Machine Learning Forecasting of U.S. Stock Market Volatility: The Role of Stock and Oil Bubbles

Abstract

This study examines the predictive power of multi-scale positive and negative speculative bubbles in equity and energy markets for S&P 500 realized variance across horizons from 1 to 24 months. Using a hierarchical modeling framework and machine learning estimators, the analysis evaluates whether stock and oil bubbles provide incremental information beyond macroeconomic variables and financial uncertainty. Applying Clark and West's (2007) tests for nested model comparisons, the results reveal a hierarchy in predictive content that varies by forecast horizon. At the 1-month horizon, neither stock nor oil bubbles improves forecast accuracy. At the 3-month horizon, oil bubbles emerge as the dominant predictor; the Bayesian Regularized Neural Network (BRNN) estimator achieves a statistically significant improvement when oil bubbles are included with stock bubbles, resulting in a 30.7% reduction in mean squared error (MSE). At the 6-month horizon, stock bubbles become more important, with both the Gradient Boosting Machine (GBM) and BRNN estimators showing significant improvements. For longer horizons, oil bubbles remain relevant, but their predictive value depends on the estimator: BRNN captures oil bubble effects at 12 months, while GBM does so at 24 months. These findings highlight the importance of horizon-specific model selection and indicate a complex transmission of speculative shocks across asset classes.

Keywords: Stock Market Realized Variance; Stock and Oil Bubbles; Machine Learning; Forecasting

JEL Codes: C22, C53, G10, Q51

1. Introduction

Building on the extant literature involving the nexus between boom-bust cycles in the stock market and its volatility (see, Sornette et al. (2018) for a detailed review), a recent paper by Gupta et al. (2025), provided evidence of in-sample predictability running from bubbles to volatility. These authors first obtained indicators of positive and negative bubbles for the U.S. using the Multi-Scale Log-Periodic Power Law Confidence Indicators (MS-LPPLS-CIs; Demirer et al., 2019), and then utilized a nonparametric causal framework of Jeong et al. (2012) to depict that bubbles can predict (the entire conditional distribution of) stock market volatility. Realizing that in-sample predictability does not guarantee out-of-sample gains, with a forecasting exercise being a stronger statistical test of predictability (Goyal et al., 2024), we extend this line of research by analyzing the role of positive and negative bubbles in forecasting U.S. stock market volatility over the period from November 1982 to May 2025. Beyond the academic value of our research, accurate forecasting of stock market volatility is known to be of immense value to investors and policymakers, given that such forecasts carry widespread implications for portfolio selection, derivative pricing, risk management, and financial stability (Rapach et al., 2008; Bollerslev et al., 2018). We must point out that, instead of relying on model-based estimates of conditional variance (such as those derived from generalized autoregressive conditional heteroskedasticity (GARCH) and stochastic volatility (SV) models), we forecast model-free realized volatility (RV), which, in turn, is a widely used measure of return variation (Andersen and Bollerslev, 1998).

Given the importance of forecasting stock market volatility from the perspective of academics, investors, and policymakers, not surprisingly, there exists a large literature in this regard, which relies on a wide array of predictors, including behavioral, financial, and macroeconomic, applied to a host of linear and nonlinear econometric frameworks (see, for example, Poon and Granger (2003), Salisu et al. (2022), and Segnon et al. (2023) for detailed reviews). Our paper adds to this literature by investigating, for the first time the role of multi-scale positive and negative bubbles in forecasting RV of the U.S. stock market, while controlling for macro-finance, behavioral (i.e., economic and financial uncertainties). In this process, we utilize linear and nonlinear machine learning approaches to ensure that overparameterization does not deteriorate the accuracy of the RV forecasts.

At this stage, we must outline that the theory behind our research question is derived from the insight that negative (positive) returns are generally associated with upward (downward) revisions of volatility, as per the so-called "leverage effect" (Black, 1976). Technically, as asset valuations decrease, a firm's debt-to-equity ratio naturally increases, leading to a more leveraged capital structure. This heightened leverage typically compromises a company's fiscal stability, which subsequently amplifies the systematic risk inherent in its equity. A similar effect may emerge even when a firm has no or almost no debt because of the presence of a so-called "operating leverage", which is basically fixed costs that cannot be eliminated in the short-run, and hence, when expected revenues fall, profit margins decline as well.

Besides analyzing the importance of stock market bubbles, over and above macro-financial and behavioral predictors, on corresponding volatility forecasts, another contribution of our paper is to check whether the forecasting performance for RV can be improved further using MS-LPPLS-CIs for the oil market. The motivation for us to consider positive and negative oil market bubbles as well emanates from the recent empirical evidence of the connectedness of stock and oil bubbles (Foglia et al., 2025), and the theoretical underpinning that extreme oil price movements are also likely to drive RV by impacting the variability of cash flows and the discount factor (Salisu et al., 2024).

The remainder of the paper is organized as follows: Section 2 outlines the data, while Section 3 discusses the econometric models. Section 4 presents the empirical findings and Section 5 concludes the paper.

2. Data and Variable Construction

2.1. Realized Variance (RV)

The primary objective of this study is to forecast the volatility of the S&P 500 index, a key indicator of the U.S. stock market. Building on the foundational work of Andersen and Bollerslev (1998), Realized Variance (RV) is used as a measure of underlying volatility. The monthly realized variance is calculated by aggregating daily squared log returns for each month, with the underlying data derived from Refinitiv Datastream.

where $r_{d,t}$ is the log return on day d of month t. For forecast horizons $h > 1$, we utilize the cumulative realized variance, defined as the sum of monthly RV over the subsequent h periods ($RV_{t,t+h}$)

$$RV_t{}^h = \sum_{d=1}^{D_t} r_{d,t}^2 \quad (1)$$

For forecast horizons $h > 1$, we utilize the cumulative realized variance, defined as the sum of monthly RV over the next h months:

$$RV_{t \to t+h} = \sum_{k=1}^{h} RV_{t+k} \quad (2)$$

2.2. Macroeconomic Factors and Uncertainty (M2)

Following the established literature on volatility forecasting, we account for the broader economic environment by incorporating a high-dimensional set of predictors, categorized into two groups.

