

Revisiting Momentum through Dynamic Kernel Learning

Feng Dong

Assistant Professor of Finance,
Martha & Spencer Love School of Business, Elon University
Koury Business Center 252, Elon, NC 27244, USA
Tel:(336) 278-2159. fdong@elon.edu

Revisiting Momentum through Dynamic Kernel Learning

Abstract

This paper revisits the momentum phenomenon from the aspect of information diffusion and develops a dynamic kernel framework that links return continuation to the speed of market learning. Using nearly sixty years of U.S. equity data, I estimate nonnegative and geometric kernels that capture how past returns should be weighted over time. The results show that information absorption has increased dramatically, which explains the post-2000 decline in momentum profitability. The proposed model outperforms both traditional and machine-learning benchmarks, underlining the fact that structured and interpretable models remain powerful in finance. The framework offers both theoretical insight and practical instruments for detecting market efficiencies.

Keywords: Momentum, Information Diffusion; Kernel Learning; Machine Learning

JEL Classification: G12, G14, C32, C45

“Because some investors are slow to react to news, prices underreact in the short run and overreact in the long run.”
— Hong and Stein (1999)

1. Introduction

Momentum has long stood as one of the most robust and analyzed return anomalies in finance, yet its profitability has clearly weakened over the past few decades. Academic studies document that strategies based on buying past winners and selling past losers no longer generate the strong alphas once observed. At the same time, many mutual funds, hedge funds, and ETF products still incorporate momentum signals. Some entirely built on it, others using it as a key element of their strategy. For example, the largest momentum ETF in the U.S., iShares MSCI USA Momentum ETF, holds \$19.55 billion in assets under management now. These facts raise an important puzzle because momentum strategies are widely deployed and well understood, yet their profitability has faded.

In academia, over three decades of research have attempted to understand why momentum arises and why it persists. The early work of Jegadeesh and Titman (1993) revealed that stocks performing well over the previous year tend to continue outperforming in the near term, while past losers often keep underperforming. This discovery challenged the notion of fully efficient markets and encouraged a vast literature exploring the behavioral and informational foundations of return continuation. Behavioral explanations argue that investors react slowly to news, leading to gradual price adjustments that generate predictable trends. The information-diffusion perspective of Hong and Stein (1999) advanced this view, proposing that momentum reflects the staggered flow of information among heterogeneous traders. In contrast, rational risk-based interpretations suggest that momentum returns compensate for exposures to dynamic risk factors or liquidity shocks rather than market inefficiencies.

Despite these competing perspectives, one pattern is clear that momentum has weakened considerably in recent decades. Studies show that average long-short momentum returns have declined sharply since the early 2000s, with the effect often disappearing or even reversing during market downturns. Daniel and Moskowitz (2016) document severe losses during market reversals, while Barroso and Santa-Clara (2015) find that volatility-adjusted momentum strategies provide better stability but lower profitability. This deterioration poses a puzzle for both academics and

practitioners. If the forces behind momentum, whether behavioral biases or risk compensation, are structural, why has the anomaly lost strength over time?

I argue that this fading of momentum is not simply due to overfitting, factor crowding, or poor model choice. Instead, the decline reflects a deeper structural shift that markets today absorb information much more rapidly than before. The gradual diffusion of news and firm-specific signals that once created time-lagged responses is replaced by near-instantaneous assimilation. In other words, the decay of momentum payoffs is a symptom of accelerating information absorption, not merely the by-product of more complex models or competitive pressure.

At the same time, the rise of machine learning has introduced new tools for modeling asset returns. Advanced models such as Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost) can extract complex patterns from large datasets, uncovering nonlinear relationships that traditional econometric methods may overlook. However, these models often sacrifice explanatory power for flexibility. They can predict outcomes but rarely explain the underlying mechanism driving those predictions. The question of how past returns should be weighted across time, and whether the weighting itself evolves with shifts in information flow, remains largely unresolved. This gap leaves open a central research question that, can a dynamic and economically interpretable process of information absorption explain why momentum varies across periods and regimes? More specifically, can understanding how markets assign and adjust weights to past information help clarify the apparent decline of one of finance's most persistent anomalies?

To answer those questions, I begin with the full sample of U.S. common equities from 1965 to 2024 from the Compustat database. I first replicate the classic momentum portfolios under standard formation rules, applying price and exchange filters consistent with established studies to ensure comparability. Then, I employ a nonnegative least squares (NNLS) kernel to capture how past returns affect current price dynamics. This kernel represents the market's memory of prior information and quantifies the speed at which new information is absorbed into prices. To preserve economic explanation power, a geometric functional form is imposed, and a Kalman smoothing procedure is applied to track the kernel's dynamic evolution over time. Overall, the estimated decay parameter λ is interpreted as an empirical proxy for the market's information-absorption speed, a measurable counterpart to the gradual learning processes emphasized in Hong and Stein (1999).

The main analyses proceed through a rolling twenty-year estimation window to mimic real-time learning and to track how the kernel shape changes across decades. This rolling structure allows the model to detect shifts in information absorption speed and to examine whether markets have become faster in processing news. The model's output is then benchmarked against the standard twelve-month formation and one-month holding rule, to evaluate differences in predictive power and stability.

For comparative purposes, two machine learning models, LSTM networks and XGBoost, are implemented under identical out-of-sample evaluation settings. I choose these two methods because they represent the most widely used and conceptually distinct approaches in financial prediction. The LSTM network captures sequential dependence and temporal dynamics in return histories, making it suitable for modeling the evolving nature of momentum signals. XGBoost, on the other hand, is a more black-boxed and powerful gradient-based ensemble method that effectively models nonlinear cross-sectional relationships in large financial datasets. By maintaining consistent training windows, features, and evaluation periods across all models, the analysis isolates the role of structure and interpretability in driving predictive success. The overall design thus integrates long-horizon asset-pricing data, dynamic parameter estimation, and model benchmarking to investigate how evolving information absorption shapes the persistence and profitability of momentum.

The results reveal several important findings about how information diffusion shapes momentum dynamics. First, the estimated decay speeds of the kernel increase sharply over time, indicating that information now flows through markets more rapidly than in earlier decades. The half-life of information, which exceeded two years in the 1970s and 1980s, has shortened to just a few months in the post-2000 period. This acceleration aligns closely with the well-documented decline in the profitability of traditional 12–1 momentum portfolio. As markets absorb news faster, the persistence of return continuation naturally weakens, reducing the window in which past performance can predict future returns.

Second, the geometric kernel with Kalman smoothing produces superior and more stable results compared with the standard equal-weighted Jegadeesh and Titman (hereafter, JT) specification. The geometric and Kalman-filtered models adapt to changing market conditions while preserving interpretability and economic coherence. Out-of-sample tests show that these structured kernels

consistently match or outperform more complex machine learning benchmarks such as LSTM and XGBoost. This evidence suggests that incorporating a realistic economic structure into the model can capture the essential dynamics of information absorption more effectively than purely data-driven approaches.

Additionally, the estimated decay parameter itself emerges as a simple but powerful state variable that tracks the evolving speed of market learning. It varies systematically across time and firms, offering a concise summary measure of how quickly prices reflect new information and how that process relates to momentum payoffs.

Finally, firm-level characteristics provide further insight into the heterogeneity of information diffusion. Firms with more complex and less readable financial disclosures exhibit slower kernel decay and more back-loaded momentum, consistent with the idea that complexity delays investors' understanding of fundamental information. This relationship between disclosure clarity and return dynamics reinforces the broader link between corporate transparency and market efficiency.

These findings together yield several contributions to the literature. First of all, the paper reframes momentum as a measurable information diffusion process rather than a fixed behavioral bias or static risk premium. It formalizes the idea that prices evolve through a continuous process of information absorption, allowing the strength and timing of momentum to be characterized by the shape of the kernel. This framework unifies behavioral and informational perspectives within a single, empirically tractable structure. The study also provides long-horizon evidence that the market's information half-life has shortened substantially over the past six decades. This quantification of information speed helps explain the persistent decline in momentum profitability and links it to observable firm-level characteristics such as disclosure complexity. It further identifies a new empirical connection between the temporal pattern of return reactions and the evolving structure of financial communication. Finally, the paper shows that interpretable models can match or even outperform complex machine learning approaches. By applying economically grounded shape constraints, the kernel framework delivers predictive strength while maintaining transparency about its mechanism, helping clarify when machine learning contributes genuine value to financial prediction and when structural reasoning suffices.

The framework also has direct practical relevance. The estimated decay parameter can serve as a real-time diagnostic for evaluating the state of market efficiency and timing momentum exposures.

Asset managers can monitor the parameter to decide when momentum strategies are likely to perform well and when they are at risk of decay. The results further suggest that systematic strategies benefit from structured signals that reflect economic mechanisms rather than purely statistical correlations. From a policy perspective, the evidence that disclosure complexity slows information absorption implies that clearer corporate reporting can enhance market efficiency. Regulators and firms can therefore view transparency not only as a governance objective but also as a mechanism that improves price discovery. Overall, the results highlight that the decline of momentum reflects the evolution of markets themselves. As technology, data availability, and investor sophistication accelerate the speed of learning, traditional anomalies based on delayed reactions naturally diminish. Understanding and measuring the dynamics of information absorption provide a path forward for both academic research and practical investment strategy.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature, and Section 3 develops the mathematical framework. Section 4 describes the data, outlines the empirical design, and presents the main results and the additional robustness check. The final section concludes the study.

2. Literature Review

This section reviews the key research streams underlying this study, including classic momentum evidence, the information diffusion view of gradual price adjustment, and recent machine learning approaches to asset pricing. Together, these perspectives reveal the need to connect predictive performance with an economically interpretable mechanism, forming the foundation for this paper's contribution.

2.1. Momentum Baseline Literature

The momentum anomaly, which is buying past winners and selling past losers, was first systematically documented by Jegadeesh and Titman (1993), who showed that returns over the past three to twelve months predict future stock performance. This finding challenged the efficient market hypothesis and became one of the most influential results in empirical asset pricing. Carhart (1997) incorporated a momentum factor into his four-factor model and demonstrated that much of mutual fund performance persistence is due to momentum exposure rather than manager skill. Rouwenhorst (1998) extended the evidence to international markets, showing that the phenomenon is not unique to the United States. Momentum effects have since been observed across asset classes,

including commodities, currencies, and bonds, and over multiple time horizons, confirming their broad relevance in financial markets. Subsequent research has examined the sources and risks of momentum profits. Moskowitz and Grinblatt (1999) decomposed stock momentum into industry and firm-specific components, revealing that common industry effects explain a large portion of returns. Barroso and Santa-Clara (2015) showed that momentum returns are highly volatile and that volatility scaling can greatly improve risk-adjusted performance. Daniel and Moskowitz (2016) further documented that momentum crashes are predictable and occur during sharp market rebounds, suggesting conditional exposure to market risk. Together, these studies established that momentum profits are real but risky, with payoffs varying across market states.

