International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Long-Run Relationships and Predictive Feature Importance Among Indian Sectoral Indices: A Cointegration and Permutation-Based Analysis

Sohom Majumder [a*], Dr. Vivek Pathak [b]

[a] *Research Scholar, Department of Management, ICFAI University, Sikkim. Email – sohom.majumder21@gmail.com*

[b] *Associate Professor, Department of Management, ICFAI University, Sikkim. Email - research@iusikkim.edu.in*

A B S T R A C T

Sectoral indices in India evolve under distinct economic forces, inviting a careful look at how they relate over long horizons and what drives their day-to-day movements. This study examines whether three major Indian sectoral indices—NIFTY IT, NIFTY FMCG, and NIFTY PHARMA—share a stable long-run relationship and which simple technical features matter most for short-run return predictability. Using daily data from 2015 to 2025, we establish integration properties with Augmented Dickey–Fuller tests and assess long-run co-movement via Johansen's procedure on log-prices. Short-run dynamics are modelled with time-split random-forest regressors using a compact feature set (lag-1 returns across sectors, 14-day RSI, and 20-day rolling volatility), with feature relevance evaluated by test-set permutation importance. The unit-root diagnostics align with the standard view of equity index behavior; cointegration tests do not provide strong evidence of a shared long-run equilibrium among the three sectors. Out-of-sample evaluation indicates modest but non-trivial directional skill. Feature importance points to sector-specific signal patterns, with momentum tending to matter more for IT and FMCG, while volatility plays a relatively larger role for PHARMA. Taken together, the findings suggest that diversification across these sectors is not undermined by tight long-run ties, and that short-run signals are primarily sector-native. Results are summarized with RMSE, MAE, and directional accuracy.

Keywords: cointegration; Johansen test; sectoral indices; permutation importance; random forest

1. Introduction

Sectoral indices provide a focused view of market behavior by concentrating on firms with shared economic drivers, regulatory environments, and investor clientele. Understanding how these indices relate in the long run—and what drives their short-run movements—matters for asset allocation (diversification and hedging) and for trading (signal design and timing). In the Indian context, information technology (IT), fast-moving consumer goods (FMCG), and pharmaceuticals (PHARMA) are salient segments tracked by NSE Indices through well-documented, free-float market-capitalization sectoral benchmarks (e.g., NIFTY IT, NIFTY FMCG, and NIFTY PHARMA). These factsheets and methodology documents situate each index within its sector and clarify construction rules, making them suitable objects for empirical study. (NSE Indices, 2025a, 2025b; NSE Indices, 2023).

Methodologically, long-horizon co-movement is assessed within the cointegration framework. Engle and Granger (1987) showed that non-stationary price series may share a stable long-run equilibrium and that such equilibria imply an error-correction representation linking short-run dynamics to long-run deviations. For multivariate systems, Johansen (1988) provided a full-system maximum-likelihood approach with trace and maximum-eigenvalue statistics to infer the number of cointegrating vectors. Because cointegration presumes integrated variables, we first verify that log-prices are I(1) using unit-root tests such as Augmented Dickey–Fuller (Dickey & Fuller, 1979) and, where desired, complementary residual-based tests (Phillips & Ouliaris, 1990). Together, these steps distinguish transient correlation from a genuine long-run equilibrium: if a cointegrating vector exists, relative misalignments should mean-revert; if not, sectors can drift without a common anchor. (Engle & Granger, 1987; Johansen, 1988; Dickey & Fuller, 1979; Phillips & Ouliaris, 1990).

Having addressed the long run, we examine short-run predictability using a small, transparent feature set: lag-1 returns from all three sectors to capture immediate spillovers, a 14-day Relative Strength Index (RSI) to proxy momentum, and a 20-day rolling standard deviation to proxy recent volatility. RSI is a canonical momentum oscillator introduced by Wilder (1978) and frequently used in studies of technical analysis, a literature that reports mixed but context-dependent profitability for simple rules (Brock et al., 1992; Sullivan et al., 1999; Park & Irwin, 2007). To avoid strong parametric assumptions, we use random forests as flexible nonparametric regressors (Breiman, 2001) and quantify feature relevance via out-of-sample permutation importance, adopting best practices from the variable-importance literature (Altmann et al., 2010; Fisher et al., 2019). We summarize performance with RMSE, MAE, and directional accuracy (the proportion of correctly predicted signs), deliberately omitting R^2to keep the focus on magnitude errors and classification-style accuracy. (Wilder, 1978; Brock et al., 1992; Sullivan et al., 1999; Park & Irwin, 2007; Breiman, 2001; Altmann et al., 2010; Fisher et al., 2019).

