Causal Discovery in Stock Return

Xinqi Huang xih037@ucsd.edu
Ruizhe Dai rdai@ucsd.edu
Vivian Zhao vxzhao@ucsd.edu
Yishan Cai yic075@ucsd.edu
Biwei Huang bih007@ucsd.edu
Jelena Bradic jbradic@ucsd.edu

Abstract

Many people who own stocks don't have enough financial knowledge to understand the stock market. This project aims to empower investors by identifying the key drivers behind stock returns, providing actionable insights to guide decision-making and mitigate risks. To achieve this, we develop a hybrid framework that integrates sentiment analysis, historical stock data, and macro and microeconomic indicators. Public sentiment, a crucial factor in stock market dynamics, is analyzed using FinBERT on company-specific tweet data. Historical stock performance is modeled using DeepAR, while macro and microeconomic factors, such as GDP, CPI, and company reports, are processed through a Random Forest Regressor. Feature selection is performed using causal learning techniques, and the outputs of these models are synthesized via a fusion layer to produce comprehensive predictions. By combining diverse data sources and advanced modeling techniques, this project aims to offer a clear and accessible understanding of the factors influencing stock returns, supporting better investment decisions.

Code: https://github.com/VivianZhao12/CAPSTONE-stockreturn

Website: https://acai1031.github.io/DSC180B-Capstone-Website

1 Introduction

1.1 Introductory Paragraph

Daily Stock return prediction is a crucial aspect of financial analysis, heavily influenced by a complex interplay of macroeconomic factors, microeconomic trends, and investor anticipation. Since the end of 2019, global policies, economic disruptions, and socio-political tensions have introduced significant volatility into the stock market, impacting millions of investors worldwide. On April 29, 2022, Amazon shares plummeted 14% following a 5% surge the previous day after the company gave a revenue outlook for the current quarter that fell short of Wall Street's estimates. This instance highlights the need to bring in economic indicators with market sentiment to address the challenges of understanding and predicting stock movements. To better guide the decision of investors, we focus on the most impacted industries—Technology, Retail, and Health—by analyzing highly volatile stocks, including Amazon (AMZN), Alphabet (GOOG), AT&T (T), CVS (CVS), Amgen (AMGN), and Abbott Laboratories (ABT).

Traditional forecasting models, such as factor models and autoregressive approaches, primarily rely on analyzing historical daily patterns using predefined economic relationships, which limits their ability to capture sudden shifts driven by external shocks and sentiment changes. Deep learning-based approaches, such as DeepAR, a neural network-based model, have emerged as powerful alternatives, offering the ability to bring in diverse factors for complex dependencies modeling, and can adapt to rapid shifts in trends. However, its black-box nature limits interpretability, making it difficult for investors and analysts to understand the reasoning behind its predictions. This lack of transparency often overshadows its predictive performance, reducing trust in model-driven financial decision-making.

In this paper, we propose a hybrid stock return prediction framework that integrates **PCMCI+ (Peter and Clark Momentary Conditional Independence plus) with DeepAR** to better capture daily trends. Using PCMCI+, we are able to identify true causal relationships in multivariate time series data rather than relying on correlations. Our approach consists of two key steps: 1) Feature Selection, which filters out spurious correlations to retain only the most influential factors and 2) Lag Optimization, which determines the optimal time dependency to enhance transparency and provide clear decision rationales.

For economic indicators, we leverage **CD-NOD** (Causal Discovery from Nonstationary Data) to identify the causal impact of real-time macroeconomic and company-level factors on stock prices. We retained only features with direct or indirect causal links to better account for economic shocks on stock price.

To accommodate diverse features with varying granularity, we conducted hypothesis testing to identify the most representative data points, an approach widely used by central banks for real-time economic prediction. By integrating causal inference techniques and statistical analysis with deep learning methods, our framework bridges the gap between predictive power, interpretability, and data availability lag, enhancing transparency in stock forecasting for investors.

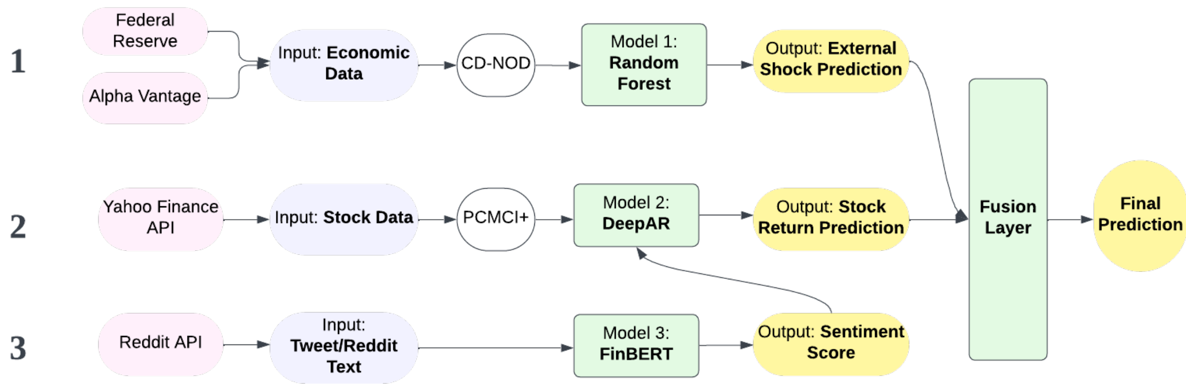


Figure 1: Overview of the Proposed Stock Return Prediction Framework

1.2 Literature Review

The challenge of identifying causal relationships in stock market returns presents unique complexities that extend beyond traditional statistical analysis. While numerous studies have examined correlations between market factors and stock performance, establishing genuine causal relationships remains elusive due to the dynamic, interconnected nature of financial markets and the presence of numerous confounding variables.

Prior research has established various approaches to understanding stock market predictability, focusing on both macroeconomic variables and media sentiment analysis. On the macroeconomic front, (2) highlighted that fundamental factors such as industrial production growth, interest rates, inflation, and unemployment are key determinants of stock market movements, with (8) demonstrating through a comprehensive study of 12 industrialized countries that interest rates serve as the most reliable predictor in an international context. Parallel to these macroeconomic studies, research has increasingly recognized the importance of media sentiment in predicting stock returns. Notably, (5) analyzed nearly one million news articles and found that daily news coverage could predict stock returns within a one to two-day window, with positive news generating immediate price responses while negative news produced delayed effects. This dual influence of macroeconomic factors and media sentiment was further reinforced by studies such as (4) and (1), who found significant relationships between media tone and future stock returns, particularly during periods of economic uncertainty.

There are also prior research works that have demonstrated the usage of DeepAR algorithm on predicting stock, (7; 9), they use the DeepAR algorithm to capture the pattern hidden in time series data, and their works have demonstrated the potency of the DeepAR algorithm in predicting under a complex environment, for example, the financial data. However they did not include the financial sentiment factor into their work.

2 Methods

2.1 Economic Impact Analysis Module

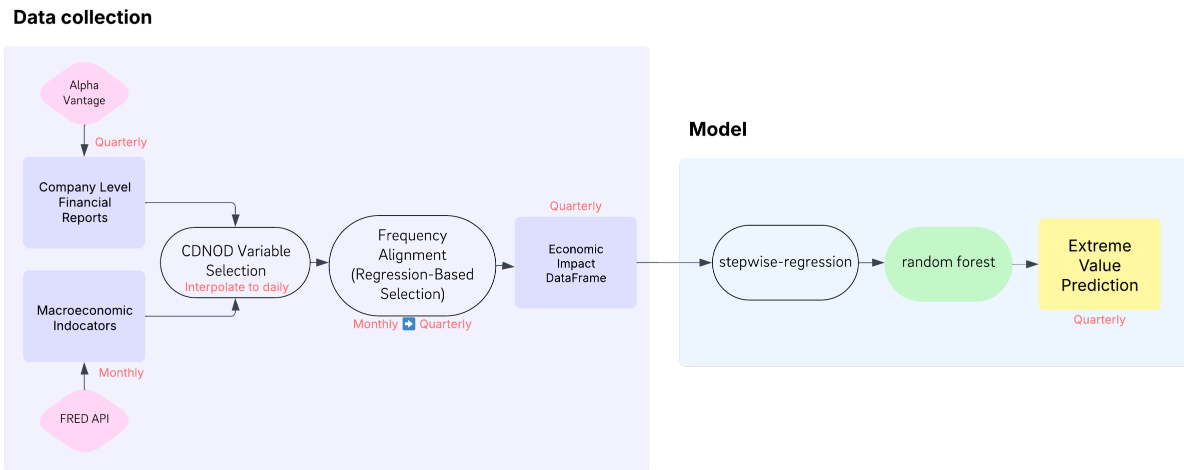


Figure 2: Overview of the Economic Impact Module Framework

2.1.1 Data Extraction

For this project, we extracted microeconomic data using the Yahoo Finance API. The microeconomic data includes features such as the number of treasury shares, ordinary shares, net debt, sales of business. These features capture company-specific financial metrics crucial for understanding stock return dynamics.

