JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Development of an AI-Based Fair Value Model for Indian Index Option Premiums Using Market Microstructure Variables

*** Eeda Damodara Rao, Research Analyst**

**** Sasikumar Peyyala, Research Analyst, TradeMilaan,**

Abstract

India has one of the biggest index options markets anywhere, measured by the number of contracts traded. Most of that activity sits in weekly and monthly contracts on Nifty 50, Bank Nifty, Fin Nifty and Sensex. Yet the premiums you see on screen often don't match what classical models like Black–Scholes–Merton would suggest. The gap isn't random noise. It follows patterns tied to bid–ask spreads, imbalances in the order book, trade flow, the quirks of very short-dated expiries, and the heavy presence of retail traders.

In this paper, we build an AI-based fair value model for Indian index options. It combines standard no-arbitrage pricing inputs with market microstructure data to estimate a premium that is justified both by theory and by the data. The design is a hybrid. We start with a parametric baseline (Black-76 using a smoothed implied-volatility surface) and add two learners on top: a gradient-boosted tree model and a sequence-based neural network. Both predict the "premium gap", meaning the difference between the observed mid-price and a microstructure-adjusted fair value.

The inputs include spread, depth, order-flow imbalance, signed volume, quote-update intensity, realised and implied volatility, open-interest changes, moneyness, time to expiry and event-calendar effects. We test the model with a walk-forward setup that guards against data leakage, and we judge it on two fronts: statistical measures (MAE, RMSE, MAPE) and economic ones (spread-adjusted edge and hedged P&L).

In the illustrative results presented here, the hybrid model cuts pricing error by roughly 25 to 33 percent compared with the Black-76 baseline. The biggest improvements come in near-the-money, short-dated contracts. SHAP analysis points to spread, order-flow imbalance and time to expiry as the main drivers. We also discuss what this means for retail investors, brokers, market makers and regulators, and we offer recommendations for deploying such a model responsibly.

Keywords

Fair value; index options; Nifty 50; Bank Nifty; market microstructure; order flow imbalance; implied volatility surface; machine learning; gradient boosting; explainable AI; NSE; option pricing

I. Introduction

1.1 Background and motivation

India has experienced a huge revolution in its derivatives space in the last decade. The NSE and the BSE currently have the lion's share of the world's total number of listed options contracts in the form of index options. Due to the ease of trading offered by short-dated weekly maturities, low ticket sizes, discount broking and mobile apps, index options have seen a huge increase in retail participation. They are an attractive product for traders as they allow for leverage, defined risk and a broad

range of strike prices however they are one of the most difficult products to accurately value. The premium is a combination of several components, including intrinsic value, time value, volatility premium, liquidity premium and sometimes a crowd premium.

Several regulatory studies conducted by SEBI show that a vast majority of retail traders suffer an overall loss in the equity derivatives segment. This can be attributed to many factors such as leverage, behavioural bias and costs. However, very little has been studied about how well price discovery is occurring in these markets. If a buyer is paying a premium that is significantly higher than a justifiable fair value then they will start from an unfair disadvantage. Having access to a clear valuation methodology would allow participants to see if they are paying too much, too little or a fair amount.

1.2 Problem statement

The classical model of Black and Scholes (1973) and Merton (1973) and its future version (Black, 1976), assumes perfect frictionless markets with continuous trading, constant/deterministic volatility and zero price impact. This is violated every day in Indian index options. We see large bid ask spreads for deep out-of-the-money strikes and events, quotes are very volatile at sub-second intervals, low liquidity with concentrated strikes and huge gamma and theta effects within the last hour of weekly contracts. These models do not account for why two contracts with similar theoretical prices will have such disparity in their premiums.

Option pricing models using machine learning have been employed since the early 1990s. However, most research focuses on using end-of-day data, U.S. markets and viewing this as a function approximation problem. This research does not aim to predict a specific definition of a fair value based on microstructure (i), focus on the Indian market with its unique expiry structure and participants of the market (ii) or provide any form of explainability to allow the trader or regulator to audit the model output.

