Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Regime Dependent Machine Learning and the Limits of Algorithmic Trading in the Dhaka Stock Exchange

Md. Shakil Sikder

shakil.stat.gstu@gmail.com

Research Article

Keywords:

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11137245/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Regime Dependent Machine Learning and the Limits of Algorithmic Trading in the Dhaka Stock Exchange

Md. Shakil Sikder

Department of Statistics, Gopalgonj Science and Technology University, Gopalgonj, Bangladesh

Sonali Bank PLC, Dhaka, Bangladesh

Corresponding author. E-mail: shakil.stat.bsmrstu@gmail.com

ORCID: 0000-0002-9280-4043

Abstract

This study tests whether Hidden Markov Model (HMM) regime-conditioning improves short-horizon return forecasting and trading profitability for equities listed on the Dhaka Stock Exchange (DSE), using three independently constructed stock panels spanning 42 securities across 17 sectors: a deliberately selected twelve-stock Fixed panel, an independently sampled six-stock Random panel, and a 24-stock Scope-Expansion panel added specifically to test these findings at a substantially broader sectoral and cross-sectional scale. Five forecasting models are evaluated under leakage-safe walk-forward validation: a naive random-walk benchmark, AR(1)/Ridge, Linear SVR, LightGBM, and an LSTM, in pooled and regime-as-feature specifications. Diebold-Mariano tests find no model significantly outperforms the naive benchmark on any panel, and find LightGBM significantly worse than naive on all three panels independently; a nested ablation study and SHAP analysis corroborate this. A regime-conditioned trading strategy, backtested under an itemized DSE cost structure and a liquidity constraint, loses the majority of capital on all three panels; a Deflated Sharpe Ratio, an out-of-sample validation split, and a stationary block bootstrap confirm no strategy demonstrates statistically defensible skill. The Hidden Markov Model also systematically misclassifies extended floor-price trading freezes as low-risk rather than as market stress. These results, replicated across three panels spanning 17 sectors and an alternative non-HMM regime detector, are consistent with weak-form market efficiency and caution against assuming regime-aware machine learning transfers cleanly to a frontier market with an active price-floor mechanism.

Keywords: Regime detection; Hidden Markov Models; Machine learning in finance; Frontier markets; Dhaka Stock Exchange; Algorithmic trading

JEL Classification: C22; C53; C58; G11; G14; G17

1 Introduction

Whether financial markets can be predicted, and whether any such predictability can be converted into trading profit, are logically distinct questions that applied machine-learning-for-finance research frequently conflates. A model can be statistically superior to a naive benchmark while failing to produce economically meaningful, cost-adjusted returns, and the reverse

is equally possible. This distinction is the organizing question of this study, examined for the Dhaka Stock Exchange (DSE) – a frontier market characterized by a regulatory floor-price mechanism, comparatively thin liquidity, historically documented extreme-volatility episodes, and retail transaction costs that are non-trivial relative to typical daily price movements.

This study extends an earlier comparative study of Support Vector Regression (SVR) and K-Nearest Neighbors (KNN) regression for DSE price-level prediction (Islam, 2021) along three dimensions: continuous return forecasting rather than price levels, formal significance testing against a naive benchmark, and a full cost-and-liquidity-realistic trading evaluation. This study further contributes, distinctively, by independently replicating its central findings on three non-overlapping stock panels spanning 42 securities and 17 sectors – a deliberately selected panel, a randomly sampled panel, and a substantially larger Scope-Expansion panel purpose-built to test these findings' robustness at a broader empirical scale – and by identifying and directly testing a specific failure mode: standard volatility-based regime detection misclassifies the DSE's floor-price-driven trading freezes as low-risk periods, since a mechanically halted price series shows zero realized volatility regardless of the underlying severity of the situation.

The remainder of this paper proceeds as follows. Section 2 reviews the relevant literature and develops the theoretical framework. Section 3 details the data, panel construction, and methodology. Section 4 presents the forecasting, trading, and risk-adjusted results. Section 5 discusses the findings, including the floor-price analysis, and states limitations. Section 6 concludes.

