Leveraging Advanced Machine Learning for Enhanced Earnings Prediction: A Comprehensive Analysis of the EarningsSurprise (ES) Trading System

Alexander Nowak, CFA. NWK

Abstract — Earnings predictions and their consequent impacts on market dynamics are central to strategic decision-making within the financial sector. Investors systematically leverage earnings data to inform their trading behaviors, aiming to profit from discrepancies between actual financial outcomes and analysts' consensus forecasts. The Earnings-Surprise initiative (ES) represents an innovative automated screening and trading system engineered to exploit these discrepancies, with a particular focus on earnings surprises that exceed +/-15% relative to analyst expectations.

This article explains the intricacies involved in the development and implementation of the ES program, which capitalizes on machine learning methodologies. At its core, the program employs a custom ensemble model that leverages gradient-boosted decision trees with neural networks to predict earnings outcomes with heightened accuracy. By conducting a systematic analysis of over 60,000 earnings reports and incorporating a comprehensive array of features, the program enhances predictive accuracy beyond the scope of probabilistic randomness. This could enable superior alpha generation.

Our empirical findings reveal that the model achieves a substantial increase (2x) in true positive predictions of earnings surprises, validating the efficacy of machine learning in refining earnings-based trading strategies. Research is now underway using this strategy for derivatives trading with a particular focus towards stock moves post-surprises, implied volatility changes, etc.

Algorithmic stock trading has become an integral component of contemporary financial markets, with a predominant share of trades now executed autonomously. Various agents have emerged towards high-frequency, quantitative-based trading applications. However, there is little public discussion around earnings-based machine learning approaches. Many of the studies reviewed present only preliminary proof-of-concept findings, often conducted in theoretical settings lacking real-world applicability. This program attempted to bridge the gap from theoretical applications to real-world use cases, by designing an end-to-end system that automatically trains, identifies, predicts and trades applicable positions. While additional research is needed to optimize this program, early work is an encouraging benchmark for future studies.

I. Introduction

Earnings per share (EPS) is a quintessential indicator in the financial analysis domain, serving as a key metric for assessing a corporation's fiscal well-being and cash flow generation capabilities. Analyses have delved into the ramifications of unanticipated earnings fluctuations and their enduring impact on market movements, thereby spurring interest in the development of predictive models for such surprises.

In recent years, the confluence of machine learning methodologies in financial markets has garnered attention. Particularly in earnings predictions, historical techniques, such as the step-wise logit regression models, have touched on the feasibility of constructing investment portfolios predicated on earnings trend forecasts. Progressing into the modern era, contemporary scholars have undertaken comparative analyses of diverse neural network frameworks to refine the precision of earnings predictions. Despite these advances, academic research in this fields remains theoretical, without tools actively developed and working for real-world market dynamics.

This manuscript introduces the ES program, a framework that leverages machine learning algorithms to predict EPS surprises within the U.S. financial markets. Diverging from the traditional focus on absolute EPS levels, this study focuses on earnings surprises—conceptualized as the deviation of actual EPS from the collective forecasts—through an ensemble of gradient-boosted decision trees and neural network models. By juxtaposing our model's performance against conventional market benchmarks, we explain the program's potential not only in forecasting but also in strategically trading on the basis of positive earnings surprises. Our findings illuminate the potential of machine learning in augmenting predictive accuracy and generating excess returns.

II. Motivation

The ES program is engineered with a strategic emphasis on predicting EPS surprises, rather than merely estimating EPS levels. The EPS surprise is quantitatively defined as the percentage deviation between the actual EPS and the consensus estimates provided by analysts:

$$EPS\ Surprise = \frac{Actual\ EPS - Expected\ EPS}{Expected\ EPS}$$

$$Expected\ EPS = \frac{\sum_{i=1}^{n} Analyst's\ EPS\ Estimate}{n}$$

$n = number\ of\ analyst\ estimates$

This metric serves as a dynamic barometer of a company's performance in relation to market expectations. Stocks are priced upon expectations – i.e. the average of the analysts' estimates. Significant deviations between actual and expected EPS can provoke swift adjustments in stock prices, with companies surpassing expectations typically experiencing upward price momentum, while those underperforming tend to see a decline. Hence, the capability to accurately forecast

Alexander Nowak, CFA. NWK Group LLC. Alex@nwk.group.