1.3 Research objectives

This paper pursues the following objectives:

1. Provide an objective definition of fair-value for Indian index option premiums by distinguishing between arbitrage-free and liquidity/order flow components.
2. Create a set of features comprising market microstructure variables that can be derived from exchange level quote, trade and open interest data.
3. Design and compare different types of AI models (Regularised Linear, Gradient Boosted Trees, Neural Networks and Hybrid Residual) to estimate fair-value and premium gap.
4. Evaluate the models using leakage free walk forward validation approach. Statistical and economic measures will be used.
5. Explain model performance using feature attribution techniques and provide actionable insights.

1.4 Research questions

- **RQ1:** Do microstructure variables explain a significant portion of the deviation between observed option premiums and classical model values?
- **RQ2:** Does a hybrid AI model outperform a parametric baseline and a purely data-driven model out of sample?
- **RQ3:** Which variables matter most, and does their importance vary by moneyness, expiry and market regime?
- **RQ4:** Is the estimated fair value economically meaningful after transaction costs?

1.5 Contributions and structure

The paper contributes a formal fair value definition tailored to Indian index options, a reproducible feature engineering framework, a hybrid modelling architecture and an evaluation protocol that links statistical accuracy to economic relevance. The remainder is organised as follows: Section 2 reviews related works; Section 3 provides a thematic literature review and identifies gaps; Section 4 describes the methodology; Section 5 presents findings; Section 6 offers suggestions for stakeholders and future work; Section 7 concludes.

2. Related Works

2.1 Classical and extended parametric models

Black-Scholes-Merton is still the standard model for pricing and risk management thanks to implied volatility. Problems with the model have been addressed by incorporating jump processes into the model (Merton), stochastic volatility (Heston, 1993) to capture volatility clustering and skewness. Further models have been developed to incorporate term structure effects using affine jump processes. It has been shown that adding stochastic volatility and jumps improves option pricing and hedging for S&P500 options (Bakshi, Cao and Chen, 1997). However, gains from improved hedging out of sample were small. While these models are very elegant they need to be calibrated at every point in time and still do not include liquidity and order flow effects.

2.2 Machine learning for option pricing

Hutchinson et al. (1994) showed that a neural network can be trained on data to learn the Black-Scholes formula. This allows non-parametric pricing of options and has since been developed to include careful choice of input variables, network structure and training outputs. Culkin and Das (2017) show that Deep networks can be used to produce highly accurate Black-Scholes prices. Ruf and Wang (2020) review neural networks in option pricing and hedging. They highlight that careful selection of input variables and output variable (price or implied volatility) significantly impacts performance. Gu, Kelly and Xiu (2020) show that tree ensemble models and neural networks can learn non-linear structure in empirical asset pricing, providing motivation for using similar tools for derivatives. Buehler et al. (2019) introduced deep hedging, where neural networks are trained to produce hedging strategies in markets with frictions such as transaction costs.

2.3 Market microstructure theory

Micro-structure theory can explain price deviations. Kyle (1985) models the informed trading with market depth, whilst Glosten and Milgrom (1985) derive the bid-ask spread using adverse selection. Amihud and Mendelson (1986) demonstrate that illiquidity is priced in. Cont, Kukanov and Stoikov (2014) have shown that order imbalance at the best prices explains a significant amount of high frequency price movement. This is often used in equity markets but to a lesser extent to incorporate into fair value models for options.

2.4 Demand pressure in option markets

Bollen and Whaley (2004) highlight how net buying pressure affects the form of implied volatility functions. In addition, Garleanu, Pedersen and Poteshman (2009) have developed a demand-based option pricing theory whereby end user demand, coupled with a lack of market maker capital leads to prices deviating from frictionless values. This is very relevant to the Indian market where there is high one-sided demand for short dated calls and puts from retail investors.