2 Literature Review and Theoretical Framework

2.1 Market efficiency and the random walk

The Efficient Market Hypothesis (Fama, 1970) holds that security prices fully reflect available information, such that abnormal risk-adjusted returns cannot be systematically earned by trading on it; its weak form, restricting the information set to historical prices and volumes, is the specific claim this study's forecasting exercise tests directly. Samuelson (1965) provides the foundational proof that properly anticipated prices fluctuate randomly, formalized as the Random Walk Hypothesis: the expected log return conditional on all past information is zero,

$$\mathbb{E}\big[\, r(t+1) \mid \mathcal{I}(t) \,\big] = 0 \tag{1}$$

the exact statistical claim embodied by this study's Naive forecasting benchmark. Lo and MacKinlay (1988) present early formal evidence of short-horizon positive autocorrelation in U.S. equity indices, a pattern this study's own exploratory analysis also documents for the DSE panel; Jegadeesh and Titman (1993) and a large subsequent literature nonetheless find that detecting statistical predictability is markedly easier than converting it into an economically exploitable strategy once costs are considered.

2.2 Machine learning for return prediction

Deep sequential and tree-based models are frequently reported as outperforming classical regressors on specific datasets, but this finding is not universal: Krauss et al. (2017) find Random Forest can outperform deep neural networks in some settings. Gu et al. (2020) find tree-based and neural-network methods generate economically meaningful out-of-sample gains over linear models for U.S. equities, but subsequent extensions to international and emerging markets

(Cakici et al., 2023; Hanauer and Kalsbach, 2023) find predictability concentrated in less liquid, harder-to-arbitrage securities and diminishing substantially once trading frictions are applied – a pattern this study's own results reproduce for a frontier market.

2.3 Regime detection

Hamilton (1989) provides the foundational regime-switching model for inferring a hidden Markov state from observed time-series data. A Hidden Markov Model characterizes the market's regime as an unobserved discrete state $S(t)$ that evolves according to the Markov property, with each state i generating observed returns from its own Gaussian emission distribution:

$$P\big(X(t) \mid S(t) = i\big) = \mathcal{N}\big(X(t);\, \mu_i,\, \sigma_i^2\big) \tag{2}$$

Given a fitted model, the filtered (point-in-time) probability of state j at time t is computed recursively via the forward algorithm, using only observations up to and including time t:

$$\alpha(t, j) = \left[\sum_i \alpha(t-1, i)\, A(i, j)\right] P\big(X(t) \mid S(t) = j\big) \tag{3}$$

Kim et al. (1998) emphasize a methodologically important distinction between this filtered probability and a smoothed (full-sample) alternative: smoothed inference incorporates information unavailable to a genuinely live decision-maker and can substantially overstate a regime-switching model's live-trading value. This study measures, on real DSE data, that filtered and smoothed labels for the same calendar day agree only 45.2% of the time, quantifying this concern directly. Only the filtered probability of Equation (3) is used for any forecast or trading decision in this study (Section 3.3).

2.4 Statistical testing and multiple-testing correction

The Diebold-Mariano test (Diebold and Mariano, 1995) formally tests whether two forecasts differ significantly in accuracy, accounting for serial correlation via a Newey-West heteroskedasticity-and-autocorrelation-consistent estimator. When multiple strategy variants are evaluated, the expected maximum observed Sharpe ratio across all variants exceeds zero purely by chance, even absent genuine skill; the Deflated Sharpe Ratio (Bailey and Lopez de Prado, 2014) corrects for this given the number of trials and their empirical variance, addressing a specific driver of the well-documented gap between backtested and live trading performance (Bailey et al., 2014). The stationary block bootstrap (Politis and Romano, 1994) provides an independent, non-parametric check by resampling contiguous blocks of the observed return series.

