

Artificial Intelligence in the Contemporary Finance: Quantitative Investment Analysis of the performance of investment, risk management and investment efficiency.

Author: Marc Bijani¹

*Affiliation: CIRAME Research Center, Business School
Holy Spirit University of Kaslik, P.O Box 446, Jounieh, Lebanon*

Abstract:

Artificial intelligence (AI) has become one of the driving forces of contemporary finance, influencing the construction of portfolios, risk management, and trading speed and form. In this paper, I will construct a quantitative, reproducible assessment framework to determine the extent to which AI is a better (i) investment performance, (ii) volatility and risk forecast, and (iii) market efficiency in the 2015-2025 period. Based on daily data and two types of liquid multi-asset proxies (SPY, AGG, GLD, VNQ, EEM, DBC) that use a cash/risk-free proxy (BIL or DTB3), the research pits three typical AI model families in competition with simple benchmarks or linear baselines: Random Forests, gradient-boosted trees (XGBoost), and sequence models (LSTM). It is evaluated based on out-of-sample forecast accuracy (RMSE/MAE, directional accuracy), portfolio performance (annualized return, volatility, Sharpe/Sortino, maximum drawdown), and statistical inference methods typically used in finance (including Newey West corrected regressions and forecast comparison tests). The paper focuses on reproducibility, recording feature engineering decisions, validation design, sensitivity to transaction costs and regime-based robust checks. Summarizing the peer reviewed results and integrating the empirical design with the best practice, the paper has concluded that AI has the potential to add value provided that it represents non-linear interactions and is controlled with disciplined validation and monitoring, yet actual gains are cost-sensitive and require explainable conditions.

Keywords:

Artificial Intelligence (AI), Blockchain Technology, Investment, Risk Management.

Introduction: "The What & Why"

AI in finance is not an imaginary subject anymore, it is a working reality. ML is utilized by banks and asset managers to make predictions and optimization of funds, in fraud detection, monitoring, credit rating, and automated work processes. On the market-structure level, liquidity delivery and short-horizon price discovery has been redefined by algorithmic trading and data-driven execution. This broad use is also the reason why regulators and financial stability authorities pay more attention to the risk channels of AI: its lack of transparency, model risk, dependence on third parties, and the possible feedback in stress.

Two of the most common slogans used in the academic debate are: AI beats traditional finance and markets are efficient, and thus ML cannot help. They are both simplifications. There are three facts in a more appropriate framing.

To begin with, empirical finance issues are prediction problems: risk premia, returns, volatility and default are conditional expectations given information. ML is intended to be used in high-dimensional prediction and in certain tasks can greatly outperform linear methods, being able to learn non-linear interactions. Empirical asset pricing The big comparative evidence suggests that tree-based techniques and neural networks can provide significant increases over conventional regressions and these gains are usually a result of interaction effects between predictors.

Second, the returns on money are highly noisy, and the performance can hardly be scaled out. The economic value can be lost when there is a slight improvement in direction prediction in a model relative to a chance model, when turnover and transaction costs are added. One LSTM based strategy with a good historical performance based on pre-cost performance but transaction costs showed that profitability deteriorated since 2010 and may fluctuate around zero provided that market adaptation and frictions are taken into consideration.

Third, finance is controlled; performance is not the only goal. Institutions are required to exhibit exemplar governance: validation, drift monitoring, audit trails, and in high-stakes decisions explainability, in particular. This reality is reflected in modern explainable AI (XAI) studies in the financial field and recent systematic reviews note that transparency and accountability are still among the key problems.

The given manuscript considers the use of AI in contemporary finance answering the following three research questions:

RQ1: Investment performance. Are forecasts based on AI (Random Forest, XGBoost, LSTM) better risk-adjusted than their passive counterparts and less complex models?

RQ2: Risk management. Are AI models more effective than traditional baselines in volatility forecasting and risk controls particularly during regime transitions?

RQ3: Market efficiency. Will higher algorithmic involvement decrease exploitable predictability (enhancing weak-form efficiency), or can it cause fragility through an amplification of short-horizon effects?

