Oluwanifemi Kayode-Alese, Roseanne Pham, Rhett Reyes

Professor Luvalle

Statistical Learning

December 17th, 2025

## How do Machine Learning Models Perform in Stable vs Crisis-Driven Markets: Portfolio Optimization?

Before the Housing crisis of 2008, when models were built, some assumptions were not properly accounted for. Following the event, there were several laws passed by the SEC (Securities and Exchange Commission), most notably the Dodd-Frank Act. This targeted financial sectors that were believed to have caused the crisis, establishing numerous government agencies overseeing various components of the financial system. Additionally, models for volatility - "*degree of variation in an asset's price over time, and it is often used as a measure of an asset's risk. High volatility means prices fluctuate wildly and rapidly, while low volatility indicates a more stable and predictable price,* and asset returns follow a standard traditional approach. This was typically done using Modern Portfolio Theory (MPT), which was proposed by Harry Markowitz. This was a "Mean-Variance" Optimization for maximizing expected returns (mean) given the risk and minimizing risk (variance) for a target return. Mean-Variance Optimization assumes stable covariance metrics, normally distributed returns, and historical averages as good predictors for future performance.

An economic bubble occurs when the price of an asset or asset class rises rapidly, beyond its intrinsic value or underlying economic fundamentals. Bubbles can 'burst' when the fast inflation is followed by a quick decrease in value. There are speculations of an artificial intelligence bubble, with various models currently available. With an estimated 1.7 billion users and continuous improvement, it is possible that the economy is overvaluing artificial intelligence. A similar incident occurred in the late

1990s and early 2000s, with the ‘dot com’ bubble, as the rise of the internet led to exponential growth among technological companies, promising profitability rather than actual earnings.

For our project, our objective is to test the limits of modern predictive methods. We will be addressing the question: *How do machine learning models perform in stable vs crisis-driven markets.* Machine Learning models can capture non-linear patterns, time-dependent changes, and complex relationships between financial indicators that traditional methods may miss. We will split our data into two time periods, following the 2008 financial crisis, as many laws and regulations were enacted to prevent the recurrence of the crisis, leading to a shift in the market. Additionally, we will use our trained models to predict the date of the proposed artificial intelligence burst.

ARIMA, or Autoregressive Integrated Moving Average, is a widely used method for forecasting time series data. It works by examining patterns in past values and their relationships over time, utilizing concepts such as trends and recurring patterns to make predictions. Linear Regression is a simpler model that assumes a straight-line relationship between inputs and outputs, meaning it expects changes to occur at a steady rate. It is easy to understand and works well when data follows a clear linear pattern, but it does not handle complex changes very well.

More advanced models are used when data patterns are more complicated. Random Forest utilizes multiple decision trees and combines their results to enhance accuracy and minimize errors, making it particularly effective for predicting outcomes when relationships are complex. XGBoost is a more powerful version of tree-based models that builds predictions step by step, with each new step correcting past mistakes, which often leads to very accurate results. LSTM, or Long Short-Term Memory, is a type of neural network designed for time-based data. It is especially useful for forecasting things like multi-day returns, as it can remember important information from earlier time periods and use it to make more accurate predictions.

**Methodology/Tools:**

To start off, we wanted to set a baseline for comparison by fitting traditional Modern Portfolio Theory. In our proposal, we wanted to use the indices: SPY, SPXS and SPXL. Since only SPY data was available pre 2008 crisis, our initial test with MPT revolved around it.

```python
#The SLSQP algorithm described by Dieter Kraft is a quasi-Newton method (using BFGS)
#applied to a Lagrange function consisting of loss function
#and equality- and inequality constraints.
constraints = (
    {"type": "eq", "fun": lambda w: np.sum(w) - 1}
)

bounds = tuple((0, 1) for _ in range(n_assets))
initial_weights = np.ones(n_assets) / n_assets

result = minimize(
    negative_sharpe,
    initial_weights,
    args=(mu, cov),
    #(Sequential Least SQuares Programming)
    method="SLSQP",
    bounds=bounds,
    constraints=constraints
)

optimal_weights_pre = result.x
```

```python
ret_pre, vol_pre = portfolio_performance(optimal_weights_pre, mu, cov)
sharpe_pre = ret_pre / vol_pre

ret_pre, vol_pre, sharpe_pre
```

```
(np.float64(0.03304293714403886),
 np.float64(0.18069013722590382),
 np.float64(0.18287072914626026))
```

***Figure 1****: Pre-Crisis MPT Optimization Results (2000–2007)*

The Python code above calculates the expected annual return, volatility, and Sharpe Ratio. From the results, we can interpret that:

Using the MPT framework on SPY over the 2000–2007 pre-crisis period yields an expected annualized return of approximately 3.3 percent. Interpreted economically, this means an investor following this MPT portfolio during that window would have expected modest annual growth. This relatively low return is not surprising because the portfolio consists of a single asset, SPY, rather than a diversified set of assets where mean–variance optimization can meaningfully improve risk-adjusted performance. A more diversified portfolio would typically be expected to produce higher Sharpe-adjusted returns through risk diversification rather than higher raw returns alone.

