Convex Optimization for Explainable Machine Learning: A Sparse Regularization Perspective

Yun Tian *

Department of Mathematics, The University of British Columbia, Vancouver, Canada

* Corresponding Author Email: joytian2003@gmail.com

Abstract. In recent years, the use of machine learning techniques for predicting the stock market has grown quickly. However, more and more people concern model explainability since the "black box" model problem has come up as a result. This essay examines the impact of convex optimization on explainable machine learning from the standpoint of mathematical optimization. It shows that Lasso (L_1) and Ridge (L_2) regularized objectives are convex, and that adding L_2 makes them strongly convex. Thus, the unique global optimal solution is obtained. This research use Apple's stock data from 2019 to 2024 to build sparse regression models like the Lasso, Ridge, and Elastic Net. Then this research compares how well those models predict and how well the model explains the feature. Finally, results show that Elastic Net strikes the best balance between explainability and generalization. It better than Lasso and Ridge in walk-forward MSE. Also, it keeps coefficients that are stable and economically meaningful. In general, using convex optimization and sparse regularization makes models so that it can be proven mathematically and explained in terms of economics. This is a principled way to make AI in finance more transparent and reliable.

Keywords: Convex optimization, explainable machine learning, Stock market forecast.

1. Introduction

The rapid growth of machine learning in financial prediction has raised serious questions about model transparency and interpretability. Explainable models can help investors trace how input factors shape predictions. From the perspective of mathematical optimization, convex models provide a rigorous foundation for interpretable learning. The convexity of the guarantee's global optimal solutions. Also, sparse regularization produces streamlined and human-readable parameter structures. In financial applications, non-transparent models cause more challenges because practitioners and stakeholders must clearly understand which variables generate predictive signals. This essay outlines the theoretical principles underlying explainability in convex regularized regression by empirical validation using Apple Computer, Inc (AAPL) return data [1, 2].

Previous research has provided the theoretical justification supporting the methodological method. A 2023 survey highlights that models are designed to be transparent. In regulated financial environments, adopting explainability is becoming increasingly important. Before, people usually relying on other explanations for inherently non-transparent systems [3]. Similarly, a 2025 empirical assessment of generalized additive models (GAMs) shows that interpretable architectures can deliver predictive performance comparable to more complex alternatives. Thus, challenging the notion that model explainability necessarily entails a loss in accuracy [1, 4].

Meanwhile, the class of sparse-regularised convex regressions continues to evolve techniques. Based on Lasso (L_1), Ridge (L_2) and Elastic Net ($L_1 + L_2$), it keeps a balance between variable selection and coefficient shrinkage. This resulted in models with fewer non-zero parameters and greater interpretability. These methods benefit from well-understood convex-optimization properties (uniqueness of solution under strong convexity). In time-series and forecasting applications, evaluation must guard against leakage. When temporal ordering matters, rolling-origin or walk-forward designs replace random K-fold splits. Empirical evidence in asset-pricing further shows that properly validated regularised linear models can perform compare with more complex architectures [5].

First, the research presents a rigorous treatment of convexity, strong convexity, and Karush-Kuhn-Tucker Conditions (KKT) conditions for L_1/L_2-regularized least squares and explains how these properties support transparent coefficient-level explanations. Second, the research describes a time-series protocol that avoids leakage. By time-series, hyperparameter selection within the training window, then walk-forward refitting at each test time. Convexity ensures global optimality, and sparsity yields concise, human-readable structures through exact zeros in coefficients. The research provides a qualitative summary of results on AAPL daily data (2019–2024). Finally, this research finds that ElasticNet balances sparsity and stability, Lasso yields sharp feature selection, and Ridge remains robust under collinearity [6, 7].

2. Related Work

Explainable machine learning emphasizes models whose structure and parameters are. Explainable models are preferable for high-stakes decisions because they make their reasoning explicit and verifiable [2]. Molnar surveys techniques for explainable learning and highlights sparse linear models as inherently transparent [1].