2.5 Prior work on the Dhaka Stock Exchange and research gap

The direct precedent for this study (Islam, 2021) compared SVR and KNN for next-day price-level prediction on three DSE companies without testing significance against a naive benchmark or evaluating trading profitability. This study's specific contribution – continuous return forecasting, regime-conditioning tested for significance across five models, a leakage-safe point-in-time regime protocol whose practical consequences are directly measured, an itemized DSE-specific cost-and-liquidity model, four independent statistical validation methods, and replication across three independently constructed panels spanning 42 securities and 17 sectors – has not, to the author's knowledge, been jointly combined in prior DSE-specific work.

3 Data and Methodology

3.1 Data and panel construction

Daily OHLCV data (2018–2025) were obtained from a corporate-action-adjusted DSE dataset, after an earlier 534-ticker candidate source was rejected for containing no adjusted-close field, thousands of duplicate rows, and a cluster of implausible market-wide price swings in 2011. Integrity checks on the adopted dataset found and corrected three OHLC violations and one market-wide zero-volume gap, and confirmed no duplicate records. The Fixed panel comprises twelve DSE companies across six sectors (Banking: EBL, NBL, DHAKABANK; Pharmaceuticals: RENATA, IBNSINA, SQURPHARMA; Telecommunications: GP, ROBI; Engineering/Steel: BSRMSTEEL, BSRMLTD; Energy: TITASGAS; Food: OLYMPIC), selected by a liquidity floor, continuous listing, and a data-cleanliness screen. The Random panel comprises six companies (ONEBANKPLC, RAKCERAMIC, IFIC, SHYAMPSUG, ILFSL, APEXTANRY) drawn by uniform random sampling (seed 42) from the broader universe of DSE securities with full 2018–2025 coverage, independently verified for the same integrity issues, used to test whether findings generalize beyond a hand-selected sample. A third, Scope-Expansion panel of 24 additional securities was subsequently added, selected to deliberately broaden sectoral coverage into eleven sectors not represented in the original eighteen – Cement, Insurance, Power/Utilities, Financial Services, Consumer Goods, Conglomerate/Tobacco, and Information Technology, plus additional depth in Banking, Pharmaceuticals, and Ceramics – bringing the combined panel to 42 securities across 17 sectors. Candidates were drawn from the 302 DSE common-equity tickers with full 2018–2025 coverage after excluding the original eighteen, liquidity-screened by average daily traded value, and verified for the same integrity issues (zero duplicate rows and zero OHLC violations across all 24).

3.2 Feature engineering

Let $P(t)$ denote the adjusted closing price on day t. The one-day log return and the two forecast targets are:

$$r(t) = \ln P(t) - \ln P(t-1), \qquad \text{target}^{(h)}(t) = \ln P(t+h) - \ln P(t), \quad h \in \{1, 5\} \tag{4}$$

Thirteen features were constructed per stock: five lagged returns (lags 1, 2, 3, 5, 10); 10- and 20-day momentum, $p(t) - p(t-n)$; 20- and 50-day price-to-moving-average ratios; a 20-day rolling volume z-score; 20- and 60-day rolling realized volatility; and the 14-day Relative Strength Index, computed from trailing average gains and losses:

$$RSI(t) = 100 - \frac{100}{1 + \text{Gain}(t)/\text{Loss}(t)} \tag{5}$$

RSI is bounded to $[0, 100]$ by construction, unlike the panel's other, unbounded technical features, and is undefined whenever $\text{Gain}(t)$ and $\text{Loss}(t)$ are simultaneously zero – traced, in this dataset, to genuine multi-day flat-price stretches and forward-filled accordingly.

3.3 Regime detection

A three-state Gaussian HMM was fit on DSEX daily returns alone, using a trailing 500-day window refit every 60 days, with states aligned across refits by emission variance. Only the forward-filtered (point-in-time) probability was used for any live label; a full-sample smoothed

decoding was computed solely for diagnostic comparison. As a robustness check, a rule-based, non-HMM regime detector (point-in-time expanding-window volatility percentile) was also constructed and substituted for the HMM in the central regime-conditioning test.