On the macroeconomic side, data was sourced from the Federal Reserve of St. Louis (FRED), including indicators like Money Supply (M1 and M2), Interest Rates, Producer Price Index (PPI), Real Dollar Index, Unemployment Rate, Consumer Price Index (CPI), and Gross Domestic Product (GDP). Together, these datasets form a comprehensive view of economic and company-specific factors impacting stock movements.

2.1.2 Feature Selection through CDNOD

For all economic indicators, we first interpolated features using forward filling to align with daily stock closing prices. We then applied causal learning for feature selection using the CD-NOD algorithm with monthly grouping and Fisher's Z-test (0.01 significance level) to identify the most relevant stock price predictors.

Define the time index C Given the volatility in stock prediction and the nature of our data, we divided the dataset into monthly time windows:

$$C = \{C_1, C_2, \ldots, C_M\}$$

where each C_m represents a distinct monthly regime.

Detection of Changing Modules and Recovery of Causal Skeleton In CD-NOD, proposed by Huang et al. (6), a complete undirected graph would be constructed on the variable set $V \cup \{C\}$. Then, for each variable V_i, it tests for marginal and conditional independence between V_i and C. If V_i is independent of C given a subset of $\{V_k \mid k \neq i\}$, the edge between V_i and C is removed. Finally, for every pair of variables (V_i, V_j) where $i \neq j$, same procedure is performed between V_i and V_j. If they are independent given a subset of $\{V_k \mid k \neq i, k \neq j\}$, the edge between them is removed.

Independence Testing Using Fisher's Z-Test Since traditional forecasting methods often assume linear relationships, we used Fisher's Z-test in CD-NOD to assess dependency for fast computation. The test statistic is computed as:

$$Z = \frac{1}{2} \ln\left(\frac{1 + r_{XY|Z}}{1 - r_{XY|Z}}\right) \sqrt{n - |Z| - 3}$$

where $r_{XY|Z}$ represents the partial correlation between stock prices and feature X_j, conditioned on other selected variables. If the p-value < 0.01, the feature is retained as a significant predictor.

Select Features on Learned Causal Graphs Once the casual graph is learned, only impactful features were kept for final prediction. Impactful features are defined as follows:

- Features that have a **direct edge** to stock price in the learned causal graph.
- Features that connect to stock price through **causal pathways** in the learned graph.

After selecting features based on the CD-NOD causal graphs, we further performed pairwise regression to quantify each predictor's direct impact on stock price movements.

2.1.3 Frequency Alignment

The key challenge arises from the inherent discrepancy in temporal granularity: macroeconomic and microeconomic indicators are typically reported on a **monthly** basis, while corporate financial statements adhere to a **quarterly** reporting cycle. In this project, the alignment of economic impact indicators with company financial performance data is achieved through a systematic transformation of different time-series frequencies.

Quarterly Stock Return Stock return data, originally recorded at the daily frequency, is aggregated into **quarterly returns** using the following compounding formula:

$$R_q = \prod_{t \in q}(1 + r_t) - 1 \tag{1}$$

where:

- R_q represents the cumulative quarterly return for a given quarter q,
- r_t denotes the daily return at time t,
- The product iterates over all trading days within the quarter.

This transformation ensures that the financial return data aligns temporally with corporate financial disclosures and macroeconomic indicators.

Temporal Alignment of Monthly Economic Data to Quarterly Structure To construct a consistent quarterly dataset, macroeconomic indicators are decomposed into three distinct monthly observations per quarter:

- **First-month observation** (M_1)
- **Second-month observation** (M_2)
- **Third-month observation** (M_3)

Each macroeconomic variable X is represented in the quarterly dataset as three separate features:

$$X_{q,M1}, \quad X_{q,M2}, \quad X_{q,M3} \tag{2}$$

where each variable corresponds to its value recorded in the first, second, and third month of the quarter, respectively. This transformation allows for systematic selection of the most predictive monthly macroeconomic indicators.

Selection of the Optimal Monthly Indicator Per Quarter To determine the most statistically relevant monthly indicator within each quarter, a **regression-based selection process** is implemented. Specifically, an **Ordinary Least Squares (OLS) regression** is fitted separately for each macroeconomic variable, using its three monthly values as independent variables:

$$Y_q = \beta_0 + \beta_1 X_{q,M1} + \epsilon \tag{3}$$

$$Y_q = \beta_0 + \beta_2 X_{q,M2} + \epsilon \tag{4}$$

$$Y_q = \beta_0 + \beta_3 X_{q,M3} + \epsilon \tag{5}$$

where:

- Y_q denotes the dependent variable (quarterly stock return),
- $X_{q,M1}$,$X_{q,M2}$,$X_{q,M3}$ represent the macroeconomic variable's values in the respective months of quarter q,
- $\beta_1, \beta_2, \beta_3$ are estimated regression coefficients,
- ϵ captures residual noise.

The model with the lowest **p-value** for its predictor is selected, identifying the most influential month for that macroeconomic variable. The chosen feature is then included in the final dataset.

Merging Company Financial Data with Aligned Macroeconomic Features The quarterly financial performance data of individual companies is merged with the processed macroeconomic dataset using the fiscal quarter as the key:

$$\text{Final Dataset} = \text{Company Financials} \bowtie_{\text{fiscal quarter}} \text{Macroeconomic Indicators} \tag{6}$$

This ensures that company-specific financial metrics are temporally synchronized with both market-wide economic conditions and stock return dynamics.

2.2 Sentiment Analysis Module

2.2.1 Data Collection and Preprocessing

We believe that incorporating sentiment factors could help our model to capture more information and have better performance on predicting stock returns. To make the numerical conversion of raw tweet data more accurate, we adopted FinBERT, a pre-trained NLP model to analyze sentiment of financial text. We have found a dataset on GitHub repository that contains tweets about famous companies including: Apple, CVS, Ebay, dating from 2020-06-01 to 2023-05-31. However, our goal is to make our model more applicable, which we need more recent tweets to make our model better capture the pattern in the recent stock market. Initially we thought about scraping more tweets, but the cost of accessing X API was too costly for our group. Instead, we decided to use Reddit posts and comments, as Reddit's API is free. We will then augment the sentiment scores with the predictions of the DeepAR algorithm, which enables the algorithm to understand factors affecting stock market returns beyond micro and macroeconomic factors. The entire flow of data collection and preprocessing in the sentiment module is shown in figure 3.