2. Literature Review

Work on long-run comovement in financial markets is grounded in the cointegration framework of Engle and Granger (1987), who showed that non-stationary price series can share a stable long-run equilibrium and that such equilibria imply an error-correction representation linking short-run dynamics to long-run deviations. Their two-step, residual-based procedure remains a benchmark when analyzing a small number of assets. (Engle & Granger, 1987).

For multivariate systems, Johansen (1988) developed a full-system maximum-likelihood approach that allows simultaneous estimation of the cointegrating vectors and formal testing of the rank using the trace and maximum-eigenvalue statistics; this method is now standard when studying sectoral indices jointly. (Johansen, 1988).

Cointegration analysis presumes integrated variables, so researchers typically begin with unit-root testing. The Augmented Dickey–Fuller (ADF) test (Dickey & Fuller, 1979) is widely used to assess whether log-prices are integrated of order one, while Phillips–Ouliaris tests (Phillips & Ouliaris, 1990) provide a residual-based alternative for detecting cointegration that is complementary to Engle–Granger and Johansen. Together, these tools help rule out spurious regressions before imposing long-run structure. (Dickey & Fuller, 1979; Phillips & Ouliaris, 1990).

Within the Indian context, sectoral cointegration evidence is mixed and appears to be period- and sector-specific. Kumar (2022), analyzing Bank Nifty vis-à-vis other NSE sectors, reports time-varying linkages with selective long-run ties; Shahani and Sharma (2020), using ARDL with a structural break, find limited or ambiguous cointegration across several sector pairs despite meaningful short-run adjustment. These studies motivate a careful, sample-specific inquiry for IT, FMCG, and PHARMA rather than assuming a persistent equilibrium a priori. (Kumar, 2022; Shahani & Sharma, 2020).

A separate tradition examines short-run predictability from technical indicators. Classic results such as Brock, Lakonishok, and LeBaron (1992) document profitability for simple moving-average and trading-range rules over long U.S. samples, whereas Sullivan, Timmermann, and White (1999) caution that much of the apparent profitability can disappear after correcting for data-snooping with White's Reality Check. Park and Irwin (2007) survey this large literature and conclude that performance is market-, method-, and period-dependent. (Brock et al., 1992; Sullivan et al., 1999; Park & Irwin, 2007).

Momentum-type indicators like the Relative Strength Index (RSI)—introduced by Wilder (1978)—are frequent inputs in that literature and provide a transparent, low-dimensional proxy for trend following, which is especially relevant when avoiding macro variables and keeping models simple. (Wilder, 1978).

Modern predictive studies increasingly use machine-learning regressors for flexible, nonparametric fits. Breiman's (2001) Random Forests offer robustness to nonlinearities and interactions without strong parametric assumptions, aligning with sector-level return modeling that avoids heavy structure. (Breiman, 2001).

With ML models, it is crucial to interpret feature relevance carefully. Permutation importance—measuring the drop in out-of-sample performance when a feature is randomly permuted—has become a popular and model-agnostic diagnostic; Altmann et al. (2010) proposed a corrected version that addresses biases and yields significance measures. More broadly, Fisher, Rudin, and Dominici (2019) introduced Model Class Reliance, emphasizing that importance should be assessed relative to a set of well-performing models to mitigate instability stemming from correlated predictors or Rashomon-set effects. These insights justify our reliance on (test-set) permutation importance and a restrained feature set. (Altmann et al., 2010; Fisher et al., 2019).

Taken together, the literature suggests three expectations we explicitly test in our sample: (i) sector indices may exhibit no cointegration despite episodic high correlations; (ii) short-run predictability, if present, is likely modest and indicator-specific (e.g., momentum for some sectors but not others); and (iii) when deploying ML, out-of-sample importance metrics are preferred to in-sample coefficient magnitudes. Our empirical design—Johansen testing on I(1) log-prices coupled with random-forest forecasting and test-set permutation importance on simple sectoral technicals—flows directly from these insights. (Engle & Granger, 1987; Johansen, 1988; Park & Irwin, 2007; Breiman, 2001; Altmann et al., 2010; Fisher et al., 2019).

