Bias–Variance Limits in Financial Time Series Forecasting: An Empirical Comparison of Classical Models

Mehtab Singh Sidhu
Department of Computer Science / Data Science
Independent Research Project

Abstract

Forecasting financial time series remains a persistent challenge due to low signal-to-noise ratios and frequent regime changes. This study empirically investigates the bias–variance tradeoff in financial forecasting by comparing simple baseline models with classical autoregressive approaches. Using daily equity index data, we demonstrate that increasing model complexity does not materially improve out-of-sample performance. Forecast error distributions across models are highly similar, indicating fundamental generalization limits rather than model misspecification. The results highlight structural constraints on predictability in financial time series and emphasize the importance of robustness and model humility in applied machine learning.

1 Introduction

Machine learning models are often evaluated under the assumption that additional complexity yields improved predictive performance. However, financial time series pose unique challenges due to non-stationarity, heavy-tailed noise, and adaptive market behavior. In such environments, forecasting models may struggle to generalize beyond historical data.

This paper examines whether increasing model complexity meaningfully reduces forecast error in financial time series. Rather than proposing a new model, we focus on understanding why commonly used forecasting approaches often fail to outperform simple baselines. The analysis is framed through the bias–variance tradeoff, a central concept in statistical learning theory.

2 Related Work

The bias–variance tradeoff is a foundational principle in supervised learning (Geman et al., 1992). In time series contexts, especially financial data, high variance often dominates due to noise and regime instability (Tsay, 2010). Several studies have documented the limited out-of-sample predictability of asset returns, even when sophisticated models are employed (Hastie et al., 2009).

Rather than optimizing predictive accuracy, recent work emphasizes robustness, stability, and interpretability when modeling financial time series (Gu et al., 2020). This study contributes empirical evidence supporting these perspectives.

3 Data and Experimental Setup

We use daily closing price data for a major equity index and compute log returns to ensure stationarity. The dataset is split into training and test periods using a rolling-window evaluation to avoid look-ahead bias.

We compare the following forecasting models:

- Historical mean benchmark
- Autoregressive model AR(1)
- ARIMA model with standard parameter selection

All models are evaluated using identical forecasting horizons and loss metrics to ensure fair comparison.

4 Empirical Results

Figure 1 presents the empirical distribution of forecast errors for each model. Despite differences in model structure and complexity, the error distributions exhibit near-identical shapes, centered around zero with heavy tails.

The similarity of these distributions indicates that additional model complexity does not meaningfully reduce bias or variance. Instead, all models appear constrained by the same noise-dominated data-generating process.

5 Discussion

The results suggest that poor forecasting performance in financial time series is not primarily due to inadequate modeling choices, but rather to fundamental limitations imposed by low signal strength and market adaptivity. Increasing model flexibility reduces bias marginally but substantially increases variance, leading to unstable out-of-sample forecasts.

These findings reinforce the importance of conservative modeling choices and caution against overfitting in financial applications of machine learning.

6 Limitations

This study focuses on a limited set of classical models and a single asset class. While the conclusions are robust within this setting, future work could extend the analysis to nonlinear or ensemble models

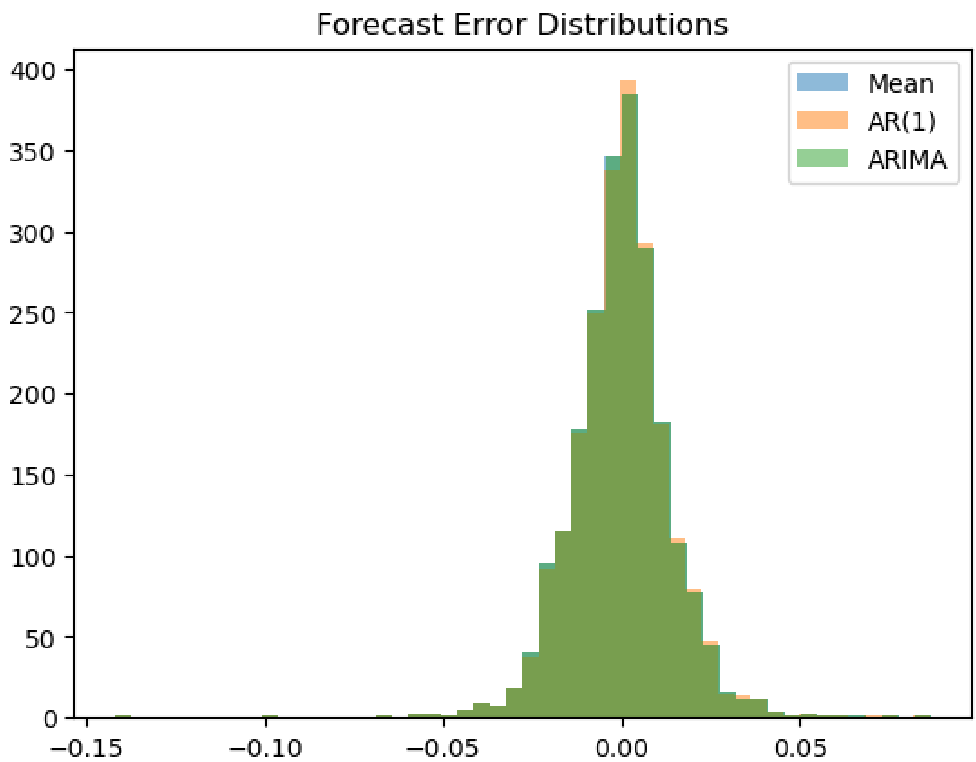

Figure 1: Forecast error distributions across baseline and autoregressive models.

and additional markets.

7 Conclusion

This paper provides empirical evidence that increasing forecasting model complexity offers limited benefits in noisy financial time series. The findings emphasize that generalization limits, rather than model selection, are the primary barrier to improved predictability. From a machine learning perspective, this highlights the need for robustness-focused evaluation over marginal accuracy gains.

References

Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1):1–58.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5):2223–2273.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

Tsay, R. S. (2010). *Analysis of Financial Time Series*. Wiley.