3.4 Forecasting models

Five models were evaluated in pooled and regime-as-feature modes, at one- and five-day horizons. The Naive model predicts a zero return for every observation. The AR(1)/Ridge model is a ridge-regularized linear regression on the lagged-return features:

$$\min_{\beta} \sum_{t} \big(y(t) - x(t)^{\top}\beta\big)^2 + \lambda\|\beta\|^2, \qquad \lambda = 1.0 \tag{6}$$

Linear SVR fits a linear function within an ε-insensitive tube on the full feature set:

$$\min_{w} \ \frac{1}{2}\|w\|^2 + C\sum_{t} \max\big(0,\ |y(t) - f(x(t))| - \varepsilon\big), \qquad C = 1.0,\ \varepsilon = 0.001 \tag{7}$$

LightGBM is a gradient-boosted ensemble of 150 shallow regression trees (max depth 4, learning rate 0.05) built additively on the negative gradient of the squared-error loss. The LSTM processes a 10-day trailing return sequence through a 16-unit recurrent layer (with regime appended as a per-timestep channel in the regime-as-feature mode), followed by an 8-unit dense layer and a linear output. Walk-forward validation used a 500-day training window, 150-day test blocks, and an embargo equal to the forecast horizon between train and test, preventing horizon-induced label leakage.

Forecast comparisons used the Diebold-Mariano statistic on the panel-pooled loss-differential series $d(t)$, with a Newey-West HAC variance estimator and Holm-corrected p-values across pre-specified comparison families:

$$DM = \frac{\bar{d}}{\sqrt{\widehat{\mathrm{Var}}(\bar{d})}}, \qquad \bar{d} = \frac{1}{T}\sum_{t=1}^{T} d(t) \tag{8}$$

A nested ablation study evaluated LightGBM on four feature sets (technical only; plus volume; plus volatility; plus regime), and a SHAP analysis using a TreeExplainer ranked feature importance.

3.5 Trading-signal construction, backtesting, and risk-adjusted evaluation

Regime-conditioned trading rules (long if predicted return exceeds a regime-specific threshold; flat in Stress) were backtested under an itemized DSE cost stack – 0.35% brokerage per side, 15% VAT on brokerage, a 0.0125% CDBL charge, and a square-root volume-dependent slippage model capped at 0.5%:

$$\mathrm{slippage}(t) = \min\left(k\sqrt{\frac{\mathrm{TradeValue}(t)}{\mathrm{DayValue}(t)}},\ 0.005\right), \qquad k = 0.10 \tag{9}$$

together with a liquidity constraint capping daily position changes at 10% of a stock's own traded value, given an illustrative BDT 5,000,000 capital base. Risk-adjusted performance

used the annualized Sharpe ratio,

$$\text{Sharpe} = \frac{\text{AnnReturn} - R_f}{\text{AnnVol}}, \qquad R_f = 7\% \text{ annually} \tag{10}$$

A hysteresis-based turnover-reduction experiment tested whether excessive turnover, rather than a weak underlying signal, explained poor net performance. The Deflated Sharpe Ratio corrected for the number of hysteresis configurations tested by comparing the observed best Sharpe ratio against its chance-expected maximum given N trials; an out-of-sample selection (2018–2022) / holdout (2023–2025) split, and a stationary block bootstrap (2,000 resamples, 20-day blocks), provided two further, independent risk-adjusted validations, repeated on both panels for the bootstrap step.

3.6 Use of AI-assisted tools

The empirical analysis reported in this study – including data cleaning, feature engineering, model estimation, statistical testing, backtesting, and risk-adjusted evaluation – was executed using an AI-based data analysis and coding assistant (Claude, Anthropic), under the author's direction and review. Portions of the manuscript text were similarly drafted with the assistance of this tool. The author reviewed, verified, and takes full responsibility for the accuracy of all methodology, code, results, and conclusions reported.

