

Quant-Safe Explainable Artificial Intelligence for Dynamic Portfolio Management

Panagiotis Karmiris
Independent Researcher
unbinder@msn.com

January 6, 2026

Abstract

Machine learning applications in finance frequently exhibit strong in-sample performance yet fail under real-world deployment due to data leakage, look-ahead bias, and a lack of interpretability. This paper introduces a *Quant-Safe* Explainable Artificial Intelligence (XAI) pipeline designed to mitigate these risks through strict temporal validation, point-in-time feature construction, and out-of-sample explainability. Building upon recent XAI frameworks applied to Small and Medium Enterprises (SMEs), we adapt the methodology to liquid public equities and validate it on the Dow Jones Industrial Average constituents from 2015 to 2025. Using an XGBoost regressor and SHAP-based interpretation, we demonstrate that the model learns economically meaningful regime dynamics, particularly the negative impact of rising interest rates during the 2022 tightening cycle. The proposed strategy achieves significant capital preservation during market stress while remaining fully reproducible and deployable in live trading environments.

1 Introduction

Machine learning techniques have become increasingly prominent in asset pricing and portfolio construction [Gu et al., 2019]. While nonlinear models such as Gradient Boosting Decision Trees (GBDT) offer superior predictive flexibility compared to linear factor models, they introduce significant challenges related to interpretability and robustness [Chen and Guestrin, 2016].

A growing body of literature highlights that many reported financial ML successes are artifacts of data leakage, overfitting, or improper validation protocols [López de Prado, 2018, Bailey et al., 2017]. Consequently, there is a growing demand for architectures that prioritize methodological rigor and transparency over raw predictive performance.

This study contributes by proposing a *Quant-Safe* pipeline that enforces strict temporal causality, integrates explainability directly into the validation loop, and bridges the gap between academic backtesting and live trading deployment.

Our contributions are threefold:

1. We formalize a Quant-Safe architecture that eliminates common sources of financial data leakage.
2. We demonstrate the use of out-of-sample SHAP values to detect macroeconomic regime shifts.
3. We provide a fully reproducible, open-source implementation suitable for live portfolio management.

2 Methodology

2.1 The Quant-Safe Architectural Principle

The Quant-Safe principle asserts that a financial ML model is valid only if all information used at prediction time was observable at that moment in history. This principle governs feature construction, model training, validation, and explainability.

2.2 Algorithmic Overview

Algorithm 1 Quant-Safe Explainable ML Pipeline

Require: Asset prices $\{P_{i,t}\}$, macro variables $\{M_t\}$, horizon H , rebalance schedule \mathcal{T}
Ensure: Out-of-sample predictions $\hat{y}_{i,t}$, portfolio weights $w_{i,t}$, SHAP explanations

- 1: Align asset and macro data on a common trading calendar
- 2: Forward-fill macro variables only (no backward fill)
- 3: **for** each asset i **do**
- 4: Compute technical features: momentum (1M, 3M, 6M), volatility (3M), RSI
- 5: Construct label (only for evaluation): $y_{i,t} \leftarrow P_{i,t+H}/P_{i,t} - 1$
- 6: **end for**
- 7: **for** each rebalance time $t \in \mathcal{T}$ (walk-forward evaluation) **do**
- 8: Train model on labeled history $\{(\mathbf{x}_{i,\tau}, y_{i,\tau}) : \tau < t\}$
- 9: Score out-of-sample predictions $\hat{y}_{i,t} \leftarrow f(\mathbf{x}_{i,t})$
- 10: Compute SHAP values on out-of-sample fold only
- 11: Form portfolio: select top N , apply volatility scaling, caps, and cash buffer
- 12: **end for**
- 13: **Live inference mode:** train once on all labeled history; score latest unlabeled rows
- 14: Persist signals, trades, positions, and mark-to-market performance logs

2.3 Feature Engineering

The model combines technical and macroeconomic predictors:

- Momentum (1M, 3M, 6M)
- Volatility (3M rolling)
- Relative Strength Index (RSI)
- Macroeconomic indicators: S&P 500, VIX, crude oil, gold, and U.S. 10-Year Treasury yield

Macroeconomic variables are transformed into rolling z-scores to capture regime deviations rather than absolute levels.