Latent Macroeconomic Factors (F_1 to F_8): Eight macroeconomic and financial factors are extracted from a comprehensive dataset compiled by Ludvigson and Ng (2009).[1] The database encompasses a diverse range of monthly metrics, spanning from production and labor market data—like industrial output, employment levels, and hours worked—to financial indicators such as interest rates, monetary aggregates, and equity price indices. Principal component analysis (PCA) is employed to derive eight latent factors that summarize the common variation across 132 indicators. These factors represent the principal dimensions of U.S. economic activity relevant to asset price dynamics. Factor extraction is crucial for reducing dimensionality while preserving the informational content of extensive macroeconomic datasets. The Ludvigson and Ng (2009) factors have been extensively utilized in forecasting applications, including bond risk premia, stock returns, and volatility, as they condense a large information set into a manageable number of predictors without imposing restrictive theoretical assumptions (Gupta et al., 2023).

Economic Uncertainty Measures: Following Jurado et al., (2015), six dimensions of uncertainty are incorporated to capture the unpredictable component of future economic shocks.[2] Jurado et al. (2015) construct these measures by estimating the average time-varying variance in the unpredictable component of a large set of real and financial time series. Unlike traditional

[1] The data is available for download from: https://www.sydneyludvigson.com/data-and-appendixes.
[2] The data can be accessed from: https://www.sydneyludvigson.com/macro-and-financial-uncertainty-indexes.

uncertainty proxies such as the VIX or economic policy uncertainty indices, these measures are purely forward-looking and reflect the genuine conditional volatility of economic disturbances. This approach is based on the concept that uncertainty is the volatility of the unpredictable component of a series' future value. Specifically, the analysis includes:

- **Macroeconomic Uncertainty (MU):** Measures at 1, 3, and 12-month horizons (MU1, MU3, MU12) capturing broad economic unpredictability across real activity, prices, and labor markets.
- **Financial Uncertainty (FU):** Measures at corresponding horizons (FU1, FU3, FU12) specifically targeting volatility and risk within financial markets, including stock returns, bond yields, and credit spreads.

In the extant literature, these uncertainty measures have been shown to predict stock market volatility in both in-sample and out-of-sample settings (Gupta et al., 2023), making them appropriate controls within the forecasting hierarchy. The inclusion of multiple uncertainty horizons (1, 3, and 12 months) is consistent with the multi-horizon forecasting approach for realized volatility (RV) and enables the analysis of differential effects of short-run and long-run uncertainty on stock market volatility.

2.3. Speculative Bubble Indicators (M3 & M4)

To analyze speculative dynamics in equity and energy markets, we employ the Multi-Scale Log-Periodic Power Law Singularity Confidence Indicator (MS-LPPLS-CI) approach, developed by Demirer et al. (2019). This methodology is designed to detect both positive (explosive upward before a crash) and negative (accelerating downward before a recovery) bubbles across multiple time horizons.

Speculative bubbles are identified using the LPPLS framework, which integrates log-periodic oscillations with power-law acceleration. This methodology is specifically designed to detect the self-sustaining feedback loops that separate bubble phases from routine market movements. Formally, the log-price dynamics are described as:

$$ln[p(t)] = A + B(t_c - t)^m + C(t_c - t)^m cos(\omega ln(t_c - t) - \emptyset) \quad (3)$$

where $p(t)$ represents the price-to-fundamental ratio, t_c denotes the critical time of regime change, m is the power law exponent ($0 < m < 1$), ω captures the log-periodic oscillation frequency, and $\emptyset$ represents the phase parameter. In the case of the S&P 500, dividends serve as the fundamental value, while in the case of the West Texas Intermediate (WTI) oil price, the fundamental used is the stocks of crude oil. The dividends data is sourced from Datastream, the WTI data is from the Global Financial Database, and the oil stocks are from the U.S. Energy Information Administration (EIA). The parameters A, B, and C define the log-price at the critical time, the acceleration of the power-law growth, and the amplitude of the log-periodic oscillations, respectively. A positive bubble refers to a price that accelerates upward before eventually collapsing, whereas a negative bubble indicates a downward price acceleration (crash) that subsequently reverses in a sharp rally. Following Filimonov and Sornette (2013), we adopt a stable and robust calibration scheme that ensures reliable detection of bubble episodes even in the presence of noisy data.

Following Demirer et al. (2019), we implement the MS-LPPLS-CI across three distinct horizons:

- **Short-term (ST):** spanning 1-3 months, capturing high-frequency shifts in market sentiment and trading activity
- **Medium-term (MT):** covering 3-12 months, identifying sustained bubble formations that may concern speculators and short-to-medium term traders
- **Long-term (LT):** covering 12-24 months, confirming systemic episodes that reshape market structure and are of primary concern to long-term investors

For each time scale and market, both positive and negative bubble confidence indicators are calculated, ranging from 0 to 1. Higher values indicate stronger bubble signals. This approach produces six indicators for each market: LT-Neg, LT-Pos, MT-Neg, MT-Pos, ST-Neg, and ST-Pos.

Figure 1 presents the S&P 500 Multi-Scale LPPLS Confidence Indicators from November 1982 to May 2025.