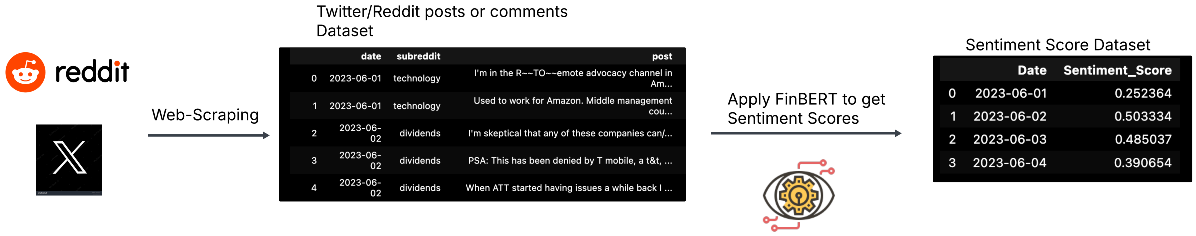


Figure 3: Work Flow of Sentiment Analysis Module

2.3 Stock Return Prediction Module

2.3.1 Data Collection and Preprocessing

We collected historical stock data for six major companies, with three from the technology sector and three from the healthcare sector using the Yahoo Finance API. The dataset spans from June 2020 to February 2025, providing 1,190 trading days per company, totaling 7,140 records.

For each company, we gathered daily metrics including opening price, closing price, adjusted closing price, daily high, daily low, and trading volume. These comprehensive metrics provide a complete picture of daily trading activities and price movements.

The raw data underwent several preprocessing steps. First, we structured the data chronologically and grouped it by company ticker to maintain consistent time series organization. To address the challenge of missing values that commonly occur in financial time series due to non-trading days or data collection issues, we implemented a two-stage imputation approach. The method begins with forward filling, which propagates the last valid observation forward to fill temporal gaps, preserving the time series characteristics. Any remaining missing values are then filled with zeros to ensure data completeness for model training.

For our prediction target, we computed daily returns using the percentage change in adjusted closing prices. The daily return at time t is calculated as:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

where R_t represents the daily return at time t, and P_t represents the adjusted closing price at time t. This metric serves as our target variable for the prediction task, capturing the daily price movements while accounting for stock splits and dividend payments through the use of adjusted closing prices.

The percentage changes of stock return in our datasets range from a -16% decline in CVS on May 1, 2024, to a 13% increase in Amazon on February 4, 2022.

For our analysis, we define changes below 4% as normal fluctuations, changes between 4% and 7% as significant fluctuations, and changes above 7% as abnormal fluctuations. Our model is designed to capture significant and abnormal fluctuations, while excluding normal fluctuations, in order to focus on more impactful price movements and avoid overfitting on minor patterns.

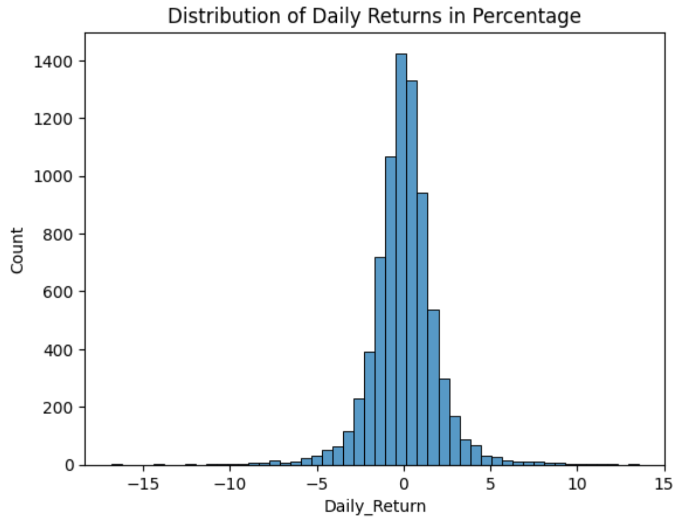


Figure 4: Distribution of Daily Returns in Percentage

2.3.2 Model Overview

For stock return prediction, we propose an enhanced DeepAR (Deep Auto-Regressive) architecture that combines deep learning with probabilistic forecasting (3). Our implementation builds upon the adaptation of DeepAR for electricity forecasting by (10), which we modified for stock return prediction. Our model is specifically designed to address three key challenges in stock return prediction: capturing long-term dependencies, handling multiple scales of price movements, and providing uncertainty estimates for risk management.

The model takes historical price data and carefully selected features as input, processes them through a deep neural network architecture, and outputs a probability distribution for future returns. This probabilistic approach not only provides point estimates but also quantifies the uncertainty of predictions, which is crucial for financial decision-making.

2.3.3 Feature Engineering and Selection

Our feature engineering process focused on creating a comprehensive set of predictive signals while avoiding redundancy and noise. We employed the PCMCI+ (Peter and Clark Momentary Conditional Independence plus) algorithm to perform causal feature selection, identifying the most relevant predictors while controlling for spurious correlations.

The selected features fall into four main categories as shown in figure below:

1. **Price-Based Features**: Including close price and trading volume, which provide fundamental market information.
2. **Technical Indicators**: Including MA5 deviation, MACD (Moving Average Convergence Divergence) with lag-2, intraday returns, and volatility measures. These features capture price momentum and market dynamics.
3. **Time-Based Features**: Calendar effects including weekday and month, which cap-

ture seasonal patterns in market behavior.

4. **External Features**: Sentiment scores derived from market news and social media, providing insights into market psychology and investor behavior.

The causal structure captured by PCMCI+ reveals important relationships between features and stock returns, with technical indicators showing particularly strong causal pathways.

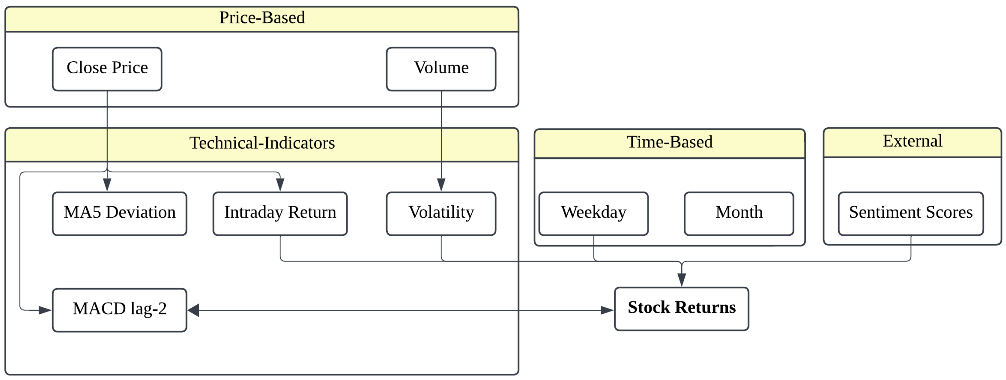


Figure 5: Causal Structure of Stock Return Predictors after PCMCI+ Analysis

2.3.4 Model Architecture

Our enhanced DeepAR architecture consists of three major components to produce accurate and reliable predictions:

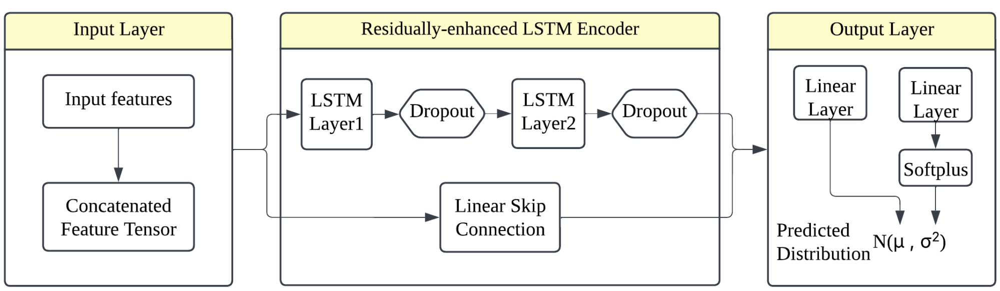


Figure 6: DeepAR Model Pipeline

Input Processing Layer The input layer of our model processes multivariate time series data through three main streams: daily historical returns, technical and market indicators, and temporal features. The input vector at each time step t is represented as $\mathbf{x_t} \in \mathbb{R}^d$, with dimensionality $d = 1 + d_{\text{cov}} + d_{\text{emb}}$ comprising:

- The previous day's return (z_{t-1})

- A set of $d_{\text{cov}} = 8$ covariate features selected through PCMCI+ causal discovery
- Learned stock embeddings (d_{emb}) that capture entity-specific patterns

To ensure consistent scale across all features, we apply z-score standardization:

$$\mathbf{x}_{\text{normalized}} = \frac{\mathbf{x} - \mu}{\sigma} \tag{7}$$

The feature tensor is constructed as:

$$\text{lstm\_input}_t = [\mathbf{x}_t; \mathbf{embedding}(\text{idx})] \tag{8}$$

where **embedding**(idx) represents the learned embedding for the specific time series.

