AUMdrive

Research paper

Improving fund flow predictions with new machine learning techniques

Rasool Taban, Vasco Laranjo, Alvaro Salazar, Richard Parkins
AUMdrive, London. May 2026

Key findings

1

One-month forecasts for US funds missed the actual net flow by 1.97% of starting TNA on average, against 2.15% for OLS and 2.84% for naive carry-forward.

2

Lower error than both benchmarks at every horizon, in the US and in Europe, confirmed by Diebold–Mariano tests ($p < 0.001$).

3

Errors are smallest for larger funds and for multi-asset funds, and stay close to in-sample levels on data the model never saw.

Abstract

For asset managers, attracting and retaining fund flows is as important as generating excess returns, as assets under management directly drive revenues, compensation, and growth. We address the problem of forecasting fund flows, defined as net capital movements into and out of investment vehicles such as mutual funds and exchange-traded funds. We introduce the AUMdrive model, a machine learning framework for fund flow prediction across multiple domiciles and prediction horizons. Fund flows exhibit partially predictable patterns but are characterised by nonlinear dynamics and interaction effects that are difficult to capture using traditional econometric models. To address these challenges, we develop a predictive pipeline integrating multi-source data, structured preprocessing, and machine learning models designed to extract complex empirical relationships from high-dimensional financial data.

The predictive performance and robustness of the AUMdrive model are evaluated across three forecasting horizons (1-month, 3-month, and 6-month) in US and European (EU) fund markets. The framework is tested using both relative (percentage-based) and absolute (USD-denominated) flow measures across fund sizes, asset classes (Equity, Fixed Income, and Multi-Asset), and periods of market stress. Results indicate that predictive accuracy is higher for larger funds, strongest for Multi-Asset and Equity strategies, and generally greater in US-domiciled funds, with performance improving at shorter forecasting horizons. While forecast errors increase at longer horizons and during periods of market stress, the model maintains stable out-of-sample performance and generalisation ability across market conditions. The AUMdrive model provides a more accurate representation of both the direction and magnitude of fund flows compared to traditional benchmarks.

Empirical results show that the AUMdrive model consistently outperforms both Ordinary Least Squares (OLS) and naive carry-forward benchmarks across all configurations. This is confirmed by pairwise Diebold–Mariano tests, which indicate statistically significant superiority of the AUMdrive model over benchmarks in predictive accuracy across specifications. Overall, the AUMdrive model offers significant utility for institutional financial applications by providing actionable predictive intelligence for asset management firms, with potential applications in

sales and distribution planning, capacity and liquidity management, risk monitoring, and portfolio allocation.

1 Introduction

Asset management firms have just as strong an incentive to attract capital inflows as they do to generate alpha, because revenues are tied to Total Net Assets (TNA), meaning that predicting investor demand is as vital to their business models as forecasting fund performance (Zanotti, 2025). Accurate prediction of fund flows for a firm's own products, as well as for competitor and broader market funds, can generate significant strategic and operational advantages. The ability to predict fund flows can help asset managers by avoiding distorted asset prices and negatively influencing fund performance. Lack of predictive or forward-looking data can lead to potential revenue loss through missed opportunities in client acquisition and retention, delayed identification of emerging trends, and slower responses relative to competitors. The absence of demand-centered market intelligence can lead to elevated client drift and inefficient distribution spending through untargeted marketing campaigns and sales outreach. As the industry faces fee compression and increased competition, it is increasingly important for asset managers to increase their TNA and market share, as cost growth continues to outpace revenue growth, and pre-tax operating margins have been declining significantly over the past decade. Anticipating investor demand for funds can, therefore, help asset managers set fee structures that better reflect underlying market demand.

To support business decisions ranging from group-level planning and budgeting to distribution, risk, and investment strategy, asset managers have historically relied on backward-looking data. This traditional data includes historical fund data, economic indicators, volatility and risk measures, and, in some cases, proprietary internal metrics. Some private market actors (Morningstar, Flowspring, Broadridge, etc) have had limited success in predicting mutual funds and exchange-traded funds (ETFs) flows in the past. Although direct comparability with existing approaches is limited by the lack of consistently reported accuracy metrics, the AUMdrive model demonstrates meaningful improvements over existing technologies based on available comparative evidence.

2 Related work

Role of fund flows in asset management. Fund flows play a central role in the mutual fund industry as they determine assets under management, shape equilibrium fund size, and affect portfolio management and market outcomes. Berk and Green (2004) show that investor flows respond to past performance and drive funds toward an equilibrium scale at which diminishing returns to skill and diseconomies of scale are fully reflected in fund size. Beyond this mechanism, flows have important real effects on trading behaviour and asset prices. Unexpected inflows and outflows can reduce performance through transaction costs and liquidity constraints (Edelen, 1999) and generate price pressure in underlying securities, leading to temporary price distortions (Coval and Stafford, 2007).

Variables affecting fund flows. While fund flow modeling remains relatively underexplored in the academic literature, mutual fund performance has been studied extensively over the past several decades. Accordingly, the selection and construction of variables for the models in this study were guided by prior literature on predicting mutual fund performance. Jones and Mo (2020) compile a comprehensive set of predictors of future mutual fund performance by systematically reviewing journal articles published between 1960 and 2015. DeMiguel et al. (2023) show that incorporating nonlinearities and interaction effects between past performance measures and fund characteristics improves predictive accuracy, highlighting the value of richer model specifications.

Nonlinear flow–performance relationship. The literature documents a well-established nonlinear relationship between fund performance and investor fund flows (Sirri and Tufano, 1998; Chevalier and Ellison, 1997). In particular, this flow–performance relationship is commonly characterised as convex, implying that investors respond disproportionately strongly to top-performing funds while being less sensitive to poor performance. More recent evidence refines this view by showing that the degree of convexity varies across the performance distribution. Mazur et al. (2017) find convexity in the upper part of the performance distribution and concavity in the lower part, confirming a heterogeneous nonlinear flow–performance pattern. These documented nonlinearities motivate our focus on modern machine learning methods capable of modeling such complex relationships, consistent with

the approach of Gu et al. (2020), DeMiguel et al. (2023), and Bali et al. (2023).

3 Dataset

The dataset used in this study aggregates information from diverse sources, combining fund-specific proprietary datasets from private vendors with publicly available financial and macroeconomic data from government agencies, intergovernmental organizations, and supranational institutions, as well as in-house market sentiment data. This integration provides a comprehensive foundation for global fund flow analysis. It allows the study to leverage both official economic records and the granular, specialized metrics typically available only through private institutional data providers.

The variables in the dataset are structured to capture different aspects of fund flow dynamics, ranging from fund-specific variables (e.g., performance, fund characteristics, historical flows) to broader financial and macroeconomic indicators. The dataset includes macroeconomic variables and composite indicators to reflect aggregate economic conditions, as well as market sentiment measures derived from dedicated sentiment analysis. This combination enables a detailed examination of how capital flows and economic trends interact with investor sentiment and market perceptions.

4 Fund flows as a target variable

The AUMdrive model's goal is to predict investment fund flows, specifically for mutual funds and exchange-traded funds (ETFs). We use monthly flow data to reduce noise associated with higher-frequency measures of performance and flows. Consistent with the existing literature (Chevalier and Ellison, 1997; Sirri and Tufano, 1998), fund flows are defined as the change in total net assets (TNA) not attributable to investment returns, including capital gains and dividend distributions, but rather to external investor capital inflows and outflows. Monthly fund flows are defined as follows:

$$\text{Flow}_{i,t} = \frac{\text{TNA}_{i,t} - \text{TNA}_{i,t-1}(1 + R_{i,t})}{\text{TNA}_{i,t-1}} \quad (1)$$