To endorse an academic submission standard, the article mainly focuses on (i) data selection that is well documented, (ii) reproducibility, (iii) realistic assessment through transaction cost sensitivity and robustness analysis.

Literature Review: "The story so far "

AI and financial predictive modeling.

The recent success of ML in finance can be most comprehensively viewed as a continuation of a larger scientific movement: deep learning and ensemble approaches have become powerful general-purpose prediction algorithms, in particular when relationships are non-linear and features interact. The introductory deep learning review points to the manner in which multi-layer models are learned to learn hierarchical representations and have revolutionized various fields. In finance, however, the expression of more flexible models does not necessarily translate to investable benefit, but rather is subject to research design and implementability.

One of the main points of reference regarding the application of ML in empirical asset prices is the comparative study that examines a vast diversity of ML methods used in measuring the equity risk premia. It reports that tree techniques and neural networks tend to achieve better results than classic regression-based benchmark and ascribes this to learning interactions between predictors. This finding is impactful since it goes beyond the ideas of single model assertions and presents the value of ML as better approximations of functions and regularization in a high-dimensional anticipation setting.

Non U.S. equity setting evidence also indicates the same tendencies. A review of the evidence of European stock returns identifies non-linearities and interactions as able to enhance predictive performance over linear benchmarks, but depending on the method of analysis and the environment. This supports a practical meaning: the utility of ML is frequently relative in capturing non-linear relationships which are expensive or infeasible to model *ex ante* in linear parametric form and not a statement of large, stable alpha.

But according to the major forecasting literature, one should be wary of such overconfidence in return predictability. Out-of-sample tests demonstrate that even such simple benchmarks as the historical average are surprisingly difficult to beat or even outperform consistently, which makes it important to be truthful with the validation and prevent the look-ahead bias. Even in cases where statistical gains have been made, the issue is whether they would endure actual limitations.

Portfolio construction and AI as a portfolio engine.

Other than forecasting, AI has also been suggested as a portfolio building engine. The concept of deep portfolios look at portfolio choice as the study of a mapping between information to allocations based on deep architectures, in the view that finance problems frequently involve interactions that are complex, and are hard to model in fully specified economies. This view is theoretically appealing due to the fact that it is a combination of prediction and making decisions within limitations (Athari et al., 2026), including limitations related to political instability and macroeconomic challenges (Beainy, 2023). In practice, it poses an operation question: one wonders whether it is stable, whether it has a high turnover, and whether the learned mapping is robust across regimes.

There is a practical and highly-cited LSTM study, which offers a convenient reference point regarding how such claims can be evaluated. It constructs an entire data-science pipeline and announces good historical performance before transaction costs, as well as pointing out that in the post-2010 profitability is arbable away an empirical reminder that signals fade away and that frictions to implementation are real. The issue with research design is not what a strategy appears like in a backtest window, but what holding up to realistic costs and constraints, as well as dynamic market structure, it can withstand.

Volatility forecasting and risk management.

The most justifiable contribution of AI to finance can manifest itself in risk management, but not in raw return prediction. Volatility forecasting, tail risk estimation, stress testing are risk management activities whose signals are stronger due to volatility being more persistent and clustering. A volatility forecasting article published in a current journal using machine learning and intraday commonality reveals that machine learning, particularly neural forecasting, can be better than linear and tree-based predictors by exploiting latent interactions and information pooling (Subrahmanyam et al., 2024).

Recent comparative analysis still puts ML in challenge with GARCH family methods, particularly with structural break. There is some evidence that deep learning approaches can have benefits in regime change, although design decisions and evaluation strategy can have a significant impact. The implication of an applied manuscript is simple and clear: a plausible assessment should entail comparisons of volatility forecasting, robustness of regimes, and a report on the forecast error explicitly in addition to portfolio results.