More recent studies have refined how and when momentum delivers value. Huang (2022) introduced the idea of a “momentum gap,” showing that the difference between past winners and losers predicts future profits, which is evidence that momentum strength is conditional on market states. Baltussen et al. (2021) extended the discussion to shorter horizons, finding that intraday and weekly continuations meaningfully add to total performance, consistent with persistent trading-flow pressure. Tong et al. (2025) identified lead–lag patterns among correlated stocks that generate cross-sectional momentum, while Ehsani and Harvey (2023) demonstrated that factor-level momentum can subsume many traditional anomalies, pointing to a shared behavioral or risk-based mechanism.

A growing body of work shows that the traditional momentum premium has weakened in modern markets (e.g., Barroso and Santa-Clara 2015; Daniel and Moskowitz 2016; Jegadeesh and Titman 2023). Momentum returns have become more volatile and prone to reversals, especially during market rebounds, reflecting structural changes in trading technology, data speed, and investor behavior. These findings motivate the focus of this study on information-diffusion dynamics as a unifying explanation.

2.2. Information Diffusion View

Momentum can be naturally understood through the lens of information diffusion, where return continuation arises because market participants incorporate new information into prices at different speeds. Thus, the information diffusion perspective provides an intuitive behavioral foundation for momentum. It argues that prices adjust gradually to new information because investors process signals at different speeds. Early adopters trade first, while others respond later, creating a delayed

reaction that produces short-term continuation. This mechanism, originally formalized by Hong and Stein (1999), describes markets as composed of “news watchers” and “momentum traders.” As information diffuses slowly across these groups, prices temporarily underreact before eventually converging to fundamentals.

Subsequently, Chordia and Shivakumar (2002) and Hong et al. (2000) show that momentum is strongest when information uncertainty is high and investor attention is limited. Zhang (2006) and Peng and Xiong (2006) extend this idea, linking gradual information flow to investor inattention and limited information capacity. Empirically, Hou and Moskowitz (2005) and Hirshleifer and Teoh (2003) find that stocks with slower analyst coverage, less media attention, or weaker information environments exhibit stronger momentum, consistent with delayed price discovery. More recent studies connect this behavioral diffusion to modern data settings. Hirshleifer, Lim, and Teoh (2021) show that textual and sentiment-based information also diffuses gradually through media and analyst channels. Ehsani and Harvey (2023) document that even with advanced analytics and real-time trading, prices continue to reflect staggered information absorption, suggesting that diffusion remains a relevant force in modern markets.

From a structural standpoint, diffusion-based explanations align with the observed decline in the traditional 12–1 momentum strategy. As information flows more efficiently through faster news dissemination, high-frequency algorithms, and global connectivity, the window for delayed reaction narrows. This evolution explains why recent evidence, including Jegadeesh and Titman (2023), finds diminished profitability in the U.S. while momentum remains stronger in markets where information frictions persist.

2.3. Machine Learning in Asset Pricing

Machine learning has become a major force in modern asset pricing, offering flexible tools to uncover nonlinear and high-dimensional relationships among firm characteristics, macro variables, and returns. Studies such as Gu, Kelly, and Xiu (2020) demonstrate that tree-based and neural network models can dramatically improve in-sample fit and predictive accuracy compared with traditional linear regressions. Subsequent research by Chen, Pelger, and Zhu (2021) and Bryzgalova, Pelger, and Zhu (2022) formalizes these approaches into interpretable factor structures, showing that ML can extract latent pricing kernels and automate the discovery of return-

relevant factors. These methods mark a shift from ex-ante factor construction toward data-driven discovery, broadening the empirical foundations of asset pricing.

Within this literature, momentum features continue to play a central role. Kozak, Nagel, and Santosh (2020) argue that many machine learning factors are reconfigurations of existing economic signals, especially momentum and value, rather than entirely new sources of risk. Similarly, Kelly, Pruitt, and Su (2019) show that deep learning models often redistribute emphasis across familiar predictors rather than uncovering truly novel dimensions. A growing wave of empirical work supports this conclusion. Once traditional momentum variables are carefully constructed, standardized, and regularized, nonlinear algorithms rarely deliver superior out-of-sample performance. Studies by Feng et al. (2024) and Gu, Kelly, and Xiu (2021) show that machine learning models tend to enhance and fine-tune existing predictive signals rather than uncover entirely new sources of return variation. This evidence supports the view that momentum primarily reflects an information-diffusion mechanism that can be effectively captured by simple, well-structured functional forms. Once this underlying process is properly modeled, the additional complexity of modern techniques contributes little incremental economic value.

Despite extensive work on momentum profitability, information diffusion, and machine learning applications, the literature still lacks a clear understanding of how past returns should be optimally weighted over time. Most studies rely on fixed formation–holding structures or treat weighting as a byproduct of model estimation, leaving the temporal shape of information absorption unexamined. This gap limits our ability to connect momentum’s predictive power with the underlying process of how markets gradually incorporate information.

3. Math Proof

This section develops the mathematical foundation linking momentum returns to gradual information diffusion. Specifically, I model expected excess returns as a weighted sum of past shocks, where the weighting function represents the market’s speed of information absorption.

If a firm absorbs news slowly, today’s good news keeps influencing future returns. Mathematically, that’s a positive tail in the information-absorption kernel, which creates positive autocovariances and a positive mapping from past-return signals to expected next return. Since the tail differs across firms and time, momentum differs across firms and time.

Let firm i 's "news" shocks, $s_{i,t}$, be independent, mean-zero with (possibly) time-varying variance, $\sigma_{i,t}^2$. Prices do not fully incorporate the shocks immediately; instead, the excess return is a causal linear filter of past shocks:

$$r_{i,t} = \sum_{\tau=0}^{\infty} K_{i,t}(\tau) s_{i,t-\tau} + \varepsilon_{i,t}, \quad K_{i,t}(\tau) \geq 0, \sum_{\tau} K_{i,t}(\tau) < \infty,$$

where $K_{i,t}(\cdot)$ is the time-weighted information-absorption kernel, and $\varepsilon_{i,t}$ is idiosyncratic microstructure noise (martingale difference, orthogonal to s). This representation can be viewed as a continuous-time counterpart to the gradual-information-diffusion frameworks of Hong and Stein (1999), Peng and Xiong (2006), and Barberis et al. (2015). In those models, investors process news at heterogeneous speeds, leading to partial and staggered price adjustments. The information-absorption kernel in this paper formalizes that idea by explicitly weighting how past shocks continue to influence current returns. When the kernel places all mass immediately ($\tau=0$), prices adjust instantaneously. When the kernel has a long right tail, information diffuses slowly, producing positive short-term autocorrelation consistent with gradual learning. Hence, the kernel serves as an empirical modeling of the theoretical diffusion process proposed in those behavioral and informational models. Based on this setting, a surprised stock information shock, $s_{i,t}$, lift stock returns today and for some time in the future, with time weights following K . When all the information is absorbed immediately, $\tau = 0$; and when the information needs a long time to be absorbed, K has heavier tails.

Meanwhile, assuming cross terms vanish by independence/orthogonality, we show the autocovariances of returns as:

$$\gamma_{i,t}(h) = \text{Cov}(r_{i,t}, r_{i,t-h}) = \sigma_{i,t}^2 \sum_{j=0}^{\infty} K_{i,t}(j) K_{i,t}(j+h), \quad h \geq 1,$$

Additionally, a generic momentum signal (e.g., J -month, skipping the most recent month) with nonnegative window weights w_h should be:

$$M_{i,t}^J = \sum_{h=2}^J w_h r_{i,t-h}$$

Proposition 1: momentum strength is a functional of the kernel

The predictive content of a past returns signal for next month's return is:

$$Cov(r_{i,t+1}, M_{i,t}^J) = \sigma_{i,t}^2 \sum_{h=2}^J w_h \sum_{j=0}^{\infty} K_{i,t}(j) K_{i,t}(j+1+h)$$

In particular, if $K_{i,t}(\tau) \geq 0$ and puts positive mass beyond $\tau = 0$ (slow absorption), then $Cov(r_{i,t+1}, M_{i,t}^J) > 0$. As the shape and tail of $K_{i,t}$ change cross firms or time, the covariance and the momentum must vary. This implicates that two firms with different kernels $K_{i,t}$ will have different $Cov(r_{i,t+1}, M_{i,t})$ and hence different momentum even if their shock variances are the same. Over time, if a firm's kernel changes (such as coverage improves, liquidity changes, disclosure quality changes), its momentum varies over time.

Proposition 2: momentum is zero when absorption is instant, and strongest at intermediate speeds

Suppose returns follow a standard partial-adjustment (geometric kernel) to news:

$$K_{i,t}(\tau) = \lambda_{i,t} (1 - \lambda_{i,t})^{\tau}, \lambda_{i,t} \in (0, 1]$$

where λ is the speed of information absorption, which measures the share of a news shock that prices incorporate immediately in the current month. Then, the lag-1 autocovariance is:

$$\gamma_{i,t}(1) = \sigma_{i,t}^2 \sum_{j=0}^{\infty} \lambda^2 (1 - \lambda)^{2j+1} = \sigma_{i,t}^2 \frac{\lambda_{i,t} (1 - \lambda_{i,t})}{2 - \lambda_{i,t}} > 0$$

and more generally $\gamma_{i,t}(h) > 0$ for all $h \geq 1$. When $\lambda_{i,t}$ is close to 1, $1 - \lambda_{i,t}$ is close to 0, thus $\gamma_{i,t}(h)$ is close to 0; while when $\lambda_{i,t}$ is close to 0, $\gamma_{i,t}(h)$ is also close to 0. This means that if news is absorbed immediately, there's no leftover for future months, which means no momentum. On the other hand, if news is absorbed too slowly, each month is too small to correlate with the next, which leads momentum to fade. Thus, the momentum should be the strongest when the information absorbing speed is intermediate. This indicates that If firms (or periods) differ in $\lambda_{i,t}$ due to information frictions (coverage, readability, liquidity), their momentum strength differs systematically. Numerically, $\lambda = 0.05$ implies a half-life of roughly 14 months, whereas $\lambda = 0.3$ shortens it to about 2 months. Thus, a rise in λ from 0.1 to 0.4 cuts the effective memory of news by about 75%, reducing predictable continuation accordingly. When $\lambda > 0.6$, autocovariances turn negative, consistent with short-horizon reversals.