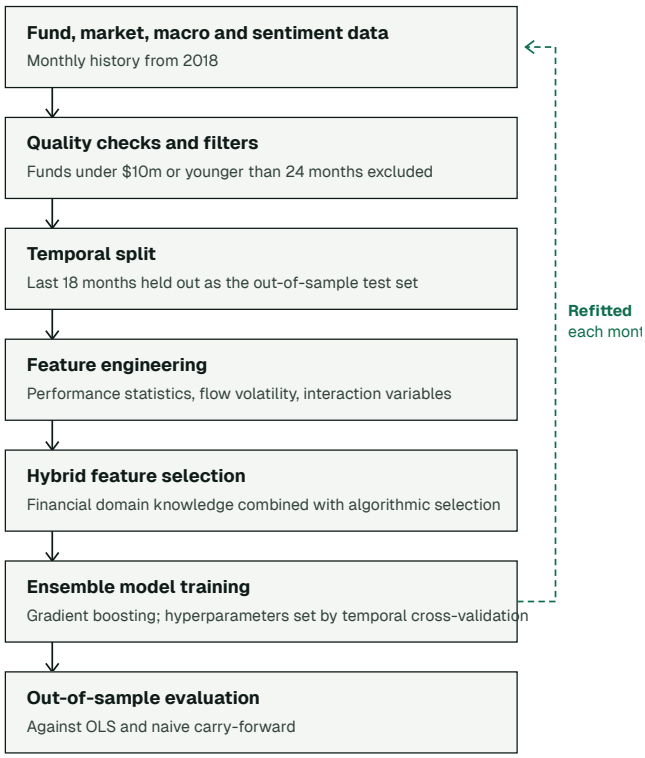
where $\text{TNA}_{i,t}$ denotes the total net assets of fund i at the end of month t, and $R_{i,t}$ denotes the return of fund i in month t. For horizons longer than one month, flows are computed as cumulative measures over desired windows. Specifically, the initial total net assets (TNA) are re-based to the start of each window, while the investment return adjustment considers cumulative returns over the corresponding horizon to ensure consistency with the cumulative flow calculation. All variables are converted into US dollars to ensure comparability across funds with different base currencies and domiciles. To mitigate the influence of extreme values, a dedicated outlier-handling procedure is applied.

5 Model architecture

The AUMdrive model is a machine learning framework designed to predict individual fund flows across a comprehensive investment fund universe. The architecture leverages a combination of rigorous data preprocessing steps with ensemble learning methods to capture underlying patterns in fund flow dynamics. Figure 1 provides a schematic overview of the high-dimensional regression architecture used in this study.

From raw data to monthly forecasts

Figure 1. The AUMdrive model's pipeline



Redrawn from the original schematic. Source: AUMdrive.

The primary dataset comprises historical time-series observations dating back to 2018. To ensure robust performance evaluation, we employ a temporal split, reserving the last 18 months of data as the out-of-sample test set. This period is excluded from training, providing an unbiased benchmark for evaluating pre-

dictive performance. Furthermore, following Morningstar Direct (2021), we exclude small-TNA funds with assets under management (AUM) less than 10 million US dollars and "young funds" with a track record of less than 24 months. This filtering step mitigates the extreme volatility often observed in the early lifecycle of funds, which can otherwise introduce substantial noise into the regression estimates.

The in-sample data is used for model estimation, validation, and hyperparameter optimisation using a temporal cross-validation approach. The modeling pipeline begins with a rigorous data quality assessment and preprocessing phase. To capture nonlinear dependencies in fund flows, we perform feature engineering by constructing additional variables derived from the raw data. These include statistical features capturing historical performance and flow volatility, as well as interaction variables designed to capture cross-sectional dependencies.

We conduct exploratory data analysis (EDA) to ensure data quality, including missing value handling, univariate and multivariate distribution analysis, correlation analysis, and outlier detection. To optimise the feature space, we implement a hybrid feature selection approach combining financial domain knowledge with algorithmic selection methods to identify the most relevant drivers of fund flows.

For the core predictive engine, the AUMdrive model utilizes ensemble learning methods specifically tailored to time-series data. Model hyperparameters are optimised via temporal cross-validation on the in-sample data. To account for shifting market regimes and potential concept drift, the model is periodically refitted using updated data. This iterative calibration allows the model to remain responsive to evolving market conditions while preserving high-quality predictive stability.

6 Model results

The performance of the AUMdrive model is evaluated in two major fund domiciles, United States (US) and Europe (EU), over three prediction horizons of 1, 3, and 6-months. To provide a comprehensive assessment of the predictive utility of the model, the analysis is divided into two distinct scenarios: an evaluation based on average TNA quintiles to examine robustness across funds of varying sizes, and a segmented analysis across three core asset classes, including Equity, Fixed Income, and Multi-Asset. This dual approach ensures that the AUMdrive model performance is evaluated across diverse market segments characterised by varying liquidity and volatility profiles. Furthermore, the structural robustness of the AUMdrive model is tested against severe exogenous shocks by evaluating its performance during three distinct market stress periods, namely the COVID-19 pandemic, the Russian invasion of Ukraine, and 'Liberation Day' tariff announcements.

In addition to the reported results of the AUMdrive model, to evaluate the predictive efficacy of the model, we conduct a comparative analysis against two established benchmarks commonly utilized in empirical financial regression tasks: the naive carry-forward model and the Ordinary Least Squares (OLS) regression model.

- **Naive carry-forward model:** This model serves as a fundamental baseline to determine if the AUMdrive model can significantly outperform a naive forecasting strategy. The naive carry-forward model assumes a random walk, where the predicted flow for the next horizon is equal to the most recent observation. Improving upon this baseline is essential to demonstrate that the model captures genuine predictive signals rather than simple temporal inertia.
- **Ordinary least squares:** Given its status as a cornerstone of empirical finance, OLS is included as the primary linear benchmark model. This comparison is designed to assess the AUMdrive model's ability to go beyond the limitations of linear models by capturing the nonlinear dependencies in fund flow dynamics.

Relative predictive performance is analysed across Framework IV and Framework V. While Framework IV compiles the performance metrics into comprehensive comparative tables, Framework V delivers the formal statistical validation by executing pairwise statistical tests against the competing baselines.

6.1 Framework I: performance across fund-size quintiles

The initial framework evaluates model performance grouped by funds' mean TNA quintiles. Fund scale is represented by the mean TNA for each fund, computed over the full sampling period. The in-sample period spans from January 2020 to April 2024, while the out-of-sample period spans from May 2024 to October 2025. Quintile rankings are then constructed based on these averages. This approach enables an assessment of model robustness across different seg-

ments of the investment universe, providing insight into predictive performance for both small and large funds. Table 1, and Table 2 presents the AUMdrive model performance results for the in-sample and out-of-sample periods, reporting Mean Absolute Error (MAE) for flow predictions in both percentage and US dollar terms.

In relative terms, errors are lowest for the largest funds

Table 1. Mean absolute error by TNA quintile, region and horizon, in sample

	Flow, % of starting TNA						Flow, USD millions					
	Europe			US			Europe			US		
TNA quintile	**1M**	**3M**	**6M**	**1M**	**3M**	**6M**	**1M**	**3M**	**6M**	**1M**	**3M**	**6M**
1 (smallest)	3.36	8.11	14.10	3.36	8.05	14.31	2.02	4.73	7.80	2.43	5.57	9.93
2	3.46	8.27	14.14	2.82	6.39	11.71	5.78	13.66	23.26	6.57	14.64	26.42
3	3.52	8.56	14.79	2.76	6.30	11.49	12.68	30.22	53.12	15.34	33.09	59.89
4	3.04	7.28	12.31	2.39	5.65	10.60	23.90	56.79	100.31	34.35	77.67	144.02
5 (largest)	2.56	6.15	10.47	2.04	4.63	8.62	71.20	173.22	306.96	225.81	471.17	838.96

In sample, January 2020 to April 2024. Quintiles of each fund's mean TNA over the full period. Source: AUMdrive.

Out of sample, the largest funds remain the easiest to forecast

Table 2. Mean absolute error by TNA quintile, region and horizon, out of sample

	Flow, % of starting TNA						Flow, USD millions					
	Europe			US			Europe			US		
TNA quintile	**1M**	**3M**	**6M**	**1M**	**3M**	**6M**	**1M**	**3M**	**6M**	**1M**	**3M**	**6M**
1 (smallest)	3.10	7.61	13.26	2.73	6.74	11.58	1.90	4.41	7.32	2.07	4.82	7.99
2	3.21	7.62	13.09	2.14	5.00	9.00	5.82	13.40	22.06	5.82	13.23	22.97
3	2.90	7.08	12.25	1.99	4.66	8.33	11.31	26.85	44.43	12.25	28.01	48.23
4	2.57	6.12	10.47	1.73	4.16	7.62	22.05	51.24	85.08	28.46	66.58	117.85
5 (largest)	2.02	4.83	8.54	1.34	3.12	5.64	67.00	159.30	279.18	204.77	458.90	793.01

Out of sample, May 2024 to October 2025. Source: AUMdrive.

The first key observation is the inverse relationship between the predictive horizon and model accuracy. The 1-month (1M) horizon yields the highest accuracy, with a gradual decline in performance as the forecast window expands to 6 months (6M). This decline is consistent with standard time-series forecasting results, where predictive accuracy typically decreases as the forecast horizon increases and the signal-to-noise ratio deteriorates. This result holds across both flow measurement units (percentage and US dollars) as well as in both in-sample and out-of-sample datasets. This pattern is particularly pronounced in financial markets due to high-frequency dynamics and exogenous shocks that are difficult to capture from historical data. Geopolitical events, fiscal policy changes, and international sanctions often arise with limited warning and are not reflected in fund-level features, introducing stochastic noise that limits long-horizon forecasting accuracy.