Sparse regularization in convex regression has a long lineage. Lasso enforces coefficient-level sparsity through an L1 penalty, whereas Ridge shrinks coefficients through L_2; ElasticNet mixes both to address multicollinearity while preserving sparsity [8, 9]. Statistical and optimization guarantees for these methods---including uniqueness, consistency under conditions, and efficient coordinate-descent solvers [8, 10].

Forecast evaluation for time series requires time-aware procedures. Random K-fold can leak future information into the past, inflating reported performance. Walk-forward (rolling-origin) evaluation better reflects deployment, training on past and testing on the next point, repeatedly [6, 7]. In asset pricing, recent work shows that regularized linear models remain competitive with more complex methods when properly validated [5].

3. Methodology

3.1. Problem Setup and Notation

Given samples $\{(x_t, y_t)\}_{t=1}^{T}$ where $X_t \in R^p$ are features constructed from past information (ensuring no-leakage), and y_t is the feature H-day cumulative return:

$$y_t = \frac{P_{t+H}}{P_t}\ (H = 5\ by\ default;\ for\ next\ day, set\ H = 1\ ) \tag{1}$$

The model is a linear regression with regularization:

$$\hat{\beta}\ \in \arg{}_{\beta\in R_p}^{min}(\frac{1}{2n} \parallel y - X\beta \parallel_2^2 +\ \lambda_1 \parallel \beta \parallel_1 + \lambda_2 \parallel \beta \parallel_2^2) \tag{2}$$

Where $\lambda_2 = 0$, the model is Lasso; where $\lambda_1 = 0$, the model is Ridge; both together is from ElasticNet.

3.2. Convexity, KKT and Sparsity (Proof Sketch)

Convexity: $\parallel y - X\beta \parallel_2^2$ is convex in β ; $\parallel \beta \parallel_1$ and $\parallel \beta \parallel_2^2$ are convex; nonnegative combination convex funtions remain convex, hence the objective is convex.

Strong Convexity and Uniqueness (ElasticNet/Ridge): When $\lambda_2 > 0$, the Hessian of the objective satisfies:

$$\frac{1}{n}X^TX + 2\lambda_2 I \succcurlyeq 2\lambda_2 I \tag{3}$$

which makes the function strongly convex. Strong convexity implies a unique global optimum and improved numerical stability (robust to noise or multicollinearity).

KKT/Subgradient Conditions (Lasso): Let $g(\beta) = \frac{1}{n}X^T(X\beta - y)$ be the smooth gradient, and $\partial \parallel \beta \parallel_1$ be the subgradient of L1. The optimal solution satisfies:

$$0 \in g(\hat{\beta}) + \lambda_1 \parallel \hat{\beta} \parallel_1 + 2\lambda_2\hat{\beta} \tag{4}$$

In coordinate form, this corresponds to a soft-thresholding update:

$$\widehat{\beta_i} = \frac{Soft(Z_j,\lambda_1)}{d_j} \ where\ Soft(Z_j,\lambda_1) = sign(z)\max(|z| - \lambda, 0) \tag{5}$$

Thresholding induces sparsity: if $|Z_j| \leq \lambda_1$, then $\widehat{\beta_j} = 0$

Statistical Meaning: Under “near-sparse truth” and suitable design conditions (compatibility or cone conditions) with bounded noise (sub-Gaussian or weak dependence), L1/EN provide theoretical guarantees for error bounds and support recovery [11, 12].

3.3. Feature Engineering (Past-only, No Leakage)

The research constructs predictors exclusively from information available at time t to prevent look-ahead bias. Hand-crafted signals cover: (i) return dynamics—lagged returns over 1, 2, and 5 days, short/medium moving-average returns (5 and 20 days), and a 20-day standardized return (volatility proxy); (ii) trend strength via a 5/20 moving-average ratio and a 14-day Relative Strength Index (RSI); (iii) momentum by Moving Average Convergence Divergence (MACD) and its 9-period signal shifted so only past values enter the model; and (iv) trading activity through lagged volume-change rates. The prediction target is the H-day ahead cumulative return, $\frac{P_{t+H}}{P_t - 1}$. All features are standardized within the training window so penalization is comparable across coefficients and to avoid contaminating the past with future statistics.