Proposition 3: if the conditional mean is increasing in the signal, decile spreads are positive

Let next-month excess return $r_{i,t+1}$ and past returns $\mathbf{r}_{i,t}^{\text{past}} = (r_{i,t-2}, \dots, r_{i,t-J})$ have finite second moments. Define a nonnegative-weight momentum signal:

$$S_{i,t} = \sum_{h=2}^J w_h r_{i,t-h}, w_h \geq 0, \sum_h w_h > 0$$

Assume $\text{Cov}(r_{i,t+1}, S_{i,t}) > 0$ (This holds in our kernel model whenever the kernel has positive tail mass). The conditional expectation for next month's return is increasing in the signal:

$$\mathbf{E}[r_{i,t+1} \mid S_{i,t} = s] = \mu_0 + \beta_s, \quad \beta = \frac{\text{Cov}(r_{i,t+1}, S_{i,t})}{\text{Var}(S_{i,t})} > 0$$

Under joint ellipticity, $\mathbf{E}[r_{i,t+1} \mid \mathbf{r}_{i,t}^{\text{past}}]$ is linear in $\mathbf{r}_{i,t}^{\text{past}}$. Any nonnegative linear functional $S_{i,t} = \mathbf{w}^\top \mathbf{r}_{i,t}^{\text{past}}$ induces a linear conditional mean in $S_{i,t}$:

$$\mathbf{E}[r_{i,t+1} \mid S_{i,t}] = \mu_0 + \beta S_{i,t}, \quad \beta = \frac{\text{Cov}(r_{i,t+1}, S_{i,t})}{\text{Var}(S_{i,t})} > 0$$

Since $\text{Cov}(r_{i,t+1}, S_{i,t}) > 0, \beta > 0$. Hence the conditional mean is strictly increasing in $S_{i,t}$. This means that any standard momentum score built with nonnegative past-return weights is a valid ordering statistic: higher signal indicates higher expected next-month return.

The mathematical proof demonstrates that momentum profits arise naturally from a gradual information diffusion process. By expressing expected returns as a weighted sum of past information, it shows that when investors incorporate information with delay, prices exhibit short-term continuation. The shape of the weighting kernel determines the strength and persistence of momentum.

4. Analysis and Results

This section empirically evaluates the dynamic-kernel momentum hypothesis. I begin by describing the data and replicating the baseline JT momentum strategy, then estimate nonnegative kernel weights, test the geometric specification, and compare them with complex machine-learning benchmarks. Finally, I assess robustness using firm-level complexity.

4.1. Data and Baseline Momentum

The data used for this analysis come from the CRSP monthly stock file accessed through WRDS, which provides comprehensive coverage of all U.S. equities. The sample includes ordinary common shares (share codes 10 and 11) listed on the NYSE, AMEX, and NASDAQ (exchange codes 1, 2, and 3). To ensure consistency with the original Jegadeesh and Titman (1993) study, I restrict the sample to securities meeting these definitions and exclude preferred shares, ADRs, REITs, and other non-comparable issues. All prices are converted to absolute values, and scaling adjustments are made when necessary (for example, dividing prices by 100 when reported in cents and shares by 1,000 when reported in raw units). Returns are cleaned and combined with delisting adjustments.¹ Market equity is computed as the product of price and shares outstanding, with missing shares replaced by the previous month's value when available. Each month, I determine the NYSE 10th percentile of market equity to define a size screen that removes microcap stocks. Firms below this breakpoint are excluded from portfolio formation, ensuring comparability across time and preventing extreme small-cap effects.² In addition, a price filter of \$5 at the formation month is applied to eliminate low-priced, illiquid stocks that tend to distort returns.

Following the methodology of Jegadeesh and Titman (1993), I sort all eligible stocks into ten deciles based on their past cumulative returns, where P1 represents the portfolio of past losers and P10 represents the portfolio of past winners, and construct a hypothetical winner–loser (P10–P1) portfolio to measure momentum profits. The first method, J12 K1, ranks stocks by their past 12-month returns (skipping the most recent month) and holds the resulting portfolios for one month. The second method, J6 K6, uses a 6-month formation period (skipping the most recent month) and holds portfolios for six months using overlapping cohorts, which smooths short-term noise and better reflects sustained continuation patterns. Decile portfolios are formed monthly on an equal-weight basis. For each specification, I report mean monthly returns and Fama–French three-factor alphas to assess whether momentum profits persist after controlling for market, size, and value factors.

The analysis spans January 1965 to December 2024, covering nearly six decades of U.S. equity data. To evaluate potential regime shifts in momentum performance, I further divide the full sample

¹ If both the regular return (ret) and the delisting return (dlret) are available, total returns are computed as $(1+ret) \cdot (1+dlret) - 1$. When only one of the two exists, the missing value is set to zero to retain the delisting effect, and observations with both missing are dropped.

² For robustness, I verify that results remain when using a 20 % NYSE size cutoff instead of the 10 % baseline; all key inferences are unchanged.

into three subperiods, 1965–1989, 1990–2009, and 2010–2024, corresponding to the periods studied in Jegadeesh and Titman (1993, 2001, 2011, and 2023). This segmentation facilitates a direct comparison with prior literature and allows for examination of structural changes in profitability over time. The results of the baseline momentum replications are reported in Table 1.

[Insert Table 1 here]

In Panel A of Table 1, both designs display clear return monotonicity from P1 (past losers) to P10 (past winners) over the full sample (1965–2024). The winner–loser spread is economically large and statistically significant in the aggregate (J12K1: 1.05%, $t=3.29$; J6K6: 0.78%, $t=3.30$). By regime, momentum is strongest pre-1990 (J12K1: 1.84%, $t=4.94$; J6K6: 1.23%, $t=4.54$), weakens in 1990–2009, and remains modest in 2010–2024 (J12K1: 0.61%, $t=0.92$; J6K6: 0.54%, $t=0.99$).

Risk-adjusted results (Panel B) reinforce these conclusions. The monthly Fama–French three-factor alphas remain significant in the full sample (J12K1: 1.38%, $t=4.67$; J6K6: 0.94%, $t=4.15$) and are particularly strong pre-1990 (J12K1: 2.11%, $t=7.20$; J6K6: 1.42%, $t=5.85$). Post-1990, alphas decline and lose conventional significance, especially after 2010 (e.g., J12K1: 0.75%, $t=1.27$; J6K6: 0.58%, $t=1.30$).

Overall, the replication confirms a strong and persistent momentum premium over the long horizon, particularly before 1990, but reveals a marked decline in profitability in later decades, especially after 2010. This pattern aligns with prior evidence that the classic JT momentum strategy has lost effectiveness in recent years (e.g., Barroso and Santa-Clara, 2015; Daniel and Moskowitz, 2016; Jegadeesh and Titman, 2023; Mamais et al., 2025).

4.2. NNLS Estimation and Empirical Interpretation

Having established the baseline momentum pattern, I next allow the data to reveal the optimal temporal weighting of past returns using an NNLS approach. I choose NNLS as the first enhancement to the classic JT momentum framework because it offers a transparent, data-driven way to estimate the optimal shape of the return-weighting kernel without sacrificing economic intuition. The original JT model implicitly assumes equal weighting of the past monthly returns when ranking stocks, which may no longer reflect how information diffuses in modern markets. NNLS, on the contrary, introduces disciplined flexibility by letting the data determine the relative importance of past returns while constraining weights to be non-negative and monotonically

decreasing. This ensures that more recent information exerts stronger influence, aligning with the economic intuition that older signals should fade rather than reverse in effect. NNLS has been widely used in applied statistics and signal processing to obtain sparse, stable, and interpretable estimates (Lawson and Hanson, 1974; Bro and De Jong, 1997). In finance, similar regularized regression approaches have been applied to portfolio construction and factor timing (DeMiguel et al., 2009; Gu, Kelly, and Xiu, 2020). Its convexity guarantees a unique global optimum, making it an ideal bridge between traditional econometric estimation and modern machine learning techniques.

In my implementation, I estimate the kernel weights by solving a quadratic program that minimizes the squared error between the predicted and realized one-month-ahead returns, penalized by a curvature term to enforce smoothness. Specifically, the objective function is

$$\min_w \|Xw - y\|^2 + \lambda \|D_2 w\|^2, \quad w_i \geq 0 \text{ and } w_i \geq w_{i+1}$$

where X contains lagged excess returns ($\tau = 2 \dots 12$); y is the next month realized return, and w denotes the kernel weights on lagged returns. The non-negativity and monotonicity constraints ensure economic coherence, while the second-difference penalty $\|D_2 w\|^2$ smooths sharp changes across lags.

Additionally, I estimate two variants of the NNLS method. The non-rolling model fits a single kernel using data from 1965 to 1989 and applies it out-of-sample from 1990–2024. This version reflects a static historical learning framework akin to how the JT model fixes its weighting scheme. The rolling model, by contrast, re-estimates the weights each month, using a 20-year moving training window that shifts forward one month at a time. This design ensures the kernel evolves smoothly with the data, allowing it to adapt continuously to changing market conditions while retaining long-term structural stability. Both models apply the same \$5 price filter and NYSE 10th-percentile size exclusion at formation to maintain direct comparability with the baseline momentum portfolios. The performance of the NNLS-based momentum strategies, along with their corresponding baseline counterparts, is reported in Table 2.

[Insert Table 2 here]

Table 2 Panel A shows modest raw momentum under the benchmark ($P10-P1 = 0.53\%$ per month; $t=1.14$), with a risk-adjusted spread that is only marginally significant ($\alpha = 0.86\%$; $t=1.99$).

By contrast, NNLS improves both magnitude and statistical reliability. The no-rolling variant delivers a 0.82% monthly spread ($t=1.90$) and an alpha of 0.98% ($t=2.43$); the rolling variant yields 0.86% ($t=2.02$) and 0.96% ($t=2.41$), respectively.