EPS surprises, without a regard to absolute EPS level, constitutes the foundation of a potentially lucrative trading strategy.

III. Machine learning models

The predictive engine of the ES program is characterized by a bespoke ensemble model that uses gradient-boosted decision trees with neural networks. This ensemble methodology facilitates the capture of intricate patterns and interdependencies within the data, thereby augmenting the model's predictive skill. Training of the model is conducted using advanced tools such as XGBoost, CatBoost, RandomForest, LightGBM, and PyTorch, all encapsulated within the AutoGluon framework to optimize hyperparameters and ensure rigorous validation.

A. *Gradient Boosted Decision Trees (GBDT)*

GBDTs are a powerful and versatile machine learning technique, renowned for their ability to handle complex datasets with high dimensionality and non-linearity. The method works by sequentially constructing an ensemble of decision trees, where each tree is built to correct the errors made by the ensemble thus far. This sequential learning process is underpinned by gradient descent, which iteratively minimizes a specified loss function. In EPS surprise prediction, models such as XGBoost and LightGBM are employed due to their capacity to model intricate interactions within financial data, capturing subtle patterns that might be overlooked by simpler models.

XGBoost, known for its speed and performance, incorporates a regularization mechanism that helps prevent overfitting, a common issue in financial modeling. LightGBM, on the other hand, excels at handling large datasets and is particularly effective with sparse data typical of financial reports. It uses a leaf-wise growth strategy that can lead to more complex models with fewer trees. The use of these models allows for the exploitation of temporal dependencies and non-linear relationships, providing a solid foundation for predicting EPS surprises with high accuracy.

B. *Random Forest Classifiers*

Random Forest is a robust ensemble learning method that generates a forest of decision trees, each constructed from a random subset of the data and features. This randomness enhances the model's ability to generalize from the training data to unseen data, thus reducing overfitting—a pervasive problem in financial prediction tasks due to market volatility and noise. In the ES program, Random Forests are utilized to provide stability and diversity in predictions, benefiting from their inherent ability to capture complex interactions between variables without requiring extensive data preprocessing.

Each tree in a Random Forest outputs a class prediction, and the mode of these classes is chosen as the final output. This process not only improves the model's robustness but also reduces the variance inherent to individual decision trees. Random Forests are adept at handling high-dimensional spaces, making them suitable for the wide array of features used in EPS surprise prediction, including technical indicators and fundamental financial ratios.

C. *Categorical Boosting Classifiers (CATBoost)*

CATBoost is designed to handle datasets with categorical features more effectively than traditional GBDT models. This capability is crucial in financial datasets where categorical variables, such as industry sectors, market indices, and analyst ratings, play a significant role. CatBoost employs a novel encoding method that reduces the risk of overfitting and ensures that the model can effectively learn from the categorical data.

Incorporating CatBoost into the ES model enhances its ability to predict EPS surprises by accurately capturing the nuances of categorical data. This model employs an efficient algorithm for handling categorical variables, which includes using a technique called Ordered Boosting. This technique mitigates the prediction shift problem and improves the model's generalization ability. CatBoost's robustness and accuracy make it an indispensable part of the ensemble, contributing to the overall predictive strength of the program.

D. *Recurrent Neural Networks (RNN)*

Neural Networks, especially those implemented with PyTorch, form the deep learning backbone of our EPS prediction system. These networks are equipped to model complex, non-linear relationships in financial datasets, thanks to their deep architectures capable of learning hierarchical representations. In the ES program, neural networks are used to complement tree-based methods by providing additional abstraction layers that capture intricate dependencies within the data.

The neural network architecture is designed to process both the temporal and cross-sectional aspects of financial data, using techniques such as recurrent layers for temporal dependencies and convolutional layers for feature extraction. PyTorch's flexibility and dynamic computation graph allow for the efficient exploration of various architectures and optimizations. Neural networks in this setup are particularly effective at learning from high-dimensional data, integrating diverse features such as historical earnings data, market sentiment indicators, and macroeconomic variables, thereby enhancing the model's overall predictive performance.

D. *Autogluon Hyperparameter Framework*

AutoGluon is employed as a sophisticated framework to streamline the model training and hyperparameter optimization process. AutoGluon supports a wide range of machine learning libraries and frameworks, enabling the ES program to leverage a diverse set of models and techniques. It employs state-of-the-art optimization algorithms to navigate the hyperparameter space, ensuring that each model component within the ensemble is finely tuned to maximize predictive accuracy. The framework's ability to adapt to different data distributions and model configurations makes it an essential tool in maintaining the program's competitive edge in predicting EPS surprises.