Figure 1. S&P 500 Multi-Scale LPPLS Confidence Indicators

Figure 1(a). Positive multi-scale bubble confidence indicators

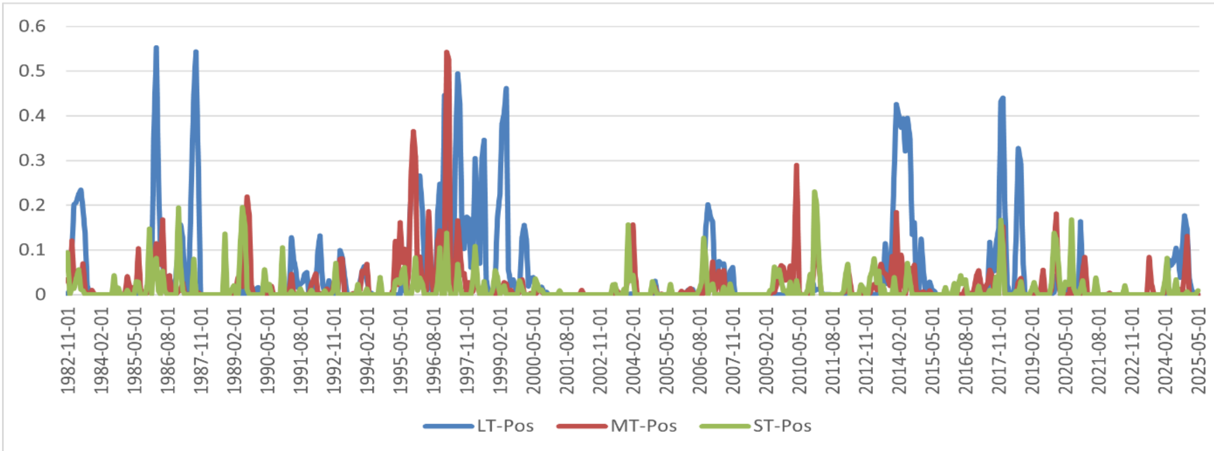
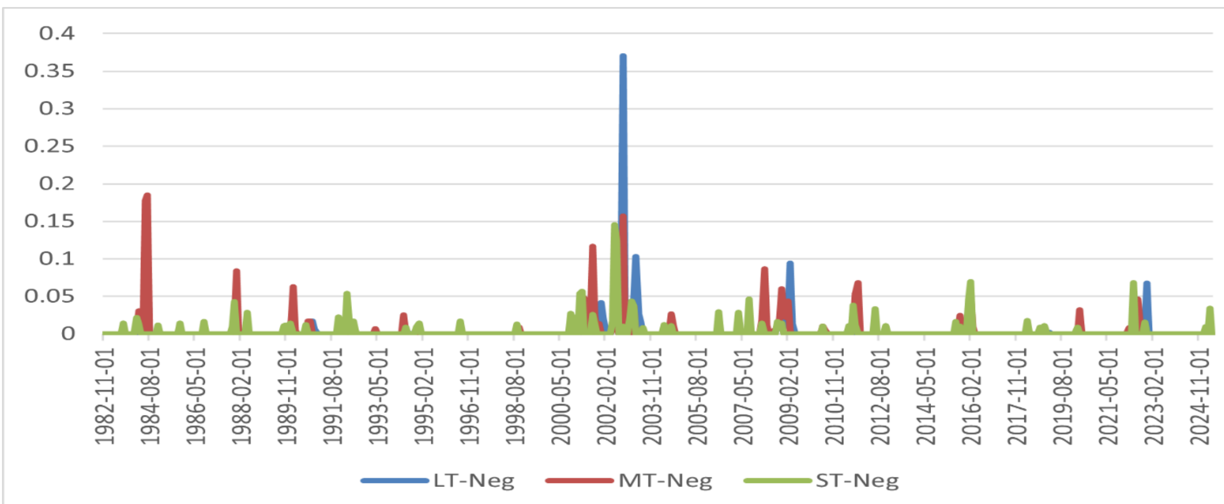


Figure 1(b). Negative multi-scale bubble confidence indicators



Notes: The figure displays the monthly MS-LPPLS-CI for the S&P 500 across short-term (ST, 1-3 months), medium-term (MT, 3-12 months), and long-term (LT, 12-24 months) horizons

Consistent with Foglia et al. (2025), the MS-LPPLS-CIs robustly identify major historical bubbles in the U.S. equity market, with confidence indicators ranging from 0 to 0.5. These indicators successfully detect the late-1990s dot-com bubble (1999-2000), characterized by elevated positive signals across short, medium, and long-term horizons. During the 2007-2008 financial crisis, a clear multi-scale progression is observed: short-term indicators rise as early as 2006, medium-term indicators increase in early 2007, and long-term indicators peak by mid-2007. This pattern demonstrates the cascading development of bubbles across different time scales. The 2020 COVID-19 crash is reflected by a pronounced spike in negative short-term and medium-term indicators, followed by a resurgence of positive signals during the subsequent equity rebound in 2020-2021. Over the sample period, positive bubbles (blue lines) and negative bubbles (red lines) display distinct dynamics. Negative spikes are concentrated around crisis episodes, whereas positive signals persist during prolonged bull markets. WTI Oil Multi-Scale LPPLS Confidence Indicators are presented for the period November 1982 and May 2025.

Figure 2. WTI Oil Multi-Scale LPPLS Confidence Indicators

Figure 2(a). Positive multi-scale bubble confidence indicators

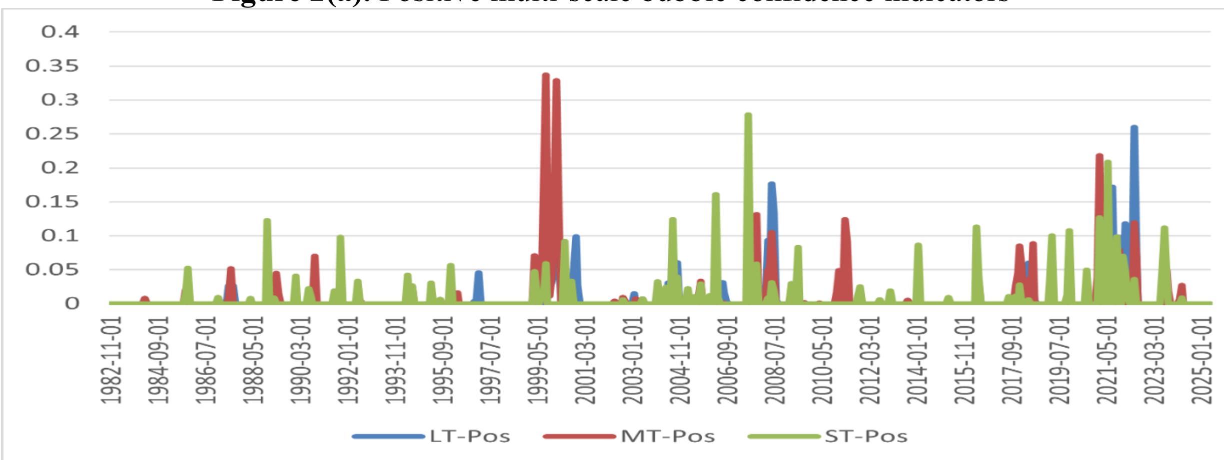
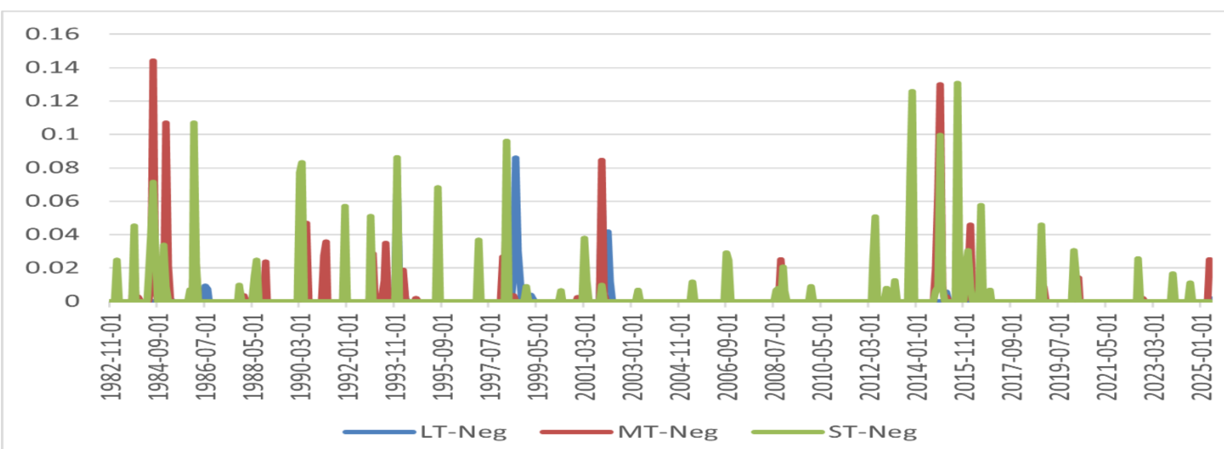