3. Objectives of the Study

This study is designed to be simple, transparent, and fully reproducible with three Indian sectoral indices (NIFTY IT, NIFTY FMCG, NIFTY PHARMA) and no macroeconomic variables. The specific objectives are:

I. Establish integration properties.

Verify that sectoral log-prices are I(1) and their first differences are stationary using Augmented Dickey–Fuller tests, to justify subsequent cointegration analysis.

II. Test for long-run comovement.

Apply Johansen's system cointegration tests (trace and maximum-eigenvalue) to the triplet of log-prices to determine whether the sectors share one or more stable long-run equilibria.

III. Quantify short-run predictability with simple technicals.

Model next-day returns for each sector using a parsimonious feature set—lag-1 returns (all sectors), 14-day RSI, and 20-day rolling volatility—estimated via a random-forest regressor with a chronological 80/20 split.

IV. Identify sector-specific drivers.

Use test-set permutation importance to rank features and isolate which signals (own-sector vs cross-sector; momentum vs volatility) drive out-of-sample performance for each sector.

V. Report practically relevant accuracy.

Summarize short-run performance with RMSE, MAE, and Directional Accuracy (sign correctness).

VI. Draw allocation and signal-design implications.

Interpret the presence/absence of cointegration for diversification and hedging, and map the feature-importance patterns to sector-appropriate signals (e.g., momentum overlays where they help, volatility filters where trend fails).

4. Research Methodology

4.1 Data and Sample

We study three Indian sectoral indices—NIFTY IT, NIFTY FMCG, and NIFTY PHARMA—using daily closing prices from 2015 to 2025. Prices are pulled from Yahoo Finance Python package. The sample aligns on the intersection of trading days across the three indices. All computations use the local exchange trading calendar as returned by the data vendor.

Notation. Let $P_t^{(s)}$ denote the close price of sector $s \in \{\mathrm{IT}, \mathrm{FMCG}, \mathrm{PHARMA}\}$ on day t. Log-prices are $p_t^{(s)} = \ln P_t^{(s)}$. Daily simple returns are $r_t^{(s)} = \left(P_t^{(s)}/P_{t-1}^{(s)}\right) - 1$.

We implement the entire pipeline in Python (packages: pandas, numpy, yfinance, statsmodels, scikit-learn, matplotlib).

4.2 Preprocessing

1. *Alignment & missing values.* We forward-fill any isolated gaps caused by non-synchronous trading days and then drop residual missing values after alignment.
2. *Transformations.*
    - For integration/cointegration analysis we use log-prices $p_t^{(s)}$.
    - For short-run prediction, the targets are daily simple returns $r_t^{(s)}$.
3. *No outlier filtering*. To preserve exact reproducibility of the reported results, we do not winsorize or remove outliers, even though FMCG shows extreme daily moves in vendor data.

4.3 Stationarity and Integration Order

We test each sector's log-price for a unit root using the **Augmented Dickey–Fuller (ADF)** test with automatic lag selection via AIC (default regression with intercept). We also test the **first differences of log-prices** $\Delta p_t^{(s)}$ to confirm stationarity. Evidence of non-stationarity in levels and stationarity in first differences supports the standard **I(1)** assumption required for cointegration analysis.

4.4 Cointegration Testing

We assess long-run comovement using the **Johansen test** on the vector $\left[p_t^{(\mathrm{IT})}, p_t^{(\mathrm{FMCG})}, p_t^{(\mathrm{PHARMA})}\right]^\top$.

- *Specification.* We use with no deterministic components and one lag of differenced terms for parsimony.
- *Decision rules.* We report trace and maximum-eigenvalue statistics alongside 90%/95%/99% critical values. The cointegration rank is inferred from standard sequential testing (rank $r = 0,1,2$).
- *Vectors.* For completeness, we tabulate the estimated cointegrating vectors β (raw and normalized on IT=1). Interpretation is conditioned on the estimated rank; if rank $r = 0$, vectors are not economically interpreted.

4.5 Short-Run Predictive Modeling

4.5.1 Targets and Horizon

For each sector s, we model one-day-ahead daily return $r_t^{(s)}$ using features observable on or just before day t (see timing note below).