Residually-enhanced LSTM Core The core of our model utilizes an enhanced LSTM architecture with the following improvements:

- **Orthogonal Weight Initialization** For recurrent weight matrices, we employ orthogonal initialization to ensure stable gradient flow during training. This technique initializes the weight matrices with orthogonal properties such that $W_{hh}W_{hh}^T = I$, maintaining consistent gradient magnitudes during backpropagation through time. The input weights use Xavier initialization to establish appropriate initial scales that help achieve faster convergence. These initialization techniques are especially important for financial time series where capturing both long and short-term dependencies is critical:
$$W_{\text{hh}} \sim \text{Orthogonal()} \tag{9}$$
$$W_{\text{ih}} \sim \text{Xavier()} \tag{10}$$
- **Forget Gate Bias Initialization** We implement a specialized initialization scheme for the LSTM forget gate bias. By setting the forget gate bias to 1 while initializing all other gate biases to 0, we address the common challenge of learning long-term dependencies. This approach allows the network to retain information by default in the early stages of training, gradually learning when to forget, rather than struggling to learn when to remember:
$$\mathbf{b}_f = \mathbf{1}, \quad \mathbf{b}_i = \mathbf{b}_o = \mathbf{b}_c = \mathbf{0} \tag{11}$$
- **Variational Dropout** Unlike standard dropout which applies different random masks at each time step, we implement variational dropout that applies the same dropout mask across the entire sequence. This temporal consistency preserves the coherence of the time series signal while still regularizing the network effectively. With a carefully tuned dropout rate of 0.15, this approach prevents overfitting without disrupting temporal dependencies:
$$\mathbf{m} \sim \text{Bernoulli}(1 - p_{\text{dropout}}), \quad p_{\text{dropout}} = 0.15 \tag{12}$$
$$\mathbf{h}_{\text{dropped}} = \mathbf{h} \odot \mathbf{m} \cdot \frac{1}{1 - p_{\text{dropout}}} \tag{13}$$

- **Hierarchical Structure** Our model employs multiple stacked LSTM layers that process information at different temporal scales. This hierarchical architecture allows lower layers to capture short-term price fluctuations while higher layers learn longer-term market trends and seasonal patterns. Each layer refines the representation before passing it to the next, creating increasingly abstract temporal features:

$$\mathbf{h}_t, \mathbf{c}_t = \text{LSTM}(\text{lstm\_input}_t, \mathbf{h}_{t-1}, \mathbf{c}_{t-1}) \tag{14}$$

- **Skip Connections** We implement residual connections that create direct pathways from the input to the output of the LSTM stack. These skip connections serve two crucial purposes: they mitigate the vanishing gradient problem by providing alternative gradient paths during backpropagation, and they allow the model to leverage both raw and processed features when making predictions. This is particularly valuable in financial forecasting where both transformed features and raw signals contain complementary information:

$$\text{skip\_connection} = \mathbf{W}_{\text{skip}} \cdot \text{lstm\_input} \tag{15}$$

$$\text{final\_hidden} = \mathbf{h}_t + \text{skip\_connection} \tag{16}$$

Probabilistic Output Layer Instead of producing single-point predictions, our model outputs a probability distribution for future returns. This is implemented as a Gaussian distribution whose parameters (mean and variance) are computed by the network:

$$p(z_t|z_{<t}, \mathbf{x}_t) = \mathcal{N}(\mu_t, \sigma_t^2) \tag{17}$$

where:

$$\mu_t = W_\mu \text{hidden\_permute} + b_\mu \tag{18}$$

$$\sigma_t = \text{softplus}(W_\sigma \text{hidden\_permute} + b_\sigma) \tag{19}$$

The hidden state is first permuted and reshaped to combine information from all LSTM layers:

$$\text{hidden\_permute} = \text{reshape}(\text{final\_hidden}^T) \in \mathbb{R}^{batch \times (\text{lstm\_layers} \cdot \text{hidden\_dim})} \tag{20}$$

This probabilistic approach provides not just predictions but also confidence intervals, enabling better risk assessment and decision-making.

2.3.5 Training Methodology

The model is trained using a carefully designed composite loss function that addresses multiple aspects of financial prediction:

$$\mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{likelihood}} + \beta \mathcal{L}_{\text{amplitude}} + \gamma \mathcal{L}_{\text{direction}} + \delta \mathcal{L}_{\text{frequency}} + \epsilon \mathcal{L}_{\text{relative}} \tag{21}$$

With the following components:

- **Maximum Likelihood Estimation** ($\alpha = 0.2$): The base component is the negative log-likelihood of the predicted distributions:

$$\mathcal{L}_{\text{likelihood}} = -\mathbb{E}[\log p(z_t|\mu_t, \sigma_t)] \tag{22}$$

encouraging accurate probabilistic forecasts.
- **Amplitude Matching Loss** ($\beta = 3.0$): A specialized component focusing on accurately capturing the magnitude of price movements:

$$\mathcal{L}_{\text{amplitude}} = \mathbb{E}[|, |\Delta\mu_t| - |\Delta z_t|, |] \tag{23}$$

where $\Delta\mu_t = \mu_t - \mu t - 1$ and $\Delta z_t = z_t - z_{t-1}$ represent the predicted and actual changes.
- **Direction Loss** ($\gamma = 2.0$): Ensures the model predicts the correct direction of movements:

$$\mathcal{L}_{\text{direction}} = \mathbb{E}[\text{ReLU}(1 - \text{sign}(\Delta\mu_t) \cdot \text{sign}(\Delta z_t))] \tag{24}$$

- **Frequency Loss** ($\delta = 1.0$): Penalizes missed fluctuations in the time series:

$$\mathcal{L}_{\text{frequency}} = \text{MSE}(\mathbf{1}|\Delta\mu_t| > \epsilon, \mathbf{1}_{|\Delta z_t|>\epsilon}) \tag{25}$$

where $\mathbf{1}$ is the indicator function and ϵ is a small threshold (typically 10^{-5}).
- **Relative Amplitude Loss** ($\epsilon = 2.0$): Focuses on the proportional accuracy of predictions:

$$\mathcal{L}_{\text{relative}} = \mathbb{E}\left[\left|\frac{|\Delta\mu_t|}{|\Delta z_t| + 10^{-6}} - 1\right|\right] \tag{26}$$

For evaluation, we use a modified Normalized Deviation (ND) metric that better captures the accuracy of predicted returns:

$$\text{ND} = \frac{\sum_t |r_{\mu,t} - r_{z,t}|}{\sum_t |r_{z,t}|} \tag{27}$$

where $r_{\mu,t} = \frac{\mu_t - \mu_{t-1}}{\mu_{t-1} + 10^{-6}}$ and $r_{z,t} = \frac{z_t - z_{t-1}}{z_{t-1} + 10^{-6}}$ represent the predicted and actual percentage returns. The training process employs the Adam optimizer with a learning rate of 10^{-3} and implements early stopping based on validation set performance (measured by the modified ND) to prevent overfitting. The best model weights are saved during training when validation performance improves.

2.4 Fusion Layer for Enhanced Prediction

To address the limitations of pure time series models and incorporate broader market context, we developed a fusion layer that combines DeepAR predictions with financial and macroeconomic indicators. This approach allows us to refine the primary model's predictions by accounting for fundamental factors that affect stock price movements but may not be fully captured in the historical price patterns alone.