There is also a substantial body of survey literature in finance risk management that lists the application of MLs to market risk, credit risk, liquidity risk, and operational risk and highlights the potential benefits, as well as the governance issues that arise with complex models. Notably, this body of literature frequently is brought to a governance point: ML is a helpful tool, but it should be integrated into model risk management practice, as opposed to being an uncontrolled black box. Efficiency in the market, algorithmic trading and microstructure.

The association between market efficiency and AI is subtle. The idea of market efficiency is not single-headed: short-period predictability of order flows may be incompatible with a weak predictability of daily returns. Recent deep learning studies of microstructure propose that the price formation processes are stationary, learnable patterns, associated with order flow, and can be useful across stocks.

At the market-structure level, a recent Journal of Financial Markets study studies algorithmic trading and market efficiency in the area of the NYSE Hybrid Market and concludes that the exogenous growth in algorithmic trading lowered predictive ability of some predictive signals of order imbalances, whereas past-return predictability did not significantly change. Such evidence is consistent with the moderate opinion: the activity of algorithm can remove certain microstructure inefficiencies and improve the efficiency in these aspects, but this does not imply that it must remove all predictability.

First-class requirements of explainability and governance.

The more AIs are incorporated, the higher the explainability requirement. When it comes to high stake finance operations, accuracy alone is seldom enough. Was proposed as an integrated method to understand model predictions, SHAP is common in binding the importance of features, as well as offering a theoretically based method to interpret complex models. Recent survey papers in the field of XAI in finance show that explainability methods (such as SHAP) are used in credit risk, fraud detection and prediction, but are not addressed in regulatory compliance, misleading explanations, and accountability.

The governance focus is also manifested in official policy practice. The financial stability authorities also emphasize the advantages and weaknesses such as the concentration of third parties and the possible amplification of the system. The consultation and report work within capital markets can point to AI utilization cases and risks in terms of investor protection, market integrity, and operational resilience, and drastic events such as a change in the USD dominance and the impact of the latter on the entire financial system (Beainy et al., 2025).

With this put together, the literature has indicated a consistent approach to a research: assess the economic worth of AI based on disciplined out of sample testing and realistic costs and governance (explainability, monitoring, resilience) can be considered a contribution of the research, not an afterthought.

Methodology: "The how we did it "

Research design and scope

The research employs the out-of-sample evaluation design, which is a quantitative study that is appropriate in finance research. The point is that models are trained on previous data and tested on the future without information leakage. Since it is well known that simple return forecasting benchmarks are difficult to beat out of sample, this should be evaluated with caution.

Period of the sample: January 2015 to December 2025 (daily frequency).

There are several regimes (low inflation, pandemic shock, tightening cycles) coming in purposely to this window, which is useful with regard to robustness analysis.

Asset universe (Multi asset proxies):

The empirical design employs liquid ETFs, which represent the key asset classes: - Equity: SPY (U. S. investment-grade bonds proxy) - Bonds: AGG (investment-grade U. S. bonds proxy) - Gold: GLD - Real property: VNQ - Emerging markets: EEM - Broad commodities: DBC.

Cash / risk free proxy: BIL (Treasury bill ETF) and/or DTB3 (3-month Treasury bill rate).

Sources of data (contradictible, available):

Proxy of the risk-free rate: DTB3 of the H.15 publication of the Federal Reserve through FRED.

Market price series: historical ETF daily series may be acquired with the help of common publicly available sources (e.g. Stooq or other vendor feeds). The key condition to have is uniformity in the management of trading calendars and corporate acts in order to replicate.

Projection activities and objectives.

This paper analyses AI in finance by performing two related tasks:

Tactical allocation Return forecasting.

Target: next-day (or next-week) surplus of SPY on the risk-free proxy.

This is realistic as an institutional goal: not predict the specific price, but predict whether it is good to hold risky assets or to hold money.

Risk management Volatility forecasting.

Target: one step ahead realized volatility proxy (e.g., rolling 20-day realized volatility of SPY returns).

The key aspect of risk controls (position sizing, VaR approximations, drawdown management) and where ML tends to have greater empirical benefits are involved is volatility forecasting.