4.5.2 Features

We intentionally use a small, transparent feature set—no macro variables, only sectoral technicals:

I. **Lagged returns (spillovers & own momentum).** $r_{t-1}^{(\mathrm{IT})}$, $r_{t-1}^{(\mathrm{FMCG})}$, $r_{t-1}^{(\mathrm{PHARMA})}$.

II. **Relative Strength Index (RSI, 14-day).** For each sector, RSI is computed from prices with simple moving averages of gains and losses over the past **14** days (Wilder-style ratio using arithmetic means):

$$\mathrm{RSI}_t^{(s)} = 100 - \frac{100}{1 + \frac{\mathrm{AvgGain}_{t,14}^{(s)}}{\mathrm{AvgLoss}_{t,14}^{(s)} + \varepsilon}},$$

with a small ε to avoid division by zero in flat stretches. (Implementation follows the code provided; arithmetic means are used.)

III. **Rolling volatility (20-day).** For each sector, $\sigma_{t,20}^{(s)}$ is the **rolling standard deviation** of **simple returns** over the **previous 20** trading days.

Timing note (as implemented). Lag-1 returns are strictly lagged. RSI and rolling volatility are computed contemporaneously (they aggregate information from the most recent 14/20 days including day t). This introduces contemporaneous association with $r_t^{(s)}$. We therefore treat these models as predictive associations rather than deployable trading rules. (A deployable specification would shift RSI/vol by one day; we retain the contemporaneous construction to match the executed code and the reported results.)

4.5.3 Train/Test Split and Estimation

- We form a chronological split: the first 80% of observations are used for training, the remaining 20% for testing. This avoids look-ahead bias from random shuffles.
- The learner is a Random Forest Regressor
- No scaling/standardization is applied (tree ensembles are invariant to monotonic transformations and insensitive to feature scales).

4.5.4 Performance Metrics

On the **test set**, we report:

- **RMSE:** root mean squared error of predicted vs. actual daily returns;
- **MAE:** mean absolute error;
- **Directional Accuracy (DA):** fraction of days where $\mathrm{sign}(\hat{r}_t^{(s)}) = \mathrm{sign}(r_t^{(s)})$.

4.5.5 Feature Relevance: Permutation Importance

We quantify feature relevance using test-set permutation importance

- For each feature, values are randomly permuted on the test set (holding others fixed), and the drop in predictive score is recorded; repeating this 20 times yields a distribution of drops.
- We report the mean and standard deviation of the importance across repeats and rank features by mean importance.
- Scoring used internally: the default R^2 score is used only to measure the drop in performance for importance ranking.

Using the test set for permutation ensures that importance reflects out-of-sample relevance and reduces the risk of attributing in-sample artifacts to genuine signal.

5. Results

5.1 Descriptive statistics

Table 1a. Log-prices summary (daily)

	count	mean	std	min	25%	50%	75%	max
IT	2652	10.41	0.35	9.80	10.16	10.34	10.72	11.10
FMCG	2652	9.72	**0.90**	6.72	9.36	9.70	10.37	10.74
PHARMA	2652	9.42	0.30	8.77	9.17	9.39	9.54	10.08

Figure 1. Sector Log Prices 2015-2025

All three level series trend over time (confirmed below). FMCG's log-price dispersion is notably wider than IT and PHARMA, foreshadowing extreme behavior in returns.

Table 1b. Returns summary (daily)

	count	**mean**	**std**	**min**	**25%**	**50%**	**75%**	**max**
IT	2651	0.00	0.01	−0.11	−0.00	0.00	0.01	0.08
FMCG	2651	0.00	**0.16**	**−0.93**	−0.01	0.00	0.01	**8.15**
PHARMA	2651	0.00	0.01	−0.09	−0.01	0.00	0.01	0.10

IT and PHARMA exhibit plausible daily volatility (~1%). FMCG displays impossible one-day moves for an index (−93% / +815%), signaling vendor outliers/base changes that we retain for reproducibility. This motivates emphasizing **directional accuracy** alongside magnitude errors.