3.4. Model Selection and Evaluation Protocol

The modelling family is linear regression with convex regularization: Lasso (L_1), Ridge (L_2), and Elastic Net ($L_1 + L_2$). Hyperparameters are tuned on a logarithmic grid for the penalty ($\alpha \in [10^{-7}, 10^{-3}]$) and, for Elastic Net, a set of $l1_{ratio} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. Tuning uses time-series cross-validation (TimeSeriesSplit, five folds) restricted to the training window to respect temporal order. Final assessment uses a walk-forward (rolling-origin) design: at each test time t, the model is refit on $\{1, \ldots, t-1\}$ and then predicts at t, producing non-constant, deployment-like, leak-free forecasts. Performance is summarized by walk-forward MSE.

4. Results

4.1. Data and Features (AIOK)

First, the research analyzes AAPL daily data from 2019–2024 (Stooq). The target is the H = 5 day-ahead cumulative. Features (and code) include lagged returns (ret 1, ret 2, ret 5), return moving averages (ret ma 5, ret ma 20), standardized volatility (ret z 20), moving-average ratio (ma ratio 5 20), RSI (rsi 14), MACD and signal (macd, macd signal), and volume change (vol chg). All features are computed from past information only and standardized within train windows to align penalty scales [1, 10].

4.2. Qualitative Findings (Walk-Forward) (AIOK)

Table 1. Comparison of MSE results of different models

Model	MSE
Lasso	0.001006
ElasticNet	0.001008
Ridge	0.001038
Linear	0.001038

Across the walk-forward evaluation, all four models succeed in capturing aspects of the medium-horizon dynamics (table 1), but their behaviours differ in meaningful ways. Elastic networks maintain competitive MSE. That shows the same idea which the L_1component isolates the most informative signals, and the L_2 component reduces variance when predictors are correlated. By contrast, Lasso often drives many coefficients to zero during volatile periods, producing highly concise but sometimes overly conservative forecasts. That is the same as aligning with its strict sparsity enforcement. Ridge retains all predictors and applies smooth shrinkage, yielding stable forecasts but spreading explanatory weight across numerous variables. Overall, these results patterns are consistent with convex regularization theory. Also, the results reinforce that hybrid penalties tend to offer the most effective compromise between sparse interpretability and predictive stability in practice [5, 9].

4.3. Coefficient-Based Explanations

Coefficient visualizations (on train CV fits) highlight a recurrent set of explanatory features. Recent return aligned with short-term momentum. Also, the average ratio of moving reflects trend alignment and expresses moving-average crossovers (trend). Standardized volatility capturing volatility clustering and captures volatility clustering and risk premium. RSI indicating overbought/oversold states. ElasticNet tends to keep small but meaningful weights on correlated momentum indicators (e.g., RSI and MACD), producing richer yet still explainable structures. Lasso yields the sharpest selection, facilitating concise narratives about which features matter most in each regime [1, 5]. If the "best predictive model" yields near-zero coefficients, a fallback strategy plots the model with the most non-zeros (often Ridge/EN).

These results align with modern quantitative finance literature [3, 11], indicating that the selected factors are economically meaningful rather than statistical noise.

4.4. Why the Protocol Matters

Earlier attempts that relied on a single fixed training set and then generated predictions for the entire test period produced nearly flat forecasts. A consequence of strong regularization combined with the weak signal-to-noise ratio at the one-day horizon. In contrast, the walk-forward protocol, with finer hyperparameter grids and the extended horizon H=5, which enhanced the detectability of time-varying structure and yielded non-flat, regime-responsive predictions. This demonstrates that explainability relies not only on the convex formulation itself but also on an evaluation procedure that reflects the conditions under which the model would be deployed [6, 7].

Among them:

Lasso achieved the lowest MSE (0.001006), thus serving as the best predictive model under walk-forward validation. ElasticNet performed comparably (MSE = 0.001008), confirming that mixed $L_1 + L_2$ regularization does not degrade performance but adds numerical stability. Ridge and Linear both showed higher MSE (0.001038), validating that unregularized or purely L_2 models overfit and fail to generalize.