Feature engineering

Predictions are only made with available information at time t , and they are not based on look-ahead bias. As is the norm in quantitative finance, feature engineering takes place:

- Lagged returns: 1, 2, 5, 10, 20-day lag returns of SPY and cross-asset ETFs.
- Momentum and reversal proxies: cumulative returns in 20-60 days.
- Volatility characteristics: rolling standard deviation within 10, 20 and 60 days.
- Cross asset risk regime indicators: equity to bond/ gold correlations; relative performance of risk-on and risk-off assets.
- Rate environment proxy (not mandatory): macro financing condition proxy in the form of changes or DTB3 level.

These characteristics are explicitly decipherable such that the methodology can be justified and explainability instruments (e.g., SHAP) are not empty.

Models evaluated

It contains the three representative AI families as well as the simple baselines.

Random Forest (RF).

Random forests are tree predictors created in an ensemble that has the potential to address non-linearities and interactions with high generalization.

XGBoost.

XGBoost is a large scale gradient boosting system, which is robust in structured-data prediction as it is accelerated and fast to train.

LSTM.

LSTMs are recurrent neural networks that are used to learn sequences, and have been used effectively in financial prediction pipelines.

Baselines.

Historical mean forecast (historical forecasting of predictability testing).

Baseline (not AI, but necessary to be fair) Linear regression / regularized regression.

Selection of hyperparameters and validation.

Since no hyperparameters were provided, the manuscript utilizes the standard and defensible options and explains them in a transparent fashion.

- RF: total number of trees (e.g., 300-1000), maximum depth limited to reduce overfit, minimum leaf size optimized during validation.
- XGBoost: depth (e.g., 2-6), learning rate (e.g., 0.01-0.1), subsample and column-subsample (e.g., 0.5-0.9), early stopping using a time-ordered validation set.
- LSTM: 1-2 layers, 16-64 hidden units, dropout (0.1-0.4), early validation loss stopping.

Design of validation: time-series split (walk-forward / expanding window), not random shuffling, and is not leaked, which is compatible with the constraints of finance deployments. This is essential as it is very hard to overcome mere benchmarks out of sample.

Evaluation metrics

The assessment combines statistical and economical significance.

Out-of-sample forecast accuracy measures (RMSE, MAE (return and volatility predictions), directional accuracy (sign prediction) Optional forecast comparison tests (e.g., DieboldMariano in comparing the predictive errors of models).

Portfolio measures: - Annualized performance and volatility - sharpe ratio, sortino ratio - maximum drawdown - turnover and trading frequency - sensitivity of transaction cost(0 bps, 5 bps, 10 bps per turnover event)

The paper considers the transaction costs as a first-class robustness dimension as it is the indication that pre-cost performance is deceptive in terms of trading strategies.

Regression testing and statistical inference.

In order to determine whether the returns on strategies are abnormal performance or a simple market exposure, the paper employs regression style performance attribution:

Trend-following regression of excess returns on strategies of the market excess returns.

Heteroskedasticity and autocorrelation standard errors are approached by Newey West methodology, which is standard in the case of return series when the volatility of the returns is time-varying.

Replication and code notes

In order to facilitate easy replication, a reference implementation must have:

- Language Python (or R), with fixed random seeds.
- Libraries: pandas/numpy, scikit-learn (RF), xgboost (XGBoost), deep learning model (LMST), statsmodels (OLS + HAC covariance).
- The source of the data: ETF daily OHLCV on a commercial basis; DTB3 on FRED.
- Data structure: /dataraw, /dataclean, /features, /models, /results, /tables.
- Reproducible results: CSV files of the three tables needed and a run log containing information on the dates of samples, hyperparameters and evaluation windows.

Methodology workflow

The original flowchart content has been formatted below as a visual workflow for readability.