4 Results

4.1 Regime detection

Within 2018–2025, 76.6% of days were classified Normal, 22.7% High-Volatility, and 0.8% Stress; Stress days concentrate at the sharp pre-freeze decline and post-reopening rebound of the COVID-19 episode, not during the intervening multi-month trading freeze (Figure 1). Filtered and smoothed labels for the same calendar day agree only 45.2% of the time (Figure 2). The rule-based alternative regime detector produces a Diebold-Mariano statistic of -0.68 ($p = 0.496$) for LightGBM's regime-as-feature-versus-pooled comparison, consistent in direction and magnitude with the HMM-based result below. The regime label is derived exclusively from the market-level DSEX index and is therefore identical across all three panels.

4.2 Forecasting and statistical significance

Table 1: Diebold-Mariano tests, 1-day horizon, all three panels

	Fixed (12 stocks)		Random (6 stocks)		Scope-Exp. (42 stocks)	
Comparison	DM	p	DM	p	DM	p
LightGBM vs. Naive	3.97	0.001	4.13	0.000	3.55	0.007
SVR vs. Naive	5.35	0.000	1.76	0.160	4.70	0.000
AR(1)/Ridge vs. Naive	1.25	1.000	0.67	0.503	-0.07	1.000
Regime-as-feature vs. pooled (all models)	all $p > 0.10$		all $p > 0.10$		all $p \geq 0.094$	

No model significantly outperforms Naive on any of the three panels. LightGBM is significantly worse than Naive on all three panels independently – the same qualitative finding recovered from three non-overlapping sets of securities, including the 42-stock, 17-sector Scope-Expansion panel added specifically to test this result's robustness at a substantially broader

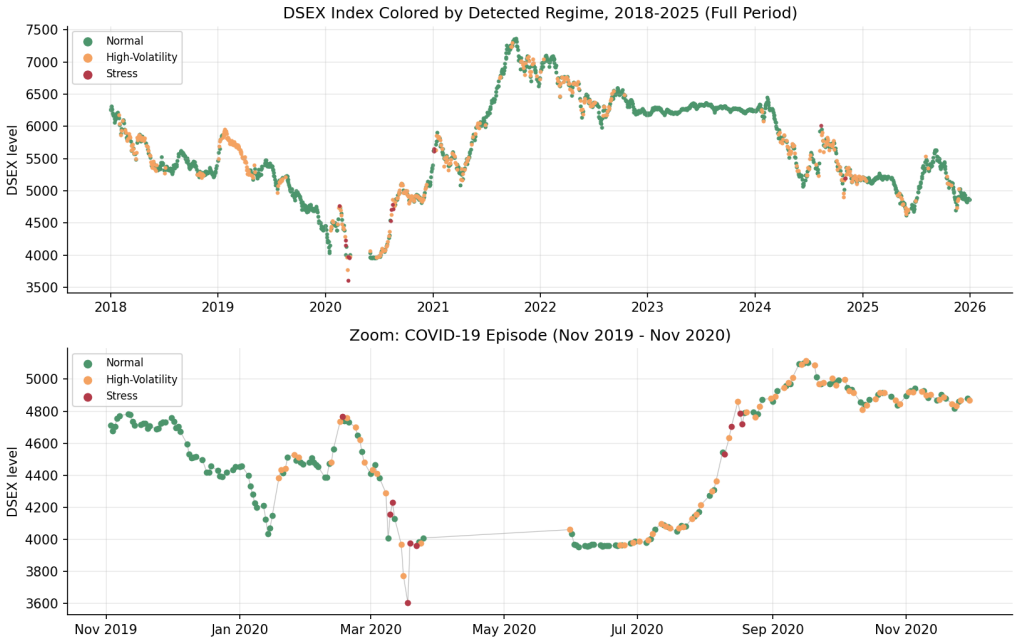


Figure 1: DSEX index colored by detected regime, full period (top) and zoomed on the COVID-19 episode (bottom)