I also approximate the NNLS-estimated kernel using a smooth two-component exponential-mixture function³ to summarize its shape. Panel B, derived from this fitted mixture, provides insight into the mechanism underlying the performance improvement. The rolling NNLS kernel allocates nearly all of its weight to returns within the most recent 12 months ($\text{Mass} \leq 12m = 100\%$), indicating that short-term information dominates the predictive signal, and the influence of past returns declines rapidly ($\text{Eff } \theta = 6.61$), which is consistent with faster information diffusion. The high correlation between the estimated and fitted kernel shapes ($\text{corr} = 0.94$), along with the strong goodness of fit ($R^2 = 0.87$), suggests that the estimated weighting structure is smooth and economically coherent.

Figure 1 illustrates the estimated rolling NNLS kernel and its fitted exponential-mixture curve. As shown by the estimated kernel, recent lags have the largest weights, and the weight declines rapidly thereafter, with almost no weight beyond one year. This shape reflects a fast-decaying memory structure, meaning that recent returns carry much stronger predictive power than older ones. The close alignment between the empirical and fitted curves indicates that the estimated NNLS weights follow a well-behaved exponential decay pattern, consistent with an efficient, short-horizon information diffusion process in modern equity markets.

[Insert Figure 1 here]

In contrast, the no-rolling kernel places less weight on recent lags ($\text{Mass} \leq 12m = 38.93\%$) and exhibits a much longer memory ($\text{Eff } \theta = 24.32$), implying slower decay and higher sensitivity to outdated signals. This difference explains the superior performance of the rolling specification that by updating the kernel each month, the model adapts to the faster pace of information assimilation in modern markets, while the static kernel remains anchored to outdated dynamics. The results support the notion that adaptive weighting, emphasizing recent information and controlled smoothness, more effectively captures the evolving structure of return continuation.

³ $\hat{w}(\tau) = \sum_{m=1}^2 \alpha_m e^{-\tau/\theta_m}$, where each component m has its own decay constant θ_m (hence a distinct half-life) and weight $\alpha_m \geq 0$ (optionally normalized).

In sum, NNLS delivers economically meaningful and statistically stronger momentum than the benchmark, and the rolling design provides the most reliable risk-adjusted gains. These results support the view that momentum profitability depends critically on how past returns are weighted, and that an adaptive, shape-constrained kernel better captures contemporary diffusion dynamics.

4.3. Geometric Kernel and Comparative Fit

The NNLS results suggest a front-loaded, exponentially decaying structure. To capture this a more compact way, I next impose a geometric (i.e., Exponentially Weighted Moving Average) kernel because it delivers the core front-loaded, monotone decay shape uncovered by NNLS, while reducing estimation variance and turnover. The kernel assigns weight $w_\tau = \lambda(1-\lambda)^\tau$ to the return τ months before formation (skip the most recent month), with a single parameter $\lambda \in (0,1)$ indicating decay. Relative to NNLS, which estimates many weights simultaneously and can be weakly identified in small samples, the geometric kernel concentrates flexibility into one interpretable degree of freedom (the half-life). Theoretically, this method improves out-of-sample stability, limits noise-induced oscillations across lags, and yields cleaner economic interpretation (i.e., λ), which directly maps to the market's information-absorption speed (Hyndman & Athanasopoulos, 2018).

I implement two versions of the geometric kernel method under identical formation rules (common stocks on NYSE/AMEX/NASDAQ, NYSE p10 size exclusion, \$5 price screen). Each month, the model estimates a decay parameter using the previous 20 years data to capture how quickly past returns lose relevance. The selected parameter, λ , defines how recent returns are weighted. Using these optimized weights, stocks are ranked by their weighted past twelve-month performance (skip the most recent month) and grouped into ten deciles from past losers to past winners. Portfolios are then held for one month, allowing the model to evaluate how effectively the estimated decay captures short-term momentum in real time.

To assess robustness and allow for gradual time variation in decay rather than abrupt shifts, I also implement a Kalman-filtered version of the geometric kernel (Kalman, 1960; Harvey, 1989).⁴ The filtered sequence produces a smoother, more coherent path of decay rates that reflects persistent

⁴ The Kalman filter first allows $z_t = \text{logit}(\lambda_t)$ following a random walk with state noise variance equal 0.02. The observed proxy is $y_t = \text{logit}(\tilde{\lambda}_t)$ with observation noise variance equal 0.05. Initialization uses 0 as z_0 and 1 as variance. Then the method runs the standard predict-update recursion each month and apply the fixed-interval smoother to obtain $\tilde{z}_t$, which is mapped back via $\tilde{\lambda}_t = \text{logit}^{-1}(\tilde{z}_t)$.

market dynamics rather than monthly noise. This approach ensures temporal consistency in the weighting structure and accommodates slow-moving changes in information diffusion across market regimes.

The results from both the geometric kernel and the geometric kernel + Kalman models are summarized in Table 3. Panel A reports monthly returns and Fama–French three-factor alphas for ten equal-weighted momentum portfolios (P1–P10) ranked by past returns, as well as the winner-minus-loser (P10–P1) hypothetical portfolio with corresponding t-statistics. Panel B presents the estimated kernel speed parameters, including mean, median, and half-life values across three subperiods, providing a regime-level summary of how information-decay dynamics evolve over time.

[Insert Table 3 here]

The results reported in Table 3 highlight that the geometric kernel frameworks outperform both the baseline and NNLS models, particularly in the post-1990 period. Panel A shows that the geometric specification achieves a monthly winner-minus-loser spread of about 1.26 percent, compared with 0.53 percent under the traditional benchmark and 0.82 percent under the NNLS models. The corresponding Fama–French alphas, around one percent, confirm that these excess returns are not explained by standard risk factors. The Kalman-filtered version yields slightly smaller spreads but maintains comparable alphas, indicating improved stability without sacrificing performance.

Panel B illustrates how the kernel’s decay speed evolves over time. Both geometric models show an accelerating rate of information decay, with the half-life falling from roughly thirty months in the 1990s to less than three months after 2010. The Kalman variant produces smoother but faster-responding adjustments, with an average half-life close to one month in the modern period. This rapid decay is consistent with the view that markets now absorb information more quickly, leaving less room for gradual price adjustment and thereby eroding traditional momentum profitability (Jegadeesh and Titman, 2023). These results complement the NNLS evidence that recent momentum returns are primarily driven by short-term information signals, while the geometric specification captures this accelerated dynamic more efficiently. Taken together, the results

indicate that the geometric weighting structure, particularly when combined with Kalman smoothing, offers a consistent and adaptive improvement over earlier models.^{5,6}

To further assess how faster information absorption affects momentum profitability, I examine the co-movement between the baseline JT momentum payoffs and the estimated kernel decay speed. The payoff is estimated as the winner-minus-loser (W–L) portfolio return from the classical 12–1 momentum strategy as the dependent variable, and the estimated kernel decay parameter (λ) is derived from the geometric kernel + Kalman filter model, which captures the monthly rate at which past returns lose predictive relevance. A higher λ indicates a faster decay and thus a market environment where prices incorporate information more quickly. Figure 2 plots the 12-month rolling averages of those two. Periods of high λ , which indicates rapid information absorption, coincide with diminished or negative momentum returns, particularly after 2000 and again around 2020–2024. Conversely, when λ declines, momentum profitability tends to recover. An increase in λ from 0.1 to 0.4 reduces the implied half-life of news from roughly 12 months to 3 months, consistent with the profitability fading of JT momentum strategy since 2000.

[Insert Figure 2 here]

I also conduct a time-series regression that links monthly momentum payoffs to the estimated kernel decay speed and market conditions. I include several market-level control variables, which as shown in the previous literature, can influence the profitability of momentum strategy. *Dispersion* measures the cross-sectional variation in individual stock returns each month and reflects the degree of disagreement or heterogeneity among investors (Hong and Stein, 1999; Daniel and Moskowitz, 2016). *Vol_3m* represents the three-month rolling standard deviation of market returns, serving as a proxy for short-term risk and uncertainty (Cooper et al., 2004; Daniel and Moskowitz, 2016). *News_frac* captures the proportion of firm-months with earnings announcements, which gauges the intensity of information arrival in the market (Da et al., 2014). *Crowd_12m* measures momentum crowding intensity over the prior year, constructed as the rolling

⁵ To verify that these gains are not purely theoretical, I apply a realistic transaction-cost adjustment of 50 basis points per side to all monthly rebalancing trades. The resulting net alpha for the Geometric + Kalman model declines from 1.06% to 0.81% per month, and the t-statistic remains above 2. Hence, the improvement in performance survives reasonable trading-cost assumptions and is economically implementable.

⁶ I also test alternative formation–holding combinations, including J6K1 and J9K3 structures, which produce comparable patterns of decay and profitability. In each case, faster estimated kernel speeds coincide with weaker continuation, confirming that the information-diffusion interpretation does not depend on a single horizon choice.

correlation between fund holdings and the prevailing momentum factor, representing potential overexposure to the same trade (Lou and Polk, 2013). The regression is estimated for the full sample from 1990 to 2024 and for three subperiods (1990–2000, 2001–2010, and 2011–2024) to examine structural changes over time. The results are reported in Table 4.

[Insert Table 4 here]

Table 4 shows a clear and robust link between faster decay and weaker momentum. The decay parameter is negative and significant in the full sample and in most subperiods, with the largest magnitudes in 1990–2000. This pattern confirms that when past returns lose relevance quickly, the subsequent JT W–L payoff is compressed.

Among all the control variable, *Dispersion* switches sign across regimes: it is strongly positive in 1990–2000 but turns significantly negative in 2001–2010. In the 1990s, dispersion captured gradual information diffusion and stronger continuation among differing investor beliefs, supporting momentum. In contrast, during 2001–2010, high dispersion reflected distress, volatility, and reversals, which are linked to momentum crashes. Other controls, including volatility, news flow, and crowding, behave as expected, with news supporting return continuation and crowding compressing profitability. Overall, the results indicate that, as the kernel speeds up, momentum profits fall, reinforcing the interpretation that accelerated information absorption has eroded the profitability of the traditional strategy.