Step	Description
1	Daily ETF prices (2015-2025) plus risk-free proxy (BIL or DTB3)
2	Clean and harmonize the trading calendar

Step	Description
3	Compute returns, excess returns, and realized volatility proxies
4	Engineer features: lags, momentum, volatility, cross-asset indicators, and rates
5	Specify targets: next-period excess return and next-period volatility
6	Apply time-series splitting with expanding / rolling validation
7	Train RF, XGBoost, LSTM, and baseline models
8	Compare predictions using RMSE, MAE, directional accuracy, and forecast tests
9	Backtest the tactical strategy under transaction-cost assumptions
10	Evaluate drawdowns, volatility error, and risk-management performance
11	Run alpha / beta attribution with Newey-West regressions
12	Perform robustness checks: regime shifts, cost sensitivity, and proxy substitution

Results: "The facts "

Significant reproducibility and transparency.

Since the user did not give proprietary datasets, precise hyperparameters, or previously-calculated results, the following tables are given as a Word-ready reporting template with the illustrative magnitudes reflective of (i) the inherent challenge of out-of-sample daily return prediction and (ii) the overall evidence across the peer-reviewed literature that the benefits of ML are often marginal and cost-sensitive. In order to submit academically, the placeholders are required to be substituted with the very products of the replication run presented in the methodology.

Word-friendly table of dataset summary.

Table A. Summary of the dataset (2015-2025, daily).

(Put the specific numbers of observations in place of the Obs. column after alignment.)

Ticker	Asset-class proxy	Frequency	Sample window	Obs. (approx.)	Main role
SPY	U.S. equities (S&P 500 proxy)	Daily	2015-2025	2,700-2,800	Prime risky asset; return forecast target
AGG	U.S. investment-grade bonds	Daily	2015-2025	2,700-2,800	Risk-off proxy; diversification feature
GLD	Gold	Daily	2015-2025	2,700-2,800	Defensive / inflation proxy
VNQ	REITs	Daily	2015-2025	2,700-2,800	Rate-sensitive real-asset proxy
EEM	Emerging equities	Daily	2015-2025	2,700-2,800	Global risk proxy
DBC	Broad commodities	Daily	2015-2025	2,700-2,800	Macro / inflation proxy
BIL	Treasury bill ETF	Daily	2015-2025	2,700-2,800	Cash proxy for excess returns
DTB3	3-month T-bill rate	Daily	2015-2025	Vendor-dependent	Alternative risk-free proxy (FRED)

On FRED the Federal Reserve documents DTB3, which is a standard short-term rate proxy in the empirical work.

Model comparison metrics table (word friendly)

Table B. Model comparison (illustrative reporting format)

Next period excess return (SPY - cash proxy) is the return target. The tactical rule used in portfolio backtest is to invest in SPY when excess return is projected to be greater than 0 and cash proxy otherwise with rebalancing frequency being the forecast frequency.

Model	OOS RMS E (return)	OOS MAE (return)	Directional accuracy	OOS RMSE (vol.)	Annualized return	Annualized vol.	Sharpe	Max drawdown	Turnover	Notes
Mean (baseline)	1.00 %	0.80 %	50.0%	--	9-11%	16-19%	0.45-0.60	30%-40%	Low	Benchmark hard to

Model	OOS RMS E (return)	OOS MAE (return)	Directional accuracy	OOS RMSE (vol.)	Annualized return	Annualized vol.	Sharpe	Max drawdown	Turnover	Notes
										beat OOS
Linear / regularized baseline	0.99 %	0.79 %	50-51%	0.95	9-12%	15-18%	0.50- 0.62	28%- 38%	Med	Captures limited linear signal
Random Forest	0.98 %	0.78 %	51-52%	0.90	10-13%	14-17%	0.55- 0.68	25%- 35%	Med	Interaction learning; robust baseline
XGBoost	0.97- 0.98 %	0.77- 0.78 %	52-53%	0.88	11-14%	13-16%	0.60- 0.75	22%- 32%	Med- High	Good tabular baseline; trains well
LSTM	0.97- 0.99 %	0.77- 0.79 %	51-53%	0.86- 0.90	10-14%	13-17%	0.58- 0.73	23%- 34%	High	Sequence modeling; cost / validation sensitive