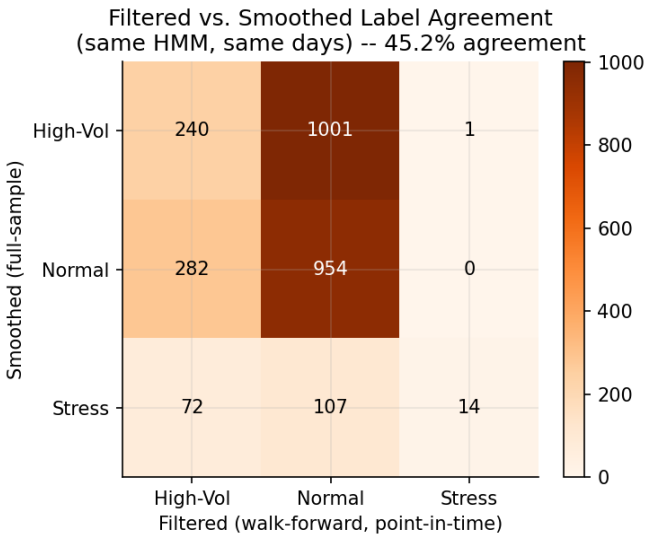


Figure 2: Confusion matrix: smoothed (full-sample) label versus filtered (point-in-time) label, same HMM, same calendar days

scale. SVR is significantly worse than Naive on the Fixed and Scope-Expansion panels but falls short of significance on the smaller Random panel, consistent with a power explanation given the Random panel's smaller cross-section: significance is restored once the cross-section is enlarged. The regime-as-feature-versus-pooled comparison, the paper's central test, is non-significant after Holm correction for every model at both horizons on all three panels, including for the LSTM model added to the Scope-Expansion panel's comparison family. A nested ablation study shows RMSE worsening monotonically as volume, volatility, and regime are added (0.015951 to 0.016006), and a SHAP analysis ranks regime 11th of 14 features by importance – two independent corroborations of the same conclusion.

4.3 Trading strategy and backtesting

Table 2: Liquidity-aware, itemized-cost backtest results, all three panels

	Fixed (12 stocks)		Random (6 stocks)		Scope-Exp. (42 stocks)	
Strategy	CAGR	Sharpe	CAGR	Sharpe	CAGR	Sharpe
LightGBM Regime-Cond.	-71.0%	-8.49	-53.9%	-3.34	-78.5%	-8.52
SVR Regime-Cond.	-63.2%	-7.44	-45.5%	-3.85	-74.8%	-9.39
AR(1)/Ridge Regime-Cond.	-67.8%	-7.81	-51.8%	-4.18	-77.9%	-9.04
MA-Crossover (10/50)	-31.1%	-2.13	-27.7%	-1.20	-35.6%	-2.17
Buy-and-Hold	-5.8%	-0.87	-15.2%	-0.87	-6.0%	-0.79

Every machine-learning regime-conditioned strategy loses the substantial majority of capital on all three panels, driven by 15–65% average daily turnover; a hysteresis-based turnover-reduction experiment cuts turnover roughly thirty-fold but leaves gross return near zero on every panel, including the Scope-Expansion panel (best configuration Sharpe -0.22, still short of Buy-and-Hold), indicating the underlying signal itself lacks exploitable value regardless of how many securities or sectors are examined (Figure 3).

4.4 Risk-adjusted evaluation

A Deflated Sharpe Ratio of 0.0000 (best of ten hysteresis trials) and bootstrap 95% confidence intervals confirm the ML strategies and moving-average crossover are statistically significantly bad (0.0% of resamples positive for any of the three) on all three panels; buy-and-hold's interval spans zero on the Fixed panel and, closely replicating that pattern, on the Scope-Expansion panel too (statistically indistinguishable from a zero-skill outcome on both), while it falls just short of doing so on the smaller Random panel, a small-sample nuance reported transparently rather than smoothed over. The Deflated Sharpe Ratio on the Scope-Expansion panel, computed identically over its own ten-configuration hysteresis sweep, is likewise 0.0000. An out-of-sample selection/holdout split shows the two most promising full-sample configurations fail to generalize for distinct reasons: one placed zero trades in the selection period, with its entire apparent profitability concentrated in the holdout period; the other reverses from mildly positive to clearly negative across the split.