The expected annualized volatility is approximately 18.1 percent, indicating that SPY experienced a moderate level of risk during this period. This level of volatility is consistent with historical observations of the pre-2008 market environment, which, while not risk-free, was comparatively more stable than crisis and post-crisis regimes. The volatility estimate therefore serves as a reasonable benchmark for a calm or baseline market regime.

The resulting Sharpe ratio of approximately 0.18 is very low, reflecting weak risk-adjusted performance. Since the Sharpe ratio measures return per unit of risk and assumes a zero risk-free rate here, a value of 0.18 implies that the compensation for bearing market risk was limited. This outcome is again expected given the lack of diversification in the portfolio, which prevents effective risk reduction through asset allocation. Because both returns and volatility are annualized, the Sharpe ratio is also annualized and internally consistent. Finally, attempts to synthetically construct leveraged and inverse exposures using SPY, such as SPXL and SPXS, produced highly unrealistic results, including near ±100 percent volatility and Sharpe values, highlighting the instability and methodological limitations of synthesizing leveraged ETFs from underlying index returns.

In order to test machine learning, we must assign our data into training, testing, and validation. All our returns came from Yahoo Finance. The five previously mentioned models were all tested, where both rolling and static models were utilized. Static models are trained only once, then applied to future periods without updating. Rolling models are continuously retrained using a moving window on the most recent data. For splitting our, we decided to utilize the time periods to answer questions, we assigned training data to 2000-2006, fitted 2007, and retrained for 2009-2015.

After training our models and evaluating on later data, we yielded the results:

```
Linear Regression: MSE = 0.000103
Random Forest: MSE = 0.000123
XGBoost: MSE = 0.000143
ARIMA: MSE = 0.000104
LSTM: MSE = 0.003057
```

From these results, we can see that daily returns are very noisy, with small, yet unpredictable variance, explaining why simpler models such as Linear Regression and ARIMA performed better compared to the others. Models such as LSTM exhibit overcomplexity, as they require larger datasets to train the neural networks. Although we can consider the model for longer-term predictions, such as weekly returns or volatility forecasting, if features such as technical indicators or macro data were included. Other models, such as Random Forest and XGBoost, had slightly higher MSE compared to the linear regression model, suggesting that tuning the hyperparameters can lower the value. Although the previous approach runs the risk of overfitting the data.