Another notable result is the divergence in predictive accuracy between the US and EU market domiciles, particularly when using flow percentages as the prediction unit. The AUMdrive model consistently performs better in the US market across all temporal horizons. This discrepancy may be attributed to structural differences in the respective financial ecosystems:

- **Market homogeneity and data centralisation:** The US market operates under a unified regulatory framework (SEC) with highly standardised reporting requirements. This homogeneity results in a cleaner signal-to-noise ratio in the training data. In contrast, the European market is fragmented across multiple jurisdictions with varying local regulations, tax regimes, and reporting standards, introducing additional layers of complexity and noise that can impede model performance.
- **Investor behaviour and market maturity:** The US retail and institutional sectors often exhibit more synchronised responses to performance

benchmarks and macroeconomic indicators. This behavioural consistency allows the model to map the underlying momentum of fund flows more effectively. In Europe, idiosyncratic local market preferences and currency differences may lead to more heterogeneous flow patterns that are harder for a global model to generalise.

- **Benchmarking sensitivity:** US funds are often more closely benchmarked to common indices. The resulting benchmark-hugging behaviour in performance translates into more systematic flow patterns. In contrast, the European market features a higher prevalence of niche and specialized cross-border funds, which may exhibit less conventional flow dynamics.

It is important to note that the superior predictive performance observed in the US relative to the European one is sensitive to the choice of flow metric. When fund flows are denominated in USD, the US performance advantage attenuates. This is primarily driven by differences in the distribution of TNA: US-domiciled funds generally exhibit higher average TNA than their European counterparts, particularly in the third to fifth TNA quintiles. Consequently, while percentage-based flow predictions reveal a clear performance advantage for the US market, nominal dollar-based predictions are influenced by scale disparities. Nevertheless, when considering both relative (percentage) and absolute (USD) metrics, predictive performance remains higher for the US market. This stability suggests that structural and regulatory characteristics of the US financial ecosystem support more accurate flow modeling, even when accounting for variation in fund size.

An important pattern in the relative (percentage-based) flow predictions concerns the relationship between fund size and model accuracy. Across both in-sample and out-of-sample data, all markets, and all temporal horizons, predictive performance improves monotonically with increasing TNA quintiles. This pattern is closely related to the volatility profile of funds across size segments. Larger funds tend to exhibit lower prediction errors, reflecting more stable growth rates due to scale effects, greater age and maturity, and a higher likelihood of inclusion in institutional and model portfolios associated with more stable allocation patterns.

In contrast, when analysing flow predictions in nominal terms (USD), MAE exhibits a positive correlation with fund size, appearing to diverge from the pattern observed in relative terms. However, this does not reflect reduced model performance, but rather the scale-dependent nature of nominal flows: as TNA increases across quintiles, the absolute magnitude of capital flows proportionally rises, leading to higher absolute error terms. Consequently, MAE in nominal units should be interpreted in relative terms when assessing predictive performance across fund sizes. A given absolute error that is meaningful for smaller funds in the first quintile becomes negligible for large institutional funds managing billions in assets. Accordingly, the results indicate that the model maintains consistent proportional accuracy relative to fund size. B. Framework II: Model Performance across Asset Classes

6.2 Framework II: performance across asset classes

The second scenario evaluates the AUMdrive model's predictive performance across three core asset classes: Equity, Fixed Income, and Multi-Asset strategies. This categorization assesses whether the model maintains predictive performance across market segments with differing liquidity and volatility profiles.

Multi-asset funds have the lowest errors

Table 3. Mean absolute error by asset class, region and horizon, in sample

	Flow, % of starting TNA						Flow, USD millions					
	Europe			US			Europe			US		
Asset class	**1M**	**3M**	**6M**	**1M**	**3M**	**6M**	**1M**	**3M**	**6M**	**1M**	**3M**	**6M**
Multi-asset	1.68	4.15	7.96	1.84	4.81	9.31	11.77	30.58	65.32	24.52	60.55	122.89
Equity	3.15	7.68	13.23	2.51	6.32	11.50	21.82	51.85	94.03	50.53	116.16	204.98
Fixed income	3.69	8.68	14.50	3.17	6.32	11.50	30.73	71.23	119.77	84.36	147.48	271.86

In sample, January 2020 to April 2024. Source: AUMdrive.

Out of sample, multi-asset funds still have the lowest errors

Table 4. Mean absolute error by asset class, region and horizon, out of sample

	Flow, % of starting TNA						Flow, USD millions					
	Europe			US			Europe			US		
Asset class	**1M**	**3M**	**6M**	**1M**	**3M**	**6M**	**1M**	**3M**	**6M**	**1M**	**3M**	**6M**
Multi-asset	1.63	4.00	7.19	1.31	3.10	5.90	9.37	23.82	48.89	18.93	46.18	85.05
Equity	2.73	6.63	11.47	2.04	4.97	8.82	21.85	51.18	88.21	58.36	126.52	216.88
Fixed income	3.10	7.45	12.83	2.00	4.66	8.22	26.45	59.31	100.92	48.82	109.07	195.92

Out of sample, May 2024 to October 2025. Source: AUMdrive.

First key result, consistent with Framework I, is the inverse relationship between prediction horizon and model accuracy across asset classes, regions, flow prediction units, and sampling periods. The model achieves its highest accuracy at the 1M horizon across all asset classes in both US and EU markets, with performance gradually declining through the 6M horizon. This pattern is partly driven by the fact that longer-horizon forecasts are based on cumulative flow values. As flows scale with the forecasting horizon, this proportionally increases absolute prediction errors.

In addition, and consistent with the evidence documented in Framework I, predictive accuracy for relative (percentage-based) flows is generally higher for US-domiciled funds than for their European counterparts across asset classes and prediction horizons, with the exception of multi-asset funds in the in-sample period. This pattern reinforces the structural and systemic differences between the two markets discussed above. In contrast, when evaluating predictive accuracy in nominal terms (USD), errors are systematically larger for US-domiciled funds. This reflects the higher average size of US funds across asset classes and over the full sample period, consistent with the scale effects discussed previously when performance is measured in absolute terms.

A further observation, consistent across asset classes and both markets, is that predictive accuracy is generally lower in the in-sample period than in the out-of-sample period, with the exception of US-domiciled equity fund flow predictions in absolute (USD) terms. This pattern is largely driven by elevated and prolonged market volatility during the in-sample window, which includes two of the three major stress episodes examined in framework III. Periods of heightened market instability affect not only asset prices but also investor behaviour and allocation decisions, introducing additional variability into fund flow dynamics. This reduces signal persistence and makes flow patterns more difficult to predict. Despite these challenging conditions, the model maintains relatively strong accuracy, highlighting its robustness in periods of significant market stress, as further discussed in framework III.

Another observation concerns the relative ordering of predictive accuracy across asset classes, which remains broadly consistent across forecasting horizons and both in-sample and out-of-sample periods, with some variation across US and European domiciles. Overall, Multi-Asset funds exhibit the lowest prediction errors, followed by Equity funds, while Fixed Income funds generally display the highest errors. An exception arises in the US out-of-sample setting, where Fixed Income funds outperform Equity funds in terms of predictive accuracy for both relative (percentage-based) and absolute (USD) flows. Across asset classes, however, no meaningful differences are observed between the two flow metrics. This suggests that scale effects associated with fund size do not materially alter the relative ranking of predictive accuracy when flows are measured in nominal terms.

To further interpret the observed differences in predictive accuracy across asset classes, this section discusses the main characteristics and market dynamics underlying each category, providing a structural interpretation of the results by linking asset class-specific features to relative model performance.

- **Multi-asset.** Multi-Asset funds exhibit the lowest prediction errors, consistent with their relatively stable flow dynamics. These funds are frequently used as model or institutional portfolio components, where allocations follow predefined rebalancing rules and are less sensitive to short-term market fluctuations. This results in smoother and more predictable flow patterns driven primarily by strategic allocation decisions rather than tactical shifts in investor sentiment. From a size perspective, multi-asset funds are the smallest across both European and US domiciles, with

average TNA of $454 million and $2 billion, respectively, over the full sample period. This smaller scale may further contribute to their lower volatility in flows, as smaller strategic allocation programs tend to be more rule-based and less reactive to market noise.