2.4 Model Specification

We employ an XGBoost regressor with a squared-error objective. Hyperparameters are selected conservatively to avoid excessive model complexity. Importantly, retraining occurs only within the validation loop and never during live inference.

2.5 Explainability via SHAP

Shapley Additive Explanations (SHAP) are computed strictly on out-of-sample predictions [Lundberg and Lee, 2017]. This avoids attribution bias and enables a historical record of how the model's decision logic evolves across regimes.

2.6 Data Leakage Failure Modes and Mitigations

Look-Ahead Bias: Prevented by excluding unlabeled rows from training.

Temporal Feature Leakage: Prevented through forward-only macro alignment.

In-Sample Explainability Bias: Prevented by computing SHAP exclusively on test folds.

Validation–Production Mismatch: Walk-forward retraining is used only for validation, while live inference uses a single model trained on all available history.

Universe Selection Bias: This study uses current Dow Jones constituents, introducing survivorship bias. As such, the focus is on methodological robustness rather than absolute historical return estimates.

3 Portfolio Construction

Assets are ranked by predicted 6-month returns. The top $N = 5$ assets are selected and weighted using inverse-volatility scaling:

$$w_i = \frac{1/\sigma_i}{\sum_{j=1}^N 1/\sigma_j}$$

Position caps and a cash buffer are applied. Transaction costs of 15 basis points per turnover are assumed.

4 Results

4.1 Performance Evaluation

The strategy was evaluated on Dow Jones constituents from 2015–2025.

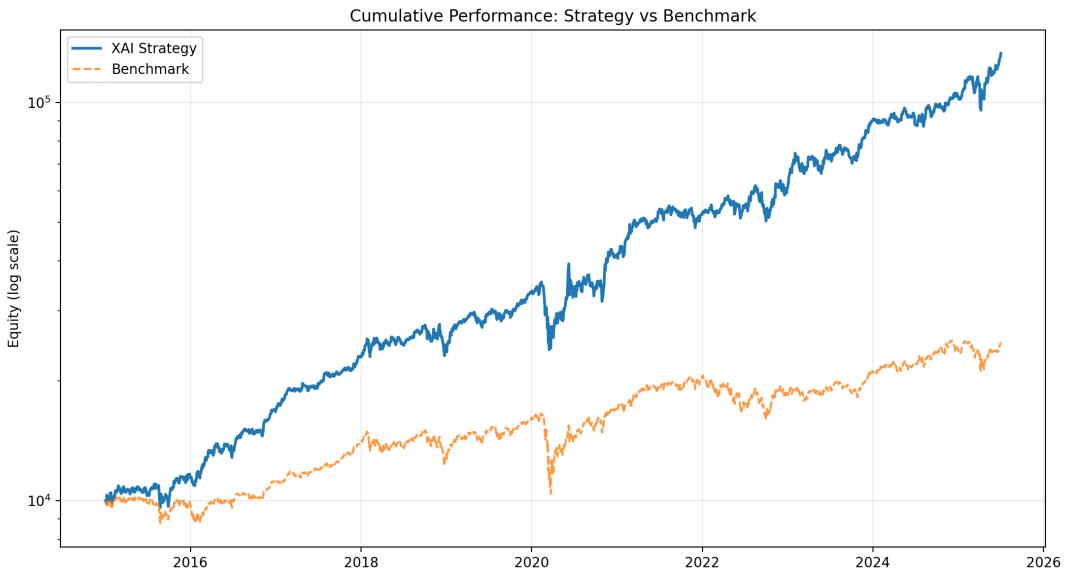


Figure 1: Cumulative equity curve (log scale) versus Dow Jones benchmark.