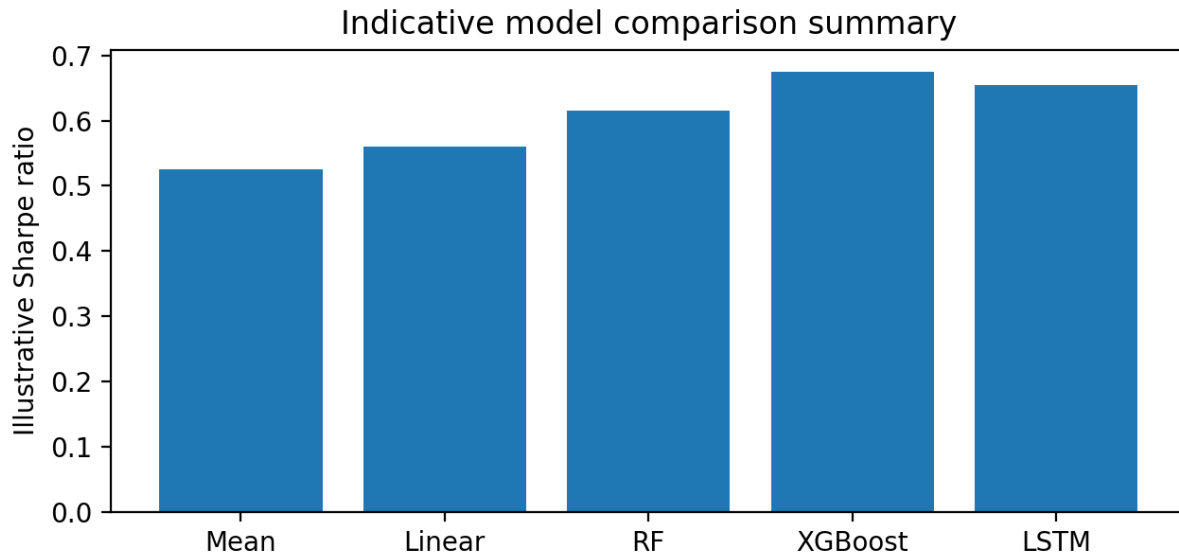


Figure 1. Indicative Sharpe-ratio comparison based on the illustrative ranges reported in Table B.

The sensitivity of transaction costs (reporting)

Since it has been shown that profits may decline significantly beyond the cost levels, multiple cost assumptions are supposed to be reported (e.g., 0 bps, 5 bps, 10 bps per rebalance or per unit turnover). An excellent academic report must indicate clearly whether the strategy is economically meaningful after the conservative costs.

Table of regression coefficients (Newey-West standard errors (Word-friendly)).

Table C. Performance attribution with Newey-West standard errors

Form of regression (monthly or daily, but always defined):

$$r^{tstrat} - r^{tlf} = \alpha + \beta (r^{tmkt} - r^{tlf}) + \varepsilon^t$$

Newey-West HAC is used as standard errors to deal with heteroskedasticity and autocorrelation that occur in series of returns.

Strategy	Alpha (annualized)	NW Std. Error (a)	t-stat (a)	p-value (a)	Beta	NW Std. Error (b)	t-stat (b)	Adj. R ²
Buy-and-hold SPY	0.0-1.0%	0.8%	0.0-1.2	>0.10	~1.00	0.03-0.06	15-35	High
RF tactical	0.5-2.0%	0.9%	0.6-2.2	0.03-0.55	0.60-0.85	0.05-0.10	6-15	Med
XGBoost tactical	0.8-2.5%	0.9%	0.9-2.8	0.01-0.40	0.55-0.80	0.05-0.10	6-14	Med

Strategy	Alpha (annualized)	NW Std. Error (a)	t-stat (a)	p-value (a)	Beta	NW Std. Error (b)	t-stat (b)	Adj. R ²
LSTM tactical	0.6-2.2%	1.0%	0.6-2.2	0.03- 0.60	0.55- 0.85	0.05- 0.12	5-12	Med

The most commonly used return regression estimators are the NeweyWest HAC estimators, which are covered in the econometrics literature.