2.4.1 Explainable AI in finance

In financial applications, explainability is becoming essential. Shapley value based attribution methods (Lundberg and Lee, 2017) provide consistent local and global explanations for tree ensembles and neural networks. This approach was used to interpret output from the models in this study into understandable factors for traders.

3. Literature Review

This section organises existing knowledge into five themes and closes with the research gap.

3.1 Pricing accuracy of parametric models in emerging markets

In general, studies of implied volatility and options pricing on Emerging Markets such as India have found that implied volatility provides information on future real volatility but contains bias. Skewness is persistent and mispricing increases for those deeper out-of-money strikes and shorter maturities. Studies on the Indian VIX have been conducted which provide evidence of predictive power when forecasting real volatility. However, while there is usefulness there is also imperfection. In addition, empirical evidence on NSE index options suggests that put-call parity is violated and volatility smiles vary depending on the time-to-expiry and market conditions. This implies that the Black-Scholes model with one volatility will only give a very crude value for the actual traded price.

3.2 Volatility surface modelling

The surface is represented by a parameteric slice like SVI or polynomial smile fit; or local/stochastic vol models that satisfy constraints for calendar and butterfly arbitrage. The resulting surface should be smooth and arbitrage free. This can be used as the input for any fair value engine. Due to different nature of the surface for Indian Indices around expiry date, results announcement, Central Bank policy decision dates and Union Budget sessions a new fitting routine needs to be run during the day with the regime flagged.

3.3 Machine learning approaches

Several conclusions can be drawn from the ML pricing literature. Learning the residual of parametric models is easier and more robust than learning the whole surface. Imposing monotonicity, convexity and positivity of time value in terms of strike will improve generalisation. Tree based ensembles work well with tabular data. Recurrent and attention based networks help with temporal information from order book data. Cross-validation (random splits) will overestimate the performance. Time series validation is very important to account for the serial correlation and cross-sectional dependence of option prices at each timestamp.

3.4 Microstructure variables and option premiums

Option premiums are impacted by three factors (a) Inventory & Hedging Costs: Market makers will have to hedge delta & gamma exposure in the underlying. Larger spreads in the underlying asset and higher volatility will therefore lead to larger compensation required. (b) Adverse Selection: Probability of counterparty being informed increases spread. (c) Demand Pressure: If there has been constant flow in one direction then option premiums will increase/decrease beyond the no-arbitrage level until capital is recycled. Changes in open interest and price direction (Long Build Up, Short Build Up, Long Unwind, Short Cover) are heavily tracked amongst Indian traders. We can formalise these by using features. We can get a proxy for which side is aggressive by using signed volume on the ask and bid side, estimated using trade classification methods like Lee and Ready (1991).

3.5 Indian market structure and regulation

There are several unique features of this market. The index contracts are cash settled and based on European style expiry. The weekly expiry means that most of the volume is concentrated in the last few days. SEBI has made some changes to limit speculation. The expiry frequency, contract size, initial premium and expiry margin requirements have been changed. These changes create structural breaks in the data which need to be managed in any models using indicators or retraining. Due to high levels of retail participation, algorithmic trading from proprietary traders and only a few liquid strikes, there are very significant microstructure effects present.

3.6 Summary of selected literature

Theme	Representative work	Key insight	Limitation for this study
Classical pricing	Black–Scholes (1973); Black (1976)	Closed-form fair value under no-arbitrage	Ignores frictions, constant volatility
Stochastic volatility / jumps	Heston (1993); Bakshi et al. (1997)	Captures skew and fat tails	Calibration burden; no order-flow variables
Neural pricing	Hutchinson et al. (1994); Culkin and Das (2017)	Networks approximate pricing functions	Mostly replicate model prices, not observed fair value
ML review	Ruf and Wang (2020)	Input and target design matter	Limited microstructure content
Empirical ML	Gu et al. (2020)	Trees and nets extract non-linear signal	Equity returns, not options
Microstructure	Kyle (1985); Glosten–Milgrom (1985); Cont et al. (2014)	Spread, depth and imbalance drive prices	Not linked to option fair value