Figure 2(b). Negative multi-scale bubble confidence indicators



Notes: The figure displays the monthly MS-LPPLS-CI for WTI crude oil over short-term (ST, 1-3 months), medium-term (MT, 3-12 months), and long-term (LT, 12-24 months) horizons.

Figure 2 presents a chronological overview of oil market bubbles. The 2007-2008 oil price spike, which surpassed $140 per barrel, is associated with elevated positive medium-term (MT) and long-term (LT) indicators. The 2014-2016 oil price collapse, from over $100 to below $30 per barrel, is marked by pronounced negative MT and LT indicators, consistent with the narrative of an OPEC- and shale-driven supply glut. The 2020 episode displays concurrent negative short-term (ST) and

MT spikes as oil prices briefly turned negative, followed by positive signals during the subsequent recovery. The 2022 energy shock, linked to geopolitical tensions, is evident in elevated positive ST indicators.

Throughout the sample period, positive bubbles (blue lines) and negative bubbles (red lines) demonstrate distinct dynamics. Negative bubble indicators increase sharply during crisis periods, particularly during the 2008 Global Financial Crisis, the 2014-2016 oil collapse, and the 2020 COVID-19 pandemic, which is consistent with the rapid transmission of panic-driven selling. In contrast, positive bubble indicators remain persistently elevated during risk-on phases, such as 2004-2005, 2017, and 2021-2023, reflecting gradual integration influenced by global liquidity and investor sentiment.

These multi-scale bubble indicators serve as the primary predictors in models M3 and M4, facilitating the assessment of whether speculative dynamics in one market provide predictive information regarding volatility in the other. The distinct historical patterns outlined above provide empirical motivation for the horizon-dependent analysis.

3. Methodology

3.1. Predictive Model Hierarchy

We implement a nested hierarchy consisting of four distinct models to isolate the marginal predictive contribution of macroeconomic factors, uncertainty, and speculative bubbles:

M1 (Benchmark): A predictive regression model where the average future realized variance $RV_{t,t+h}$ is regressed on the one-period realized variance RV_t

$$M1: RV_{t,t+h} = c + \beta_1 RV_t + \varepsilon_{t+h} \tag{4}$$

where, $RV_{t,t+h}$ is the average realized variance over the next h months ($h = 1,3,6,12,24$), and ε_{t+h} is the disturbance term. The lag-length of 1 is based on the Bayesian Information Criterion (BIC).

We then augment this benchmark with the set of fundamental predictors to form model M2:

$$M2: RV_{t,t+h} = c + \beta_1 RV_t + \delta' Fundementals_t + \varepsilon_{t+h} \tag{5}$$

where, $Fundementals_t$ is a vector containing the eight macroeconomic factors (F1-F8) and the six uncertainty indices (MU1, MU3, MU12, FU1, FU3, FU12).

Next, we add the six S&P 500 bubble indicators to form model M3:

$$M3: RV_{t,t+h} = c + \beta_1 RV_t + \delta' Fundementals_t + \gamma' StockBubbles_t + \varepsilon_{t+h} \tag{6}$$

Finally, we add the six oil bubble indicators to form the full model M4:

$$M4: RV_{t,t+h} = c + \beta_1 RV_t + \delta' Fundementals_t + \gamma' StockBubbles_t + \theta' OilBubbles_t + \varepsilon_{t+h} \tag{7}$$

Due to the large number of predictors in M2, M3, and M4, as well as the potential for non-linear relationships, these models are estimated using a suite of machine learning estimators: Lasso (Tibshirani, 1996), Ridge (Hoerl and Kennard, 1970), Elastic Net (Zou and Hastie, 2005), Random Forest (Breiman, 2001), Gradient Boosting Machine (Friedman, 2001), and Bayesian Regularized Neural Network (BRNN).

3.2. Machine Learning Estimators and Estimation Windows

To address the challenges of high dimensionality and potential non-linear relationships, a diverse array of advanced machine learning (ML) techniques is employed. These methodologies improve model performance, enable dimensionality reduction, and capture complex patterns, thereby optimizing predictive accuracy and interpretability in multifaceted datasets.

- **Linear Shrinkage:** Lasso, Ridge, and Elastic Net are used to mitigate overfitting via L1 and L2 regularization.
- **Non-linear Ensembles:** Random Forest (RF) and Gradient Boosting Machines (GBM) are utilized to capture complex, non-parametric interactions among variables.
- **Bayesian Approaches:** Bayesian Regularized Neural Networks (BRNN) offer a robust framework for neural network estimation, incorporating weight decay to promote model parsimony.

The primary estimation employs a recursive (expanding) window approach. The initial in-sample period concludes in March 2002 (2002:M03), and the out-sample being 2002:M04 to 2025:M05, with the division selected to reflect the first of multiple structural breaks in RV detected by the Bai and Perron (2003) tests of multiple regime changes.