- **Equity.** Equity funds benefit from relatively rich and informative dynamics over the sample period, despite the presence of multiple market correction episodes, particularly in-sample. The overall positive trend in global Equity markets provides sufficient signal variation for the model to learn underlying relationships, contributing to relatively strong predictive performance in this asset class. From a size perspective, Equity funds are the largest across both European and US domiciles, with average TNA of $667 million and $3.22 billion, respectively, over the full sample period.
- **Fixed income.** Fixed Income funds were affected by pronounced shifts in interest rate regimes over both the in-sample and out-of-sample periods, driven by alternating monetary policy cycles of easing and tightening. These regime changes increased the variability of fund flows, at times exceeding that observed in equity markets, thereby reducing the stability of predictive relationships and contributing to higher prediction errors. From a size perspective, Fixed Income funds occupy an intermediate position between equity and multi-asset funds across both European and US domiciles, with average TNA of $621 million and $2.31 billion, respectively, over the full sample period.

6.3 Framework III: performance in periods of market stress

Investor behaviour is frequently influenced by behavioural biases that can lead to deviations from fully rational decision-making (Thaler, 1980). In particular, loss aversion implies that losses tend to have a stronger psychological impact than equivalent gains (Kahneman and Tversky, 1979), with this asymmetry often becoming more pronounced during periods of market stress (Ouzan, 2020).

In addition, investor overreaction is a well-established empirical phenomenon (De Bondt and Thaler, 1985), which can affect asset prices and, in turn, fund allocation decisions and aggregate fund flows. During periods of heightened uncertainty, such behaviour may contribute to stronger flow responses, with increased net inflows into perceived safe-haven assets (e.g., Money Market funds and selected Fixed Income segments) and elevated outflows from riskier asset classes, such as Equities and Alternative investments.

Our study covers the period from January 2020 to October 2025. Within this sample, we identify three major market stress events characterised by elevated volatility and macroeconomic or geopolitical uncertainty. For each event, the month of the initial shock is used as the reference point for the event study analysis. The considered events and respective month of occurrence are:

> **COVID-19 pandemic (March 2020)**
> **Russian invasion of Ukraine (February 2022)**
> **“Liberation Day” tariff announcements (April 2025)**

This analysis focuses on one-month-ahead (1M) predictions, which provide a concise assessment of model performance during periods of market stress. Results are reported over a window ranging from two months prior to the stress event, and three months after each shock, as defined previously, to evaluate the stability of predictive performance across pre-shock, shock, and post-shock regimes. Table 5 reports MAE results around the three identified stress events for both European and US markets, in terms of both relative (percentage-based) and absolute (USD) flows.

Errors peaked in the month of the COVID-19 shock

Table 5. One-month-ahead mean absolute error around three market stress events

	Flow, % of starting TNA						Flow, USD millions					
	Europe			US			Europe			US		
Month	**COVID-19**	**Ukraine**	**Tariffs**	**COVID-19**	**Ukraine**	**Tariffs**	**COVID-19**	**Ukraine**	**Tariffs**	**COVID-19**	**Ukraine**	**Tariffs**
t–2	3.20	3.07	2.81	2.40	2.77	2.02	34.01	38.60	30.08	58.76	101.17	66.41
t–1	3.57	3.02	3.29	2.16	2.65	2.19	34.11	36.21	36.50	61.52	97.76	86.93
t_0, shock	5.68	3.43	3.27	4.56	2.45	2.57	73.35	42.18	36.91	185.93	73.43	100.51
t+1	4.37	3.94	2.83	4.56	2.66	1.91	37.09	47.32	33.78	106.20	85.12	69.52
t+2	3.71	2.82	2.83	3.24	2.55	1.86	30.21	34.80	33.03	83.62	95.82	65.62

Shocks: COVID-19, March 2020; Russian invasion of Ukraine, February 2022; “Liberation Day” tariffs, April 2025 (out of sample). Source: AUMdrive.

A first observation is that the effects of market stress events on predictive accuracy are heterogeneous, both in terms of their magnitude and persistence over time. The COVID-19 shock exhibits the most pronounced and persistent impact on model performance, while the effects of the Russia–Ukraine war and the "Liberation Day" tariff shock are more moderate and shorter-lived.

A second observation is that predictive accuracy during market stress periods is generally lower than that observed over the broader in-sample and out-of-sample samples. However, the magnitude of this decrease is more moderate than might be expected given the severity of the underlying shocks. This effect is most pronounced during the COVID-19 episode, where prediction errors increase markedly relative to the in-sample averages in Table 6 (US: 4.56% in the month of the shock vs. 2.66%; EU: 5.68% vs. 3.18%). The impact is more limited during the Russia–Ukraine war (US: 2.45% vs. 2.66%, peaking at 2.77% a month later; EU: 3.43% vs. 3.18%, peaking at 3.94% a month later), and remains contained during the "Liberation Day" tariff shock, measured against the out-of-sample averages in Table 7 (US: 2.57% vs. 1.97%; EU: 3.27% vs. 2.75%). It is also important to note that the COVID-19 and Russian–Ukraine episodes fall within the in-sample period, whereas the "Liberation Day" tariff shock is evaluated in an out-of-sample setting. Overall, these results suggest that while market stress periods are associated with a deterioration in predictive accuracy, the magnitude of this effect varies across events and is not uniformly amplified during more severe shocks.

A third observation concerns the relative resilience of predictive performance across domiciles during stress periods. With the exception of the Russian–Ukraine war, which had a more pronounced impact on European markets and European-domiciled funds, changes in predictive accuracy during stress episodes are broadly comparable between the US and EU models. This is notable given that European fund flows are generally more difficult to predict over the full sample period. Nevertheless, the increase in prediction errors during stress periods occurs at a similar magnitude across both regions rather than disproportionately affecting European funds. Overall, these results suggest that market stress episodes induce broadly comparable disruptions to investor allocation dynamics across domiciles, despite the structural differences between the US and European fund markets documented earlier.

6.4 Framework IV: comparison with benchmarks

In this framework, we conduct a formal benchmarking analysis, comparing the predictive performance of the proposed model against two standard financial benchmarks: the naive carry-forward and Ordinary Least Squares (OLS) models. The naive carry-forward model serves as a fundamental baseline, assuming a random walk in which future flows are equal to the most recent observation, and is used to assess whether the model captures meaningful predictive signals beyond temporal persistence. OLS provides a standard linear benchmark, enabling an evaluation of the model's ability to go beyond linear dependencies and capture the nonlinear structure underlying fund flow dynamics.

The empirical performance of the proposed AUMdrive model is evaluated against two baseline models using both in-sample (Table 6) and out-of-sample (Table 7) data across three predictive horizons (1M, 3M, and 6M). Predictive performance is assessed in terms of both relative (percentage-based) and absolute (USD) flows. The evaluation is organized along five dimensions, as discussed below.