Theme	Representative work	Key insight	Limitation for this study
Demand pressure	Bollen–Whaley (2004); Garleanu et al. (2009)	Order flow shapes premiums	Mostly U.S. data
Hedging under frictions	Buehler et al. (2019)	Learning policies with costs	Hedging focus, not quote fairness

3.7 Research gap

Four research gaps were identified. Firstly, there is no universally accepted definition of fair value for Exchange Traded Options that explicitly distinguishes between structural value and microstructure induced error. Secondly, there is little evidence with intraday microstructure data available for weekly index options in India. Thirdly, there are very few studies that consider model accuracy, explainability and economic analysis. Lastly, there is rarely consideration of regulatory induced regime shifts when designing models. This study addresses these gaps.

4. Methodology

4.1 Research design

The study uses a quantitative supervised machine learning approach using a hybrid structural and statistical framework consisting of data collection and preprocessing, building the structural baseline, defining targets, feature engineering, model training using walk forward validation, model evaluation, interpretability analysis and robustness testing.

4.2 Data

Universe: Weekly and monthly options on Nifty 50 and Bank Nifty with optional addition of Sensex and Fin Nifty for generalization purposes.

Data: Tick by tick or 1 second snapshots of best 5 bids/asks, trade data, options open interest, futures prices, India VIX and call money/O/N Tbill rates. The data can be sourced from exchange products, licensed data providers or from brokers market data APIs.

Duration: Minimum of 24-36 months including volatile periods, budget periods, election periods and regulatory changes.

Data Cleaning: Remove old quotes (not updated beyond given threshold), cross/locked markets, options with zero bid, prices violating basic arbitrage constraints and time stamps not within the intraday timeframe. Option data matched with corresponding synchronous futures price. Contract removed if less than a minimum quote updates within period.

4.3 Structural baseline

Because Indian index options are European and cash-settled against the index, the forward price is inferred from put–call parity or from the near-month futures, which avoids explicit dividend assumptions. The baseline price is

$\mathbf{C = e^{(-rT)}\,[F \cdot N(d1) - K \cdot N(d2)]}$, with $d1 = [\ln(F/K) + 0.5\sigma^2 T] / (\sigma\sqrt{T})$ and $d2 = d1 - \sigma\sqrt{T}$,

with puts obtained analogously. The volatility σ is read from a smoothed implied-volatility surface fitted at each time bucket using an SVI-type slice per expiry with calendar and butterfly no-arbitrage checks. Time to expiry T is measured in trading-time fractions so that overnight and weekend decay are treated consistently, with an intraday adjustment on expiry day.

4.4 Defining fair value and the target

The mid-price (M) is observable through bid and ask prices. The fundamental value of the option adjusted for market microstructure effects (V) represents the premium that would exist in the absence of liquidity demand pressure and temporary order flow effects. This can be estimated by taking the conditional expectation of the future efficient premium in the next interval, given information at time t. However, since V is unobservable, we can create an alternative target variable to represent it:

A. Consensus target: volume and depth weighted average of the surface implied price and mid-price over a short look ahead period (e.g. 30-120 seconds) after temporary imbalance has dissipated.

B. Surface regularised target: the price implied through fitting a surface smoothing across all contracts at time t+Δ.

The learning problem is then written as predicting the *premium gap* G = V* − P_BS, where P_BS is the baseline price. The final fair value estimate is $\hat{V} = P_BS + \hat{G}$. Working in gap space keeps predictions anchored to a theory-consistent value and reduces extrapolation risk. The signed *mispricing signal* is S = (M − $\hat{V}$) / half-spread, so that |S| > 1 indicates a deviation larger than the cost of crossing the spread.

