 [-0.384, 0.124]	-0.046 [-0.668, 0.576]	1.0000	1.0000
BW	Loser	-0.129 [-0.457, 0.199]	-0.030 [-1.424, 1.364]	1.0000	1.0000
RIS	MOM	2.486 [0.336, 4.635]	-0.885 [-3.559, 1.789]	0.2826	1.0000
RIS	Winner	0.680 [-0.073, 1.434]	-0.026 [-0.904, 0.853]	0.8431	1.0000
RIS	Loser	-1.805 [-3.805, 0.194]	0.859 [-1.428, 3.146]	0.8431	1.0000

Common-sample estimates changed the BW comparison. Its UP momentum slope became 0.582 percentage points, with a pointwise interval from -0.469 to 1.632. Figure 2 displays both indices on the same 323 outcome months, including 56 DOWN months. The common-sample BW DOWN momentum slope was positive and its pointwise interval excluded zero. This secondary result prevents a claim that every estimated relationship was null. It remains distinct from the primary full-sample family, and the two indices still represent shared US outcomes.

4.2 Forecast Comparisons and Stability

The interaction model did not achieve the lowest MSPE in either full sample. Table 2 compares all five models on matching forecast months within each index. BW generated 581 forecasts from August 1975 to December 2023; RIS generated 203 from February 2006 to December 2022. BW's interaction forecast improved squared error over the historical mean by 0.745%, compared with 1.050% for state only. For RIS,

the interaction yielded -3.615% out-of-sample R^2 , whereas sentiment only yielded 4.597% . These percentages describe loss reductions, not portfolio returns.

Figure 2: Common-sample sentiment slopes by market state. Points show return slopes per within-sample sentiment standard deviation; whiskers show unadjusted 95% HAC intervals. Both indices use 323 months, including 56 DOWN months. Regressions include state, lagged volatility, the sentiment-state interaction and contemporaneous three-factor controls; winner and loser returns are excess returns. These secondary intervals are descriptive; the primary Holm tests appear in Table 1

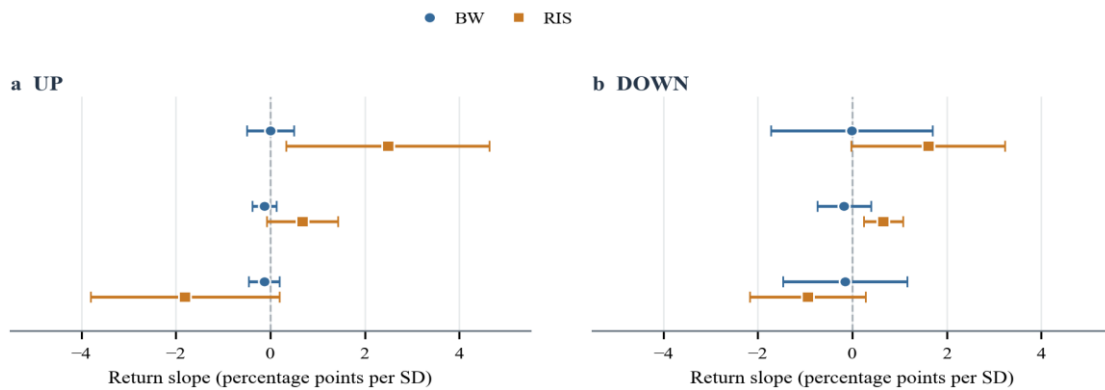


Table 2. Full-sample forecast performance. Lower MSPE and mean absolute error (MAE) are better; higher out-of-sample R^2 is better. MSPE is multiplied by 10^4 ; MAE and R^2 are percentages. BW and RIS contain 581 and 203 forecast months, respectively

Model	MSPE (BW)	MAE (BW)	R^2 (BW)	MSPE (RIS)	MAE (RIS)	R^2 (RIS)
Historical Mean	56.490	5.037	0.000	78.160	5.912	0.000
Sentiment Only	56.559	5.040	-0.122	74.567	5.876	4.597
Market State Only	55.897	5.086	1.050	76.775	5.977	1.772
Additive	55.984	5.091	0.895	75.128	5.942	3.879
State Interaction	56.069	5.092	0.745	80.985	6.086	-3.615

Figure 3 shows paired monthly squared-error differences and HAC uncertainty. BW's small improvement over the mean did not establish a general advantage over its comparators. RIS's interaction model had larger mean squared error than each alternative. For BW, the interaction-minus-state-only loss was 0.172 in squared percentage-point units, with a HAC interval from 0.026 to 0.318. Both HAC and bootstrap intervals favored state only in this full sample. Against the additive model, BW's HAC interval crossed zero while the bootstrap lower endpoint was slightly positive. This borderline disagreement does not establish an interaction benefit. RIS intervals crossed zero for every comparison, leaving the extent of its point-estimate disadvantage uncertain.