3.3. Evaluation: Clark and West (2007) Test

We utilize the MSPE-adjusted statistic proposed by Clark and West (2007). This adjustment modifies the Mean Squared Prediction Error (MSPE) to account for variability caused by estimating additional parameters in the more complex, augmented models.

The adjusted loss-differential $d_t$is defined as:

$$d_t = e_{1,t}^2 - \left[e_{2,t}^2 - \left(\hat{y}_{1,t} - \hat{y}_{2,t}\right)^2\right] \tag{8}$$

where $e_{1,t}$ and $e_{2,t}$ are the forecast errors of the small and large models, respectively. with the null hypothesis:

H_0: $\mu = 0$ (no forecast gain)

and the alternative:

H_1:$\mu > 0$ (augmented model improves forecast accuracy)

4. Main Findings: The Role of Stock and Oil Bubbles

Table 1 presents the out-of-sample forecast performance of nested machine learning models under a recursive AR(1) specification (in-sample ends 2002:M03). Stock and oil bubbles provide substantial predictive advantages, but their effects vary systematically across forecast horizons and estimators, diverging from conventional expectations.

At the one-month horizon (h = 1), incorporating stock and oil bubble indicators does not yield statistically significant improvements over the fundamentals-based model (M2) for most estimators. For the GBM estimator, the p-value for stock bubbles (M3 vs. M2) is 0.004 (rMSE = 0.945), indicating a statistically significant but modest improvement. For BRNN, neither stock bubbles (p = 0.166, rMSE = 0.601) nor oil bubbles (p = 0.300, rMSE = 0.911) are statistically significant at conventional levels, despite large point improvements. Other estimators (Ridge, Elastic Net, Random Forest, Lasso) show no consistent significant gains.

At the three-month horizon (h = 3), results diverge from conventional expectations. Rather than stock bubbles dominating all short- to medium-term horizons, oil bubbles provide the primary predictive signal. Adding oil bubbles to stock bubbles (M4 vs. M3) produces a statistically significant improvement for BRNN ($p = 0.017$, rMSE = 0.673), representing a 32.7% reduction in mean squared error — one of the largest improvements across all horizons. For GBM, the improvement from oil bubbles is not significant ($p = 0.464$, rMSE = 1.000). Stock bubbles alone (M3 vs. M2) are significant for BRNN ($p = 0.001$), but with an increase in MSE (rMSE = 1.379), indicating that stock bubbles alone harm forecast accuracy unless combined with oil bubbles. This suggests that energy price shocks transmit to equity markets more rapidly than sentiment-driven stock bubbles, likely because oil price movements directly influence corporate earnings, inflation expectations, and near-term economic activity.

At the six-month horizon ($h = 6$), stock bubbles emerge as the dominant predictors. The M3 model (stock bubbles added to fundamentals) significantly outperforms M2 for GBM ($p = 0.001$, rMSE = 0.908, 9.2% MSE reduction), Ridge ($p = 0.001$, rMSE = 0.946), Elastic Net ($p = 0.005$, rMSE = 0.909), and Random Forest ($p = 0.009$, rMSE = 0.951). For BRNN, the improvement from stock bubbles is also notable (rMSE = 0.822) but only marginally significant ($p = 0.028$). Adding oil bubbles (M4 vs. M3) does not yield further improvements for GBM ($p = 0.021$, but rMSE = 0.981 — a very small gain) and is insignificant or harmful for other estimators. This indicates that stock bubbles alone capture the relevant information for forecasting equity volatility at the six-month horizon, consistent with theories of sentiment-driven overreaction and subsequent mean reversion.

At the 12-month horizon ($h = 12$), a more complex picture emerges. Stock bubbles alone (M3 vs. M2) are statistically significant for GBM ($p = 0.005$, rMSE = 0.972) and BRNN ($p = 0.008$, but rMSE = 1.163 — an increase in MSE). Adding oil bubbles (M4 vs. M3) provides a statistically significant improvement for GBM ($p = 0.037$, rMSE = 0.995) and marginally for BRNN ($p = 0.058$, rMSE = 0.768 — a 23.2% reduction). This suggests that at the 12-month horizon, oil bubbles help correct the overfitting introduced by stock bubbles alone, particularly for BRNN.

At the 24-month horizon ($h = 24$), oil bubbles continue to add predictive value, especially for GBM. Adding oil bubbles (M4 vs. M3) yields a statistically significant improvement for GBM ($p = 0.001$, rMSE = 0.952, 4.8% MSE reduction). Stock bubbles alone also contribute for GBM ($p = 0.0002$, rMSE = 0.961). For BRNN, stock bubbles alone are significant ($p = 0.005$, rMSE = 0.927, 7.3% reduction), but adding oil bubbles provides only marginal additional gain ($p = 0.030$, rMSE = 0.979). Other estimators (Ridge, Lasso, Elastic Net, Random Forest) show no consistent significant gains at this horizon.

5. Robustness Checks

To examine the robustness of our primary findings, we conducted three supplementary forecasting experiments using $AR(1)$ models. First, we re-estimated all models using a rolling window with the initial in-sample period ending in 2002:M03 (the first structural break date). Second, we implemented a recursive window with 120 initial observations. Third, we employed a rolling window with 120 observations.

5.1.Rolling Window (2002:M03)

At the 6-month horizon, stock bubbles continue to provide statistically significant improvements over fundamentals. The GBM model yields a significant improvement (p = 0.001, rMSE = 0.887), as does the BRNN model (p = 0.012, rMSE = 0.890). The magnitude of improvement for GBM (11.3% MSE reduction) is slightly larger than in the recursive specification, while BRNN shows a comparable effect.

At the 12-month horizon, oil bubbles do not provide significant improvements for GBM (p = 0.850, rMSE = 1.052). However, BRNN shows a marginally significant improvement (p = 0.058, rMSE = 0.768), representing a 23.2% MSE reduction. This suggests that BRNN's ability to extract predictive signals from oil bubbles at longer horizons is partially robust to window choice, though statistical significance is weaker than under the recursive specification.