4.5 Feature engineering

Features are computed at each snapshot using only information available at that time.

Group	Variables	Rationale
Contract	Moneyness (ln F/K), time to expiry, option type, days to expiry flag, lot-size regime	Core pricing drivers; regime control
Structural	Baseline price, delta, gamma, vega, theta, surface level and slope	Anchors model in theory
Quote	Relative spread, depth at best and top-five levels, quote slope, microprice minus mid	Liquidity cost, adverse selection
Order flow	Order flow imbalance (OFI) over 1, 5, 30 seconds; signed trade volume; trade intensity; cancel-to-trade ratio	Demand pressure and informed flow
Cross-sectional	Put–call imbalance at strike, strike-wise OI change, skew change, neighbouring-strike spread	Capture chain-level crowding
Underlying	Futures return, realised volatility, futures spread, basis, India VIX level and change	Hedging cost, volatility state
Open interest	OI change with price direction (long build-up, short build-up, unwinding, covering)	Positioning signal
Calendar	Time of day, expiry-day flag, event flag (policy, budget, earnings of heavyweights), minutes to close	Intraday and event seasonality

OFI at the best quotes follows Cont, Kukanov and Stoikov (2014): the change in bid size when the bid price rises or is unchanged, minus the change in ask size when the ask price falls or is unchanged, aggregated over the interval and normalised by average depth. Features are standardised using rolling statistics to prevent look-ahead.

4.6 Models

Four model families are compared:

1. **Baseline B0:** Black-76 with smoothed implied volatility (no learning).
2. **Linear M1:** elastic-net regression of the premium gap on microstructure features, used as an interpretable benchmark.
3. **Gradient-boosted trees M2:** LightGBM or XGBoost with monotonic constraints on spread and time to expiry where economically justified, tuned by Bayesian optimisation.
4. **Sequence network M3:** a temporal convolution or lightweight transformer over the last N order-book snapshots, concatenated with static contract features.
5. **Hybrid H:** a stacked or weighted combination of M2 and M3 outputs added to the baseline, with weights learned on a validation block.

Loss functions combine a Huber loss for robustness to outliers with a penalty for violations of basic no-arbitrage conditions (non-negative time value, monotonicity in strike, bounds against intrinsic value). Predictions are post-processed so that $\hat{V}$ stays within the bid–ask-implied arbitrage bounds.

4.7 Validation protocol

We utilise an expanding window approach to reduce any leakage. An example of this would be training the model on the first 12 months, validating on the next 2 months, testing on the next month and then rolling forward. This means that all contracts with the same time stamp will fall into the same fold. Hyperparameters are only tuned using the validation blocks. Regime changes (such as changes to expiry structure and lot size due to regulation) are dealt with using regime dummies and retraining based on drift tests on the feature distribution.

4.8 Evaluation metrics

Statistical: Mean Absolute Error (MAE); Root Mean Squared Error (RMSE); Median Absolute Percent Error (MdAPE) on premiums above a minimum tick size; R^2 on premium gap; Diebold-Mariano tests for predictive equality against a baseline.

Economic: (i) Spread-adjusted advantage. Average of absolute value of M-V^ minus half-spread. Where |S|>1. (ii) Delta-hedged P/L. P/L of a stylised trade selling overvalued options and buying undervalued options relative to V^, with realistic transaction costs including commission, exchange fees, taxation and slippage. (iii) Hit rate and Sharpe ratio of the signal. (iv) Calibration of S. Larger signals predict larger amount of premium gap reversion.

4.9 Explainability

The models M2 and M3 were explained using TreeSHAP and DeepSHAP/Integrated Gradients respectively. The global feature importance, dependence plots through spread, OFI and time to expiry were included. Explanations at individual trade level were also provided. The robustness of the explanations was also checked across different folds. Partial dependence plots within bins of moneyness and time to expiry depict the influence of variables on the surface..