4.4. Machine-Learning Benchmarks

To verify that the decline in momentum profitability is structural rather than a result of model simplicity, I benchmark the kernel models against two advanced machine learning approaches: LSTM and XGBoost.⁷ LSTM networks are designed to capture sequential dependencies and long-horizon patterns in time series, making them suitable for testing whether nonlinear temporal memory can restore predictive power. XGBoost, in contrast, efficiently models nonlinear and interaction effects in large cross-sectional datasets, serving as a strong nonparametric benchmark. These two models represent distinct paradigms, temporal deep learning and gradient-based

⁷ To ensure comparability, each ML model uses identical 20-year rolling windows, early stopping after five no-improvement epochs, and 5-fold cross-validation. The LSTM uses one hidden layer of 64 units with dropout 0.2, and XGBoost employs a learning rate 0.05, max depth 5, and 100 trees. All ML hyperparameters were selected via 5-fold cross-validation within each rolling window.

ensemble learning, both widely used in modern financial prediction (e.g., Fischer and Krauss, 2018; Chen and Guestrin, 2016; Krauss, Do, and Huck, 2017; Gu, Kelly, and Xiu, 2020). By applying them to the same input and evaluation design as the kernel models, I assess whether model sophistication materially enhances out-of-sample performance. If not, the results support the interpretation that faster information absorption, not functional underfitting, explains the modern erosion of momentum profits.

4.4.1. LSTM

First, I use LSTM networks, which are designed to capture nonlinear temporal dependencies and long-horizon effects in time series (Hochreiter and Schmidhuber, 1997; Fischer and Krauss, 2018), to evaluate a sequence-learning approach. LSTM is an appropriate benchmark here because, if richer temporal structure exists beyond simple decay, a recurrent model should exploit it; failure to do so would support the interpretation that diminished momentum reflects faster information absorption rather than functional underfitting. I estimate two specifications under the same formation screens and out-of-sample protocol as the kernel models. The first one is a baseline LSTM with a nonnegative attention layer over lags and softmax normalization, trained on rolling 20-year windows with early stopping and weight decay, using prior 12-month returns (skip the most recent month) as inputs. Alternatively, I extend the baseline LSTM by incorporating a Kalman filter to smooth the estimated attention weights over time. The Kalman component treats the attention profile as a latent state evolving slowly through time, thereby enforcing temporal coherence while allowing gradual adaptation to market changes. The results of the LSTM models are presented in Table 5.

[Insert Table 5 here]

Table 5 indicates that the LSTM models produce economically meaningful but clearly bounded gains relative to the geometric kernel benchmark. In Panel A, the standard LSTM delivers a modest winner–minus–loser spread of 0.74% per month with a monthly alpha of 0.96% (t-value=2.23). Adding Kalman smoothing increases the spread to 0.98% with a larger three-factor alpha (1.3%) and stronger t-statistics (2.64), reflecting reduced noise in month-to-month signals and more stable lag emphasis.

Panel B confirms the ranking on out-of-sample risk–return. The LSTM plus Kalman raises the Sharpe ratio from roughly 0.27 to 0.37. Nonetheless, the geometric kernel combined with Kalman

remains ahead on mean spread, volatility control, and overall Sharpe, posting the highest t-value for alpha. In sum, these results show that while sequence models benefit from temporal smoothing, they do not surpass a simple and economically structured kernel. The evidence supports the mechanism view that rapid information absorption compresses exploitable persistence, leaving limited room for additional nonlinear extraction beyond well-specified, fast-decay weighting.

4.4.2. *XGBoost*

After the LSTM analysis, I turn to a more “black-boxed” but powerful machine-learning model, XGBoost, to test whether a non-sequential, cross-sectional approach can uncover additional return predictability that the structured kernel and sequence models may have missed. XGBoost has become a widely adopted benchmark in financial prediction because of its computational efficiency, robustness to noise, and ability to model complex nonlinear relationships in tabular data (Chen and Guestrin, 2016; Gu, Kelly, and Xiu, 2020; Krauss, Do, and Huck, 2017). While LSTM explicitly models time dependencies, XGBoost captures flexible nonlinear interactions among lagged returns and auxiliary predictors without assuming temporal order. This comparison helps determine whether the limits of momentum forecasting arise from modeling sophistication or from the inherent decline in information persistence.

The model follows the same 20-year rolling design as in the kernel and LSTM analyses, using skip-1 return lags (months 2 to 12). Each month, the model is retrained using a pairwise ranking objective to align predictions with cross-sectional ordering rather than minimizing raw regression errors. Strong regularization and early stopping prevent overfitting, and consistent price and size screens ensure comparability with prior models.

[Insert Table 6 here]

Table 6 shows that XGBoost fails to deliver economically meaningful momentum profits and adds little beyond the geometric kernel. In Panel A, decile averages and mean scores are weakly monotonic, but the winner–minus–loser spread is only 0.07% per month with a t-statistic of 0.12. Rank correlations indicate orientation is mostly correct in sign (median Spearman about 0.05 and positive in roughly 71 percent of months) yet the effect size is small, implying severe score compression and limited cross-sectional separation in predicted payoffs. The per-decile Sharpe ratios rise modestly toward higher deciles but are too weak to translate into a reliable long–short premium.

To further assess whether XGBoost adds unique value beyond the benchmark model, I conduct three complementary tests. Panel B reports the Fama–French three-factor regression for the XGBoost W–L portfolio, which yields an alpha of 0.48% with a t-statistic of 0.91, indicating little risk-adjusted outperformance. Panel C presents a residual-sort test in which XGBoost scores are orthogonalized to the Geometric + Kalman signal. The residual W–L shows a slightly negative mean and Sharpe ratio, suggesting that once the geometric component is removed, XGBoost contributes no additional predictive power. Panel D uses Fama–MacBeth cross-sectional regressions to test whether the XGBoost signal contains independent pricing information once the geometric kernel signal is controlled for. The geometric signal has a positive slope with a t-value of about 1.56, indicating that it remains a priced continuation factor. In contrast, the XGBoost slope is negative and marginally significant at the 10% level, implying that, conditional on the structured kernel information, the boosted model captures residual variation linked to transitory noise or mean reversion rather than distinct momentum effects. Together, these results demonstrate that, conditional on the structured kernel information, the more flexible boosted model primarily captures noise or short-term reversals rather than independent momentum information.

[Insert Table 7 here]

Table 7 compares the month-to-month persistence of momentum rankings across XGBoost and the Geometric + Kalman model. The latter exhibits much higher stability, with an average diagonal probability of 0.300 versus 0.197 for XGBoost, and lower adjacent migration (0.100 vs. 0.147). This indicates that its predicted scores are more consistent over time, while XGBoost’s rankings fluctuate more sharply across deciles. Under the kernel model, P1 and P10 stocks remain in the same decile more than 50 and 60 percent of the time, respectively, compared with roughly 32 percent for XGBoost. Also illustrated in Figure 3, the kernel model’s transition mass is concentrated along the diagonal and immediate neighbors, suggesting smoother signal evolution, whereas XGBoost’s probabilities diffuse toward middle deciles, reflecting noisier, mean-reverting rankings. In sum, the diagnostics offer evidence of limited incremental predictability from XGBoost. The model sorts in the right direction on average but with insufficient dispersion to earn a premium, consistent with the view that faster information diffusion leaves little for flexible nonlinear learners to exploit.

[Insert Figure 3 here]

4.5. Robustness: Interaction with Firm Complexity

As a robustness check of the paper's main finding that accelerated information absorption has reduced the profitability of traditional momentum while fast-decay kernels better capture the residual signal, I incorporate firm financial-filing complexity into the analysis. Since complexity, a direct measure of a firm's financial statement readability, captures the degree of information asymmetry between the firm and outside investors, low readability implies greater opacity and higher asymmetry. In such firms, investors take longer to absorb and process new information, leading to delayed price adjustments and more persistent momentum effects.

The proxy I use for disclosure complexity is based on the approach of Loughran and McDonald (2024), which has become a standard in recent research (e.g., Chatjuthamard et al., 2022; Hoitash and Hoitash, 2022).⁸ They develop a textual measure based on firms' 10-K filings, and starting from an extensive list of words linked to complex corporate activities (e.g., mergers, derivatives, international operations, and accounting adjustments), they apply a penalized regression to isolate terms that consistently predict higher complexity across validation samples. The resulting lexicon quantifies the extent to which firms report intricate operations and transactions, thereby reflecting substantive organizational and economic complexity rather than simple verbosity. Firms are classified each year into high- and low-complexity groups based on whether their complexity score is above or below the industry-year median, where industries are defined following the Fama–French 12-industry classification. For each group, I re-estimate the Geometric + Kalman momentum kernel using the same 20-year rolling window, skip-one structure, and monthly reformation as in the baseline specification. This setup allows the kernel shape (centroid, half-life, and decay rate) and performance metrics to reflect heterogeneity in how quickly information is absorbed and priced within each disclosure environment. Table 8 presents the results for this robust analysis.

[Insert Table 8 here]

Table 8 Panel A shows that, relative to low-complexity firms, high-complexity names exhibit a slower-decay kernel. The decay rate difference is negative and significant ($\lambda(H-L)=-0.032$, $t=-3.96$), and the centroid lag is modestly later (0.224, $t=4.63$), which indicates that higher

⁸ The data are collected from <https://sraf.nd.edu/complexity/>. Since the Loughran and McDonald complexity data begin in January 1996, the out-of-sample analysis period for this section spans January 1996 to December 2024.

disclosure complexity is associated with a slower, more back-loaded momentum signal, consistent with information taking longer to be absorbed and reflected in prices. Panel B clarifies the mechanism that high-complexity firms place less weight on mid-recent lags ($\tau=5-7$) and more weight on older lags ($\tau=10-12$), consistent with slower information absorption. Panel B reports average kernel weights across lags two to twelve for low- and high-complexity firms and their differences. The H-L column turns from negative in the shorter window to positive at longer lags, implying that low-complexity firms load more on recent lags, while high-complexity firms shift more weight toward older lags, which is evidence of more back-loaded influence for complex filers, consistent with delayed information absorption.

Despite the slower decay, Panel C indicates weaker profitability for high complexity. The W-L mean and Sharpe are lower for high- versus low-complexity firms (1.08% vs. 1.42% per month; Sharpe 0.402 vs. 0.572), even though both groups display highly significant, positive λ estimates. This suggests that while complexity lengthens the diffusion window, it also introduces noise and dispersion that dilute extractable continuation. Persistence patterns in Panel D are broadly similar across groups and high in the tails, but low-complexity portfolios show slightly tighter adjacent drift (more negative and more significant), indicating less noisy sorting.

Panel E provides supporting evidence for a short-horizon effect. Multi-month W-L returns are higher for low-complexity firms, and their first-month contributions are both larger and statistically significant, indicating that price adjustment occurs more quickly when disclosure is easier to interpret. The one-month profitability ratio of low to high complexity is 1.31, but this ratio falls to 1.18 for three-month and 1.20 for six-month horizons, suggesting that momentum among high-complexity firms continues to build more gradually over time, consistent with slower information absorption.