IV. Data

To train ES, we captured earnings data on a wide universe of U.S.-listed companies. This list comprises of the S&P 500, large Nasdaq stocks, and $1B+ market cap companies across consumer, defense, technology, oil & gas and healthcare. For each company, earnings from the last 10 years were calculated, and so too was the EPS Surprise. All-in, the training dataset comprised 60,000 earnings results from around 1,000 companies. For each earnings report, 400+ features were created at T-1 (one day before the earnings report). The features are categorized into six primary groups:

1. Analyst Recommendation Changes: Tracks modifications in analysts' buy, hold, or sell recommendations, which reflect shifts in market sentiment and potential future performance.

2. Insider Trades: Monitors the buying and selling activities of company insiders, offering insights into internal expectations and confidence levels.

3. Peer Earnings: Considers the earnings results of comparable companies, providing contextual information about industry trends and competitive positioning.

4. Technical Indicators: Utilizes historical price and volume data to identify patterns and trends that may indicate future price movements.

5. Quantitative Measures: Includes mathematical and statistical metrics that capture various aspects of financial performance and market behavior.

6. Fundamental Ratios: Encompasses key financial ratios that assess a company's operational efficiency, profitability, and financial health.

The target of the data is three discrete outcomes based on the level of surprise: a negative surprise (<-15%) is categorized as -1, a neutral range surprise (-15% to 15%) is marked as 0, and a positive surprise (>+15%) is represented as +1. The program's primary objective is to accurately forecast positive EPS surprises exceeding 15%, triggering strategic stock purchases followed by a two-day holding period—one day dedicated to reacting to the earnings announcement and the subsequent day for optimized selling. By shifting the distribution of predictions towards more accurate detections of positive surprises, ES aspires to consistently outperform random guessing, thereby generating alpha and enhancing the program's profitability in market conditions.

V. Model Performance

A. V1 Model

ES' V1 model performance is evaluated using a robust test dataset comprising 12,645 earnings results. The dataset is characterized by a distribution where 6% of the results were a -15% earnings surprise (truth -1), 69% fell between -15% and 15% (truth 0), and 25% were +15% (truth 1). This distribution provides a diverse testing ground for assessing the model's predictive accuracy and robustness.

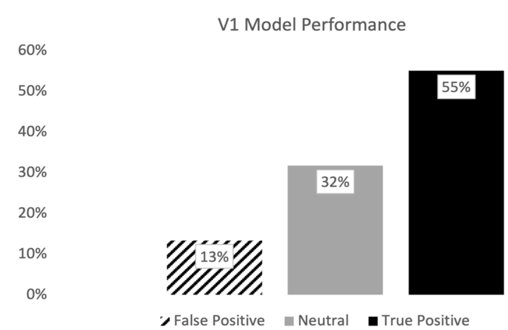
The V1 model achieves an accuracy of 74% on the test data, with a Matthews Correlation Coefficient (MCC) of 36%. The MCC provides a balanced measure of the model's predictive quality by considering true and false positives and negatives, offering a more comprehensive view than accuracy alone. Given the skewed distribution of earnings surprises, detailed analysis of the prediction outcomes is essential. Among the 12,645 earnings results, the V1 model made a prediction of +15% (1) 2,029 times. The outcomes of these +1 predictions are as follows:

- True Positive (TP): Cases where both the actual and predicted earnings surprises are +15% (truth is 1, prediction is 1). The V1 model achieved a True Positive Rate of 55%, more than doubling the base rate from 24%.

- False Positive (FP): Instances where the actual surprise is -15% but the model predicts +15% (truth is -1, prediction is 1). The model maintained a relatively low False Positive Rate of 13%.

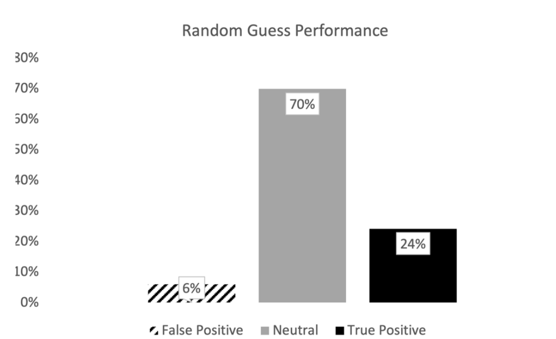
- Neutral (N): Occasions where the actual surprise falls between -15% and 15%, but the model predicts +15% (truth is 0, prediction is 1). The Neutral prediction rate was 32%.