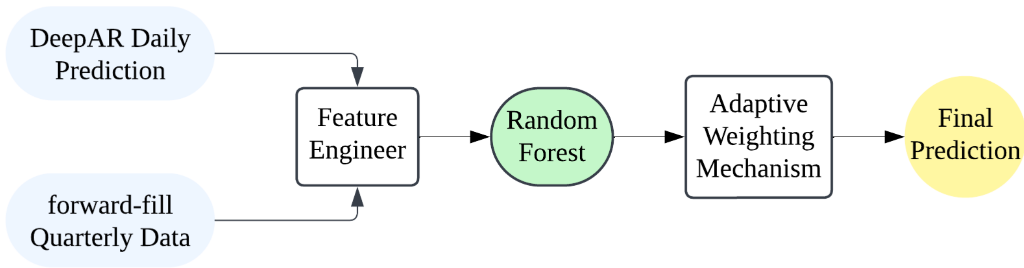


Figure 7: Fusion Layer Architecture

Data Integration The fusion layer integrates two primary data sources:

- **Time Series Predictions**: Outputs from our DeepAR model, containing predictions, actual values, and error metrics.
- **Quarterly Financial Data**: Company-specific financial statements including revenue, profit, and balance sheet metrics.

We map quarterly financial data to daily predictions using an **asof-merge** approach, where each trading day is assigned the most recently available quarterly data. This creates a rich multivariate dataset that combines high-frequency market data with lower-frequency fundamental indicators.

Feature Engineering The fusion layer creates several categories of features to capture different aspects of market behavior:

- **Technical Features**: Including rolling averages of predictions and historical prediction errors over 3, 7, and 14-day windows to capture recent model performance patterns.
- **Fundamental Features**: Key financial ratios derived from quarterly statements, such as profit margins, year-over-year growth rates, and revenue-to-cost ratios.
- **Temporal Features**: Calendar-based variables (day of week, month, quarter) to capture seasonal patterns in market behavior.
- **Lagged Features**: Historical predictions and errors from previous time steps to provide sequence information.
- **Volatility Indicators**: Recent market volatility measurements to adjust prediction confidence.

Missing values are handled using domain-specific methods, with time-based features forward-filled and technical indicators imputed using relevant statistical approaches.

Implementation Details The fusion model uses a Random Forest regressor with 100 estimators, chosen for its robustness to outliers and ability to capture non-linear relationships without overfitting. Model training follows a chronological train-test split (80/20) to preserve the temporal nature of financial data.

Feature importance analysis is performed to identify the most significant factors influencing prediction refinement. This provides interpretability and insight into which external factors most effectively complement the DeepAR predictions.

For forward-looking predictions, where future quarterly data is unavailable, we implement a simple time series projection of financial metrics based on recent growth trends, enabling the fusion layer to make predictions beyond the latest available financial reporting period.

Adaptive Prediction Fusion Rather than simply replacing DeepAR predictions, our fusion layer implements an adaptive blending approach:

$$\hat{y}_{\text{fusion}} = \alpha \cdot \hat{y}_{\text{RF}} + (1 - \alpha) \cdot \hat{y}_{\text{DeepAR}} \tag{28}$$

where $\hat{y}_{\text{fusion}}$ is the final prediction, $\hat{y}_{\text{RF}}$ is the prediction from the fusion model, $\hat{y}_{\text{DeepAR}}$ is the original DeepAR prediction, and α is a dynamic weighting factor.

Crucially, α varies based on market conditions:

$$\alpha = \begin{cases} 0.6 & \text{for normal market conditions} \\ 0.4 & \text{during periods of high volatility} \end{cases} \tag{29}$$

This adaptive approach acknowledges that deep learning models often perform better during volatile periods, while ensemble methods tend to excel in more stable markets.

These fusion layer outputs refined daily stock return predictions that incorporate both time series patterns and fundamental market factors, providing more robust and contextually aware forecasts compared to pure time series approaches. These enhanced predictions can be directly used for trading decisions or as inputs to downstream portfolio optimization processes.

3 Results

3.1 Economic Impact Analysis Module

From the resulting CDNOD causal graphs, we observed that

AT&T (T): Treasury Stock directly impacts AT&T's closing price, driven by Inventory, Property, Plant & Equipment, and Capital Lease Obligations. These factors shape free cash flow, enabling stock buybacks that reduce outstanding shares and boost the stock price. Since AT&T relies heavily on infrastructure to deliver services, efficient capital allocation and liquidity management are key to sustaining shareholder value.

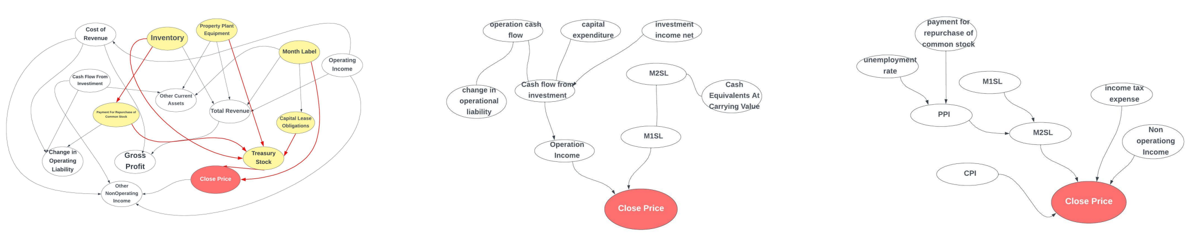
Amazon (AMZN): Operating Income and M1 Money Stock directly impact Amazon's closing price, driven by M2 Money Stock, Cash Equivalents, Investment Income, Capital Expenditure, and Cash Flows. These factors shape free cash flow. Given Amazon's infrastructure-heavy model, efficient capital allocation and liquidity management are crucial for sustaining shareholder value.

Google (GOOG): M2 Money Stock, Non-Operating Income, Income Tax Expense, and CPI directly impact Google's closing price, driven by M1 Money Stock, PPI, stock repurchases, and the Unemployment Rate. These factors affect Google's ability to invest, control expenses, and execute buybacks. With its reliance on infrastructure and ad revenue, efficient liquidity management is essential for sustaining shareholder value.

CVS (CVS): The Unemployment Rate, Inventory, Total Non-Current Liabilities, and Current Debt directly impact CVS's stock price, driven by Capital Expenditure, Debt-Related Metrics, and Security Expenditure. With substantial debt, these factors reflect CVS's ability to manage financial obligations, sustain operations, and invest in growth, ultimately influencing its stock performance. As a retail and healthcare company, CVS heavily relies on human resources, making labor market conditions a key factor in its operational efficiency and profitability.

Abbott Laboratories (ABT): Changes in Operating Cash Flow, Interest, and Debt Expense directly impact ABT's stock price, originating from Changes in Receivables, Current Debt, Operating Cash Flow, and Depreciation, Depletion, and Amortization (DDA). As a healthcare company reliant on drug innovation, these factors influence cash availability, borrowing costs, and profitability, all of which are critical to Abbott's ability to fund Research and Development, expand operations, and return capital to shareholders.

Amgen (AMGN): Changes in Operating Assets and Month Labels directly impact Amgen's stock price, as fluctuations in inventory, receivables, and seasonal trends in drug sales can affect revenue recognition and investor sentiment.