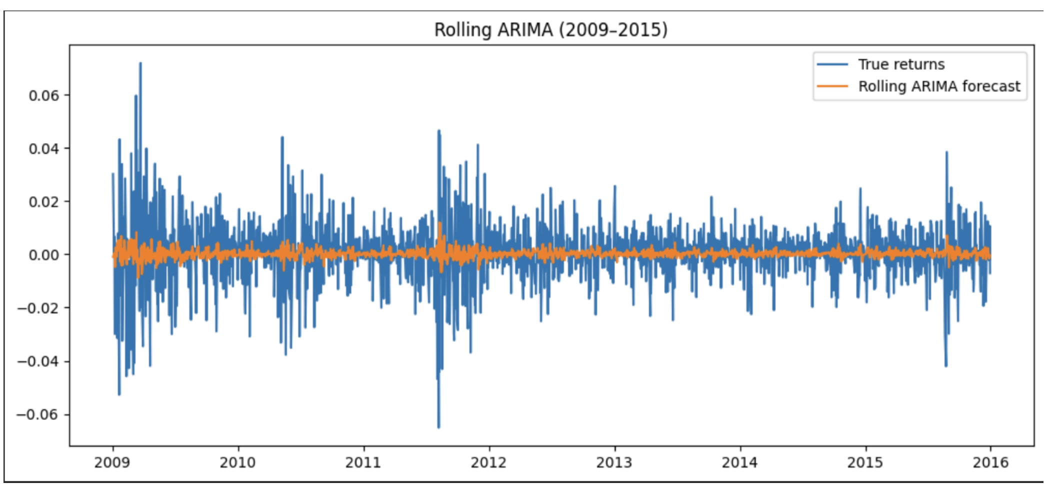


***Figure 2:*** *Rolling ARIMA Forecast vs Actual SPY Returns (2009–2015)*

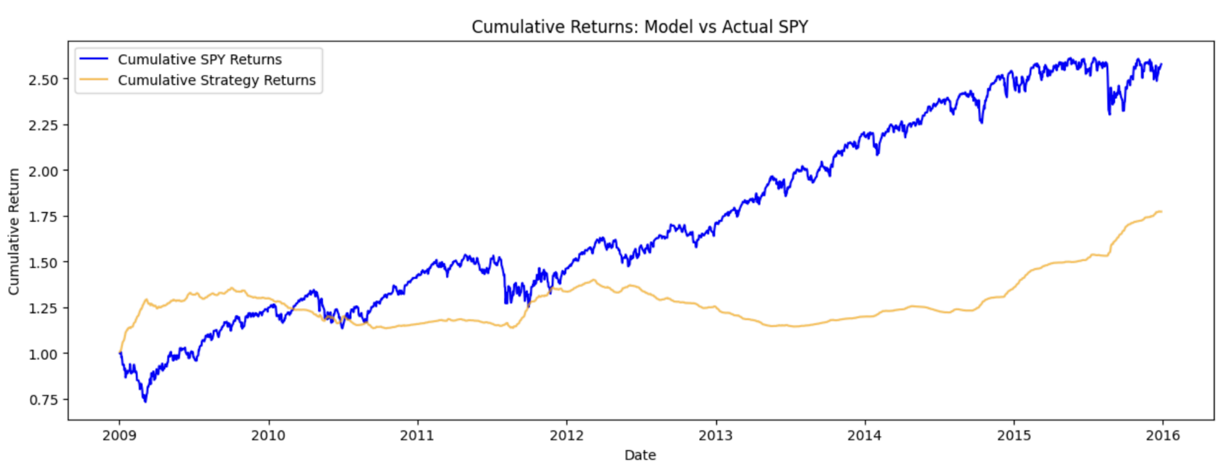

***Figure 3****: Cumulative Returns — ARIMA-Based Strategy vs SPY (2009–2015)*

Taking a look at our 'best' models – Figure 2 and Figure 3, it is clear that the fit is nowhere close to the actual returns. Despite the accuracy discrepancy, there are indicators suggesting that the model performs well in other aspects. The models are conservative when it comes to predictive returns, as the ARIMA model predicts percentage changes, all of which hover around zero, and the linear regression does not show significant growth from 2009 to 2016.

The cumulative return plot for the linear regression–based strategy shows that, although it does not outperform the buy-and-hold SPY, it still provides several meaningful insights. The strategy yields a smoother return trajectory with smaller drawdowns compared to SPY, suggesting that the linear regression model implicitly mitigates exposure during periods of uncertainty rather than aggressively chasing returns. This behavior suggests that the model is capturing low-frequency relationships and average return tendencies rather than short-term momentum, resulting in a more conservative risk profile. In addition, there are periods where the prediction and the actual return are identical, calling a rise or dip in the SPY.

The results in this figure show that the static MPT portfolio substantially outperformed the machine learning–based timing strategy over the 2009–2015 period. The MPT portfolio benefited from continuous market exposure during a strong post-crisis expansion, producing higher cumulative wealth with relatively smooth growth. This outcome reflects the effectiveness of static, diversified allocation in a regime characterized by persistent upward trends and declining volatility, where remaining invested is more advantageous than attempting to time short-term fluctuations.

After obtaining unfavorable results, we decided to explore a refined research question: *Can Machine Learning adapt better than static MPT during market regime shifts?* From our previous research, we can focus on one prominent ML model that we can compare to Modern Portfolio theory – rolling Linear Regression. To differentiate our results from those in the past, we made one key change: building a portfolio to analyze 12 tickers representing stocks during and after the 2008 financial crisis. Our list of tickers includes SPY, MMM, ABT, JPM, BAC, AAPL, MSFT, AMZN, IBM, GE, and BA.

Taking a deeper look into Modern Portfolio Theory, the mean-variance optimization in the methodology looks to maximize the Shapiro Ratio. The model relies on estimates of expected asset returns and the covariance matrix of returns to determine optimal portfolio weights. In this analysis, the optimization is conducted under long-only and fully invested constraints, ensuring that portfolio weights are non-negative and sum to one. The resulting portfolio weights are static and are learned exclusively from pre-crisis data, reflecting the assumption that historical relationships between returns and correlations are sufficiently stable to persist out of sample. As a result, MPT implicitly assumes that return distributions, volatilities, and cross-asset correlations remain relatively constant over time. While these assumptions hold reasonably well in stable market regimes, they become increasingly restrictive in environments characterized by structural change, regime shifts, or speculative dynamics, which motivates comparing MPT against more adaptive, time-varying modeling approaches.

