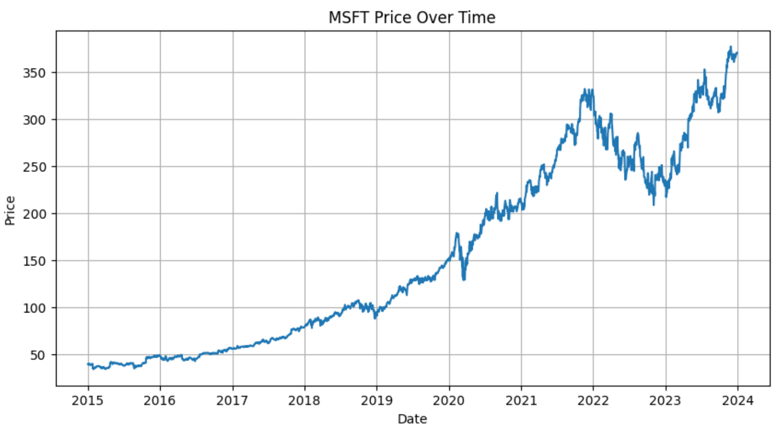
Assignment 2

WiDS Kalman Filtered RL Agent

Aboli Malshikare
23B1211

1 Data and Price Selection

Daily MSFT data from 2015 to 2024 was used. The closing price was selected as it represents the final consensus market value after all intraday information is incorporated. Opening prices are influenced by overnight news, while high and low prices reflect intraday extremes rather than representative value.



2 Stationarity Analysis

The Augmented Dickey–Fuller test showed that the raw price series is non-stationary, while price ratios and return-based features are stationary. Therefore, ratios and returns were used for modeling to avoid spurious regression.

```
ADF Test on MSFT Price
p-value       : 0.9889
Conclusion    : Series is non-stationary (Fail to Reject H0)
--------------------------------------------------
ADF Test on MSFT Price ratio
p-value       : 0.0000
Conclusion    : Series is stationary (Reject H0)
--------------------------------------------------
```

Table 1: Feature Engineering Description

Feature Name	**Definition / Formula**
Price (P_t)	Adjusted closing price at day t
Log Return	$\log\left(\frac{P_t}{P_{t-1}}\right)$
Lag Return	$\log\left(\frac{P_{t-1}}{P_{t-2}}\right)$
MA_5	$\frac{1}{5}\sum_{i=0}^{4} P_{t-i}$
MA_20	$\frac{1}{20}\sum_{i=0}^{19} P_{t-i}$
MA_60	$\frac{1}{60}\sum_{i=0}^{59} P_{t-i}$
ROC_5	$\frac{P_t - P_{t-5}}{P_{t-5}}$
MA Momentum	$MA_5 - MA_{20}$
Volatility (20-day)	σ_{20}(log returns)
Volume MA_20	20-day moving average of trading volume
Volume Change	$\frac{V_t - V_{t-1}}{V_{t-1}}$
Price Ratio (Target)	$\frac{P_{t+1}}{P_t}$
Current Ratio	$\frac{P_t}{P_{t-1}}$
Volatility Ratio	$\frac{\sigma_{20}}{\mathrm{MA}_{60}(\sigma_{20})}$
Trend_60	$\frac{P_t}{MA_{60}} - 1$
Z-score Price	$\frac{P_t - MA_{20}}{\sigma_{20}(P)}$

3 Kalman Filter Model

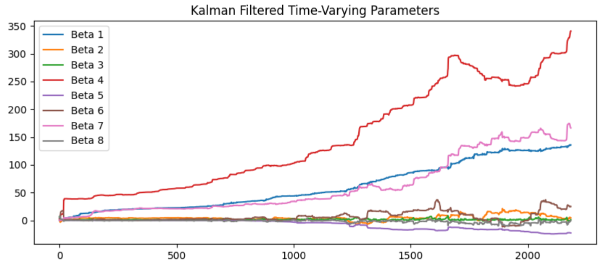
A state-space model was formulated in which the latent state represents **time-varying regression coefficients** linking market features to the observed price.