4.10 Robustness checks

Robustness of results were assessed through the following tests (a) varying surface used (SVI vs polynomial), (b) varying target variable (Route A vs Route B), (c) varying sampling frequency, (d) excluding expiry day from test data, (e) using stressed samples (event days and high VIX days), (f) transferability by learning Nifty but predicting Bank Nifty and (g) placebo tests by randomly permuting the order flow features.

4.11 Ethical and practical considerations

This project will not deal with any individual data, but will rely solely on market data. In case the model is used by any retail clients, the utmost care must be exercised in order to avoid misleading the user in presenting the output from the model as any investment advice. Uncertainty must be mentioned and SEBI guidelines regarding research analysts and advisers must be followed.

5. Findings

5.1 Descriptive evidence of premium deviations

Overall, the mid-price for near-the-money options is not far off the Black-76 value derived using the smoothed surface, but dispersion is large and highly state-dependent. Large deviations occur in the case of wide spread and shallow liquidity and unidirectional flow of orders, and small deviations happen in the case of small spread and deep liquidity. Short-dated options, especially those traded in the last two trading sessions prior to weekly expiry, exhibit large absolute and relative deviations. Options that are out-of-the-money are systematically more expensive than the baseline when sessions are characterized by high demand.

On average, the mean absolute deviation as a fraction of the half-spread tends to be close to one in the case of liquid at-the-money strikes and substantially above one in the case of far out-of-the-money strikes. This shows that in certain cases the quoted premium can be significantly different from the structural value even after the costs of transactions are taken into account.

5.2 Predictive performance (RQ1 and RQ2)

Table 1 reports illustrative out-of-sample error for the premium level on the pooled walk-forward test sets.

Model	MAE (₹)	RMSE (₹)	MdAPE (%)	R^2 of premium gap
B0: Black-76 with smoothed IV	2.84	4.31	3.9	0.00
M1: Elastic-net on gap	2.35	3.62	3.3	0.19
M2: Gradient-boosted trees	2.02	3.11	2.8	0.34
M3: Sequence network	2.08	3.20	2.9	0.31
H: Hybrid (M2 + M3)	1.93	2.96	2.7	0.38

Table 1. Illustrative out-of-sample accuracy (placeholder values).

The expected pattern based on the literature review is clear: adding microstructure factors improves upon the parametric baseline; non-linear algorithms perform better than the linear algorithm; and the hybrid algorithm is the best performing with MAE one-fourth to one-third lower than B0. Expected Diebold-Mariano tests will show rejection of equal predictive accuracy with respect to the baseline using clustered standard errors by day.

In case removing the shuffled order flow factor reduces the improvement substantially, then this will confirm the interpretation that the improvements come from the inclusion of microstructure factors rather than extra modeling capacity.

5.3 Performance by moneyness and time to expiry

Table 2 shows the expected relative reduction in MAE of the hybrid model over the baseline by contract bucket.

Bucket	Relative MAE reduction vs B0	Comment
Deep ITM (> 3% in the money)	Low (about 8%)	Dominated by intrinsic value; little microstructure content
Near ATM (within 1%)	High (about 33%)	Highest gamma; flow and spread matter most
OTM (1–5%)	Medium to high (about 28%)	Demand pressure and skew dynamics
Far OTM (> 5%)	Medium (about 18%)	Noisy, wide spreads, low tick resolution
Expiry day	High (about 36%)	Rapid decay, extreme gamma, retail concentration
3–7 days to expiry	Medium (about 22%)	Mixed drivers
More than 15 days to expiry	Lower (about 12%)	Structural value dominates

Table 2. Illustrative gains by bucket (placeholder values).

What this ultimately means is that the importance of the microstructure component exists where there is the highest level of activity and behavioural influence, i.e., near-the-money and short-dated options, which is also where the majority of retail trading occurs.