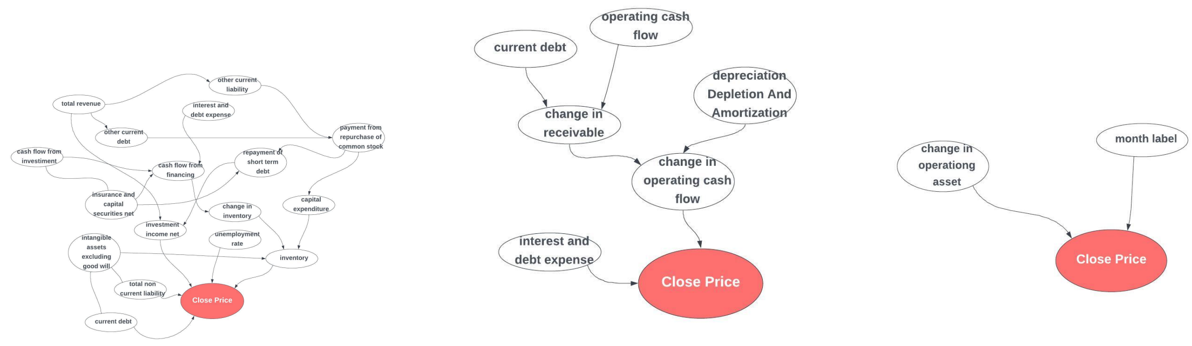


(a) T

(b) AMZN

(c) GOOG

Figure 8: Impactful features relationship for T, AMZN, and GOOG



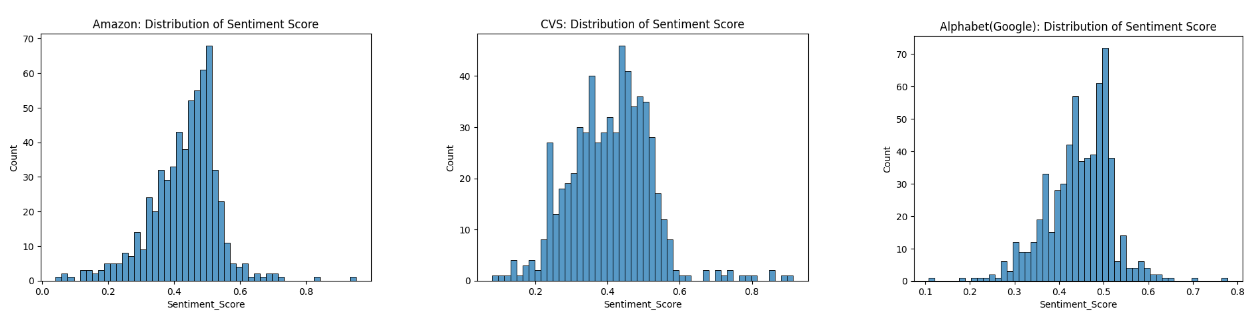
(a) CVS (b) ABT (c) AMGN

Figure 9: Impactful features relationship for CVS, ABT, and AMGN

3.2 Sentiment Analysis Module

We incorporated the FinBERT model to transform raw posts/comments regarding certain companies from X and Reddit social platforms to quantifiable sentiment scores. Our sentiment data spans from June 2020 to February 2025. We observed that these sentiment data seem to be sporadic, so we assumed that posts/comments within a short period would have similar sentiment strength. Then we employed a linear interpolation method that is scaled by time differences to realistically fill in missing values. The sentiment score dataset is poised to capture nuanced factors not easily shown in micro or macro factors of the stock market, thus enhancing our model's stock return prediction performance.

Using FinBERT's sentiment labels and their associated confidence levels, we computed our final sentiment score using the weighted sum of the sentiment labels (positive, negative, and neutral), where each label was weighted by its corresponding confidence level. These sentiment scores were then normalized to a range between 0 and 1 with an approximately normal distribution.



(a) AMZN (b) CVS (c) GOOG

Figure 10: Sentiment Score Distribution of Stocks

3.3 Stock Return Prediction Module

3.3.1 Overall Model Performance Metrics across All Companies

Based on the performance metrics comparison between DeepAR and Fusion Layer models across six companies, we observe mixed results. DeepAR demonstrates superior error metrics (RMSE, MAE, MAPE) for ABT, AT&T, AMGN, and CVS, while the Fusion model achieves better error metrics for AMZN and GOOG. Direction accuracy shows varied performance, with Fusion Layer achieving perfect accuracy (100%) for ABT and AMZN, while DeepAR maintains perfect accuracy for AMGN and CVS. Notably, for CVS, DeepAR significantly outperforms the Fusion model across all metrics, with a direction accuracy of 100% compared to the Fusion model's 25%. These performance differences highlight the variable effectiveness of both approaches across different companies and market conditions.

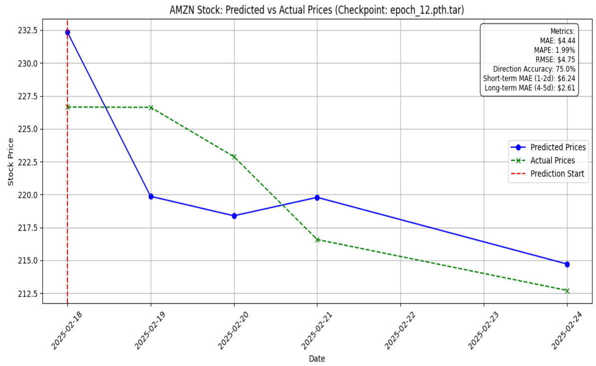
Table 1: Performance Comparison between DeepAR and Fusion Layer Models

Company	RMSE		MAE		MAPE (%)		Direction Accuracy (%)	
	DeepAR	Fusion	DeepAR	Fusion	DeepAR	Fusion	DeepAR	Fusion
ABT	**2.47**	3.09	**2.20**	2.87	**1.65**	2.14	75.0	**100.0**
AT&T	**0.49**	0.84	**0.47**	0.81	**1.79**	3.06	75.0	75.0
AMGN	**3.75**	9.85	**2.77**	9.63	**0.92**	3.20	100.0	100.0
AMZN	4.75	**4.42**	4.44	**3.41**	1.99	**1.57**	75.0	**100.0**
CVS	**0.99**	4.62	**0.93**	3.77	**1.44**	5.91	**100.0**	25.0
GOOG	3.46	**3.19**	2.96	**2.59**	1.61	**1.42**	**75.0**	50.0

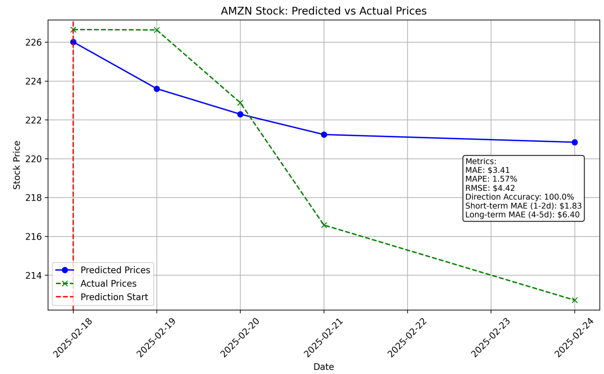
Note: Better performance values for each company are highlighted in **bold**.
Lower values are better for RMSE, MAE, and MAPE; higher values are better for Direction Accuracy.

3.3.2 Forward Prediction Comparisons

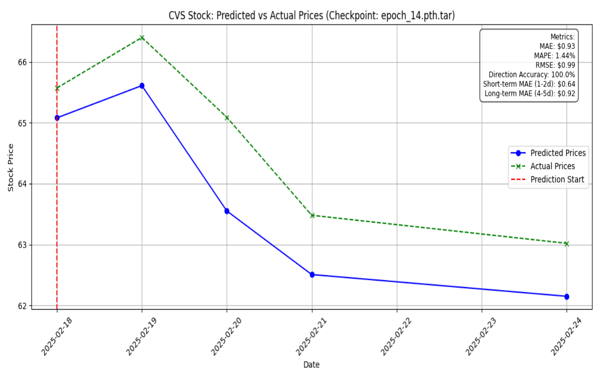
The prediction performance comparison reveals distinct patterns across both models and companies. For AMZN, both models predict a general downward trend, but the Fusion Layer model tracks closer to actual values in the early prediction days before diverging later. In contrast, the CVS comparison shows dramatically opposing predictions, with DeepAR accurately forecasting a downward price movement while the Fusion Layer model incorrectly predicts a sustained upward trend. These visualizations directly support the performance metrics in the table, where DeepAR significantly outperforms for CVS (100% vs 25% direction accuracy) while the Fusion Layer shows better metrics for AMZN.