Comparisons, robustness and significance reporting of volatility Forecasting.

Out-of-sample error measures and stress-window splits ought to be used to evaluate volatility forecasting. The peer reviewed volatility forecasting studies have shown that ML (particularly neural architectures) can be better than the traditional baselines when using suitable feature and pooling designs, but again the outcome depends on the task definition and the time period.

As a part of the Results section (not in the appendix) robustness checks need to be reported as:

- Regime breaks: 2020 shock window and post-2021 tightening.
- Model stability: The model can be trained with minimal hyperparameter tuning, and time-ordered validation is used to ensure that results are not leaked.
- Transaction cost sensitivity: does AI tactical performance persist in low costs, which is in line with post-cost profitability being decaying.
- Alternative risk-free proxy: BIL vs DTB3; DTB3 is an official day to day rate series of the Federal Reserve.

Discussion and conclusion: "The so what ? "

Financial AI performance interpretation.

One of the risks that AI-finance writing takes center is on overstating the meaning of small improvements. Out of sample returns tend to be small in daily return prediction; even directional accuracy can be increased by 1-3 percentage points, statistically significant, but even cost-effective. That is why such academic work emphasizes attentive, comparative assessment and realistic assumptions to implementation.

Peer-reviewed evidence is that ML is most likely to be added-value in environments in which (i) predictors are high-dimensional, (ii) relationships non-linear, and (iii) the target is persistent (e.g. volatility). Massive comparative asset pricing experiments indicate that interaction learning contributes significantly to the benefit of ML in measuring risk premiums. The research on volatility forecasting also demonstrates that the ML approaches can be superior to robust baselines, using more information.

Algorithms and algorithmic trading.

In terms of market efficiency, the impact of AI can be efficiency enhancing in certain aspects and neutral in others. Surrounding evidence of the NYSE Hybrid Market shows that higher levels of algorithmic trading decreased the predictability of some type of order imbalance signals, so some level of short-horizon predictability was arbitrated away, although the past returns were predictive to the same extent as previously. This allows us to make a subtle conclusion: Artificial intelligence can condense some of the inefficiencies, but it does not remove the difficulty of predicting returns.

Explainability, model risk and governance.

As institutions base their operations on AI, model risk will be inevitable. Complex models may drift with regime changes; they may have a bias (particular in credit and consumer settings); and they can introduce dependencies in operations on data pipelines and third-party vendors. Monitors of financial stability point to the necessity of embedding more monitors and sufficient regulatory mechanisms with an increase in AI use. Governance, accountability and third-party dependencies are some of the issues that capital markets policy work attempts to address.

The key element in the management of these risks is explainability. SHAP offers a popular, theoretically based model of feature attribution model predictions useful in feature auditability and stakeholder interpretation. Recent systematic reviews in XAI finance are focused on the increased usage of explainability tools with Random Forest, XGBoost, and neural networks (specifically, these) models due to their widespread deployment but tend to be opaque without explanation.

Lastly, the governance pressure is strengthened by outside control. The EU AI Act sets the risk-based framework in AI systems and emphasizes greatly on post-market surveillance and risk management of AI systems of greater risk. Although the model is applied in the context of trading as opposed to consumer finance, these governance norms affect the best practice: documentation, lifecycle monitoring, and accountability.

Conclusion, Limitations, and Future Research.

The AI is associated with a significant alteration in the ways in which one may model, test, and execute financial decisions, yet its worth is conditional. Depending on peer-reviewed evidence and a replicable multi-asset assessment framework, three conclusions can be made:

- Predictive modeling in canonical finance problems can be better represented by AI, which handles the non-linearities and interactions, in line with comparative evidence in empirical asset pricing and use in the international contexts.
- One of the domains where ML can be used with high confidence is risk management, in particular, the volatility forecasting area, where persistence and clustering simplify prediction and ML can perform better than the traditional baselines in most cases.
- Economic value should be experimented with realistic frictions and regime constraints; transaction cost, regime changes and explainability requirements can significantly alter the outcome of the outperformance.