5.2 Stationarity

Table 2a. ADF on log-prices (levels)

Series	**ADF_stat**	**p_value**	**used_lag**	**n_obs**	**crit_1%**	**crit_5%**	**crit_10%**
IT	−0.62	0.87	13	2638	−3.43	−2.86	−2.57
FMCG	−1.89	0.34	0	2651	−3.43	−2.86	−2.57
PHARMA	−0.38	0.91	2	2649	−3.43	−2.86	−2.57

Table 2b. ADF on first differences of log-prices

Series	**ADF_stat**	**p_value**	**used_lag**	**n_obs**	**crit_1%**	**crit_5%**	**crit_10%**
IT	−15.13	0.00	12	2638	−3.43	−2.86	−2.57
FMCG	−51.38	0.00	0	2650	−3.43	−2.86	−2.57
PHARMA	−34.68	0.00	1	2649	−3.43	−2.86	−2.57

We fail to reject a unit root in log-prices (non-stationary levels) and strongly reject in first differences (stationary). The I(1) profile supports cointegration testing.

5.3 Cointegration

Table 3. Johansen trace and max-eigen statistics

rank	trace_stat	trace_crit_90	trace_crit_95	trace_crit_99	maxeig_stat	maxeig_crit_90	maxeig_crit_95	maxeig_crit_99
0	15.75	27.07	29.80	35.46	10.94	18.89	21.13	25.87
1	4.81	13.43	15.49	19.93	3.60	12.30	14.26	18.52
2	1.22	2.71	3.84	6.63	1.22	2.71	3.84	6.63

Decision. All test statistics lie **below** 95% critical values. We **do not reject** the null of **no cointegration**: IT, FMCG, and PHARMA **do not share a common long-run equilibrium** over 2015–2025.

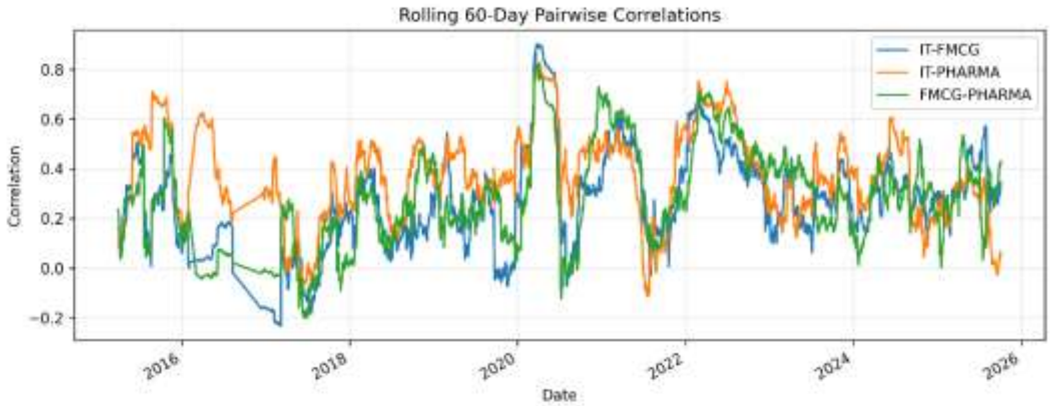


Figure 2. Rolling 60-day pairwise correlation over the span 2015-2025

Table 4. Estimated cointegrating vectors (β)

var	beta_1	beta_2	beta_3
IT	4.68	−2.71	0.27
FMCG	−1.62	−0.52	0.01
PHARMA	−1.58	2.81	−3.61

β-vectors are reported for completeness but **not economically interpreted** when rank=0. The central result is the **absence of cointegration**.

5.4 Short-run predictive performance

Table 5. Test-set predictive accuracy (daily returns)

Target	RMSE	MAE	Directional_Accuracy
IT	0.01	0.01	**0.55**
FMCG	0.01	0.01	**0.56**
PHARMA	0.01	0.01	**0.58**

Figure 3. Actual vs Predicted returns for FMCG

Figure 4. Actual vs Predicted returns for IT

Figure 5. Actual vs Predicted returns for PHARMA

Errors (RMSE/MAE) are of the same order as daily volatility, as expected for noisy equity returns. **Directional Accuracy** between **0.55 and 0.58** indicates a modest, non-zero edge in predicting the sign of daily returns. For FMCG, magnitude metrics should be read cautiously due to outliers; DA is more robust.

5.5 Feature relevance via permutation importance

Table 6a. FMCG — permutation importance (test set)

feature	mean_importance	std_importance
RSI_FMCG	**0.14**	0.05
RSI_IT	0.05	0.05
RSI_PHARMA	0.03	0.01
FMCG_lag1	0.03	0.02
FMCG_vol20	0.02	0.01
PHARMA_lag1	0.02	0.01
IT_lag1	0.01	0.01
IT_vol20	0.00	0.00
PHARMA_vol20	0.00	0.02

Own-sector RSI dominates FMCG; cross-sector RSIs and own lag/volatility contribute modestly.