The **state vector** consists of regression coefficients that are allowed to change over time, capturing evolving market dynamics. The **observation equation** models the observed price as a linear combination of the feature vector and these latent coefficients, with additive observation noise representing market randomness. The **state transition equation** assumes that coefficients follow a random walk, allowing them to evolve smoothly rather than remain fixed.

A small process noise covariance was used to enforce gradual parameter changes, while observation noise accounts for short-term price fluctuations. This formulation enables the model to adapt to structural changes in the market.

In addition, a separate Kalman Filter was applied directly to prices to extract a smooth latent **fair value** and **trend** component, which helps distinguish long-term movement from short-term noise.

Only Beta 1, Beta 4, and Beta 7 show significant variation and meaningfully influence the model, while the remaining betas stay close to zero and have negligible impact. These less influential betas can therefore be ignored without affecting overall performance.



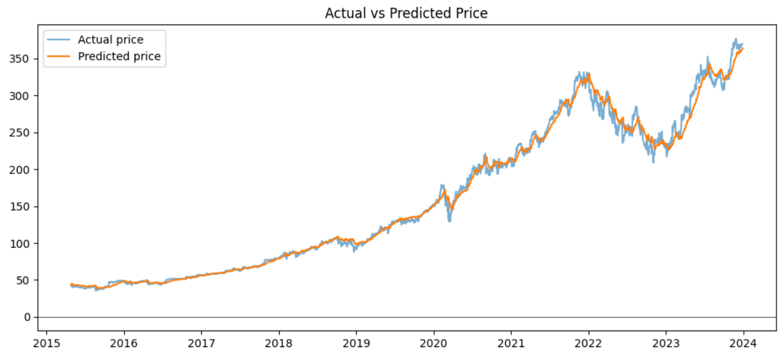
4 Prediction Model

The Kalman-filtered time-varying coefficients were used as inputs to supervised learning models to predict the next-period price. This approach leverages the Kalman Filter to reduce noise and stabilize feature relationships before prediction.

Three models were evaluated: Linear Regression, Ridge Regression, and Random Forest. Model performance was assessed using Root Mean Squared Error (RMSE) on a time-based train–test split to prevent look-ahead bias.

Ridge Regression achieved the lowest prediction error and was therefore selected as the final model. Its regularization helps control overfitting while preserving the interpretability of a linear relationship between filtered features and future prices. . Linear Regression RMSE: 14.577466 Ridge Regression RMSE: 14.577461 Random Forest RMSE: 27.589538

Best Model: Ridge Regression Best RMSE : 14.577461094159522

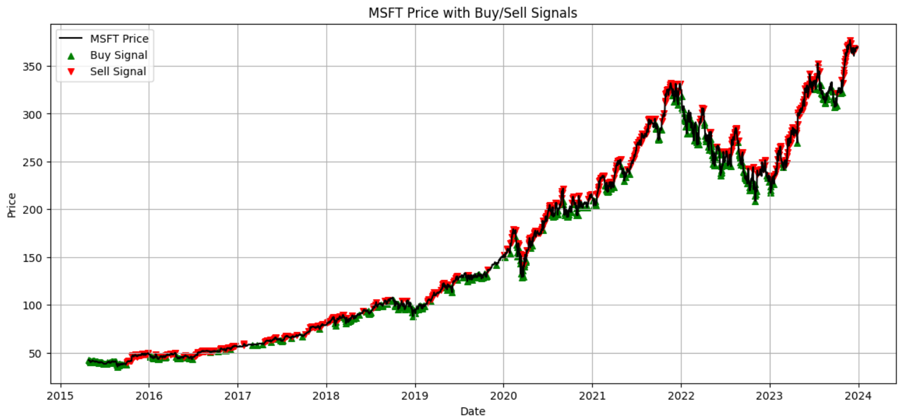


5 Trading Strategy

Trading signals were generated by comparing the predicted price ratio with the current price ratio.

- Buy signal: predicted ratio significantly higher than current ratio
- Sell signal: predicted ratio significantly lower than current ratio

A threshold based on prediction volatility was applied, and signals were shifted to ensure no look-ahead bias. Transaction costs were included.