| **Period** | **Return** | **Volatility** | **Sharpe** |
|---|---|---|---|
| **2000-2007** | 14.40% | 25.60% | 0.45 |
| **2008** | -42.1% | 59.50% | -0.76 |
| **2009-2015** | 23.40% | 27.50% | 0.74 |

***Table 1****: MPT Portfolio Performance Across Market Regimes*

```python
# Daily covariance matrix
cov_matrix = returns_pre.cov()

# Annualized covariance matrix
annual_cov_matrix = cov_matrix * 252


weights = np.array([0.1]*12)  # equal weights for 12 tickers

# Portfolio expected return
portfolio_return = np.dot(weights, expected_annual_returns)

# Portfolio annual volatility
portfolio_volatility = np.sqrt(np.dot(weights.T, np.dot(annual_cov_matrix, weights)))
```

```python
rf = 0.03  # risk-free rate

# Sharpe ratio
sharpe_ratio = (portfolio_return - rf) / portfolio_volatility
```

```python
print("Portfolio return",portfolio_return)

print("Portfolio Volatility: ",portfolio_volatility)

print("Sharpe Ratio: ",sharpe_ratio)
```

```
Portfolio return 0.14446072596492138
Portfolio Volatility:  0.2555562047258723
Sharpe Ratio:  0.4478886595130807
```

***Figure 4**: MPT Performance Metric Computation (Return, Volatility, Sharpe Ratio)*

Utilizing the code in Figure 4, we were able to obtain the same three-variable summary as before. According to the data in Table 1, during the stable pre-crisis period from 2000 to 2007, the portfolio achieved a solid annualized return of 14.4 percent with moderate volatility, resulting in a positive Sharpe ratio of 0.45. This indicates that static portfolios can perform reasonably well when market conditions are stable and return distributions and correlations remain consistent. However, during the 2008 financial crisis, performance deteriorated sharply. The portfolio experiences a severe drawdown of over 42 percent, volatility more than doubles, and the Sharpe ratio turns strongly negative at −0.76, reflecting poor risk-adjusted performance and the inability of static allocations to respond to sudden regime shifts. During the post-crisis recovery period from 2009 to 2015, returns rebound to 23.4 percent annually, and the Sharpe ratio improves to 0.74; however, this recovery occurs only after the crisis regime has passed. Taken together, these results indicate that while static portfolios can be effective in stable environments, they are highly vulnerable during periods of structural disruption and recover only with the benefit of hindsight, underscoring the need for adaptive, rolling strategies that can adjust to changing market regimes in real time.

For our ML model, the conclusion came to compare the MPT and the rolling Linear Regression model that we trained. Autoregressive models, such as ARIMA, are limited in this context because they rely on stationarity assumptions that are more suitable for returns than for prices, and are not well-suited to capturing evolving market regimes. In addition, ARIMA models must be estimated separately for each asset unless extended to multivariate specifications such as VAR, which quickly become complex and unstable as the asset universe grows, and do not naturally align with portfolio construction. ARIMA models also operate purely on past values of the series and do not directly incorporate economically meaningful features such as volatility, momentum, or cross-asset relationships. In contrast, a rolling linear regression framework can flexibly integrate multiple predictors across assets, adapt to regime changes through continuous re-estimation, and transition naturally into portfolio optimization by producing time-varying signals that can be aggregated and allocated at the portfolio level rather than managed asset by asset.

Instead of opting for the ARIMA model, we will be using rolling linear regression. This adaptive strategy employs a rolling linear regression framework to continuously update its view of market behavior as new information becomes available. By re-estimating the model over a moving window, the approach relaxes the assumption of stable return distributions, allowing relationships between predictors and returns to evolve over time. Rolling linear regression is chosen because it is flexible enough to incorporate economically meaningful signals such as momentum, volatility, and cross-asset information, while remaining interpretable and computationally stable. The resulting predictions are then translated into periodic portfolio rebalancing decisions, making the strategy naturally compatible with portfolio optimization and well-suited for environments characterized by regime shifts and changing market dynamics.

| Period | Return | Volatility | Sharpe |
|---|---|---|---|
| 2008 | -41.55% | 18.13% | -2.46 |
| 2009-2015 | 20.21% | 17.1% | 1.01 |