Lower mean error than both benchmarks in every in-sample setting

Table 6. Performance metrics by model, region and horizon, in sample

	Europe			US		
Metric	**AUMdrive**	**Naive**	**OLS**	**AUMdrive**	**Naive**	**OLS**
Panel A. Flow, % of starting TNA						
1M horizon						
MAE, mean	**3.18**	4.93	3.62	**2.66**	3.51	5.59
MAE, median	**2.57**	3.63	2.97	**1.98**	2.28	4.51
MAE, IQR	2.53	3.86	**2.19**	**2.13**	2.88	4.13
RMSE	**8.28**	109.51	8.50	**7.01**	15.83	10.12
R^2	**0.03**	−168.20	−0.02	**0.02**	−3.98	−1.04
Directional accuracy, %	63.82	**64.08**	60.97	67.38	**68.97**	62.73
3M horizon						
MAE, mean	**7.67**	10.79	8.51	**6.20**	8.06	11.30
MAE, median	**6.05**	7.89	6.85	**4.31**	5.10	9.77
MAE, IQR	6.20	8.75	**5.49**	5.22	6.80	**3.26**
RMSE	**17.00**	111.12	17.49	**15.59**	21.97	19.73
R^2	**0.05**	−39.42	0.00	**0.08**	−0.82	−0.47
Directional accuracy, %	64.47	**65.23**	61.43	69.82	**71.03**	63.53
6M horizon						
MAE, mean	**13.15**	17.75	15.21	**11.33**	14.44	17.49
MAE, median	**10.35**	12.43	12.49	**7.80**	8.86	15.04
MAE, IQR	10.56	14.74	**8.03**	9.85	12.18	**8.65**
RMSE	**25.35**	191.17	27.21	**25.38**	35.07	30.69
R^2	**0.08**	−51.59	−0.07	**0.11**	−0.70	−0.30
Directional accuracy, %	60.33	**65.14**	61.17	67.87	**70.48**	62.05
Panel B. Flow, USD millions						
1M horizon						
MAE, mean	**23.70**	34.13	28.57	**58.76**	69.76	176.63
MAE, median	**10.36**	14.51	12.71	**11.72**	14.27	30.16
MAE, IQR	**23.71**	34.81	27.99	**37.32**	43.57	91.59
RMSE	**81.73**	193.09	93.97	**341.85**	517.10	1,234.68
R^2	**0.14**	−3.82	−0.14	**−0.19**	−1.73	−14.55
Directional accuracy, %	63.82	**64.08**	60.97	67.38	**68.97**	62.73
3M horizon						
MAE, mean	**55.94**	75.18	65.13	**121.23**	148.35	361.55
MAE, median	**23.23**	30.92	27.83	**25.41**	30.59	59.37
MAE, IQR	**55.32**	76.33	63.99	**79.66**	96.96	177.06
RMSE	**172.91**	250.64	189.86	**607.65**	776.76	2,657.65
R^2	**0.03**	−1.03	−0.17	**0.04**	−0.56	−17.29
Directional accuracy, %	64.47	**65.23**	61.43	69.82	**71.03**	63.53
6M horizon						
MAE, mean	**99.37**	133.73	125.99	**218.40**	264.69	494.86
MAE, median	**40.22**	50.83	52.27	**45.59**	54.12	88.41
MAE, IQR	**98.08**	132.19	122.34	**144.72**	177.29	265.88
RMSE	**273.24**	447.45	365.22	**1,017.88**	1,298.93	2,938.25
R^2	**−0.03**	−1.75	−0.83	**0.04**	−0.56	−6.98
Directional accuracy, %	60.33	**65.14**	61.17	67.87	**70.48**	62.05

In sample, January 2020 to April 2024. Best value in each row and region in bold; lower is better except for R^2 and directional accuracy. Naive: last month's flow carried forward. OLS: ordinary least squares. Source: AUMdrive.

Lower mean error than both benchmarks in every out-of-sample setting

Table 7. Performance metrics by model, region and horizon, out of sample

	Europe			US		
Metric	**AUMdrive**	**Naive**	**OLS**	**AUMdrive**	**Naive**	**OLS**
Panel A. Flow, % of starting TNA						
1M horizon						
MAE, mean	**2.75**	4.21	2.88	**1.97**	2.84	2.15
MAE, median	**1.95**	2.74	2.07	**1.25**	1.65	1.48
MAE, IQR	2.40	3.68	**2.35**	1.60	2.28	**1.50**
RMSE	**7.77**	21.55	7.80	**6.01**	12.89	6.04
R^2	**0.02**	−6.52	0.01	**0.03**	−3.47	0.02
Directional accuracy, %	**67.32**	66.47	63.62	**71.39**	70.23	68.03
3M horizon						
MAE, mean	**6.65**	9.35	6.90	**4.73**	6.43	5.35
MAE, median	**4.62**	6.04	4.90	**2.95**	3.68	3.64
MAE, IQR	5.77	8.22	**5.53**	3.94	5.25	**3.30**
RMSE	**15.66**	21.64	15.68	**12.17**	16.95	12.36
R^2	**0.04**	−0.84	0.03	**0.06**	−0.82	0.03
Directional accuracy, %	**68.36**	67.63	65.17	**73.88**	72.84	69.78
6M horizon						
MAE, mean	**11.50**	15.28	12.04	**8.42**	10.35	9.06
MAE, median	**7.88**	9.52	8.49	**5.30**	5.85	6.13
MAE, IQR	9.82	13.15	**9.67**	7.14	8.67	**5.79**
RMSE	**23.80**	32.37	23.88	**18.63**	23.72	18.90
R^2	**0.05**	−0.75	0.05	**0.09**	−0.47	0.07
Directional accuracy, %	68.33	**69.14**	64.44	72.91	**75.16**	73.84
Panel B. Flow, USD millions						
1M horizon						
MAE, mean	**22.16**	32.42	24.46	**52.34**	73.28	63.05
MAE, median	**8.17**	11.58	8.82	**8.01**	10.70	9.55
MAE, IQR	**21.28**	30.90	23.03	**27.27**	36.83	31.50
RMSE	**76.22**	213.93	89.24	**725.70**	1,002.57	734.03
R^2	**0.21**	−5.19	−0.08	**0.05**	−0.81	0.03
Directional accuracy, %	**67.32**	66.47	63.62	**71.39**	70.23	68.03
3M horizon						
MAE, mean	**51.24**	70.16	55.27	**115.08**	148.24	159.65
MAE, median	**18.35**	24.36	19.46	**17.69**	22.31	22.89
MAE, IQR	**49.09**	65.87	52.05	**62.19**	77.01	73.20
RMSE	**164.24**	244.46	180.24	**1,261.74**	1,369.60	1,431.98
R^2	**0.29**	−0.58	0.14	**0.12**	−0.03	−0.13
Directional accuracy, %	**68.36**	67.63	65.17	**73.88**	72.84	69.78
6M horizon						
MAE, mean	**88.58**	113.85	95.15	**200.43**	226.99	287.33
MAE, median	**30.28**	37.07	33.19	**31.75**	35.57	38.20
MAE, IQR	**81.14**	100.79	89.15	**108.37**	124.39	126.18
RMSE	**265.73**	372.49	280.15	**1,649.45**	1,689.55	2,327.69
R^2	**0.36**	−0.26	0.29	**0.19**	0.16	−0.60
Directional accuracy, %	68.33	**69.14**	64.44	72.91	**75.16**	73.84

Out of sample, May 2024 to October 2025. Best value in each row and region in bold; lower is better except for R^2 and directional accuracy. Source: AUMdrive.

1. Temporal horizon dynamics and generalisation. A consistent inverse relationship is observed between forecasting horizon and predictive accuracy across all evaluation metrics. Accuracy performance declines monotonically as the forecast window expands from 1 month to 6 months, consistently across all models, market domiciles, and performance measures. This pattern aligns with standard results in financial time-series forecasting, where longer horizons are associated with increased exposure to compounding market noise and exogenous shocks, leading to a reduction in predictive signal strength.

A notable property of the AUMdrive model is its strong out-of-sample generalisation performance. In financial machine learning applications, out-of-sample errors often diverge substantially from in-sample errors due to overfitting to historical noise. In contrast, a comparison of Table 6 and 7 shows that the model's out-of-sample performance is broadly comparable to, and in some cases slightly better than, its in-sample performance. For instance, in the 1M European (EU) percentage flow setting, MAE decreases from 3.18% in-sample to 2.75% out-of-sample. This behaviour suggests that the ensemble learning architecture successfully circumvents overfitting paradigms, capturing stable structural patterns in capital flows rather than fitting sample-specific noise.

2. Distribution of errors: central tendency and dispersion. An analysis of absolute residual distributions using the mean MAE, median MAE, and interquartile range (IQR) reveals substantial differences between the linear benchmarks and the nonlinear AUMdrive model:

- **Central tendency reduction:** the AUMdrive model consistently achieves lower average prediction errors across all configurations. In the 1M out-of-sample US predicted flow percentage unit, the AUMdrive model attains the MAE (Mean) of 1.97%, outperforming both the naive carry-forward model (2.84%) and the OLS baseline (2.15%). This advantage remains evident at the extended 6M out-of-sample prediction horizon, where the AUMdrive model's MAE (Mean) of 8.42% continues to exceed the predictive accuracy of both the naive carry-forward baseline (10.35%) and the OLS baseline (9.06%).
- **Skewness and residual tracking:** Across all models and experimental settings, MAE (Mean) consistently exceeds MAE (Median), indicating right-skewed residual distributions with occasional large forecasting errors with fat-tailed asymmetric tails. The AUMdrive model demonstrates tighter error concentration, as reflected in consistently narrower interquartile ranges (IQRs). For example, in the 3M out-of-sample EU predicted flow percentage setting, the AUMdrive model records an IQR of 5.77%, compared with 8.22% for the naive carry-forward model. The simultaneous reduction in both IQR and median error suggests that the AUMdrive model more consistently captures the underlying distribution of fund flows, resulting in greater predictive stability across both standard and elevated volatility regimes.

3. Structural stability and outlier mitigation. Root Mean Square Error (RMSE) provides an informative diagnostic of structural model stability due to its quadratic penalisation of large residuals. The empirical results indicate substantially greater sensitivity to extreme prediction errors in the benchmark models, whereas the AUMdrive model demonstrates greater robustness to such outlier-driven deviations, and exhibits the best RMSE performance in all settings.