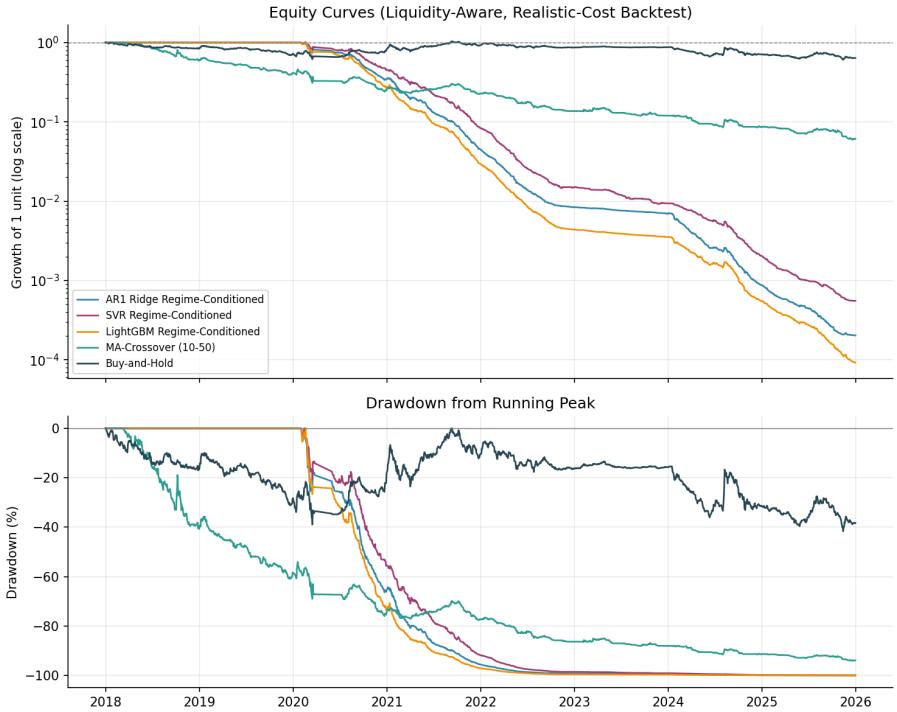


Figure 3: Equity curves under the liquidity-aware, itemized-cost backtest, Fixed panel

5 Discussion

5.1 Interpretation

Across five models, three independently constructed panels spanning 42 securities and 17 sectors, and multiple robustness checks, the evidence converges: regime-conditioning does not significantly improve forecasting accuracy, and no trading strategy examined converts any panel's return signal into a statistically defensible economic edge over passive exposure. This is consistent with weak-form market efficiency holding, for practical purposes, across this broad cross-section of DSE-listed securities and this study period.

5.2 The DSE floor-price mechanism

A verified analysis of the DSE's floor-price mechanism – prompted by, and correcting, an initially inaccurate external claim about its specifics – finds one panel stock's longest flat-price episode spans 374 trading days, and that a volume-aware definition (flat price and volume below 10% of trailing median) reduces this to a genuinely low-volume spell of only four days, indicating flat price alone is an incomplete proxy for a frozen market. A directly backtested floor-day-exclusion rule performs worse than plain buy-and-hold on every dimension (-10.08% versus -4.49% annualized return), since the exclusion itself introduces rebalancing costs exceeding any benefit. The regime detector's Stress-conditioned flat rule does provide a narrow, genuine protective benefit during the fifteen days it identifies as acute stress, but this is vastly outweighed by the strategy's overall underperformance.