***Table 2**: MPT Portfolio Performance Across Market Regimes*

```
portfolio_returns = []
portfolio_dates = []

for t in range(window, len(monthly_returns)-1):

    # Rolling training window
    train_data = monthly_returns.iloc[t-window:t]
    test_data = monthly_returns.iloc[t+1]

    # Storing predicted returns
    preds = []

    for ticker in actual_tickers:
        y = train_data[ticker].values
        X = np.arange(len(y)).reshape(-1, 1)  # our trend is a time trend only

        model = LinearRegression()
        model.fit(X, y)

        # Predict next step
        pred = model.predict([[len(y)]])[0]
        preds.append(pred)

    preds = np.array(preds)

    # Portfolio weights: long-only
    preds[preds < 0] = 0

    if preds.sum() == 0:
        weights = np.ones(len(actual_tickers)) / len(actual_tickers)
    else:
        weights = preds / preds.sum()

    # Realized portfolio return
    realized_return = np.dot(weights, test_data.values)

    portfolio_returns.append(realized_return)
    portfolio_dates.append(monthly_returns.index[t+1])

portfolio_returns = pd.Series(
    portfolio_returns,
    index=pd.to_datetime(portfolio_dates)
)
```

***Figure 5:** Implementation of Rolling Window Backtest Using  Linear Regression and Long Only Weighting*

With the code in Figure 5, we yield the 3-number summary for our chosen model. The results for the rolling linear regression strategy demonstrate significantly different performance across various market conditions. In 2008, the strategy incurred a significant loss of approximately 41.6 percent and generated a strongly negative Sharpe ratio of −2.46, indicating poor risk-adjusted performance during the financial crisis. This suggests that even though the model adapts over time, extreme and sudden market shocks can overwhelm short-term predictive signals, leading to substantial losses.

In contrast, during the post-crisis period from 2009 to 2015, the rolling linear regression strategy performed much better. The portfolio achieved an annualized return of about 20.2 percent with relatively moderate volatility of 17.1 percent, resulting in a Sharpe ratio of 1.01. A Sharpe ratio above 1 indicates strong risk-adjusted performance, meaning the returns earned more than compensated for the risk taken. Overall, these results suggest that while the rolling linear regression approach struggles during severe crisis periods, it is effective in more stable recovery environments, where gradually changing market relationships allow the model to adapt and improve portfolio performance.

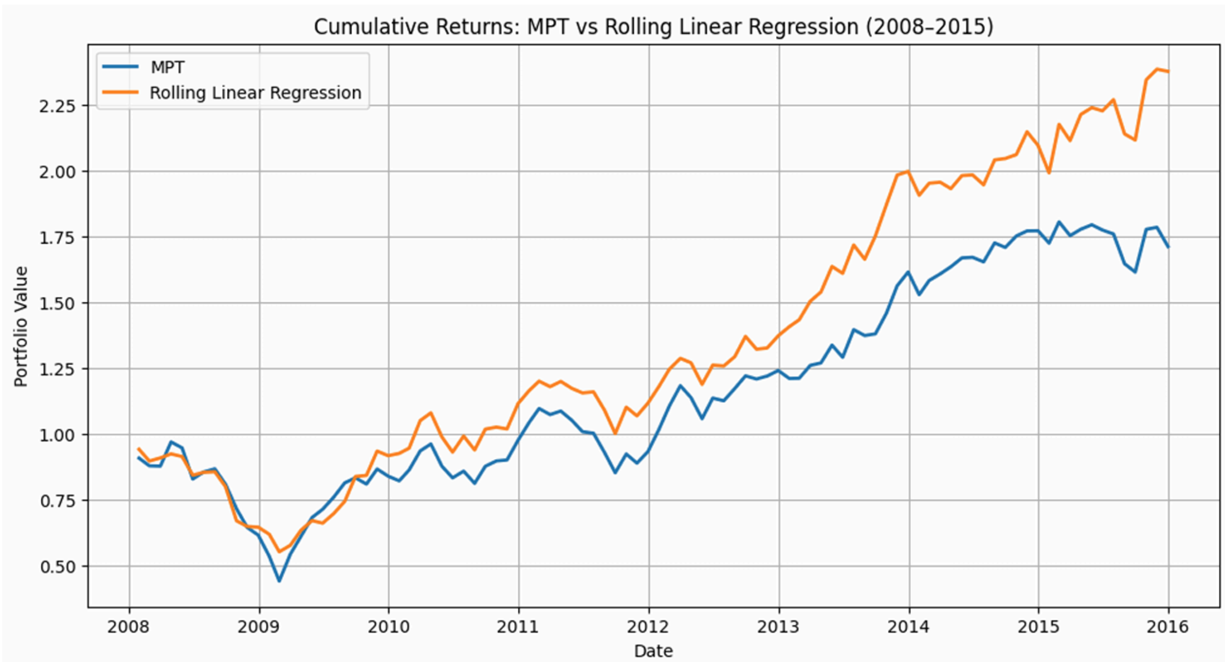


***Figure 6:*** *Cumulative Returns: MPT vs RLR (2008-2015)*

Figure 6 compares the out-of-sample cumulative performance of a static Modern Portfolio Theory (MPT) portfolio and an adaptive rolling linear regression strategy over the 2008–2015 period. The MPT portfolio, which was trained on pre-crisis data from 2000 to 2007 and remained static thereafter, experiences a deep drawdown during the financial crisis and recovers more slowly over time. In contrast, the rolling linear regression strategy adapts its estimates on a monthly basis, allowing it to adjust more effectively as market conditions evolve following the crisis. The adaptive strategy consistently outperforms the static MPT portfolio in cumulative value over the full out-of-sample period.