The model exhibits strong drawdown control during the 2022 tightening cycle, a period during which traditional momentum strategies underperformed.

4.2 Explainability and Regime Detection

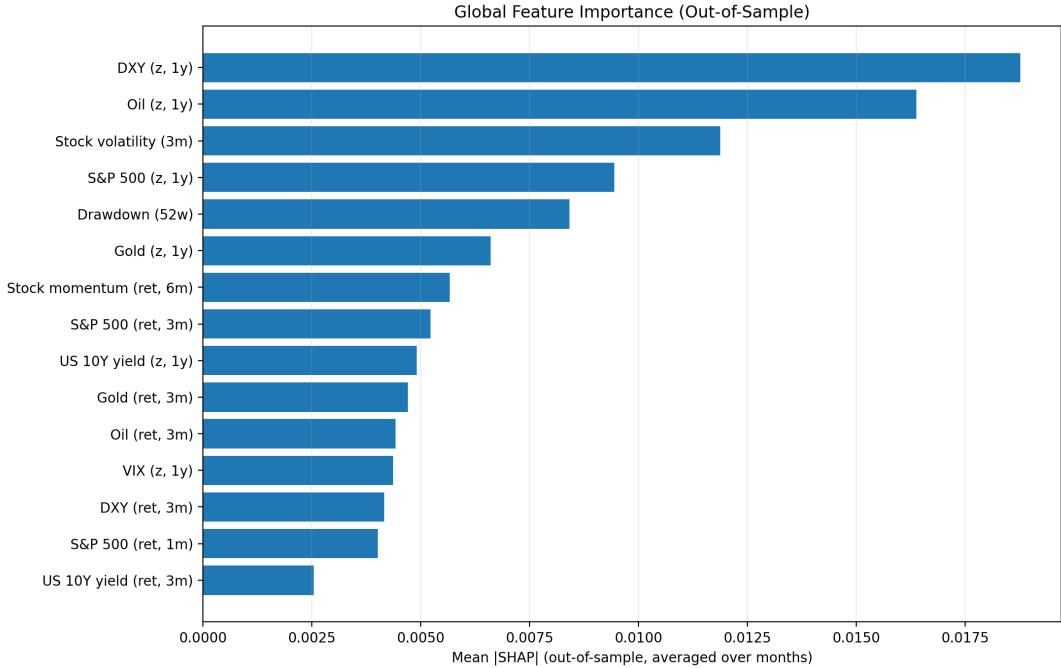


Figure 2: Global out-of-sample SHAP feature importance. Interest rate variables dominate during stress regimes.

The dominance of U.S. 10-Year yield SHAP values during 2022 indicates that the model internalized macroeconomic valuation effects rather than relying solely on price trends.

4.3 Comparison with Naïve ML Backtests

Table 1: Quant-Safe Pipeline vs Naïve ML Backtests

| Aspect | Naïve ML | Quant-Safe |
|-----------------------|----------------------|--------------------|
| Temporal validation | Random / K-fold | Walk-forward |
| Feature timing | Implicit future data | Point-in-time only |
| Explainability | In-sample | Out-of-sample |
| Portfolio constraints | None | Volatility-scaled |
| Live deployability | No | Yes |

5 Conclusion

This paper demonstrates that robust, explainable financial ML systems are achievable when strict temporal discipline is enforced. The Quant-Safe architecture provides a reproducible blueprint for bridging academic modeling and real-world trading, emphasizing interpretability and capital preservation over fragile headline returns.

Future work will extend the framework to other markets and incorporate point-in-time index membership data.

References

- David H Bailey, Jonathan Borwein, and Marcos López de Prado. The probability of backtest overfitting. *Journal of Computational Finance*, 2017.
- Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD*, pages 785–794, 2016.
- Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *Journal of Financial Economics*, 131(2):335–360, 2019.
- Marcos López de Prado. *Advances in Financial Machine Learning*. Wiley, 2018.
- Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, 2017.