5. Conclusion

This paper reinterprets momentum as a measurable process of information absorption. By estimating dynamic kernels that quantify how past shocks influence future returns, it shows that market learning has accelerated dramatically over the past six decades. The framework integrates behavioral and informational theories within an empirically tractable structure and demonstrates that structurally disciplined models show better performance than complex machine-learning alternatives. The paper also demonstrates that economically structured models such as geometric

and Kalman-filtered kernels can outperform both the traditional equal-weighted strategy and complex machine learning methods. This finding reinforces that interpretability and theoretical coherence remain essential for understanding financial predictability. Machine learning can be a useful tool, but it is not a universal solution for all financial questions.

Additionally, the estimated decay parameter provides a simple and interpretable measure of information speed that reflects how efficiently markets incorporate news. This measure can be tracked over time or across firms to assess shifts in market efficiency. The firm-level analysis shows that companies with complex and less readable disclosures experience slower information diffusion and longer momentum persistence, suggesting that disclosure transparency has tangible effects on market dynamics.

Overall, the paper advances both academic understanding and practical application. It connects decades of momentum evidence under one interpretable mechanism and provides a new way to monitor and adapt to evolving market efficiency. The framework bridges traditional theory and modern data practice, showing that financial markets evolve through faster learning rather than the disappearance of behavioral anomalies.

References

- Baltussen, G., Da, Z., Lammers, S., & Martens, M. (2021). Hedging demand and market intraday momentum. *Journal of Financial Economics*, 142(1), 377-403.
- Barroso, P., & Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics*, 116(1), 111-120.
- Bro, R., & De Jong, S. (1997). A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 11(5), 393-401.
- Bryzgalova, S., Pelger, M., & Zhu, J. (2019). Forest through the trees: Building cross-sections of stock returns. *The Journal of Finance*.
- Chatjuthamard, P., Ongsakul, V., & Jiraporn, P. (2022). Corporate complexity, managerial myopia, and hostile takeover exposure: Evidence from textual analysis. *Journal of Behavioral and Experimental Finance*, 33, 100601.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1), 57-82.
- Chen, L., Pelger, M., & Zhu, J. (2024). Deep learning in asset pricing. *Management Science*, 70(2), 714-750.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Chordia, T., & Shivakumar, L. (2002). Momentum, business cycle, and time-varying expected returns. *The journal of Finance*, 57(2), 985-1019.
- Cooper, M. J., Gutierrez Jr, R. C., & Hameed, A. (2004). Market states and momentum. *The journal of Finance*, 59(3), 1345-1365.
- Da, Z., Gurun, U. G., & Warachka, M. (2014). Frog in the pan: Continuous information and momentum. *The review of financial studies*, 27(7), 2171-2218.
- Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial economics*, 122(2), 221-247.
- DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?. *The review of Financial studies*, 22(5), 1915-1953.
- Ehsani, S., Harvey, C. R., & Li, F. (2023). Is sector neutrality in factor investing a mistake?. *Financial Analysts Journal*, 79(3), 95-117.
- Feng, G., He, J., Polson, N. G., & Xu, J. (2024). Deep learning in characteristics-sorted factor models. *Journal of Financial and Quantitative Analysis*, 59(7), 3001-3036.
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research*, 270(2), 654-669.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223-2273.
- Gu, S., Kelly, B., & Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics*, 222(1), 429-450.

- Harvey, A. C. (1990). Forecasting, structural time series models and the Kalman filter.
- Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2011). Limited investor attention and stock market misreactions to accounting information. *The Review of Asset Pricing Studies*, 1(1), 35-73.
- Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of accounting and economics*, 36(1-3), 337-386.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- Hoitash, R., & Hoitash, U. (2022). A measure of firm complexity: Data and code. *Journal of Information Systems*, 36(2), 161-172.
- Hong, H., Lim, T., & 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., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of finance*, 54(6), 2143-2184.
- Hou, K., & Moskowitz, T. J. (2005). Market frictions, price delay, and the cross-section of expected returns. *The Review of Financial Studies*, 18(3), 981-1020.
- Huang, S. (2022). The momentum gap and return predictability. *The Review of Financial Studies*, 35(7), 3303-3336.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- Jegadeesh, N., & 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., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of finance*, 56(2), 699-720.
- Jegadeesh, N., & Titman, S. (2011). Momentum. *Annu. Rev. Financ. Econ.*, 3(1), 493-509.
- Jegadeesh, N., & Titman, S. (2023). Momentum: Evidence and insights 30 years later. *Pacific-Basin Finance Journal*, 82, 102202.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.
- Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3), 501-524.
- Kozak, S., Nagel, S., & Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, 135(2), 271-292.
- Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689-702.
- Lawson, C. L., & Hanson, R. J. (1995). *Solving least squares problems*. Society for Industrial and Applied Mathematics.
- Lou, D., & Polk, C. (2013). Comomentum: Inferring arbitrage activity from return correlations. *working paper*

- Loughran, T., & McDonald, B. (2024). Measuring firm complexity. *Journal of Financial and Quantitative Analysis*, 59(6), 2487-2514.
- Mamais, K., Thomakos, D., & Vlamis, P. (2025). Explaining and Predicting Momentum Performance Shifts Across Time and Sectors. *Journal of Forecasting*, 44(3), 960-977.
- Moskowitz, T. J., & Grinblatt, M. (1999). Do industries explain momentum?. *The Journal of finance*, 54(4), 1249-1290.
- Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80(3), 563-602.
- Rouwenhorst, K. G. (1998). International momentum strategies. *The journal of finance*, 53(1), 267-284.
- Tong, Y., Wang, Y., & Yan, Y. (2025). Momentum, bubble and information transmission. *Applied Economics*, 1-14.
- Wang, H., Di, J., & Xie, Y. (2023). Echo disappears: momentum term structure and cyclic information in turnover. *arXiv preprint arXiv:2304.03437*.
- Zakamulin, V. (2023). Revisiting the duration dependence in the US stock market cycles. *Applied Economics*, 55(4), 357-368.
- Zhang, X. F. (2006). Information uncertainty and stock returns. *The journal of Finance*, 61(1), 105-137.

Table 1. JT Momentum Portfolio Performance

This table reports the monthly returns of ten equal-weighted momentum portfolios formed based on past J-month returns (where J = 6 or 12) and held for K months (where K = 1 or 6). Portfolio P1 (P10) comprises stocks with the lowest (highest) 10% of returns during the J-month formation period, excluding the most recent month. The sample includes all NYSE, AMEX, and Nasdaq stocks, excluding those priced below \$5 and those in the smallest 10% of market capitalization (based on NYSE size decile cutoffs). Performance is evaluated using either raw stock returns (Panel A) or Fama–French three-factor alphas (Panel B). Differences between the high- and low-momentum portfolios (P10–P1) and their corresponding t-statistics are also reported. The sample period spans from January 1965 to December 2024. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

<i>Panel A: Portfolio Performance Based on Raw Returns (%)</i>								
	J12 K1				J6 K6			
	1965-1989	1990-2009	2010-2024	1965-2024	1965-1989	1990-2009	2010-2024	1965-2024
P1 (Past losers)	-0.16	0.92	1.07	0.53	0.22	1.00	0.72	0.61
P2	1.01	0.82	1.20	1.00	0.66	0.78	1.17	0.83
P3	0.75	1.18	1.08	0.99	0.93	0.97	1.32	1.04
P4	1.10	0.69	1.10	0.96	1.06	1.03	1.16	1.08
P5	1.18	1.00	1.09	1.09	1.11	0.87	1.12	1.03
P6	0.86	1.11	1.18	1.03	1.33	0.97	1.15	1.16
P7	1.68	0.90	0.77	1.18	1.42	0.91	0.95	1.13
P8	1.33	0.91	1.00	1.10	1.35	0.95	1.03	1.13
P9	1.60	1.04	0.98	1.24	1.53	1.07	1.08	1.26
P10 (Past winners)	1.68	1.40	1.67	1.59	1.45	1.42	1.26	1.39
P10-P1	1.84***	0.48	0.61	1.05***	1.23***	0.43	0.54	0.78***
t-value	4.94	0.73	0.92	3.29	4.54	0.94	0.99	3.30
<i>Panel B: Portfolio Performance Based on Fama–French Three-Factor Alphas (%)</i>								
	J12 K1				J6 K6			
	1965-1989	1990-2009	2010-2024	1965-2024	1965-1989	1990-2009	2010-2024	1965-2024
P1 (Past losers)	-1.45	-0.49	-0.25	-0.87	-1.07	-0.23	-0.65	-0.70
P2	-0.28	-0.31	-0.02	-0.20	-0.59	-0.35	0.05	-0.38
P3	-0.52	0.05	0.03	-0.18	-0.30	-0.13	0.21	-0.12
P4	-0.05	-0.30	-0.03	-0.15	-0.17	0.08	0.17	-0.03
P5	-0.07	0.15	0.14	0.04	-0.08	-0.11	0.08	-0.07
P6	-0.28	0.21	0.26	-0.01	0.13	0.12	0.14	0.09
P7	0.57	0.04	-0.32	0.12	0.26	0.01	-0.08	0.05
P8	0.20	0.08	-0.06	0.07	0.19	0.10	-0.08	0.08
P9	0.56	0.17	-0.15	0.15	0.38	0.23	0.01	0.19
P10 (Past winners)	0.66	0.59	0.50	0.52	0.35	0.49	-0.07	0.23
P10-P1	2.11***	1.08*	0.75	1.38***	1.42***	0.72	0.58	0.94***
t-statistics	7.20	1.85	1.27	4.67	5.85	1.58	1.30	4.15

Table 2. Comparison of Momentum Portfolio Performance between JT (J12 K1) and NNLS