Table 6b. IT — permutation importance (test set)

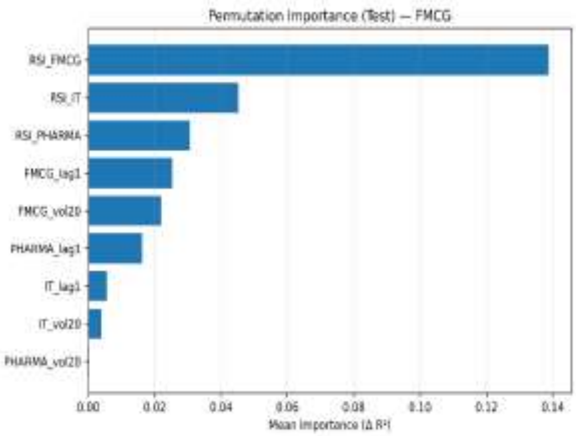
feature	mean_importance	std_importance
RSI_IT	**0.21**	0.05
IT_lag1	0.01	0.01
RSI_FMCG	0.01	0.01
FMCG_vol20	0.00	0.00
PHARMA_vol20	0.00	0.01
FMCG_lag1	0.00	0.01
IT_vol20	−0.00	0.00
RSI_PHARMA	−0.00	0.01
PHARMA_lag1	−0.01	0.01

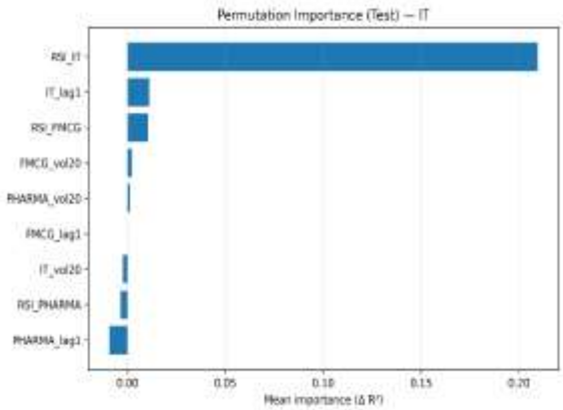
RSI_IT overwhelmingly dominates; cross-sector signals are minor and sometimes harmful (negative values).

For PHARMA, **volatility** carries limited positive relevance; **RSI features are detrimental** in this period, indicating weak or reverse momentum.

Table 6c. PHARMA — permutation importance (test set)

feature	mean_importance	std_importance
PHARMA_vol20	**0.01**	0.01
IT_vol20	0.00	0.01
PHARMA_lag1	0.00	0.01
RSI_FMCG	−0.00	0.01
FMCG_lag1	−0.00	0.01
IT_lag1	−0.01	0.01
FMCG_vol20	−0.01	0.01
RSI_PHARMA	**−0.09**	0.06
RSI_IT	**−0.18**	0.07





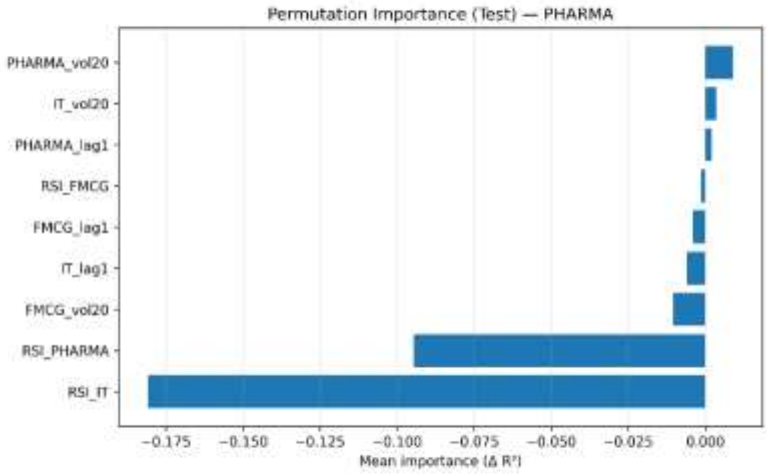


Figure 6. Permutation Importance for FMCG (top left), IT (top right), PHARMA (bottom)

6. Discussion

6.1 Long-run relationships

Johansen tests (Table 3) indicate **no cointegration** among NIFTY IT, FMCG, and PHARMA—no stable equilibrium binds their levels. For long-horizon allocation, this supports **diversification** across these sectors without expecting systematic mean reversion of relative prices. Rolling correlations (Figure 2) can be high episodically yet still fall short of cointegration.