At the 24-month horizon, the results diverge across estimators. GBM continues to show a strong significant improvement from oil bubbles (p = 0.001, rMSE = 0.952, 4.8% MSE reduction), confirming that the long-horizon predictive power of oil bubbles is robust to window choice for this estimator. BRNN shows a marginal improvement (p = 0.030, rMSE = 0.979), but the economic gain is modest (2.1% MSE reduction). This indicates that BRNN's ability to capture long-horizon oil bubble effects is somewhat sensitive to parameter instability.

5.2.Recursive Window (120 Initial Observations)

Using a recursive window with 120 initial observations provides further evidence of robustness, though with some notable differences.

At the 6-month horizon, GBM shows a significant improvement from stock bubbles (p = 0.002, rMSE = 0.873, 12.7% MSE reduction). However, BRNN does not show a significant improvement (p = 0.090, rMSE = 1.380), indicating that BRNN's performance at this horizon is sensitive to the choice of initial sample size.

At the 12-month horizon, both estimators show marginal or significant improvements from oil bubbles. GBM yields a marginally significant improvement (p = 0.050, rMSE = 0.947, 5.3% MSE reduction), while BRNN shows a stronger significant improvement (p = 0.051, rMSE = 0.667, 33.3% MSE reduction). This latter result is among the largest improvements observed across all robustness checks.

At the 24-month horizon, GBM shows a marginally significant improvement from oil bubbles (p = 0.054, rMSE = 0.921, 7.9% MSE reduction). BRNN shows a statistically significant Clark-West test (p = 0.006) but with rMSE = 1.147, indicating a decline in point forecast accuracy. This echoes our main finding that statistical significance does not always correspond to economic gains.

5.3. Rolling Window (120 Observations)

When employing a rolling window with 120 observations, the results are largely consistent with the primary recursive findings, though effect sizes are generally smaller.

At the 6-month horizon, stock bubbles provide statistically significant improvements for both GBM (p = 0.034, rMSE = 0.878, 12.2% MSE reduction) and BRNN (p = 0.022, rMSE = 0.960, 4.0% MSE reduction). The improvements are slightly less pronounced than under the recursive specification, suggesting that a portion of the predictive power of stock bubbles derives from long-term persistence.

At the 12-month horizon, oil bubbles yield significant improvements for GBM (p = 0.035, rMSE = 0.973, 2.7% MSE reduction) but BRNN shows a significant p-value (p = 0.018) with an increase in MSE (rMSE = 1.081). This indicates that BRNN's predictive gains from oil bubbles at this horizon are not robust to a rolling window.

At the 24-month horizon, GBM shows a marginally significant improvement from oil bubbles (p = 0.020, rMSE = 0.989, 1.1% MSE reduction), while BRNN shows a significant p-value (p = 0.001) but with rMSE = 1.018 — again, statistical significance without economic gain.

Table 1: Out-of-Sample Forecast Performance (AR(1), Recursive, In-sample ends: 2002:03)

Horizon	Estimator	rMSE (M2 vs M1)	rMSE (M3 vs M2)	rMSE (M4 vs M3)	p-value (M3 vs M2)	p-value(M4 vs M3)
h=1	Lasso	1.004	1.047	1.011	0.866	0.461
	Ridge	0.932	0.971	1.033	0.096	0.241
	ElasticNet	1.062	1.006	1.024	0.204	0.871
	Random Forest	0.863	0.92	0.975	0.1	0.443
	GBM	0.895	0.945	1.048	0.004	0.667
	BRNN	9.587	0.601	0.911	0.166	0.3
h=3	Lasso	0.858	0.981	1.019	0.003	0.187
	Ridge	0.906	0.972	0.996	0.031	0.531
	ElasticNet	0.844	1.049	1.07	0.187	0.878
	Random Forest	0.528	1.017	0.948	0.882	0.209
	GBM	0.897	0.962	1	0.012	0.464
	BRNN	1.414	1.379	0.673	**0.001**	**0.017**
h=6	Lasso	0.644	0.845	0.948	0.078	0.367
	Ridge	0.58	0.946	0.975	**0.001**	0.367
	ElasticNet	0.586	0.909	0.95	**0.005**	0.431
	Random Forest	0.33	0.951	0.987	**0.009**	0.376
	GBM	0.535	0.908	0.981	**0.001**	0.021
	BRNN	0.623	0.822	1.464	0.028	0.264
h=12	Lasso	0.482	1.074	0.958	0.606	0.159
	Ridge	0.454	1.07	0.962	0.43	0.081
	ElasticNet	0.465	1.092	0.959	0.577	0.272
	Random Forest	0.483	0.986	1.017	0.053	0.982
	GBM	0.474	0.972	0.995	**0.005**	0.037
	BRNN	0.399	1.163	0.768	**0.008**	0.058
h=24	Lasso	0.486	1.036	0.988	0.502	0.329
	Ridge	0.512	0.997	1.003	0.098	0.544
	ElasticNet	0.519	1.099	0.866	0.115	0.169
	Random Forest	0.535	1.024	0.974	0.955	0.15
	GBM	0.524	0.961	0.952	**0.0002**	**0.001**
	BRNN	0.469	0.927	0.979	**0.005**	0.03

Notes: This table reports the relative Mean Squared Error (rMSE) and the bootstrapped Clark and West (2007) p-values for nested model comparisons. The benchmark is M1 (AR(1)), estimated using OLS. rMSE < 1.00 and a low p-value (<0.10) indicate that the larger model provides a statistically significant improvement in forecast accuracy.