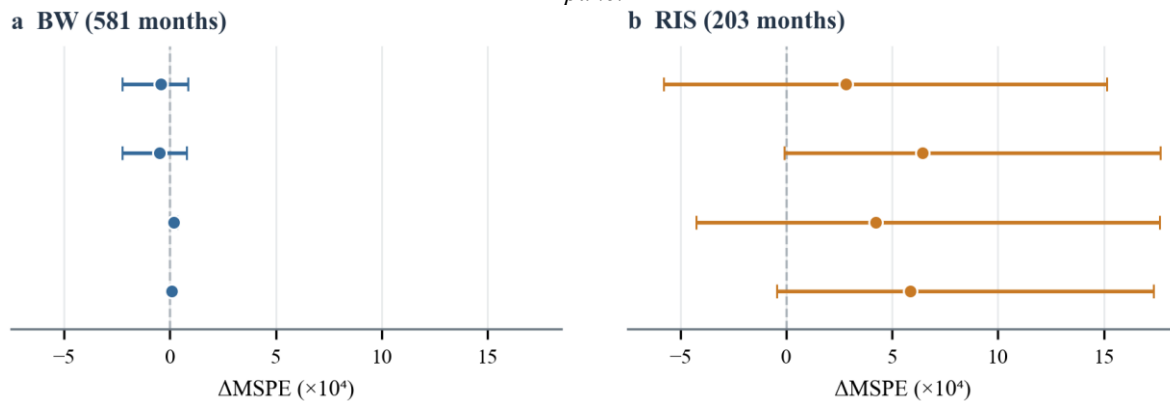
The full-sample index comparison also mixed measurement and calendar coverage. Restricting estimation to the common sample raised BW interaction R^2 to 2.695%, while RIS remained at -3.615% . BW's interaction then had the lowest MSPE among its five common-sample models, reversing its ranking against state only and additive forecasts. No separate paired-loss intervals were computed for this aligned comparison, so the ranking is descriptive. The change under the documented specification makes historical-window sensitivity visible without isolating its cause. Both predictors still forecast identical US outcomes rather than independent markets.

Predetermined periods revealed additional instability. BW's state-difference estimate was negative in 1966–1989 and positive in both later partitions; its 1990–2006 pointwise interval excluded zero. RIS's UP slope was larger in 2007–2022 than in its earlier available partition. These exploratory estimates are not formal structural-break tests. The saved sensitivity matrix covers 45 regression settings and 60 forecast cells, retaining unfavorable results. An additional precision check used five bootstrap streams at draw budgets from 200 to 2,000. It assessed Monte Carlo variation in interval endpoints without changing primary estimates, predictors or monthly forecasts. Deterministic least squares has no analogous training-seed variation.

The symmetric case selection also identified large losses outside a simple DOWN-state narrative. The three worst spread months were April 2009 at -45.18% , January 2001 at -41.84% and March 2009 at -39.19% . January 2001 was classified UP by the lagged market rule. The three best months were February 2000 at

25.34%, December 1999 at 22.22% and March 2020 at 21.33%. Both sentiment samples selected the same extremes because they share these return months. These observations illustrate the outcome distribution; they neither estimate crisis probabilities nor validate an investor-level explanation.

Figure 3: Incremental forecast loss of the interaction model. Points show its MSPE minus each comparator's MSPE; negative values favor the interaction. Whiskers are 95% HAC-six-lag intervals for paired monthly squared errors, using 581 BW and 203 RIS forecasts. All models begin with 120 training observations. Horizontal scales differ between panels. The revised source data support pseudo-out-of-sample comparison, with identical target months within each panel



5. Discussion and Limitations

The results lead to three behavioral-finance observations. First, the sentiment-momentum relation changes with the measure and the calendar: RIS gives the largest primary UP estimate, while BW moves more when the samples are aligned. Second, the winner/loser signs point toward a possible loser contribution, but multiplicity-adjusted tests do not identify a particular psychological mechanism. Third, adding the state interaction does not improve full-sample forecasts over the best simple comparator. Because no primary result survives the multiplicity adjustment, the slope magnitude remains uncertain; this does not make sentiment irrelevant or overturn the broader behavioral literature.

The portfolio-leg evidence sharpens the behavioral interpretation but does not uniquely identify the underlying mechanism. The RIS pattern is consistent with a larger loser-leg contribution, while portfolio-level returns cannot distinguish psychological responses from market frictions such as borrowing constraints or changing exposure. These constraints define the scope of the mechanism evidence rather than overturn the conditional-return results.