5.3 Limitations

This study is scoped to 42 DSE securities across 17 sectors over 2018–2025 – a substantially broadened but still deliberate scope choice, not a claim of full-market representativeness; a small number of sectors (e.g., jute, paper/printing) remain unrepresented even in this expanded panel. Cost, liquidity, and capital-base parameters are explicit, illustrative assumptions rather than directly observed execution data, held constant across all three panels. With only fifteen HMM-identified Stress days across the full window – a market-level regime label identical across all three panels, so unaffected by the cross-sectional expansion – conclusions specific to that state rest on a small sample. Conclusions rest on the Deflated Sharpe Ratio and bootstrap-corrected results rather than any single best-looking configuration among those examined. Finally, the Fixed panel's twelve tickers use a custom back-adjustment for corporate actions, while the Random and Scope-Expansion panels' 30 tickers use the vendor-supplied adjusted series directly; this is not expected to materially affect the return-based analysis but is noted for completeness.

6 Conclusion

This study finds no statistically significant evidence that HMM-based regime-conditioning improves short-horizon return forecasting for equities listed on the Dhaka Stock Exchange, across five forecasting models, multiple robustness checks, and three independently constructed panels spanning 42 securities and 17 sectors – a twelve-stock Fixed panel, a six-stock Random panel, and a 24-stock Scope-Expansion panel added specifically to test these findings at a substantially broader empirical scale. No trading strategy examined converts any panel's return signal into a statistically significant, cost-and-liquidity-realistic economic edge over passive buy-and-hold exposure, a conclusion independently replicated across all three panels and corroborated by four separate statistical validation methods. A specific, previously unquantified interaction between volatility-based regime detection and the DSE's floor-price mechanism is identified, and a plausible remedy for it is directly tested and found not to work. These findings are reported as a rigorously validated, three-times-replicated null result across a broad cross-section of the DSE, with direct relevance to frontier-market practitioners and to the broader machine-learning-for-finance literature.

Statements and Declarations

Funding: No funding was received to assist with the preparation of this manuscript.
Competing interests: The author has no relevant financial or non-financial interests to disclose.
Data, materials, and code availability: The raw secondary data underlying this study are publicly available in the Mendeley Data repository, "Dhaka Stock Exchange End-of-Day Financial Dataset with Coverage Metadata," https://data.mendeley.com/datasets/23553sm4tn/4 (DOI: 10.17632/23553sm4tn.4). The cleaned panel datasets, the full analysis pipeline, and key result tables that support the findings of this study are available from the corresponding author upon reasonable request.
Author contributions: Md. Shakil Sikder is the sole author and is responsible for all aspects of this work: conceptualization, methodology, software, formal analysis, investigation, data curation, writing, and visualization.

Clinical Trial Number: Not applicable. This study did not involve a clinical trial.
Ethics declaration: Not applicable. This study used only publicly available, anonymized, exchange-level historical stock-price data for companies listed on the Dhaka Stock Exchange; it did not involve human participants, human data, or animals, and therefore did not require ethics committee approval.
Consent to Participate declaration: Not applicable.
Consent to Publish declaration: Not applicable.

References

Bailey, D. H., Borwein, J., Lopez de Prado, M., and Zhu, Q. J. (2014). Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance. *Notices of the American Mathematical Society*, 61(5):458–471.

Bailey, D. H. and Lopez de Prado, M. (2014). The deflated sharpe ratio: Correcting for selection bias, backtest overfitting, and non-normality. *Journal of Portfolio Management*, 40(5):94–107.

Cakici, N., Fieberg, C., Metko, D., and Zaremba, A. (2023). Machine learning goes global: Cross-sectional return predictability in international stock markets. *Journal of Economic Dynamics and Control*.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383–417.

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

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Hanauer, M. X. and Kalsbach, T. (2023). Machine learning and the cross-section of emerging market stock returns. *Journal of Asset Management*, 24(5):419.

Islam, e. a. (2021). Predicting the daily closing price of selected shares on the dhaka stock exchange using machine learning techniques. *SN Business and Economics*.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.

Kim, C.-J., Nelson, C. R., and Startz, R. (1998). Testing for mean reversion in heteroskedastic data based on gibbs-sampling-augmented randomization. *Journal of Empirical Finance*, 5(2):131–154.

Krauss, C., Do, X. A., and Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the s&p 500. *European Journal of Operational Research*, 259(2):689–702.

Lo, A. W. and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, 1(1):41–66.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6(2):41–49.