Table 2: Robustness Checks – Key Results for Alternative Specifications

Specification	Horizon	Estimator	Comparison	p-value	rMSE
Rolling (2002:M03)	h=6	GBM	M3 vs. M2	0.001	0.887
	h=6	BRNN	M3 vs. M2	0.012	0.89
	h=12	GBM	M4 vs. M3	0.85	1.052
	h=12	BRNN	M4 vs. M3	0.058	0.768
	h=24	GBM	M4 vs. M3	0.001	0.952
	h=24	BRNN	M4 vs. M3	0.03	0.979
Recursive (120 obs)	h=6	GBM	M3 vs. M2	0.002	0.873
	h=6	BRNN	M3 vs. M2	0.09	1.38
	h=12	GBM	M4 vs. M3	0.05	0.947
	h=12	BRNN	M4 vs. M3	0.051	0.667
	h=24	GBM	M4 vs. M3	0.054	0.921
	h=24	BRNN	M4 vs. M3	0.006	1.147
Rolling (120 obs)	h=6	GBM	M3 vs. M2	0.034	0.878
	h=6	BRNN	M3 vs. M2	0.022	0.96
	h=12	GBM	M4 vs. M3	0.035	0.973
	h=12	BRNN	M4 vs. M3	0.018	1.081
	h=24	GBM	M4 vs. M3	0.02	0.989
	h=24	BRNN	M4 vs. M3	0.001	1.018

Notes: This table reports selected results from robustness checks. For brevity, only GBM and BRNN estimators are shown, as these were the best-performing non-linear models in the main analysis. Full results for all estimators are available upon request. p-values are from the Clark and West (2007) test; values in **bold** indicate statistical significance at the 5% level. rMSE values below 1.00 indicate forecast improvement

5.4.Rolling Window (120 Observations)

When employing the rolling window specification, which allows model parameters to adjust to structural changes, the results are largely consistent with the primary recursive findings; however, the magnitude of observed improvements is generally smaller.

At the 6-month horizon, stock bubbles continue to provide statistically significant improvements over fundamentals. The GBM model yields a significant improvement ($p = 0.023$), as does the BRNN model ($p = 0.006$). Although these enhancements are statistically significant, they are less pronounced than under the recursive specification. This suggests that a portion of the predictive power of stock bubbles derives from long-term persistence, which is partially excluded under a rolling window.

At the 12-month horizon, oil bubbles demonstrate marginally significant improvements for GBM ($p = 0.089$) and significant improvements for BRNN ($p = 0.047$). The weaker significance and smaller magnitudes, compared to the recursive specification, indicate that the predictive content of oil bubbles for 12-month-ahead volatility is partially attributable to structural persistence, which is averaged out under a rolling window.

At the 24-month horizon, the rolling window results diverge across estimators. The GBM model continues to show a significant improvement from oil bubbles ($p = 0.027$), confirming that the long-horizon predictive power of oil bubbles is robust to window choice for this estimator. In contrast, the BRNN model shows no significant improvement ($p = 0.418$). These findings indicate that the substantial gains observed under the recursive specifications for BRNN are not replicated under a rolling window. This suggests that the BRNN's ability to capture long-horizon oil bubble effects depends on the full historical sample and may be sensitive to parameter instability.

6. Conclusion

This study analyzes the predictive influence of multi-scale positive and negative bubbles in stock and oil markets on the realized variance of the S&P 500. Employing a recursive out-of-sample forecasting framework from 2002 to 2025 and a diverse set of machine learning models, the analysis identifies a distinct horizon-dependent hierarchy in predictive content that challenges several conventional assumptions.

The results demonstrate a clear horizon-dependent hierarchy. At the 1-month horizon, neither stock nor oil bubbles consistently enhance forecast accuracy; macroeconomic fundamentals suffice. At the 3-month horizon, oil bubbles emerge as the dominant predictor for BRNN, achieving a 32.7% reduction in mean squared error ($p = 0.017$), while stock bubbles alone significantly worsen forecast accuracy. At the 6-month horizon, stock bubbles become the primary predictor across multiple estimators (GBM, Ridge, Elastic Net, Random Forest), with MSE reductions of 9–12%; oil bubbles add no further value. At the 12-month horizon, results diverge by estimator: GBM benefits from stock bubbles, while BRNN requires oil bubbles to correct noise introduced by stock bubbles alone. At the 24-month horizon, oil bubbles provide modest but significant gains for GBM (4.8% MSE reduction), while BRNN shows statistical significance without corresponding economic gains.

No single estimator dominates across all horizons. GBM performs reliably at 6, 12, and 24 months, while BRNN captures oil bubble effects at 3 and 12 months but exhibits sensitivity to window choice and occasional divergence between statistical significance and economic gain. The fundamentals-based model delivers the largest forecast improvements, with bubbles providing incremental gains of 5–32%.

Robustness checks utilizing a variety of window specifications, including rolling windows and recursive windows with 120 initial observations, affirm the primary horizon-dependent hierarchy while illuminating significant nuances. The effects of stock bubbles remain consistently significant at the 6-month horizon across all specifications, particularly for the GBM model. In contrast, the effects of oil bubbles at extended horizons exhibit greater sensitivity to the chosen methodological framework. Notably, the substantial gains observed with the BRNN at the 12-month horizon under recursive estimation, which indicate a 33.3% reduction in Mean Squared Error (MSE), do not fully replicate under rolling window analysis, suggesting that some of the predictive content is derived from long-term structural persistence rather than transient dynamics. These findings underscore the critical importance of validating forecasting results across multiple window designs and caution against an excessive reliance on any singular specification

For investors, the horizon-dependent hierarchy implies monitoring oil bubbles for 3-month risk management, stock bubbles for 6-month horizons, and a combination for longer horizons. For policymakers, the predictive power of oil bubbles at the 3-month horizon underscores the importance of energy market surveillance for financial stability. As oil increasingly functions as a financial asset, understanding cross-market propagation of speculative dynamics remains critical.

Future research should extend this analysis to other economies and investigate the economic sources of estimator-specific heterogeneity.

References

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review, 39(4), 885–905.

Bai, J., and Perron, P. (2003). Computation and analysis of multiple structural change models. Journal of Applied Econometrics, 18(1), 1–22.

Black, F. (1976). Studies of stock price volatility changes. In Proceedings of the 1976 Meeting of the Business and Economic Statistics Section; American Statistical Association: Washington, DC, USA, 177–181.

Bollerslev, T., Hood, B., Huss, J., & Pedersen, L.H. (2018). Risk everywhere: Modeling and managing volatility. Review of Financial Studies, 31(7), 2729−2773.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Burden, F., & Winkler, D. (2008). Bayesian regularization of neural networks. In *Artificial Neural Networks: Methods and Protocols* (pp. 23–42). Springer.

Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics, 138(1), 291–311.

Demirer, R., Demos, G., Gupta, R., & Sornette, D. (2019). On the predictability of stock market bubbles: Evidence from LPPLS confidence multi-scale indicators. Quantitative Finance, 19(5), 843–858.

Filimonov, V., & Sornette, D. (2013). A stable and robust calibration scheme of the log-periodic power law model. Physica A: Statistical Mechanics and its Applications, 392(17), 3698–3707.