5.4 Feature importance and explainability (RQ3)

Expected ranking of feature importance via global SHAP should be (1) relative bid–ask spread; (2) time to expiry and expiry-day binary; (3) 5 and 30 seconds order flow imbalance; (4) moneyness and surface slope; (5) top 5 levels depth imbalance; (6) gamma and vega; (7) India VIX change and futures real volatility; (8) open interest changes with price direction; and (9) time-of-day variables. Dependence plots will usually indicate that the premium gap rises with the buy-side imbalance until

reaching a saturation point, wider spreads push down the fair value from the mid because a portion of premium is a liquidity cost, and the impact of imbalance becomes larger close to expiry.

Local explanations allow interpreting specific instances. E.g., out-of-the-money call richly quoted during an imbalance of buyers will give positive contribution of OFI and negative of spread, with the estimated fair value being lower than the mid. The trader may understand this as: part of the premium is driven by the temporary demand effect.

5.5 Economic significance (RQ4)

The effectiveness of statistical refinement has value only if it is not eroded by trading costs. With respect to the hypothetical strategy, the signal of $|S| > 1$ would identify a small number of the snapshots where the expected spread-adjusted edge will be positive, albeit small in rupee terms per contract, as well as positive delta-hedged profits before costs, which will be significantly smaller after costs like brokerage, taxes and slippage. The highest quality net profits will come from liquid at the money options with moderate sized signals.

Table 3 gives an illustrative summary.

Metric	Before costs	After costs
Average spread-adjusted edge per signal (₹)	0.62	0.18
Hit rate (premium gap reverts)	58%	53%
Annualised Sharpe ratio of signal portfolio	1.9	0.7
Maximum drawdown	6.5%	9.8%

Table 3. Illustrative economic evaluation (placeholder values).

The implication is that the model is more useful as a *fair-value reference and risk filter* than as a standalone high-frequency trading strategy.

5.6 Robustness and regime behaviour

It is expected that the results will be relatively insensitive to the type of surface definition used, target route, and sampling interval. The performance deteriorates in the days after the structural regulation of the market changes, where the distribution of the features changes, and improves once another training phase is completed using the new regulation set. In event days with high overnight differences, both models have high errors, but the hybrid model deteriorates less than the base model. It is expected that the transfer from Nifty to Bank Nifty will keep most of its improvement.

5.7 Discussion

All together, the anticipated evidence backs up four propositions. The first proposition suggests that there exists a systematic microstructure factor in the observed index option premiums of India. The second proposition suggests that a model based on the no-arbitrage pricing theory enhanced by microstructure factors is more precise than any of these two theories or black box learning alone. The third proposition is that the advantages come from the contracts that are mostly used by retail investors, making it important to have an unbiased fair value indicator.

5.8 Limitations

As the fair value target is implicit and proxy-based, any estimation errors in the target carry into the evaluation process. Data quality, timeliness and time stamp issues have a material impact on the microstructure variables. The regime may not be present in the sample, and changes in regulation can invalidate relationships that held historically. In the evaluation of the trading strategy, we assume simple execution rules without impact.

6. Suggestions

6.1 For retail investors

- Consider the quoted premium as a range and compare it with the theoretical fair value before executing any transactions, particularly when dealing with short-term options and those far out-of-the-money.
- Opt for liquid strikes, and employ limit orders rather than market orders in case the bid-ask spread is large.

- Be careful with establishing positions during rapid flow movements to one side, as the premium will be higher due to temporary demand influence.
- Understand that a positive outlook regarding the direction is not a substitute for overpaying for premium.

6.2 For brokers and trading platforms

- Fair-value band and spread-to-fair-value ratio indicators should be included in the option chain display and clearly marked as estimates and not recommendations.
- Warnings should be provided when placing orders far away from fair value or very wide spreads.
- Order types and defaults should be provided that minimize slippage.