The forecasting results also define a clear boundary of the analysis. The interaction does not outperform the strongest simple benchmark in the full samples, and relatively few DOWN observations reduce the precision of state-dependent estimates. In addition, revised sentiment histories prevent a strict real-time interpretation. The study is therefore best viewed as a historical test of measurement sensitivity and forecast incrementality. Future work using vintage sentiment releases, independent markets and implementable portfolios with observed trading costs would provide stronger evidence on external validity and real-time applicability.

The use of revised histories limits direct real-time interpretation; vintage releases and independent-market tests are needed to assess external validity.

6. Conclusion

Public US portfolio data yield three qualified conclusions about sentiment and momentum across market states. First, the two sentiment measures produce different slope patterns, and no primary test survives Holm correction for the full test family. Second, portfolio-leg signs remain compatible with several behavioral mechanisms without identifying one of them. Third, the interaction forecast fails to outperform the strongest simple full-sample comparator for either index. The documented comparison records variation across

measurement, portfolio legs and time while separating explanatory evidence from forecast evaluation. Original-release data and independent-market tests remain necessary for stronger predictive claims.

References

- [1] Jegadeesh, N., and Titman, S., “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency,” *The Journal of Finance*, 48, (1), 65–91, 1993, <https://doi.org/10.1111/j.1540-6261.1993.tb04702.x>.
- [2] Cooper, M. J., Gutierrez, R. C., Jr., and Hameed, A., “Market States and Momentum,” *The Journal of Finance*, 59, (3), 1345–1365, 2004, <https://doi.org/10.1111/j.1540-6261.2004.00665.x>.
- [3] Antoniou, C., Doukas, J. A., and Subrahmanyam, A., “Cognitive Dissonance, Sentiment, and Momentum,” *Journal of Financial and Quantitative Analysis*, 48, (1), 245–275, 2013, <https://doi.org/10.1017/S0022109012000592>.
- [4] Hong, H., and Stein, J. C., “A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets,” *The Journal of Finance*, 54, (6), 2143–2184, 1999, <https://doi.org/10.1111/0022-1082.00184>.
- [5] Shleifer, A., and Vishny, R. W., “The Limits of Arbitrage,” *The Journal of Finance*, 52, (1), 35–55, 1997, <https://doi.org/10.1111/j.1540-6261.1997.tb03807.x>.
- [6] Baker, M., and Wurgler, J., “Investor Sentiment and the Cross-Section of Stock Returns,” *The Journal of Finance*, 61, (4), 1645–1680, 2006, <https://doi.org/10.1111/j.1540-6261.2006.00885.x>.
- [7] Baker, M., and Wurgler, J., “Investor Sentiment in the Stock Market,” *Journal of Economic Perspectives*, 21, (2), 129–151, 2007, <https://doi.org/10.1257/jep.21.2.129>.
- [8] Stambaugh, R. F., Yu, J., and Yuan, Y., “The Short of It: Investor Sentiment and Anomalies,” *Journal of Financial Economics*, 104, (2), 288–302, 2012, <https://doi.org/10.1016/j.jfineco.2011.12.001>.
- [9] Stambaugh, R. F., Yu, J., and Yuan, Y., “The Long of It: Odds That Investor Sentiment Spuriously Predicts Anomaly Returns,” *Journal of Financial Economics*, 114, (3), 613–619, 2014, <https://doi.org/10.1016/j.jfineco.2014.07.008>.
- [10] Daniel, K., and Moskowitz, T. J., “Momentum Crashes,” *Journal of Financial Economics*, 122, (2), 221–247, 2016, <https://doi.org/10.1016/j.jfineco.2015.12.002>.
- [11] Gao, X., Koedijk, K., Walther, T., and Wang, Z., “Relative Investor Sentiment,” *International Review of Economics & Finance*, 100, 104105, 2025, <https://doi.org/10.1016/j.iref.2025.104105>.
- [12] Aksoy-Yurdağul, D., Buchner, A., and Zareei, A., “The persistence of news sentiment: Implications for return predictability,” *Economics Letters*, 260, 112803, 2026, <https://doi.org/10.1016/j.econlet.2025.112803>.
- [13] Fama, E. F., and French, K. R., “Common Risk Factors in the Returns on Stocks and Bonds,” *Journal of Financial Economics*, 33, (1), 3–56, 1993, [https://doi.org/10.1016/0304-405X\(93\)90023-5](https://doi.org/10.1016/0304-405X(93)90023-5).
- [14] Fama, E. F., “Efficient Capital Markets: A Review of Theory and Empirical Work,” *The Journal of Finance*, 25, (2), 383–417, 1970, <https://doi.org/10.2307/2325486>.
- [15] French, K. R., *Detail for 10 Portfolios Formed Monthly on Momentum*, [Official portfolio documentation](#), 2026.
- [16] Newey, W. K., and West, K. D., “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, (3), 703–708, 1987, <https://doi.org/10.2307/1913610>.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