Foglia, M., Gupta, R., Caraiani, P., & Pacelli, V. (2025). Time-varying spillover of multi-scale positive and negative bubbles in stock and oil markets. Finance Research Letters, 88(C), 109179.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.

Goyal, A., Welch, I., & Zafirov, A. (2024). A Comprehensive 2022 Look at the Empirical Performance of Equity Premium Prediction, Review of Financial Studies 37(11), 3490–3557.

Gupta, R., Nel, J., Nielsen, J., & Pierdzioch, C. (2025). Stock market volatility and multi-scale positive and negative bubbles. North American Journal of Economics and Finance, 75(Part A), 102300.

Gupta, R., Nel, J., & Pierdzioch, C. (2023). Investor Confidence and Forecastability of US Stock Market Realized Volatility: Evidence from Machine Learning. Journal of Behavioral Finance, 24(1), 111–122.

Jeong, K., Härdle, W. K., & Song, S. (2012). A consistent nonparametric test for causality in quantile. Econometric Theory, 28(4), 861–887.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.

Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. American Economic Review, 105(3), 1177–1216.

Ludvigson, S. C., & Ng, S. (2009). Macro factors in bond risk premia. Review of Financial Studies, 22(12), 5027–5067.

MacKay, D. J. C. (1992). Bayesian interpolation. Neural Computation, 4(3), 415–447.

Poon, S-H., & Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. Journal of Economic Literature, 41(2), 478–539.

Rapach, D. E., Wohar, M. E., & Strauss, J. (2008). Forecasting stock return volatility in the presence of structural breaks. Forecasting in the Presence of Structural Breaks and Model Uncertainty, Edited by: David E. Rapach and Mark E. Wohar, Vol. 3 of Frontiers of Economics and Globalization, Bingley, United Kingdom: Emerald, 381–416.

Salisu, A. A., Gupta, R., Cepni, O., & Caraiani, P. (2024). Oil shocks and state-level stock market volatility of the United States: a garch-midas approach. Review of Quantitative Finance and Accounting, 63(4), 1473–1510.

Salisu, A.A., Gupta, R., & Ogbonna, A.E. (2022). A moving average heterogeneous autoregressive model for forecasting the realized volatility of the US stock market: Evidence from over a century of data. International Journal of Finance and Economics, 27(1), 384−400.

Segnon, M., Gupta, R., & Wilfling, B. (2023). Forecasting stock market volatility with regime switching GARCH-MIDAS: The role of geopolitical risks. International Journal of Forecasting, 40(1), 29−43.

Sornette, D., Cauwels, P., & Smilyanov, G. (2018). Can we use volatility to diagnose financial bubbles? Lessons from 40 historical bubbles. Quantitative Finance and Economics, 2(1), 486–590.

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

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301–320.

Appendix A: Technical Description of Machine Learning Estimators

This appendix provides concise technical details of the six machine learning estimators used in the hierarchical forecasting framework. All estimators are implemented using a recursive expanding window.

A.1. Lasso (Tibshirani, 1996)

Lasso *(Least Absolute Shrinkage and Selection Operator)* performs L_1-regularized linear regression. The estimator minimizes:

$\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - x'_i \beta \right)^2 + \lambda \|\beta\|_1 \right\}$ where $\|\beta\|_1 = \Sigma_{j=1}^{p} |\beta_j|$. The L_1 egularization term, Lasso promotes model sparsity by forcing redundant coefficients to zero, thereby integrating feature selection directly into the estimation process.

A.2. Ridge (Hoerl and Kennard, 1970)

Ridge regression employs L_2-regularization:

$$\min_{\beta} \{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - x'_i\beta)^2 + \lambda \|\beta\|_2^2 \}$$

where $\|\beta\|^2_2 = \Sigma_{j=1}^{p} \beta^2_j$. Unlike Lasso, Ridge shrinks coefficients toward zero but does not set them exactly to zero, which is optimal when many predictors have non-zero but small effects.

A.3. Elastic Net (Zou and Hastie, 2005)

Elastic Net combines L_1 and L_2 penalties:

$$\min_{\beta} \{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - x'_i\beta)^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \lambda \|\beta\|_2^2 \}$$

This hybrid approach inherits the sparsity of Lasso while maintaining the grouping effect of Ridge, making it effective when predictors are correlated.

A.4. Random Forest (Breiman, 2001)

As an ensemble learning technique, Random Forest builds an extensive array of independent decision trees to improve predictive stability. For each tree:

- A bootstrap sample of the data is drawn.
- At each node, a random subset of predictors is selected for splitting.
- Trees are grown to full depth without pruning.

The final forecast is the average predictions across all trees. This approach reduces variance relative to a single tree and captures complex nonlinear interactions.

A.5. Gradient Boosting Machine (GBM) (Friedman, 2001)

GBM builds trees sequentially, where each new tree attempts to correct the errors of the previous ensemble. Formally, starting with an initial model $F_0(x)$, at each iteration $m = 1, \ldots M$:

$$F_m(x) = F_{m-1}(x) + \nu \cdot h_m(x)$$

where $h_m(x)$ is a weak learner (typically a shallow tree) fitted to the pseudo-residuals, and ν is a shrinkage parameter (learning rate). This stage-wise approach often achieves state-of-the-art predictive performance.

A.6. Bayesian Regularized Neural Network (BRNN) (MacKay, 1992; Burden and Winkler, 2008)

A BRNN is a feedforward neural network trained using Bayesian regularization. Consider a single hidden layer network:

$$f(x) = \sum_{j=1}^{H} w_j^{(2)} \cdot \varphi(\sum_{i=1}^{p} w_{ji}^{(1)} x_i + b_j^{(1)}) + b^{(2)}$$

where $\varphi(\cdot)$ is a sigmoidal activation function, H is the number of hidden units, and w denotes all weights and biases. Bayesian regularization modifies the standard sum-of-squares error by adding a penalty term proportional to the sum of squared weights:

$$F = \beta \cdot E_D + \alpha \cdot E_W$$

where E^D is the data error, $E^W = \|w\|^2$ is the weight penalty, and α, β are hyperparameters optimized using Bayesian inference (MacKay, 1992). This approach automatically penalizes overly complex networks, promoting parsimony and improving generalization in high-dimensional settings (Burden and Winkler, 2008).