Figure 1. V1 Model Performance – A Significant TP Shift



A benchmark assessment involved a random guess model, which returned 3,147 +1 results (approximately 33% of the test sample). This random model replicated the distribution observed in the test data, confirming that the V1 model's prediction capability is not due to chance alone but rather an informed adjustment towards true positives.

Figure 2. Random Guess – Matches Random Distribution



B. Experimental Cut Model

To explore alternative classification strategies, an experimental cut model was tested on a subset of 6,323 earnings results. This model divides the classification buckets into -15%-0% (truth -1), 0%-15% (truth 0), and 15%+ (truth 1). Here, the distribution is 25% for -15%-0%, 56% for 0%-15%, and 19% for 15%+.

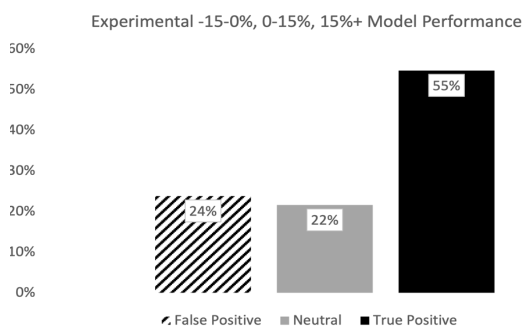
The experimental cut model achieved an accuracy of 63% with an MCC of 33%. While slightly lower than the V1 model, it demonstrates a nuanced understanding of the new classification boundaries. In this experimental setup, the model predicted +15% (1) 816 times out of 6,323 results. Of the +1 predictions:

- True Positive (TP): The model accurately predicted a +15% surprise in 55% of cases, significantly improving from the base rate of 18%.

- False Positive (FP): The rate of false positives remained stable at 24%.

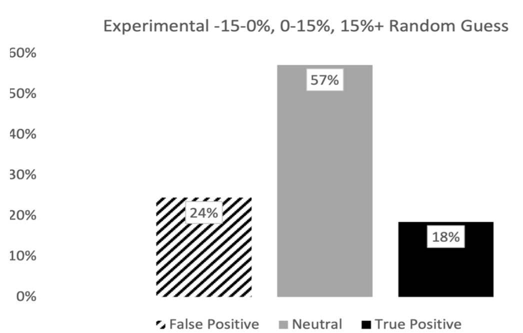
- Neutral (N): The Neutral prediction rate was 22%.

Figure 3. Experimental Model Cut Performance



A benchmark random guess model yielded 2,106 +1 results, or roughly 33% of the test sample. This outcome mirrored the distribution in the test data, underscoring the experimental cut model's capacity to meaningfully shift prediction capabilities towards true positives without disproportionately increasing false positives.

Figure 4. Experimental Cut Random Guess



In summary, the ES program's models demonstrate strong predictive capabilities, with each iteration offering unique insights and strategic advantages in forecasting earnings surprises. Future refinements will continue to build on this solid foundation, optimizing for accuracy and market impact.

VI. Trading Strategy Implementation

ES was designed for real-world applicability, and the program is designed to be repeatedly ran in a live trading environment. Specifically, on each business day at 3:30 PM ET, the program identifies firms scheduled to release earnings the following day, filtering them through a curated list of tradable stocks, primarily from the S&P 500 and major Nasdaq listings. Predictions are generated for each firm, categorizing potential EPS surprises into +15% (1), -15% (-1), or neutral (0). In instances where a +15% surprise is anticipated, the program executes a purchase order during the end-of-day auction, maintaining the stock position for two days to exploit the anticipated post-announcement price movement. After two days elapses, the stock is sold. Additional research is currently underway evaluating ES' potential in derivatives trading, specifically investing its use in implied volatility crush scenarios (calendar spreads), neutral trades (iron condors), stock upside (spreads), among others.

The entire system is containerized via Docker and deployed through Google Cloud Run, ensuring a seamless, automated operation. Essential operations such as trade execution and logging are managed through integrations with the broker API and Google Firestore.