The Lasso coefficients (rsi 14 = 0.0020, ma ratio 5 20 = 0.00056, ret z 20 = -0.00030, macd signal = -0.00251) reveal that short-term momentum (rsi 14) and moving-average crossovers contribute positively. Also, volatility and smoothed momentum (macd signal) have slight negative effects. All features were engineered using only lagged information and scaled within each training window, ensuring there was no look-ahead bias.

5. Conclusion

Starting from convex optimization and sparse regularization, this paper provides a unified framework combining provable properties (convex/strongly convex structure, KKT, sparsity thresholding, uniqueness) with practical implementation (TimeSeriesSplit + walk-forward). In the empirical study of AAPL (2019–2024):

ElasticNet shows more stable MSE. Lasso offers the strongest interpretability (zero coefficients) but may yield all-zero solutions under weak signals, which is a rational shrinkage. Ridge is stable but dense, weaker in interpretability. Walk-forward produces time-varying, leak-free predictions closer to real deployment. Coefficient plots align with economic intuition, proving that the model captures economically interpretable signals.

Convex optimization with sparse regularization provides a structured and principled foundation for explainable machine learning in financial settings, and strong convexity guarantees globally optimal and unique solutions. The KKT conditions on the other hand explicitly characterize why particular coefficients become zero or remain active, and it enables explanations that map directly onto underlying features. Moreover, in the AAPL daily return analysis, Elastic Net consistently strikes the most effective balance between predictive accuracy and interpretability. Notably, Lasso delivers sharp and aggressive feature selection, and Ridge offers stability in the presence of correlated predictors. In the future, a natural extension is to explore kernelized convex models or Bayesian convex formulations. Those kinds of models satisfy global optimality, interpretability, and formal uncertainty quantification.

References

[1] Molnar, C. Interpretable machine learning: A guide for making black box models explainable (2nd ed.). Lulu Press, 2022.

[2] Rudin C. Stop explaining black box machine learning models for high-stakes decisions and use explainable models instead. Nature Machine Intelligence, 2019, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

[3] Tiddi I, Schlobach S. Explainable artificial intelligence (XAI): A systematic review of methods and applications in finance. Artificial Intelligence Review, 2023, 57(2), 1679–1711. https://doi.org/10.1007/s10462-024-11077-7

[4] Roscher R, Bohlender J, Kieseberg P. Balancing accuracy and interpretability: Empirical evaluation of generalized additive models for explainable AI in finance. Business & Information Systems Engineering, 2025, 67(1), 45–58. https://doi.org/10.1007/s12599-024-00922-2

[5] Gu S, Kelly B, Xiu D. Empirical asset pricing via machine learning. The Review of Financial Studies, 2020, 33(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

[6] Bergmeir C, Hyndman R J, Koo B. A note on the validity of cross-validation for evaluating time series forecasting models. Journal of Time Series Analysis, 2018, 39 (6), 1262–1264. https://doi.org/10.1111/jtsa.12412

[7] Cerqueira V, Torgo L, Mozetič I. Evaluating time series forecasting models: An empirical study on performance estimation methods. Machine Learning, 2020, 109(11), 1997–2028. https://doi.org/10.1007/s10994-020-05950-0

[8] Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series *B*, 1996, 58(1), 267–288.

[9] Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series *B*, 2005, 67 (2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

[10] Hastie T, Tibshirani R, Wainwright M. Statistical learning with sparsity: The lasso and generalizations. CRC Press, 2015.

[11] Hyndman R J, Athanasopoulo G. Forecasting: Principles and practice (3rd ed.). OTexts, 2021.

[12] Xu Y, Chen X, Wang L. Convex optimization methods for explainable sparse modeling. *IEEE* Transactions on Neural Networks and Learning Systems, 2021, 32(11), 4975–4990. https://doi.org/10.1109 /TNNLS.2020.3043497