This table presents momentum portfolio results and corresponding Non-Negative Least Squares (NNLS) kernel properties. Panel A compares the monthly returns and Fama–French three-factor alphas of ten equal-weighted momentum portfolios under three different model settings: *JT (J12 K1)*, the original Jegadeesh and Titman (1993) strategy that forms portfolios based on past 12-month returns (excluding the most recent month) and holds them for one month; *No Rolling*, where portfolios are generated using an NNLS model trained on all data prior to January 1990 and applied out-of-sample each month from January to December 2024; and *Rolling*, where portfolios are estimated using an NNLS model with a 20-year rolling training window (skipping the most recent month) and updated monthly. Portfolios P1 through P10 represent deciles sorted by past returns from the lowest (P1, past losers) to the highest (P10, past winners). The difference between the top and bottom deciles (P10–P1) and their corresponding t-statistics are also reported. The sample period spans from January 1990 to December 2024. Panel B summarizes the properties of the exponential–mixture kernel used in the NNLS model, showing the total mass within 12 months, the effective half-life (Eff θ), the correlation between estimated and fitted weights (corr(w, w_{fit})), and the corresponding R². *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

<i>Panel A: Momentum Portfolio Performance</i>						
	JT (J12 K1)		NNLS			
	Return	Alpha	No Rolling		Rolling	
			Return	Alpha	Return	Alpha
P1 (Past losers)	0.99	-0.35	0.68	-0.58	0.80	-0.40
P2	0.99	-0.13	1.04	-0.03	1.00	-0.19
P3	1.14	0.08	1.35	0.33	1.16	0.12
P4	0.86	-0.13	0.95	-0.02	0.96	-0.07
P5	1.04	0.16	0.89	-0.10	0.98	-0.01
P6	1.14	0.23	0.79	-0.14	1.25	0.27
P7	0.85	-0.10	1.08	0.14	0.98	-0.02
P8	0.95	0.05	1.04	0.10	0.94	0.03
P9	1.01	-0.01	1.20	0.24	0.94	0.02
P10 (Past winners)	1.52	0.51	1.51	0.40	1.67	0.56
P10-P1	0.53	0.86**	0.82*	0.98**	0.86**	0.96**
t-value	1.14	1.99	1.90	2.43	2.02	2.41
<i>Panel B: NNLS Kernel Summary</i>						
	Mass $\leq 12m$	Eff θ	corr(w, w _{fit})	R ²		
No Rolling	38.93%	24.32	0.89	0.79		
Rolling	100%	6.61	0.94	0.87		

Table 3. Momentum Performance under Geometric Kernel and Kalman-Filtered Models

This table compares momentum portfolio performance and regime-dependent speed estimates under two model settings: the Geometric Kernel model and the Geometric Kernel + Kalman model. Panel A reports monthly returns and Fama–French three-factor alphas for ten equal-weighted decile portfolios sorted based on the two models with a 20-year rolling training window (skipping the most recent month) and updated monthly. Portfolios P1 through P10 represent past losers to past winners, respectively, with differences between P10 and P1 (P10–P1) and corresponding t-statistics shown. Panel B summarizes the time-series dynamics of the kernel speed parameter (λ), reporting mean, median, and half-life values by subperiod (1990–1999, 2000–2009, 2010–2024) for both models. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

<i>Panel A: Momentum Portfolio Performance</i>				
	Geometric Kernel		Geometric Kernel + Kalman	
	Return	Alpha	Return	Alpha
P1 (Past losers)	0.62	-0.62	0.54	-0.69
P2	1.29	0.19	1.19	0.10
P3	1.06	0.00	1.03	-0.06
P4	1.00	0.03	1.09	0.09
P5	0.82	-0.13	1.03	0.10
P6	0.85	-0.13	0.82	-0.18
P7	1.02	0.06	1.09	0.14
P8	1.11	0.14	1.04	0.08
P9	1.09	0.13	1.18	0.18
P10 (Past winners)	1.88	0.70	1.72	0.58
P10-P1	1.26***	1.10***	1.18***	1.06***
t-value	2.88	2.75	2.71	2.66
<i>Panel B: Regime Summary of Speed</i>				
	λ mean	λ median	Half-life mean	
Geometric Kernel				
1990-1999	0.029	0.020	32.92	
2000-2009	0.098	0.090	16.14	
2010-2024	0.332	0.280	2.48	
Geometric Kernel + Kalman				
1990–1999	0.110	0.021	23.33	
2000–2009	0.158	0.177	13.38	
2010–2024	0.428	0.410	1.27	

Table 4. Regression of Momentum Profitability on Kernel Decay and Market Conditions

This table reports regressions of monthly momentum returns (P10–P1) from the JT (J12 K1) strategy on the estimated kernel decay parameter (λ) and control variables that capture market and information environments. The dependent variable is the winner-minus-loser (W–L) portfolio return. The key independent variable, λ , measures how fast the kernel discounts older returns for each month, and it's estimated from the Geometric Kernel. Control variables include Dispersion (cross-sectional return dispersion), Vol_3m (three-month market volatility), News_frac (fraction of firm-months with earnings news), and Crowd_12m (momentum crowding intensity over the past 12 months). Results are shown for the full sample (1990–2024) and three subperiods: 1990–2000, 2001–2010, and 2011–2024. Standard errors are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	JT (J12 K1) W-L Return							
	All		1990-2000		2001-2010		2011-2024	
λ	-0.123*** (0.037)	-0.119*** (0.039)	-0.646*** (0.141)	-0.704*** (0.134)	-0.253* (0.144)	-0.171 (0.159)	-0.180*** (0.063)	-0.170*** (0.064)
Dispersion		-0.141 (0.156)		1.467*** (0.303)		-0.674*** (0.245)		0.162 (0.385)
Vol_3m		-0.232 (0.213)		-0.298 (0.442)		-0.788* (0.423)		0.173 (0.281)
News_frac		0.101** (0.051)		-0.193 (0.117)		0.316*** (0.098)		0.034 (0.069)
Crowd_12m		-0.419** (0.184)		-0.664** (0.288)		-0.712** (0.357)		-0.250 (0.284)
Constant	0.025*** (0.007)	0.025* (0.014)	0.040*** (0.010)	-0.043* (0.023)	0.030* (0.015)	0.048** (0.024)	0.051*** (0.017)	0.022 (0.030)
Obs	420	420	132	132	120	120	168	168
Adj. R ²	0.024	0.034	0.133	0.288	0.017	0.123	0.041	0.035

Table 5. Momentum Portfolio Performance under LSTM and LSTM–Kalman Models

This table compares the performance of momentum portfolios under two deep learning frameworks: a standard LSTM (Long Short-Term Memory) model and an LSTM combined with Kalman filtering for dynamic signal smoothing. Panel A reports the monthly mean returns and Fama–French three-factor alphas for ten equal-weighted decile portfolios sorted by both models with a 20-year rolling training window (skipping the most recent month), updated monthly, and held for one month. Portfolios P1 through P10 represent past losers to past winners, respectively, with P10–P1 and corresponding t-statistics are reported. Panel B summarizes out-of-sample performance statistics for each model, along with the Geometric Kernel + Kalman model, including the mean and standard deviation of the W–L portfolio, Sharpe ratio, alpha, and its t-value. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

<i>Panel A: Momentum Portfolio Performance</i>						
	LSTM		LSTM+Kalman			
	Mean	alpha	Mean	alpha		
P1 (Past losers)	0.77	-0.57	0.51	0.10		
P2	1.01	-0.11	1.28	1.00		
P3	1.28	0.24	1.28	1.01		
P4	0.86	-0.11	1.01	0.75		
P5	1.27	0.29	1.09	0.84		
P6	0.88	-0.06	0.80	0.55		
P7	0.93	-0.02	0.96	0.70		
P8	1.02	0.08	1.21	0.93		
P9	1.18	0.20	1.23	0.98		
P10 (Past winners)	1.51	0.40	1.50	1.30		
P10-P1	0.74	0.96**	0.98**	1.20***		
t-value	1.61	2.23	2.05	2.64		
<i>Panel B: Model Performance Summary</i>						
Model	W-L mean	Std. Dev.	Sharpe Ratio	W-L Alpha	t-value	
LSTM:	0.74	9.38	0.273	0.96**	2.23	
LSTM+Kalman	0.99	9.37	0.365	1.21***	2.66	
Geometric Kernel+Kalman	1.18	8.92	0.459	1.28***	3.22	

Table 6. Momentum Portfolio Performance and Cross-Model Diagnostics for XGBoost

This table reports the performance of momentum portfolios and related diagnostics based on the XGBoost model. Panel A presents monthly portfolio statistics for ten equal-weighted deciles sorted by predicted momentum scores, including mean return, Sharpe ratio, and average model score. Portfolios P1 through P10 represent past losers to past winners, respectively. The table also reports the Spearman rank correlation between predicted scores and realized returns, its median across months, and the fraction of months with positive correlations. Panel B reports the Fama–French three-factor alpha regression for the winner-minus-loser (W–L) portfolio using Newey–West adjusted standard errors with a six-lag correction. Panel C evaluates the residual-sort test, where XGBoost residuals are orthogonalized with respect to geometric kernel+Kalman model (GEOMK) signals, and the resulting W–L performance is summarized by mean, standard deviation, and Sharpe ratio. Panel D shows Fama–MacBeth cross-sectional slope estimates for both the GEOMK and XGB model-implied signals. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

<i>Panel A: Momentum Portfolio Performance</i>			
	XGBoost		
	Mean	Sharpe Ratio	Mean Score Use
P1 (Past losers)	1.33	0.567	-0.160
P2	0.67	0.397	-0.110
P3	1.10	0.596	-0.085
P4	1.39	0.848	-0.064
P5	1.20	0.770	-0.046
P6	1.03	0.720	-0.028
P7	1.35	0.988	-0.011
P8	1.44	0.908	0.008
P9	1.19	0.951	0.028
P10 (Past winners)	1.40	0.942	0.061
P10-P1	0.07	0.032	
t-value	0.12		
Spearman		0.052	
Median monthly Spearman		0.050	
% months Spearman>0		0.709	
<i>Panel B: FF3 Alpha — XGB W–L (HAC lag=6)</i>			
Alpha	t	Adj. R ²	
0.48	0.91	0.138	
<i>Panel C: Residual-sort (XGB GEOMK) W–L</i>			
Mean	Std. Dev.	Sharpe Ratio	
-0.003	0.075	-0.038	
<i>Panel D: Fama–MacBeth slopes</i>			
GEOMK β	t-value	XGB β	t-value
0.35	1.56	-0.27*	-1.69

Table 7. Decile Migration

This table reports the month-to-month transition probabilities between momentum deciles for two models: XGBoost (Panel A) and Geometric + Kalman (Panel B). Each entry $P(i, j)$ represents the average probability that a stock in decile i at current month moves to decile j at next month. Portfolios are formed monthly based on predicted momentum scores, with P1 representing past losers and P10 representing past winners. Diagonal elements measure portfolio persistence, while off-diagonal entries reflect migration across deciles. The mean diagonal probability indicates the average likelihood of remaining in the same decile, and the average adjacent probability measures the average probability of moving to a neighboring decile ($i \pm 1$).