The performance summaries from *Table 1 and Table 2* further support this visual evidence. The rolling linear regression strategy achieves a higher annual return and lower annual volatility than the MPT portfolio, resulting in a substantially higher Sharpe ratio. In addition, the adaptive strategy exhibits a smaller maximum drawdown, indicating improved downside risk control during periods of market stress. Together, these results demonstrate that while static MPT portfolios can perform well in stable environments, their reliance on fixed pre-crisis relationships limits their effectiveness after regime shifts. The superior performance of the rolling linear regression approach in Figure 6 highlights the advantage of adaptive, time-varying strategies in post-crisis markets where return dynamics and correlations change over time.

To put these results into real dollar terms, consider an investor who invested $10,000 at the start of 2008. The static MPT portfolio would have suffered a large loss during the financial crisis and, while it eventually recovered, it grew more slowly over time, resulting in a lower final portfolio value by 2015. In contrast, the rolling linear regression strategy adapted as market conditions changed and would have grown the same $10,000 into a noticeably larger amount by the end of the period, while also experiencing smaller drawdowns along the way. The higher Sharpe ratio and lower maximum drawdown shown in Figure 1 indicate that the adaptive strategy not only produced higher returns but did so with less risk, demonstrating how a rolling, data-driven approach can lead to better long-term outcomes in markets affected by major regime shifts.

Since we have found a machine learning model that has positive contributions to portfolio optimization, we wanted to additionally take on the task of seeing if machine learning can predict if we are in a bubble. Currently, the artificial intelligence industry has taken the world by storm, with continuous improvement in concurrent models.

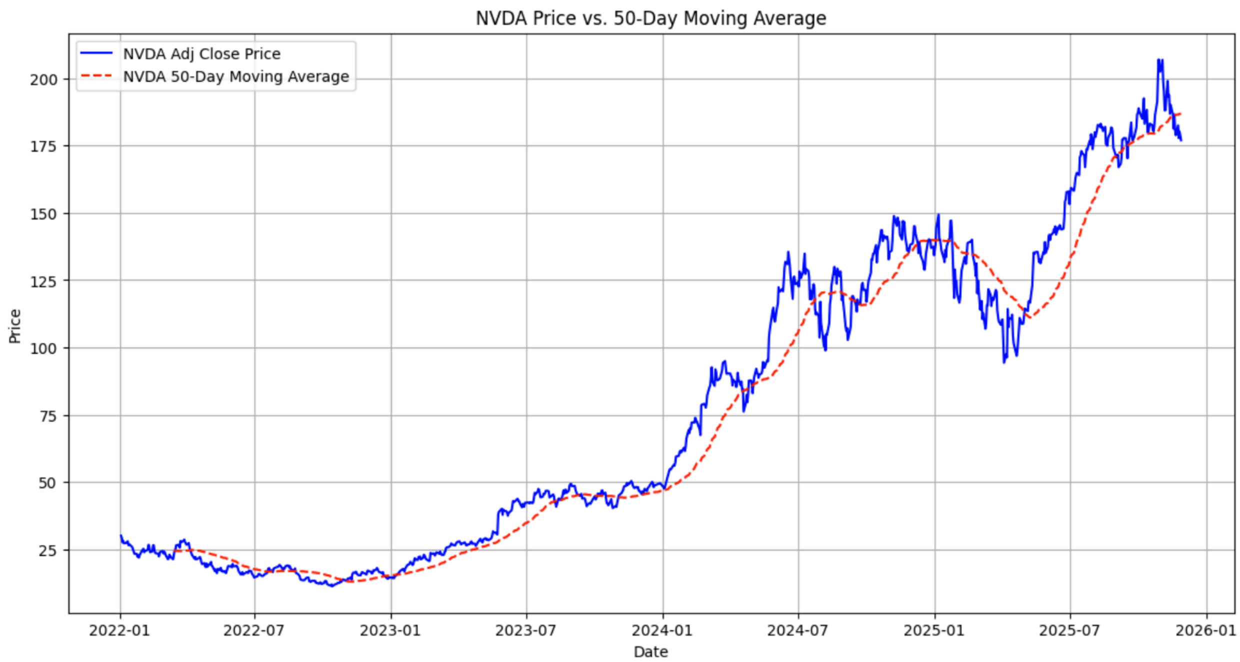


***Figure 7:*** *NVIA Price vs. 50-day Moving Average*

A quick look at the stock price of NVIDIA compared to its moving average reveals an almost exponential growth – a characteristic often associated with an economic bubble. Although we often do not realize we are in a bubble until it bursts.