VII. Conclusion

In this exploration, we delve into the intricate workings of the ES program, a sophisticated financial model developed to predict earnings per share (EPS) surprises and leverage these insights for strategic trading within the U.S. stock market. This model primarily targets companies listed in the S&P 500 and large Nasdaq stocks, ensuring a focus on entities with substantial market influence and robust financial data. Through this focus, the program taps into a dataset comprising approximately 60,000 earnings results from around 1,000 companies, creating a substantial foundation for predicting EPS surprises.

In testing two models, ES achieved a significant improvement in true positives vs. benchmark, while keeping false negative relatively stagnant. In both models the true positive more than doubled vs. the benchmark.

ES underscores the potential of machine learning in refining earnings-based trading strategies. Future work will focus on model optimization, additional features and derivatives strategies from live trading. This methodical approach not only provides a structured mechanism for predicting earnings surprises but also implements a trading strategy to capitalize on these forecasts, thereby aspiring to generate substantial alpha by systematically enhancing the prediction curve towards more accurate true positives.

This research contributes to the broader discourse on the application of machine learning in finance, offering a sophisticated framework for earnings prediction and strategic trading that leverages the latest advancements in artificial intelligence. Ultimately, while the program demonstrates robust predictive capabilities and strategic advantages, it acknowledges the challenges of real-world implementation, emphasizing the need for continuous adjustment and enhancement to maintain an edge in the dynamic landscape of financial markets.

References

[1] Ball, R., & Brown, P. An Empirical Evaluation of Accounting Income Numbers. Journal of Accounting Research. 1968.

[2] Foster, G., Olsen, C., & Shevlin, T. Earnings Releases, Anomalies, and the Behavior of Security Prices. The Accounting Review. 1984.

[3] Bernard, V. L., & Thomas, J. K. Evidence that Stock Prices Do Not Fully Reflect the Implications of Current Earnings for Future Earnings. Journal of Accounting and Economics. 1990.

[4] Livnat, J., & Mendenhall, R. Comparing the Performance of Value and Growth Portfolios in an Earnings-Surprise Strategy. Financial Analysts Journal. 2006.

[5] Jegadeesh, N., & Titman, S. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. The Journal of Finance. 1993.

[6] Goodfellow, I., Bengio, Y., & Courville, A. Deep learning. MIT Press. 2016.

[7] Hastie, T., Tibshirani, R., & Friedman, J. The elements of statistical learning: data mining, inference, and prediction. Springer science & business media. 2009.

[8] Lipton, Z. C. The mythos of model interpretability. Communications of the ACM. 2018.

[9] Dixon, M., Klabjan, D., & Bang, J. H. Machine Learning in Finance: From Theory to Practice. Springer. 2020.

[10] Khandani, A. E., & Lo, A. W. What Happened to the Quants in August 2007?. Journal of Investment Management. 2011.

[11] Lopez de Prado, M. Advances in Financial Machine Learning. John Wiley & Sons. 2018.

[12] Chen, T., & Guestrin, C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016.

[13] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems. 2017.

[14] Breiman, L. Random forests. Machine learning. 2001.

[15] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. CatBoost: unbiased boosting with categorical features. Advances in neural information processing systems. 2018.

[16] Hochreiter, S., & Schmidhuber, J. Long short-term memory. Neural computation. 1997.

[17] Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., & Hutter, F. Efficient and robust automated machine learning. Advances in neural information processing systems. 2015.

[18] Erickson, D., Mueller, J., Shirkov, A., Zhang, H., Rohde, P., Li, M., ... & Smola, A. Autogluon: A python package for auto machine learning. arXiv preprint arXiv:2003.06505. 2020.

[19] Fama, E. F., & French, K. R. The Cross-Section of Expected Stock Returns. The Journal of Finance. 1992.

[20] Lev, B. On the usefulness of earnings and earnings research: A review and critique. Journal of Accounting Research. 1989.

[21] Chan, L. K., Hameed, A., & Tong, W. H. Profitability and stock returns in emerging markets. Journal of financial economics. 2000.

[22] Jegadeesh, N., & Titman, S. Momentum. Review of financial studies. 2011.

[23] Goyal, A., & Wahal, S. The Role of Institutional Investors in Equity Markets: An Overview. Foundations and Trends® in Finance. 2015.

[24] Bouchaud, J. P., Farmer, J. D., & Lillo, F. How markets slowly digest changes in supply and demand. Handbook on Systemic Risk. 2009.