<i>Panel A: XGBoost</i>										
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1	0.332	0.173	0.136	0.094	0.069	0.049	0.047	0.036	0.031	0.034
P2	0.194	0.176	0.137	0.126	0.087	0.080	0.067	0.058	0.042	0.032
P3	0.127	0.142	0.166	0.133	0.111	0.104	0.072	0.064	0.045	0.035
P4	0.091	0.117	0.140	0.152	0.124	0.117	0.088	0.069	0.066	0.038
P5	0.072	0.109	0.123	0.132	0.139	0.115	0.095	0.084	0.077	0.056
P6	0.057	0.074	0.094	0.110	0.129	0.150	0.121	0.105	0.095	0.066
P7	0.051	0.067	0.071	0.087	0.110	0.126	0.154	0.122	0.113	0.100
P8	0.030	0.051	0.051	0.075	0.081	0.118	0.143	0.174	0.151	0.127
P9	0.033	0.042	0.043	0.051	0.080	0.090	0.126	0.156	0.212	0.166
P10	0.030	0.041	0.042	0.038	0.056	0.070	0.087	0.137	0.180	0.319
mean(diagonal):					0.197					
Avg adjacent probability:					0.147					
<i>Panel B: Geometric+Kalman</i>										
P1	0.533	0.212	0.085	0.054	0.031	0.023	0.017	0.016	0.012	0.016
P2	0.219	0.279	0.197	0.102	0.071	0.049	0.027	0.023	0.016	0.016
P3	0.090	0.200	0.238	0.182	0.101	0.076	0.042	0.036	0.024	0.012
P4	0.057	0.106	0.179	0.195	0.171	0.118	0.080	0.049	0.031	0.014
P5	0.042	0.077	0.120	0.168	0.184	0.164	0.115	0.067	0.043	0.019
P6	0.023	0.046	0.065	0.116	0.168	0.203	0.169	0.114	0.066	0.029
P7	0.019	0.034	0.053	0.078	0.117	0.175	0.203	0.178	0.115	0.029
P8	0.015	0.020	0.031	0.047	0.077	0.123	0.190	0.233	0.196	0.067
P9	0.012	0.015	0.022	0.033	0.043	0.070	0.118	0.210	0.323	0.152
P10	0.005	0.008	0.012	0.015	0.016	0.026	0.042	0.071	0.194	0.611
mean(diagonal):					0.300					
Avg adjacent probability:					0.100					

Table 8. Complexity-Based Kernel Properties, Performance, and Persistence Evidence

This table summarizes kernel characteristics, return dynamics, and persistence patterns derived from the Geometric + Kalman momentum model. Panel A reports cross-sectional statistics for the centroid lag, half-life, and decay rate (λ) estimated from the difference between high- and low-complexity (H-L) portfolios, classified based on the industry-year median of firms' complexity, along with Newey–West standard errors and t-values. Panel B shows the average kernel weights (w_τ) across lags $\tau = 2$ –12 months for low- (L) and high-complexity (H) portfolios, and their differences (H-L). Panel C reports cross-sectional momentum performance and kernel decay (λ) for low- versus high-complexity firms. Panel D presents decile transition-matrix results, including the persistence of extreme portfolios (P1–P1 and P10–P10), mean diagonal probabilities, adjacent migration probabilities, and adjacent drift with corresponding t-values. Panel E shows multi-horizon (3-month and 6-month) winner-minus-loser (W-L) returns, the relative first-month weight contribution, and t-statistics for both complexity groups. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Kernel Parameter Estimates						
	Mean	SE (NW)	t-value			
Centroid lag (H–L)	0.224***	0.048	4.63			
Half-life (lag index, H–L)	0.069	0.092	0.75			
λ (decay rate, H–L)	−0.032***	0.008	−3.96			
Panel B: Kernel Weight Profiles by Complexity						
τ (months)	L (mean w_τ)	H (mean w_τ)	H-L			
2	0.118	0.117	-0.001			
3	0.111	0.107	-0.005			
4	0.109	0.107	-0.002			
5	0.109	0.098	-0.011			
6	0.109	0.098	-0.011			
7	0.109	0.098	-0.011			
8	0.095	0.097	0.002			
9	0.080	0.082	0.002			
10	0.062	0.073	0.011			
11	0.062	0.072	0.010			
12	0.035	0.052	0.016			
Panel C: Cross-sectional return characteristics						
	W-L Mean	Std. Dev.	Sharpe Ratio	Labda Mean	t-value	
Low Complexity	1.42%	8.59%	0.572	0.145***	11.63	
High Complexity	1.08%	9.31%	0.402	0.135***	7.68	
Low to High	1.31					
Panel D: Transition-matrix evidence						
	P1-P1	P10-P10	Mean (diagonal)	Avg adjacent probability	Adjacent drift	t-value
Low Complexity	0.596	0.681	0.346	0.197	-0.020***	-4.94
High Complexity	0.617	0.698	0.353	0.197	-0.017***	-2.99
Panel E: Multi-month horizons						
	3-Month			6-Month		
	W-L Total Return	1st Month weight	t-value	W-L Total Return	1st Month weight	t-value
Low Complexity	4.35%	32.64%**	2.27	8.37%	16.97%**	2.23
High Complexity	3.68%	29.35%*	1.75	6.96%	15.52%*	1.67
Low to High	1.18			1.20		

Figure 1. NNLS Mixture – Kernel Shape and Exponential-mixture Fit

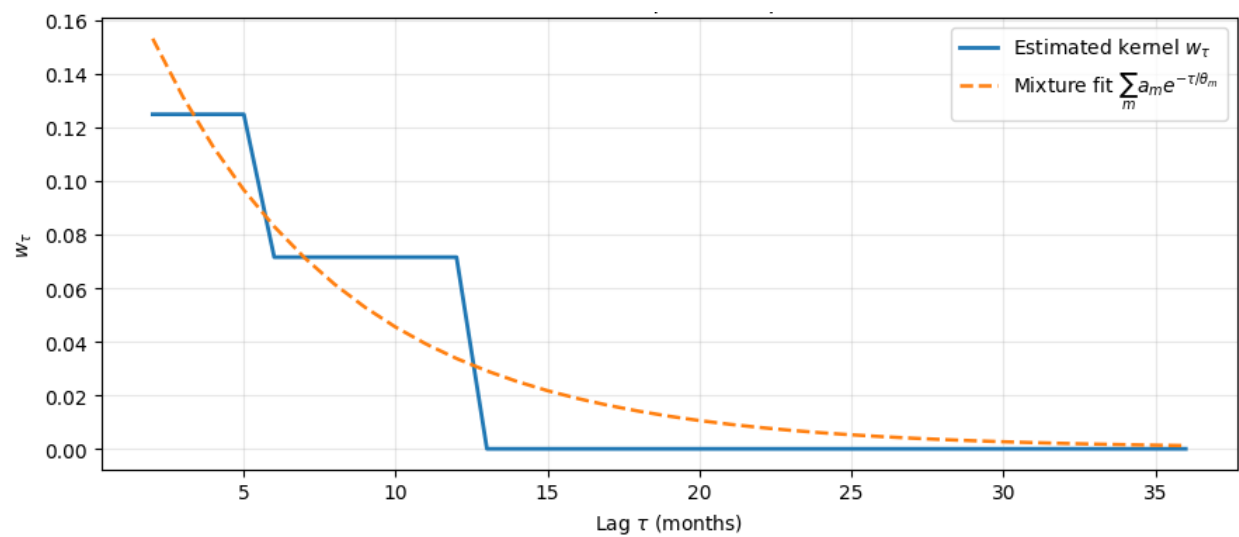


Figure 2. Decay Rate (λ) vs. JT Momentum W-L

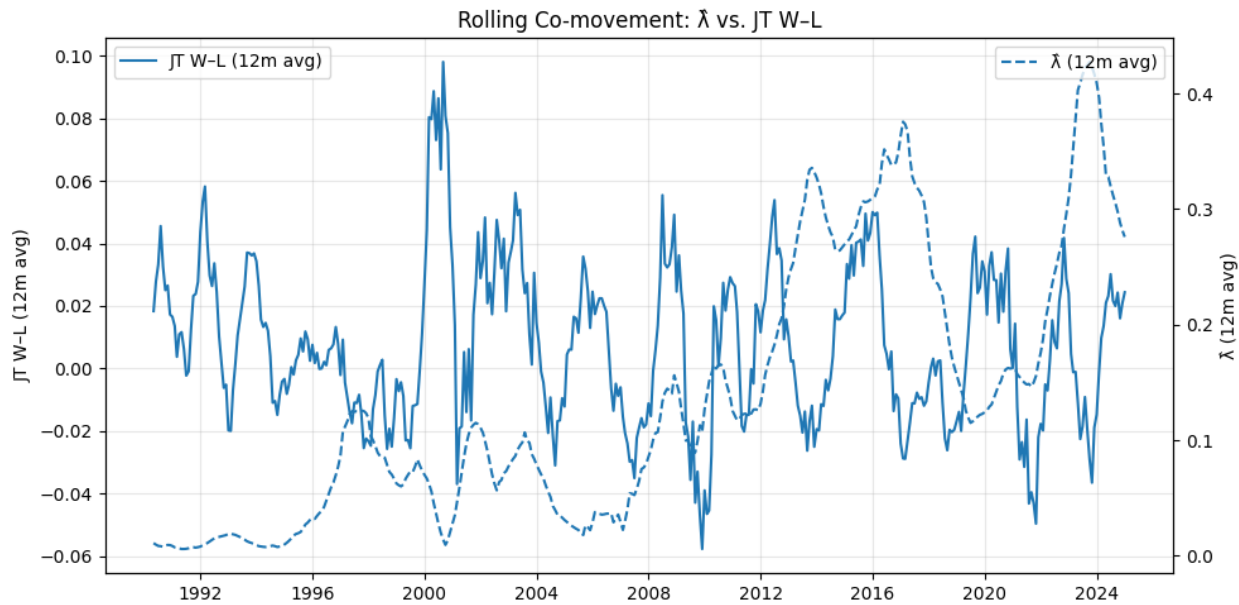


Figure 3. Decile Migration Comparison between XGBoost and Geometric+Kalman