The naive carry-forward model exhibits pronounced sensitivity to regime-specific volatility, which reflects a key limitation of persistence-based forecasting: sensitivity to abrupt and non-persistent flow reversals. When a fund experiences a discrete capital reallocation or temporary flow shock, the naive strategy propagates the full magnitude of that disturbance into the subsequent prediction period, leading to amplified residual errors under RMSE's quadratic penalisation.

While the OLS estimator exhibits greater stability than the naive carry-forward baseline in percentage-based settings, its performance deteriorates when modeling high-dimensional absolute (USD) flows. For example, in the 1M in-sample US absolute flow setting, OLS records an RMSE of 1, 234.68, compared with 341.85 for the AUMdrive model. This reflects an inherent limitation of global linear estimation, which can overreact to large-scale capital movements in institutional funds. By contrast, the AUMdrive model's adaptive nonlinear partitioning enables it to accommodate localized heterogeneity more effectively, thereby reducing sensitivity to extreme flow realizations.

4. Explanatory power and the directional accuracy trade-off. The interaction between the coefficient of determination (R^2) and directional accuracy (DA) provides complementary insights into the models' ability to capture both the magnitude and directional structure of fund flow dynamics:

- **Explanatory robustness (R^2):** Fund flows exhibit high stochastic variability, often driving traditional benchmarks into heavily negative R^2 regimes that underperform a simple historical mean (e.g., a 1M in-sample EU percentage R^2 of –168.20 for the naive carry-forward model and –1.04 for the US OLS model). Conversely, the AUMdrive model consistently establishes a superior explanatory frontier. Although the framework records marginally negative R^2 values in exactly two isolated in-sample configurations—the 1M and 3M US nominal dollar settings (Table 6)—this reflects a deliberate regularization trade-off to suppress localized institutional noise rather than a structural flaw. The validity of this architecture is robustly confirmed by out-of-sample evaluation (Table 7), which provides the definitive, uncompromised test of real-world predictive utility as new monthly data arrives. In this critical forward-looking domain, the AUMdrive model generates uniformly positive R^2 values across all temporal horizons and regions. This transition from minor in-sample anomalies to comprehensive out-of-sample success provides strong empirical proof of generalisation, confirming that the model discards historical noise to capture invariant structural relationships. The AUMdrive model uniformly outperforms both benchmark models, consistently yielding superior R^2 metrics across all test configurations.
- **The directional accuracy trade-off:** A notable divergence emerges at longer horizons. In the 6M out-of-sample US setting, for example, the naive carry-forward model achieves a directional accuracy of 75.16%, marginally surpassing the AUMdrive model's 72.91%. However, an examination of the corresponding error metrics reveals substantially weaker magnitude calibration for the naive baseline, with mean MAE and RMSE of 10.35 and 23.72, respectively, compared with 8.42 and 18.63 for the AUMdrive model. This divergence suggests that while historical momentum may allow a persistence-based model to correctly anticipate the direction of capital movements over extended horizons, it remains less effective at estimating their magnitude. For practical applications such as liquidity planning and risk management, directional correctness alone is insufficient without accurate scale calibration. The AUMdrive model addresses this trade-off by maintaining competitive directional accuracy while delivering substantially stronger magnitude precision.

5. Cross-regional dynamics and scale effects. The interaction between relative (percentage-based) and absolute (USD) performance metrics highlights important structural differences between the US and European fund ecosystems:

- **Percentage-based predictability:** When evaluated in relative (percentage-based) terms with respect to starting total net assets (TNA), the US market exhibits higher predictability than the European (EU) market. In the 1M out-of-sample setting, the AUMdrive model achieves an MAE (Mean) of 1.97% and directional accuracy of 71.39% in the US, compared with 2.75% and 67.32% in the European market. This pattern is consistent with the US market's greater regulatory centralisation and more standardised reporting conventions, which produce a cleaner signal-to-noise ratio.
- **The nominal scale paradox:** When shifting the analysis to absolute (USD) flows, the pattern reverses, with the US market exhibiting higher absolute prediction errors. For example, under the 1M out-of-sample horizon, the AUMdrive model records an MAE (Mean) of 52.34 for the US compared to 22.16 in the European market. This difference primarily reflects scale effects rather than differences in model performance. Because US-domiciled funds operate at significantly higher average levels of total net assets, particularly in the upper size quintiles, even small percentage deviations translate into larger nominal dollar movements.

The empirical results across all prediction horizons, and for both in-sample and out-of-sample settings, indicate that the AUMdrive model consistently outperforms traditional benchmarks. Across a range of error-based and directional metrics, the AUMdrive model delivers more accurate and stable predictions of fund flows in both US and EU markets, under both relative (percentage-based) and absolute (USD) specifications.

To complement the benchmark performance analysis, we extend the evaluation to a formal pairwise comparison framework. This includes statistical tests of predictive accuracy differences as well as direct loss differentials between models. In particular, we assess whether the observed predictive gains of the AUMdrive model over the benchmark models are statistically meaningful across the different market environments considered. The results are reported in Tables 8 and 9.

6.5 Framework V: pairwise comparison statistics

While the preceding analysis focuses on standalone predictive performance, evaluating differences in predictive accuracy requires a formal pairwise comparison framework. We therefore compare the predictive errors of the AUMdrive model against the naive carry-forward model and the Ordinary Least Squares (OLS) estimator using pairwise comparison statistics.

More specifically, we evaluate loss differentials based on the Mean Absolute Error (MAE). For each observation, the loss differential is defined as the difference between the absolute error of the AUMdrive model and that of the benchmark model. Negative values therefore indicate lower predictive error for the AUMdrive model and, accordingly, superior predictive accuracy.

To quantify the magnitude and statistical significance of these differences, we report the pairwise difference in MAE (ΔMAE) along with corresponding 95% confidence intervals. The confidence intervals are estimated using two-way clustered standard errors by fund and month of prediction to account for the panel structure of the dataset, including both serial dependence within funds and cross-sectional dependence across time.

We complement this analysis with a one-sided Diebold–Mariano (DM) test to assess whether the observed differences in predictive accuracy are statistically significant. The Diebold–Mariano statistic (DM_{stat}) is computed from the mean loss differential, while inference is obtained using a pooled panel regression framework with two-way clustered standard errors by fund and month of prediction, where a negative DM_{stat} indicates lower predictive error for the AUMdrive model.

The null hypothesis (H_0) for the one-sided DM test states that the benchmark competitor exhibits equal or superior predictive accuracy relative to the AUMdrive model, implying that the expected loss differential is greater than or equal to zero ($\mu_{diff} \geq 0$). Conversely, the directional alternative hypothesis (H_a) posits that the AUMdrive model is strictly superior to its competitors, which is mathematically characterised by a negative mean loss differential ($\mu_{diff} < 0$). Under this one-sided framework, a negative DM_{stat} coupled with a small p-value provides definitive evidence to reject the null hypothesis in favor of the AUMdrive model's predictive superiority, confirming that the error reductions are not attributable to stochastic variation.

Table 8 and 9 summarise the pairwise comparison results between the AUMdrive model and the benchmark models across all forecasting horizons for the in-sample and out-of-sample data partitions, respectively. The reported statistics include ΔMAE, 95% confidence intervals (CI_{lower} and CI_{upper}), DM_{stat}, and the corresponding p-values.

Every in-sample error reduction is statistically significant

Table 8. Pairwise comparison with each benchmark, in sample

	Flow, % of starting TNA				Flow, USD millions			
	Europe		US		Europe		US	
Statistic	**vs naive**	**vs OLS**	**vs naive**	**vs OLS**	**vs naive**	**vs OLS**	**vs naive**	**vs OLS**
1M horizon								
ΔMAE	−1.74	−0.44	−0.86	−2.94	−10.43	−4.87	−11.00	−117.87
CI, lower	−2.16	−0.57	−1.03	−3.89	−12.04	−6.40	−17.19	−158.27
CI, upper	−1.33	−0.30	−0.68	−1.99	−8.82	−3.34	−4.82	−77.46
DM statistic	−8.28	−6.22	−9.74	−6.06	−12.69	−6.24	−3.49	−5.72
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
3M horizon								
ΔMAE	−3.12	−0.84	−1.86	−5.10	−19.25	−9.19	−27.12	−240.32
CI, lower	−3.57	−1.20	−2.14	−6.67	−21.80	−12.91	−35.17	−318.07
CI, upper	−2.66	−0.49	−1.58	−3.52	−16.69	−5.47	−19.06	−162.57
DM statistic	−13.40	−4.67	−13.04	−6.34	−14.77	−4.84	−6.60	−6.06
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
6M horizon								
ΔMAE	−4.60	−2.07	−3.11	−6.16	−34.37	−26.63	−46.29	−276.46
CI, lower	−5.43	−3.06	−3.52	−7.74	−41.13	−39.20	−56.35	−346.34
CI, upper	−3.78	−1.07	−2.69	−4.59	−27.60	−14.05	−36.23	−206.59
DM statistic	−10.92	−4.08	−14.62	−7.68	−9.96	−4.15	−9.02	−7.75
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

In sample, January 2020 to April 2024. ΔMAE: the AUMdrive model's mean absolute error minus the benchmark's; negative favours the AUMdrive model. CI: 95% confidence interval, two-way clustered by fund and month. DM: one-sided Diebold–Mariano test. Source: AUMdrive.