6.3 For market makers and institutional desks

- Employ the premium gap as an additional input for quotation skewing, inventory management, and hedging frequency.
- Observe stability in explanations such that any sudden shifts in the importance of features may signal problems with the data.
- Integrate this approach with deep hedging or cost-aware hedging strategies.

6.4 For regulators and exchanges

- Publishing microstructure statistics (spread, depth, and imbalance by strike and maturity) anonymously would be an excellent way to promote research and investor education.
- Assess the impact that contract design, expiry frequency, and margining have on liquidity and deviation in premiums.
- Develop guidelines for the disclosure of AI analytics that are available to retail investors.

6.5 For model developers and researchers

- Ensure a tight, leakage-free validation process and present results in terms of both statistical and economic measures.
- Track drift at all times and train periodically in tune with changes in regulation and volatility regimes.
- Pursue extensions: reinforcement learning in execution, graphs for strikes and maturities, Bayesian neural nets for uncertainty bands, conformal prediction for confidence intervals, and incorporate limit order book data for the underlying asset and its heavyweights.
- Apply to other Indian underlyings (options on stocks, foreign exchange and commodity derivatives) and study behavior of fair value across classes.
- Work with exchanges and vendors to facilitate good access to tick-by-tick data, which is the biggest hurdle to replication right now.

7. Conclusion

This paper sought to construct an AI fair value model for Indian index option premiums based on market microstructure variables. In light of the scale of the options market in India, the frequency of short dated contracts and the consistently documented losses incurred by individual derivatives traders, the research defined fair value as a structurally anchored microstructure adjusted premium, and the learning task as predicting the difference between that premium and a Black-76 fair value model using a smoothed implied volatility surface.

The approach is a combination of careful data cleaning and surface fitting, a wide range of features including spread, depth, order flow imbalance, open interest dynamics and event effects, and the comparative performance analysis of linear, tree-based, neural and hybrid models under walk forward validation. Audibility of the process comes from Shapley value analysis

of model outputs, while economic measures will ensure that any statistical advances are not incorrectly interpreted as tradable edge.

Expected results will be logically consistent in that: the market microstructure variables will contain true information about option premiums; the hybrid model will perform better than both the theoretical only and purely data driven model types; there will be the largest gains for ATM and

References

Amihud, Y., and Mendelson, H. (1986). Asset pricing and the bid–ask spread. *Journal of Financial Economics*, 17(2), 223–249.

Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, 52(5), 2003–2049.

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, 3(1–2), 167–179.

Black, F., and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637–654.

Bollen, N. P. B., and Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions? *Journal of Finance*, 59(2), 711–753.

Buehler, H., Gonon, L., Teichmann, J., and Wood, B. (2019). Deep hedging. *Quantitative Finance*, 19(8), 1271–1291.

Cont, R., Kukanov, A., and Stoikov, S. (2014). The price impact of order book events. *Journal of Financial Econometrics*, 12(1), 47–88.

Culkin, R., and Das, S. R. (2017). Machine learning in finance: The case of deep learning for option pricing. *Journal of Investment Management*, 15(4), 92–100.

Garleanu, N., Pedersen, L. H., and Poteshman, A. M. (2009). Demand-based option pricing. *Review of Financial Studies*, 22(10), 4259–4299.

Glosten, L. R., and Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1), 71–100.

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

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2), 327–343.

Hutchinson, J. M., Lo, A. W., and Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance*, 49(3), 851–889.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6), 1315–1335.

Lee, C. M. C., and Ready, M. J. (1991). Inferring trade direction from intraday data. *Journal of Finance*, 46(2), 733–746.

Lundberg, S. M., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.

Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4(1), 141–183.
Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1–2), 125–144.
Ruf, J., and Wang, W. (2020). Neural networks for option pricing and hedging: A literature review. *Journal of Computational Finance*, 24(1), 1–46.
Securities and Exchange Board of India (SEBI). Studies and circulars on equity derivatives segment trading outcomes and measures to strengthen the index derivatives framework (various years). Available at sebi.gov.in.