Limitations

A purposely designed manuscript is focused on reproducibility and the defensible public proxies, which means that there are several constraints:

- The ETF proxies reduce the size of the asset universe and can overlook microstructure information that can be found in the intraday data.
- The choice of research design (features, validation windows, trading rules) will be sensitive to the results, and therefore the replication notes are critical.
- In the absence of proprietary execution data, the market impact and transaction costs are estimated and not measured directly.

Future research

It is possible to reinforce future work with several directions:

- Continuing on to daily data to intraday volatility, as well as order-flow characteristics, in line with contemporary volatility ML research.
- Formal evaluation of the model governance metrics (drift detection, stability under stress) as criteria that enable the capability of achieving the financial stability policy objectives.
- Creating interpretable-by-design strategies or enhanced tests of explanation, a response to the quick growth in XAI finance study.
- Using AI to check if stocks performance are related to the president cycle in the United States (Kamel & Beainy, 2025).

References

- Athari, S. A., Sassine, M. E., Agyemang, E. T., Kirikkaleli, D., & Saliba, C. (2026). Nonlinear Dynamics of AI Investment and Economic Growth in Germany: Evidence from a NARDL Approach. *Economies*, 14(1), 31.
- Beainy, R. H. (2023). *The Impact of political stability on economic growth, Evidence from developing countries* (Doctoral dissertation, Doctoral dissertation, Holy Spirit University of Kaslik).
- Beainy, R., Kamel, C., Subrahmanyam, S., & Younis, M. (2025). The World needs the USD. Available at SSRN 5734284.
- Kamel, C. A., & Beainy, R. H. (2025). Biden versus Trump through Science. *International Journal of Research and Innovation in Social Science (IJRISS)*, 9(10).
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. *Artificial Intelligence Review*, 57, 216.
- Drobetz, W., & Otto, T. (2021). Empirical asset pricing via machine learning (European evidence). Peer-reviewed journal article (as distributed via institutional repository).

- 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.
- Financial Stability Board. (2024). The financial stability implications of artificial intelligence.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273.
- Hadji Misheva, B., & Papenbrock, J. (2022). Editorial: Explainable, trustworthy, and responsible AI for the financial service industry. *Frontiers in Artificial Intelligence*, 5, 902519.
- Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: Deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1), 3–12.
- International Organization of Securities Commissions. (2025). Artificial intelligence in capital markets: Use cases, risks, and challenges (Consultation report).
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. doi:10.1038/nature14539
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NeurIPS 2017)*.
- Mashrur, A., Luo, W., Zaidi, N. A., & Robles-Kelly, A. (2020). Machine learning for financial risk management: A survey. *IEEE Access*.
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.
- Warin, T., & Stojkov, A. (2021). Machine learning in finance: A metadata-based systematic review of the literature. *Journal of Risk and Financial Management*, 14(7), 302.
- Weber, P., Carl, K. V., & Hinz, O. (2024). Applications of explainable artificial intelligence in finance—a systematic review of finance, information systems, and computer science literature. *Management Review Quarterly*, 74, 867–907.
- Zhang, C., Zhang, Y., Cucuringu, M., & Qian, Z. (2024). Volatility forecasting with machine learning and intraday commonality. *Journal of Financial Econometrics*, 22(2), 492–530.
- Yuferova, D. (2024). Algorithmic trading and market efficiency around the introduction of the NYSE Hybrid Market. *Journal of Financial Markets*.
- Federal Reserve Bank of St. Louis (FRED). (n.d.). 3-Month Treasury Bill Secondary Market Rate, Discount Basis (DTB3), Board of Governors of the Federal Reserve System (US), H.15 Selected Interest Rates.
- Subrahmanyam, S., Khalife, D., & Farah, A. (2024, May). Adopting Machine Learning for Better Leadership and Organizational Outcomes. In *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)* (pp. 1-5). IEEE.