Every out-of-sample error reduction is statistically significant

Table 9. Pairwise comparison with each benchmark, out of sample

	Flow, % of starting TNA				Flow, USD millions			
	Europe		US		Europe		US	
Statistic	**vs naive**	**vs OLS**	**vs naive**	**vs OLS**	**vs naive**	**vs OLS**	**vs naive**	**vs OLS**
1M horizon								
ΔMAE	−1.47	−0.13	−0.87	−0.18	−10.26	−2.29	−20.93	−10.71
CI, lower	−1.65	−0.15	−0.98	−0.21	−11.90	−3.07	−26.67	−13.77
CI, upper	−1.28	−0.10	−0.77	−0.15	−8.63	−1.52	−15.20	−7.64
DM statistic	−15.79	−11.01	−16.12	−10.80	−12.29	−5.82	−7.15	−6.85
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
3M horizon								
ΔMAE	−2.70	−0.25	−1.70	−0.62	−18.93	−4.03	−33.15	−44.56
CI, lower	−2.91	−0.36	−1.86	−0.84	−21.32	−5.74	−40.21	−64.42
CI, upper	−2.49	−0.14	−1.53	−0.40	−16.54	−2.33	−26.10	−24.71
DM statistic	−25.25	−4.44	−19.99	−5.45	−15.53	−4.64	−9.21	−4.40
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
6M horizon								
ΔMAE	−3.78	−0.54	−1.93	−0.65	−25.28	−6.58	−26.56	−86.90
CI, lower	−4.14	−0.71	−2.17	−1.00	−29.46	−9.00	−34.76	−127.28
CI, upper	−3.41	−0.38	−1.69	−0.29	−21.09	−4.15	−18.36	−46.53
DM statistic	−20.40	−6.42	−15.86	−3.57	−11.83	−5.31	−6.35	−4.22
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Out of sample, May 2024 to October 2025. ΔMAE: the AUMdrive model's mean absolute error minus the benchmark's; negative favours the AUMdrive model. CI: 95% confidence interval, two-way clustered by fund and month. DM: one-sided Diebold–Mariano test. Source: AUMdrive.

Overall, the pairwise comparison statistics provide consistent, unambiguous evidence of the AUMdrive model's superior predictive capacity across all forecasting horizons, regions, and sampling periods. The estimated ΔMAE is uniformly negative, and the associated confidence intervals remain strictly below zero, indicating statistically robust and stable reductions in forecasting error for the AUMdrive model. This directional outperformance is formally validated by the one-sided DM test; the framework yields consistently negative DM_{stat} values alongside p-values that are strictly below 0.001 across all pairwise comparisons. Taken together, these results confirm highly statistically significant predictive gains for the AUMdrive model under a rigorous, two-way clustered inference framework.

7 Practical applications

While the preceding section establishes the statistical superiority of the AUMdrive model relative to traditional benchmarks, the practical value of these predictive gains is best illustrated through representative institutional use cases. Improved forecast accuracy does not directly alter investor behavior or prevent flow events. Instead, it enhances forward-looking visibility into fund flow dynamics, enabling more informed operational and strategic decision-making within asset management firms.

In institutional asset management, earlier and more accurate identification of large inflows and outflows can support a range of commercial applications, including sales and distribution planning, liquidity and capacity management, risk monitoring, and portfolio allocation decisions. To illustrate these practical implications, this section presents three representative case studies focused on large redemption risk, missed inflow opportunity detection, and false alarm reduction.

The case studies draw on representative examples from the out-of-sample 1-month ahead predictions for US-domiciled funds, with flows evaluated in absolute terms (USD) to provide a commercially relevant perspective on predictive performance and business impact. Each case study presents a direct comparison between the AUMdrive model and the OLS benchmark against actual fund flows. For each example, the reported statistics include the predicted fund flow (USD), the corresponding absolute error, and the resulting error reduction achieved by the AUMdrive model relative to OLS. This structure provides a transparent assessment of the incremental predictive accuracy delivered by the proposed framework and its potential operational relevance for institutional decision-making.

7.1 Case study I: large redemption risk

This case examines a large redemption event for Vanguard Total Stock Market Index Fund observed in May 2024, as represented in Table 10. The actual fund flow reflected a substantial net outflow of $2.43 billion, correctly identified by the AUMdrive model as a redemption event (–$2.27 billion). In contrast, the OLS benchmark incorrectly predicted a net inflow of $4.10 billion, failing to capture both the direction and magnitude of the actual net flow.

A $2.4bn redemption that the linear model read as an inflow

Table 10. Case study I: Vanguard Total Stock Market Index Fund, May 2024

USD millions	AUMdrive	OLS	Actual
1-month fund flow	−2,268	4,101	−2,432
Absolute error	**164**	6,533	–
Error reduction against OLS	**6,369**	–	–

Out of sample, May 2024 to October 2025; one-month-ahead forecasts. Source: AUMdrive.

In absolute terms, the AUMdrive model's prediction reduced prediction error by approximately $6.37 billion relative to OLS. Considering an annual management fee of 4 basis points and assuming that the realized TNA reduction persists over a one-year horizon, this improvement corresponds to approximately $2.55 million in annualized fee-revenue exposure for which the model provided materially improved forward-looking visibility.

This figure should not be interpreted as revenue directly preserved through the prediction itself. Rather, it reflects the incremental informational advantage created by more accurate flow prediction. Earlier identification of a materially larger redemption risk could provide asset managers with additional time to assess client concentration exposure, prioritize retention outreach, review liquidity positioning, and escalate potential outflow risks through internal distribution and risk-management channels.

From an operational perspective, the OLS forecast may have suggested a manageable or even favorable flow environment, whereas the AUMdrive model would have flagged the event as a material redemption requiring proactive commercial and liquidity-related review.

7.2 Case study II: missed inflow opportunity

This case study examines a positive flow event observed for iShares Core US Aggregate Bond ETF in September 2025, as represented in Table 11. The actual flow outcome corresponds to a net inflow of $411 million, which was correctly anticipated by the AUMdrive model ($480 million). In contrast, the OLS benchmark incorrectly predicted a net outflow of $796 million, thereby misclassifying both the direction and the nature of the flow movement.

A $411m inflow that the linear model read as an outflow

Table 11. Case study II: iShares Core US Aggregate Bond ETF, September 2025

USD millions	AUMdrive	OLS	Actual
1-month fund flow	480	−796	411
Absolute error	**69**	1,207	-
Error reduction against OLS	**1,138**	-	-

Out of sample, May 2024 to October 2025; one-month-ahead forecasts. Source: AUMdrive.

The AUMdrive model reduced prediction error by approximately $1.14 billion relative to OLS. Under the same assumption that the realized TNA effect persists over a one-year horizon, and applying an annual management fee of 3 basis points, this corresponds to approximately $342,000 in annualized fee-revenue visibility.

While this effect does not represent realized revenue, it reflects the incremental informational advantage provided by more accurate identification of inflow dynamics. In practical terms, an erroneous outflow signal may lead distribution teams to de-prioritize client engagement, while portfolio management and liquidity functions prepare for defensive positioning and unnecessary cash buffers. This can result in inefficient capital allocation, reduced sales responsiveness, and suboptimal capacity planning.

By correctly identifying the inflow event, the AUMdrive model enables earlier recognition of positive TNA momentum, supporting more effective prioritization of sales outreach, liquidity optimization, and strategic allocation of internal resources.

7.3 Case study III: false alarm reduction

This case study examines a redemption event for Fidelity Magellan Fund in April 2025, as represented in Table 12. The actual outcome corresponds to a net outflow of $425 million, which was correctly identified in direction by both the AUMdrive model (–$287 million) and the OLS benchmark (–$5.91 billion). However, the two models differ materially in terms of magnitude calibration.