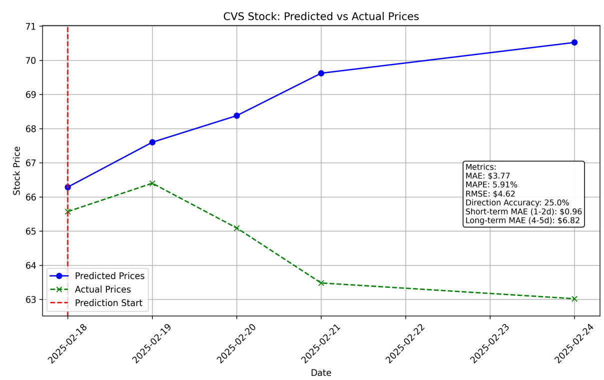
(a) AMZN: DeepAR



(b) AMZN: Fusion Layer



(c) CVS: DeepAR



(d) CVS: Fusion Layer

Figure 11: DeepAR vs Fusion Layer 5-Day Forward Prediction Performance Comparison

4 Discussion

Our project contains several innovative elements as compared to previous works. First, our system integrates multiple data sources, including stock market data, quantifiable sentiment data derived from the FinBERT model, and real-time macroeconomic indicators, which form a comprehensive probabilistic model. Second, we improved predictive accuracy and model interpretability, through refining the DeepAR architecture with hierarchical LSTM layers, and causal feature selection through PCMCI+ and CDNOD algorithms. Transparency can foster higher investor trust and encourage more machine-learning applications in stock return prediction. Our system could also play the role of catalyst for research directions on the creative fusion of probabilistic forecasting and traditional econometric methods.

Our project still has several limitations that need to be addressed in future works. The data quality and frequency alignment issues may enthrall further improvements on our model's prediction accuracy, without access to X API due to its expensive cost, we were not able to scrape sentiment data from the richest sentiment pool, and this led to the usage of interpolation methods to fill missing data. Therefore, the interpolation methods we adopted may introduce noise and uncertainty to our model's predictions.

A significant limitation in our fusion layer approach was the temporal misalignment between quarterly financial data and daily stock returns. Our method of mapping quarterly statements to daily prediction intervals using asof-merge techniques likely contributed to decreased prediction accuracy, as it created artificial data continuity where abrupt changes might occur following earnings announcements or financial disclosures. This frequency mismatch may explain why the fusion layer underperformed compared to the pure DeepAR model for several companies despite incorporating richer fundamental data.

Moreover, our model's reliance on historical trends made it unable to predict results during unprecedented events very well. We hope these limitations can be addressed by future works to improve the model's reliability and applicability.

5 Conclusion

In conclusion, our projects integrate advanced deep learning, sentiment analysis, and causal inference to enhance stock return predictions. The system combines multiple data sources to improve the accuracy and transparency of the model. This comprehensive approach enables us to capture complex market shifts more effectively and also addresses some limitations of traditional predictive models. Through incorporating causal inference analysis within our project, we display clearer insights into relationships of market factors affecting stock returns, which makes the model more transparent to investors. Essentially, this work could help to provide more reliable and transparent investment suggestions.

6 Contributions

Xinqi primarily focused on collecting historical stock stock, company-level data, and sentiment data using the API. Additionally, she collaborated with Vivian on the Economic Impact Analysis Module. Specifically, she conducted CDNOD on features and utilized the selected features for external shock prediction. In the report, Xinqi was responsible for writing the introduction paragraph, the CDNOD section under Methods, and the CDNOD section in the Results.

Jason mainly worked on collecting sentiment data and interpolating for missing data. He also contributed in running and debugging the DeepAR algorithm and exploring possible design of the fusion layer. In the report, he was responsible for writing the literature review, sentiment analysis module, discussion, and conclusion.

Vivian mainly worked on the economic impact module. She contributed to verifying the usability of the DeepAR algorithm, collecting macroeconomic data, and working on the data frequency mismatch issues, transforming monthly macroeconomic data into quarterly datasets using a dynamic factor model. Vivian also worked on the random forest model to predict the most extreme values of the quarter from economic impact data.

Yishan was responsible for model design of the FinBERT sentiment analyzer, PCMCI+ feature selection process, the implementation and optimization of the DeepAR prediction model, as well as the design of fusion layer.

7 Acknowledgments

We would like to express our sincere gratitude to our mentors, Biwei Huang and Jelena Bradic, for their guidance and insights on causal discovery algorithms and our overall framework. We also extend our appreciation to Professor James Hamilton for his valuable advice on addressing frequency mismatches in our economic features. This project has been a collaborative effort, and we are grateful for the support and expertise that have contributed to its success.

References

[1] Hailiang Chen, Prabuddha De, Yu (Jeffrey) Hu, and Byoung-Hyoun Hwang. Wisdom of crowds: The value of stock opinions transmitted through social media. volume 27, pages 1367–1403, 02 2014.

[2] Robert F. Engle, Eric Ghysels, and Bumjean Sohn. Stock market volatility and macroeconomic fundamentals. *The Review of Economics and Statistics*, 95(3):776–797, 2013.

[3] Valentin Flunkert, David Salinas, and Jan Gasthaus. Deepar: Probabilistic forecasting with autoregressive recurrent networks. volume abs/1704.04110, 2017.

[4] DIEGO GARCÍA. Sentiment during recessions. volume 68, pages 1267–1300, 2013.

[5] Steven L. Heston and Nitish Ranjan Sinha. News vs. sentiment: Predicting stock returns from news stories. volume 73, pages 67–83. Routledge, 2017.

[6] Biwei Huang, Kun Zhang, Clark Glymour, and Bernhard Schölkopf. Causal discovery under nonstationary environments. In *Proceedings of the 26th international joint conference on artificial intelligence (IJCAI)*, pages 1347–1353, 2017.

[7] Jiacheng Li, Wei Chen, Zhiheng Zhou, Junmei Yang, and Delu Zeng. DeepAR-Attention probabilistic prediction for stock price series. Springer Science and Business Media LLC, May 2024.

[8] David E. Rapach, Mark E. Wohar, and Jesper Rangvid. Macro variables and international stock return predictability. volume 21, pages 137–166, 2005.

[9] QingLin Xie, Qi Lang, and Xiaodong Liu. Stock price forecasting based on feature fusion deepar model. In *2023 35th Chinese Control and Decision Conference (CCDC)*, pages 4242–4248, 2023.

[10] Yunkai Zhang, Qiao Jiang, Shurui Li, Xiaoyong Jin, Xueying Ma, and Xifeng Yan. You may not need order in time series forecasting. volume abs/1910.09620, 2019.

Appendices

A.1 Project Proposal

Causal Discovery in Stock Return

Xinqi Huang xih037@ucsd.edu
Ruizhe Dai rdai@ucsd.edu
Vivian Zhao vxzhao@ucsd.edu
Yishan Cai yic075@ucsd.edu
Biwei Huang bih007@ucsd.edu
Jelena Bradic jbradic@ucsd.edu

1 Introduction

1.1 General Theme

Many people who own stocks don't have enough financial knowledge to understand the stock market. This lack of knowledge makes it harder for them to avoid losing money, which could be important for their families. The stock market changes quickly and can be hard to predict, making it even more difficult to protect investments. We will spend 10 weeks to address this issue using causal inference algorithms on a dataset that tracks different factors and stock returns. By finding the main causes behind stock returns, we hope to give people a simple and clear explanation of what matters most in the stock market. This can help investors make better decisions and reduce the chance of losing money.

1.2 Problem Statement

The challenge of identifying causal relationships in stock market returns presents unique complexities that extend beyond traditional statistical analysis. While numerous studies have examined correlations between market factors and stock performance, establishing genuine causal relationships remains elusive due to the dynamic, interconnected nature of financial markets and the presence of numerous confounding variables.