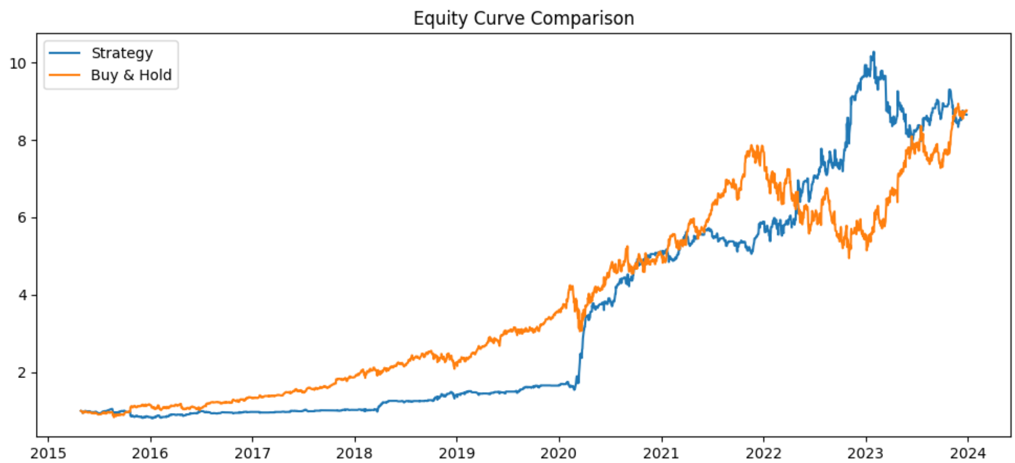


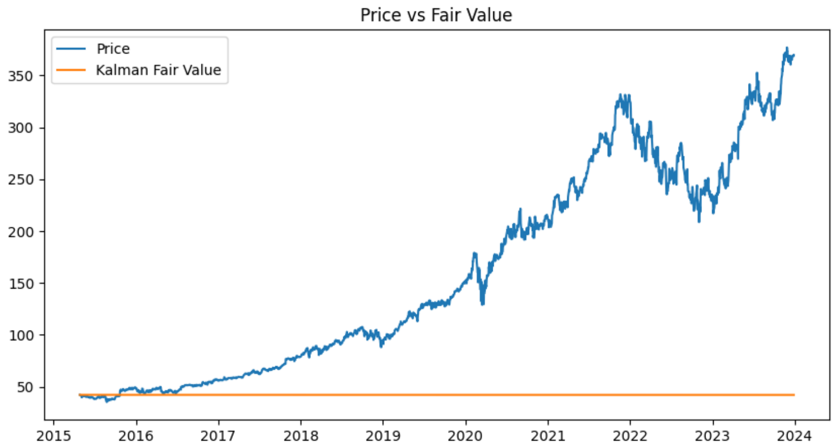
6 Backtesting and Performance

The strategy was evaluated using cumulative return, Sharpe ratio, maximum drawdown, and win/loss ratio. Performance was compared with a buy-and-hold benchmark. The strategy showed improved risk-adjusted returns with controlled drawdowns.

Cumulative Return : 765.18%
Sharpe Ratio : 1.22
Maximum Drawdown : -24.01%
Win/Loss Ratio : 0.56

The strategy demonstrates strong overall performance, achieving a cumulative return of 765.18%, indicating substantial capital growth over the evaluation period. The Sharpe ratio of 1.22 suggests that the returns are well-compensated for the level of risk taken, reflecting solid risk-adjusted performance. While the maximum drawdown of 24.01% indicates periods of notable losses, such drawdowns are generally acceptable within equity trading strategies. Despite a win/loss ratio of 0.56, the strategy remains profitable, as winning trades tend to outweigh losing ones in magnitude, highlighting effective trade management and a favorable payoff structure.





7 Conclusion

The combination of Kalman Filters and machine learning allows adaptation to changing market dynamics. Time-varying parameters improve prediction stability, resulting in a practical and causal trading strategy.