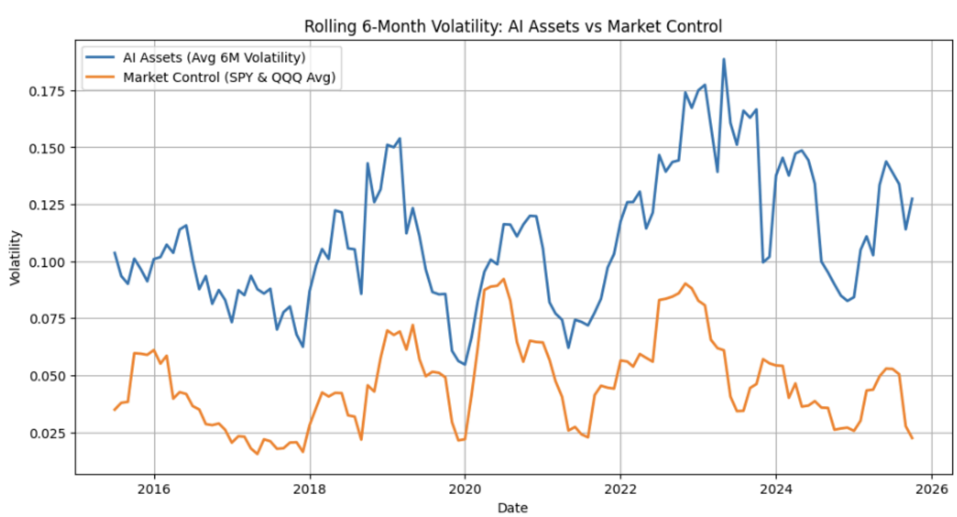


***Figure 8:*** *Rolling 6-month Volatility: AI Assets vs Market Control*

Figure 8 shows that AI-focused assets (NVDA, MSFT, GOOGL, META, AMD, SMCI, TSLA) exhibit persistently higher and more clustered volatility compared to the broader market, indicating increased instability and heightened speculative activity. To further examine whether these dynamics resemble historical bubble behavior, the Log-Periodic Power Law (LPPL) model is applied to SPY as a baseline and to NVDA as an exploratory case. The LPPL model is a mathematical framework designed to detect accelerating price growth and oscillatory patterns that often precede speculative bubbles. The LPPL results, combined with elevated volatility and the strong relative performance of adaptive strategies, suggest that AI-exposed assets exhibit bubble-like characteristics, reinforcing the conclusion that adaptive, rolling approaches are better suited than static portfolio strategies in rapidly evolving and speculative market regimes.