Prior research has established various approaches to understanding stock market predictability, focusing on both macroeconomic variables and media sentiment analysis. On the macroeconomic front, Engle, Ghysels and Sohn (2013) highlighted that fundamental factors such as industrial production growth, interest rates, inflation, and unemployment are key determinants of stock market movements, with Rapach, Wohar and Rangvid (2005) demonstrating through a comprehensive study of 12 industrialized countries that interest rates serve as the most reliable predictor in an international context. Parallel to these macroeconomic studies, research has increasingly recognized the importance of media sentiment in predicting stock returns. Notably, Heston and Sinha (2017) analyzed nearly one million news articles and found that daily news coverage could predict stock returns within a one to two-day window, with positive news generating immediate price responses while negative news produced delayed effects. This dual influence of macroeconomic factors and media sentiment was further reinforced by studies such as GARCÍA (2013) and Chen et al. (2014), who found significant relationships between media tone and future stock returns, particularly during periods of economic uncertainty.

Our Quarter 1 project demonstrated the effectiveness of causal discovery algorithms (PC, FCI, and GES) on a simulated dataset. However, applying these methods to real-world financial data presents additional challenges. Unlike simulated data, real financial markets exhibit intricate relationships between macro indicators (CPI, GDP, unemployment rate, policy interest rates), micro conditions (company financials, liability, cash flow), and market sentiment derived from social media data. These relationships often violate the assumptions of basic causal discovery algorithms. Furthermore, stock market data inherently includes temporal dependencies that weren't present in our Quarter 1 simulated dataset, necessitating the exploration of time-series-specific causal discovery methods.

To address these challenges, our research extends beyond the basic PC, FCI, and GES algorithms used in Quarter 1. We investigate more sophisticated approaches, including time-series extensions of FCI (tsFCI), neural network-based causal discovery methods, and hybrid approaches that combine traditional and modern causal inference techniques. This expansion of methodological tools allows us to better capture the complexity of real-world financial data. For the domain expert, our specific research question focuses on extending and adapting causal discovery algorithms to accurately identify the causal relationships between multiple market factors and stock returns. This investigation must account for temporal dependencies, hidden confounders, non-linear relationships, and the potential violation of causal sufficiency assumptions in financial market data. This approach addresses a critical gap in both causal inference methodology and financial market analysis, as previous work has either focused on simplified causal discovery in controlled settings or relied on traditional statistical analysis of market factors.

1.3 Output Expectation

In the next phase of our project, we aim to present our findings through a detailed and visually engaging poster. The poster will summarize the results of our causal discovery analysis, highlighting the primary drivers behind stock returns with DAGs. By combining textual explanations, visualizations of causal networks, and key data insights, the poster will serve as an accessible medium to convey our methodology and conclusions. We will focus on demonstrating the relationships between macroeconomic factors, company-specific metrics, and market sentiment, and the implications for financial decision-making. We will be validating these results with an Economics Department Professor and prior domain knowledge.

We choose the poster format for the following reasons:

1. **Visualizes Complexity**: Causal relationships are intricate, and diagrams or network graphs can convey these insights more effectively than textual descriptions.

2. **Engages Audience**: The format encourages viewers to explore sections of interest, enabling focused discussions and interactive learning.
3. **Supports Accessibility**: Summarized findings are easy to comprehend, making the content suitable for both technical and non-technical audiences.
4. **Fosters Feedback**: The conversational nature of poster sessions allows us to gather constructive criticism and perspectives, enriching our understanding and future work.

1.4 Potential Data

1.4.1 Variables Selection

To discover the causal relationship in stock return, we selected a set of variables that represent macroeconomic conditions, company-specific financial metrics, and market sentiment, as we hypothesized that these factors collectively influence stock prices.

- **Macro Conditions**:
  - Consumer Price Index (CPI): Reflects inflation trends, which are related to individual investor confidence.
  - Gross Domestic Product (GDP): Measures overall economic health and market conditions.
  - Unemployment Rate: Indicates labor market conditions and have direct affect on individual spending.
  - Interest Rate: Directly influences investment decisions.
  - Annual Growth in Interest Rate: Captures the trend of investment activity.
  - Geopolitical Events: Represents global stability, which can impact investor behavior.

  The macroeconomic condition reflects the overall health of the economy, market conditions, job market stability, and international relations which directly impact individual and institutional confidence in investing. These factors shape investor sentiment and risk tolerance, which in turn influence stock prices and returns.

- **Micro Conditions**:
  - Industry: The sector or market in which a company operates.
  - Dividend Rate: The proportion of earnings paid to shareholders as dividends.
  - Overall Risk: A composite measure of various risks.
  - Return on Equity (ROE): A measure of profitability, indicating how effectively a company uses shareholders' equity to generate profits.
  - Liabilities: The company's debts or financial obligations.
  - Cash Flow: The net amount of cash generated or used by a company.
  - Revenue Growth: The rate at which a company's sales are increasing.
  - Debt-to-Equity Ratio: A financial leverage ratio.

  The microeconomic conditions offer a detailed perspective on a company's financial health, its efficiency in generating profits for shareholders, and trends within the industry that would influence investment decisions. These factors are closely linked to stock pricing and, consequently, have a significant relationship with stock returns.

- **Market Sentiment**:
  - Tweet Sentiment: Reflects the collective perception and emotional reaction of investors, influencing short-term stock price movements.

  In economic theory, people's anticipation of future events significantly influences their investment behavior. If investors believe that a company will perform well in the future, either based on their insights on earnings forecasts, new product launches by the company, or positive news reports, they are likely to buy the stock and thus drive its price up. Moreover, when a large number of investors act on sentiment, their collective actions can create a feedback loop due to herd behavior. As a result, we see the necessity of including sentiment as one of our factors. We would include the tweet sentiment on companies to find the causal relationship to the stock return as tweets provide real-time, unfiltered insights into market sentiment, reflecting the emotional and speculative reactions of a diverse investor base.

1.4.2 Data Sources

- **Macroeconomic Variables**: These data can be easily scraped or collected from reliable institutions such as the *U.S. Bureau of Labor Statistics*, *Bureau of Economic Analysis*, and *Federal Reserve Board*. Given that these data are typically reported on a monthly, quarterly, or annual basis, they can also be obtained manually if needed.
- **Microeconomic Variables**: These data can be obtained using the *Yahoo Finance API*, which provides comprehensive company information, including equity, liabilities, and dividends paid to stockholders. The source is reliable as Yahoo Finance's API is one of the pioneers in the financial data market.
- **Market Sentiment and Stock returns**: We will conduct sentiment analysis on tweets from the following dataset: Tweet Sentiment's Impact on Stock Returns.

  This dataset contains 862,231 labeled tweets along with associated stock returns, offering a comprehensive view of how social media sentiment affects company-level stock market performance. The dataset includes the following variables:
  - **Tweet Text**: The content of the tweet.
  - **Stock Symbol**: The company ticker associated with the tweet.
  - **Date**: The date the tweet was posted.
  - **Closing Price at the Time of Tweet**: The stock's closing price when the tweet was published.
  - **10-Day Volatility**: Stock price volatility over the subsequent 10 days.
  - **30-Day Volatility**: Stock price volatility over the subsequent 30 days.

  The tweet data have been directly extracted from the platform and are ready for analysis after pre-processing. Also, the volume of data available in the dataset helps mitigate the effects of missing data and potential information loss during sentiment analysis. For any missing stock return data, we will supplement it using the Yahoo

Finance API to ensure completeness.

References

Chen, Hailiang, Prabuddha De, Yu (Jeffrey) Hu, and Byoung-Hyoun Hwang. 2014. "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media." [Link]

Engle, Robert F., Eric Ghysels, and Bumjean Sohn. 2013. "STOCK MARKET VOLATILITY AND MACROECONOMIC FUNDAMENTALS." *The Review of Economics and Statistics* 95 (3): 776–797. [Link]

GARCÍA, DIEGO. 2013. "Sentiment during Recessions." [Link]

Heston, Steven L., and Nitish Ranjan Sinha. 2017. "News vs. Sentiment: Predicting Stock Returns from News Stories." Routledge. [Link]

Rapach, David E., Mark E. Wohar, and Jesper Rangvid. 2005. "Macro variables and international stock return predictability." [Link]