An outflow the linear model overstated almost fourteenfold

Table 12. Case study III: Fidelity Magellan Fund, April 2025

USD millions	AUMdrive	OLS	Actual
1-month fund flow	−287	−5,905	−425
Absolute error	**138**	5,480	-
Error reduction against OLS	**5,342**	-	-

Out of sample, May 2024 to October 2025; one-month-ahead forecasts. Source: AUMdrive.

The AUMdrive model produces an outflow estimate that is close to the actual flow, whereas OLS overestimates the event by predicting a substantially larger redemption than actually occurred.

From an operational perspective, this type of overestimation is important because it may trigger unnecessary defensive actions, including excessive liquidity buffering, premature de-risking, and heightened internal escalation across distribution and risk-management functions. By contrast, the AUMdrive model provides a more calibrated signal, allowing institutions to respond to genuine redemption risk without overreacting to overstated signals.

This case highlights that accurate magnitude estimation is as important as directional correctness in fund flow forecasting, particularly for liquidity planning and operational decision-making in institutional asset management.

Overall, taken together, the three case studies illustrate the operational relevance of improved fund flow forecasting across distinct dimensions of institutional decision-making. Beyond directional accuracy, the AUMdrive model provides more reliable magnitude estimates and earlier identification of material flow events, reducing both missed opportunities and false alarms. These improvements translate into more effective allocation of internal resources across sales, liquidity management, and risk-monitoring functions, highlighting the practical value of enhanced predictive precision in real-world institutional operations.

8 Methodological constraints

While the AUMdrive model provides a comprehensive framework for predicting fund flows, its implementation is subject to several data-driven and sample-specific constraints. To ensure the model learns from

high-quality signals rather than statistical noise, a set of filters was applied during the preprocessing and exploratory data analysis (EDA) stages. These restrictions are primarily based on fund characteristics, namely fund type, fund age, and fund size, as well as the geographical scope of the analysis.

- **TNA-induced volatility and predictability:** During exploratory data analysis (EDA) and initial model refinement, funds with low TNA exhibit higher volatility in flow dynamics. These small-TNA funds often experience substantial short-term fluctuations in organic growth over 1–3 month horizons, which are largely driven by idiosyncratic rather than systematic factors. In addition, smaller funds may be more exposed to discretionary internal reallocations by asset managers, which are not directly observable in external market data. As a result, these dynamics introduce additional noise into the data, reducing the stability of statistical relationships between observed flows and explanatory variables.
- **Structural heterogeneity of alternative and money market funds:** Alternative funds exhibit heterogeneous and relatively unstable flow dynamics, driven by niche investment strategies, concentrated investor bases, and limited data availability and coverage. Money market funds, in contrast, are highly sensitive to monetary policy regimes and may exhibit structural breaks, particularly during periods of rapid shifts in interest rate environments, such as the transition from prolonged low rates to tightening cycles in Europe. From a statistical perspective, both categories are associated with higher volatility and reduced stability in the relationship between flows and explanatory variables, reflecting weaker and more time-varying signal structures. As a result, these fund types are excluded from the main modeling framework in order to maintain sample consistency and comparability across fund categories.
- **Stochastic dynamics of young funds:** Funds with an operational history of less than 24 months (young funds) exhibit highly stochastic flow dynamics with limited temporal persistence relative to more established vehicles. In the early stages of a fund's lifecycle, flows are often driven by seed capital, targeted launch effects, or a small number of concentrated investors, rather than stable performance-based allocation mechanisms. This results in limited regime stability and weaker time-consistent relationships between flows and explanatory variables. As a result, the inclusion of young funds may reduce the stability of estimated relationships and affect the model's ability to generalise to the broader mature fund universe.
- **Geographical scope and future expansion:** The current study is confined to US and European (EU) markets. These regions are selected due to their relatively high levels of reporting transparency and the maturity of their financial data infrastructures, which provide a reliable basis for model estimation and evaluation. Other major domiciles, including Asian markets and emerging economies in the Americas, are excluded from the present analysis due to differences in regulatory reporting frameworks and data availability. While the current implementation is geographically restricted, the modeling framework is not inherently region-specific. Future work may extend the analysis to additional jurisdictions in order to assess the model's cross-market generalisability.

9 Conclusion

This study investigates the predictability of fund flows across US and European markets over multiple forecasting horizons using the proposed the AUMdrive model machine learning framework. The empirical evidence indicates that fund flows can be predicted with meaningful accuracy using a non-linear modeling architecture designed to capture complex interaction effects in investor allocation behaviour

The predictive performance of the AUMdrive model is consistently superior to that of both the Ordinary Least Squares (OLS) and naive carry-forward benchmarks across all forecasting horizons (1M, 3M, and 6M) and geographic regions (EU and US). The model achieves lower Mean Absolute Error (MAE), reduced residual dispersion (IQR), and improved stability in error distributions, while mitigating the large error spikes and poor fit associated with linear and naive specifications, as reflected in RMSE and R^2 metrics.

These results are statistically supported by the pairwise Diebold–Mariano tests, which confirm that the AUMdrive model's improvements in predictive accuracy are significant across all models comparisons. The observed gains are therefore not attributable to random variation but reflect systematic differences in forecasting performance. The statistical tests confirm the superiority of the AUMdrive model's performance over the benchmarks in all settings.

While overall performance is robust, systematic heterogeneity persists across markets, fund characteristics, and economic regimes. Predictive accuracy is higher for US-domiciled funds relative to European funds, increases with fund size, and is strongest for Multi-Asset and Equity funds. Performance deteriorates during market stress periods, although the magnitude and persistence of these effects vary across events. Despite this, the model maintains consistent out-of-sample stability, with forward-looking performance closely aligned with in-sample results.

In conclusion, the results demonstrate that non-linear machine learning frameworks, when combined with carefully curated input data and structured evaluation design, can substantially improve the prediction of fund flow dynamics. By providing a highly calibrated balance between directional accuracy and volume scale estimation, the AUMdrive model offers significant utility for institutional financial applications. These results provide actionable predictive intelligence for asset management firms, optimising workflows in sales and distribution planning, capacity and liquidity management, risk monitoring, and portfolio allocation decision-making.

About this edition

This paper was first published in May 2026 by Evenco Analytics as “Flowmentum: A Machine Learning Framework for Accurate Fund Flow Prediction”. AUMdrive reissued it in November 2026 under its own name; Flowmentum is now the AUMdrive model, and the Persistent benchmark is called naive carry-forward. The data, model and results are unchanged. Four errors in the first edition are corrected: the stress-period figures in section 6.3 now match Tables 5 to 7; the actual flow in case study III is $425 million, as in Table 12; the left-hand side of equation (1) is the flow rather than TNA; and Figure 1 now shows the data period and the 18-month hold-out described in section 5.

References

1. M. Zanotti. On the retraining frequency of global models in retail demand forecasting. arXiv:2505.00356, 2025.
2. J. Berk and R. Green. Mutual fund flows and performance in rational markets. Journal of Political Economy, 2004.
3. R. Edelen. Investor flows and the assessed performance of open-end mutual funds. Journal of Financial Economics, 1999.
4. J. Coval and E. Stafford. Asset fire sales (and purchases) in equity markets. Journal of Financial Economics, 2007.
5. C. Jones and H. Mo. Out-of-sample performance of mutual fund predictors. The Review of Financial Studies, 2020.
6. V. DeMiguel et al. Machine learning and fund characteristics help to select mutual funds with positive alpha. Journal of Financial Economics, 2023.
7. E. Sirri and P. Tufano. Costly search and mutual fund flows. The Journal of Finance, 1998.
8. J. Chevalier and G. Ellison. Risk taking by mutual funds as a response to incentives. Journal of Political Economy, 1997.
9. M. Mazur et al. Comparing performance sensitivity of retail and institutional mutual funds’ investment flows. Finance Research Letters, 2017.
10. S. Gu et al. Empirical asset pricing via machine learning. The Review of Financial Studies, 2020.
11. T. Bali et al. Option return predictability with machine learning and big data. The Review of Financial Studies, 2023.
12. Morningstar Direct. Asset flow forecasts. Morningstar Direct research note, 23 September 2021.
13. R. Thaler. Toward a positive theory of consumer choice. Journal of Economic Behavior & Organization, 1980.
14. D. Kahneman and A. Tversky. Prospect theory: an analysis of decision under risk. Econometrica, 1979.
15. S. Ouzan. Loss aversion and market crashes. Economic Modelling, 2020.
16. W. De Bondt and R. Thaler. Does the stock market overreact? The Journal of Finance, 1985.