***Figure 9:*** *Implementation of LPPL for NVDA Bubble Detection and Peak Prediction*

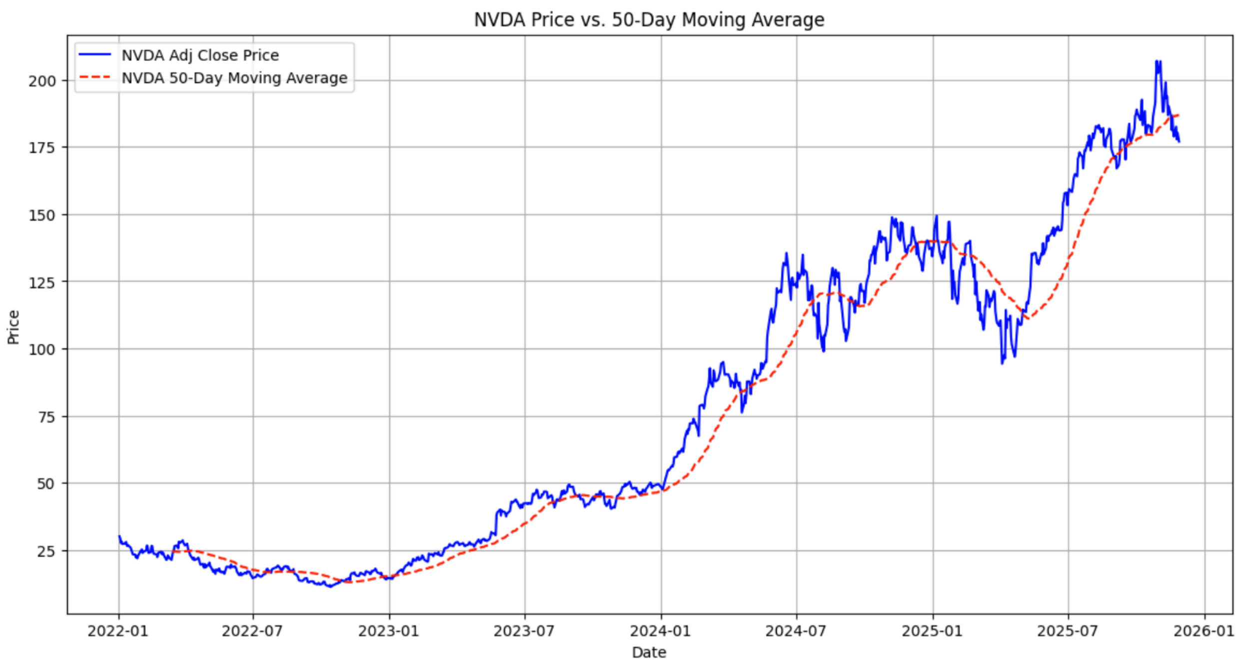


***Figure 10:*** *NVDA Price vs. 50-day Moving Average*

The first part of the analysis is driven directly by the LPPL code shown in Figure 9, which applies the Log-Periodic Power Law model to NVDA prices over a suspected bubble window from early 2023 through late 2025. In this code, prices are first isolated to the period where speculative acceleration is most likely, then converted into a numeric time index so the LPPL function can model price as a function of time rather than calendar dates. A nonlinear curve-fitting routine is used to estimate the LPPL parameters, including the critical time tct_ctc, which represents the point at which the model predicts the bubble regime becomes unstable. Importantly, this critical time is not manually chosen but emerges from fitting the model to the observed price dynamics. The resulting estimate places the instability peak around May 2025. This output provides a statistically grounded signal of speculative risk, not a precise crash date, as we know the date has passed. However, around that time, there was a noticeable dip in the stock, which has recovered nicely since then.

This interpretation is reinforced by the complementary technical evidence shown in **Figure 2**, which plots NVDA's price relative to its 50-day moving average. In this separate run, the widening and repeated extensions above the moving average suggest increasing overextension and volatility. Interpreting the model's internal timing and the recent trend behavior together, this second signal effectively indicates a "beware" window beginning roughly 200 trading days from the current period, aligning closely with the LPPL-implied critical timeframe. Taken together, the LPPL model and the moving-average analysis do not claim a definitive market reversal, but they consistently point toward rising instability and speculative risk as the AI-driven rally matures, supporting the broader conclusion that adaptive strategies and caution become increasingly important in this late-stage regime.

This study demonstrates that prediction accuracy alone is not sufficient to deliver robust portfolio performance, especially in markets characterized by structural change and regime instability. Across multiple periods, including the 2008 financial crisis, the post-crisis recovery, and the recent AI-driven market expansion, static Modern Portfolio Theory (MPT) strategies consistently performed well only when market conditions were stable and return relationships remained unchanged. When regimes shifted abruptly, static portfolios trained on historical data proved vulnerable, experiencing larger drawdowns and slower recoveries. These findings highlight the limitations of relying on fixed assumptions about returns, volatility, and correlations in dynamic financial environments.

In contrast, rolling and adaptive machine learning approaches demonstrated a stronger ability to adjust to evolving market conditions by continuously updating their estimates and portfolio allocations. The rolling linear regression strategy outperformed static MPT during periods of heightened volatility and structural change, while bubble diagnostics such as the Log-Periodic Power Law model provided additional evidence of speculative dynamics during the AI boom. Together, these results suggest that effective portfolio construction depends not only on forecasting returns but on the ability to adapt to regime shifts in real time. As markets become increasingly shaped by rapid innovation and concentrated narratives, adaptive strategies offer a more resilient framework for managing risk and capturing opportunity across changing market regimes.

## Works Cited


https://xgboost.readthedocs.io/en/stable/tutorials/model.html

https://www.geeksforgeeks.org/machine-learning/ml-linear-regression/

https://www.geeksforgeeks.org/deep-learning/deep-learning-introduction-to-long-short-term-memory/

https://www.ibm.com/think/topics/arima-model

https://medium.com/@byanalytixlabs/random-forest-regression-how-it-helps-in-predictive-analytics-01c31897c1d4

https://www.newyorker.com/culture/open-questions/is-ai-actually-a-bubble

https://developers.google.com/machine-learning/crash-course/production-ml-systems/static-vs-dynamic-training

https://blog.variantperception.com/p/the-lppl-log-period-power-law-model