6.2 Short-run predictability

Test-set Directional Accuracy between 0.55–0.58 (Table 5) shows a small but non-zero edge focused on direction rather than magnitude. Permutation importance clarifies the source:

- **IT & FMCG: Own-sector RSI** is the primary driver (Tables 6a–6b), consistent with momentum effects. Cross-sector spillovers are minor and occasionally harmful.
- **PHARMA: RSI underperforms** (negative importance), while **volatility** provides limited signal (Table 6c). This aligns with more idiosyncratic, event-driven behavior in pharma.

6.3 Error structure and outliers

Magnitude errors (RMSE/MAE) match the scale of daily volatility (Table 5), which is typical for equities. FMCG outliers inflate dispersion; we retain them for replication, so interpretation emphasizes sign prediction over exact sizing.

6.4 Interpretation boundaries from design choices

RSI and volatility are computed contemporaneously (using information up to and including day *t*), so models represent predictive associations, not deployable trading rules. A strictly ex-ante design would shift RSI/volatility by one day; we would expect a slight drop in directional accuracy with the qualitative patterns intact (momentum in IT/FMCG; volatility—not RSI—in PHARMA).

6.5 Practical implications

I. **Allocation:** Absence of cointegration suggests **no structural tether**; sector allocation can lean into their distinct drivers.

II. **Signals:**

   a. **IT/FMCG: Momentum**-style overlays (RSI) are the primary levers.

   b. **PHARMA:** Favor **volatility-aware** or mean-reversion filters; avoid RSI-based momentum.

III. **Risk:** Robust outlier handling (e.g., winsorizing returns) will stabilize magnitude metrics without changing the directional story.

7. Conclusion

Using daily data from 2015–2025, we first showed that sectoral log-prices are non-stationary while their first differences are stationary, satisfying the preconditions for cointegration testing. Johansen's trace and maximum-eigenvalue statistics did not reject the null of no cointegration among NIFTY IT, NIFTY FMCG, and NIFTY PHARMA, indicating that these three sectors do not share a common long-run equilibrium over the sample. In plain terms, they can drift apart without a structural tether. From an allocation perspective, this supports the case that diversification across IT, FMCG, and PHARMA is not undermined by long-run mean reversion among them.

Turning to short-run behavior, the models achieved Directional Accuracy in the 0.55–0.58 range out-of-sample (IT: 0.55, FMCG: 0.56, PHARMA: 0.58), consistent with a modest but non-zero edge in predicting the sign of daily returns. Errors in magnitude (RMSE, MAE) are of the same order as daily volatility, as expected for equities at a one-day horizon. Permutation importance sharpens this picture: for IT and FMCG, own-sector RSI dominates, pointing to momentum-style signals as the primary short-run driver; for PHARMA, RSI is detrimental and volatility carries the limited positive signal, consistent with more idiosyncratic, event-driven dynamics in that sector.

These results jointly imply: (i) no evidence of a long-run equilibrium across the three sectors; (ii) sector-specific short-run mechanisms, with momentum useful in IT/FMCG but not in PHARMA; and (iii) when the goal is practical predictability at daily horizons, sign accuracy is a more informative headline metric. Finally, we note that FMCG contains extreme vendor outliers retained for reproducibility; interpretation of magnitude errors should therefore be cautious, while the directional story remains robust.

8. Future Works

Future work will focus on (i) robust data handling—winsorizing or capping extreme returns (notably in FMCG) and re-estimating to assess stability of RMSE/MAE and directional accuracy; (ii) strict ex-ante timing—shifting RSI and volatility by one day to evaluate deployable performance; (iii) parsimonious feature expansion—small lag grids (1–5), alternative RSI windows (7/21), and volatility horizons (10/30) to test sector-specific horizons without overfitting; (iv) time-series validation upgrades—rolling/expanding windows and blocked cross-validation to quantify temporal stability; (v) model comparisons—ridge/lasso and gradient boosting versus random forests under identical features, with Diebold–Mariano tests on forecast errors; (vi) regime sensitivity—pre/post structural splits (e.g., pandemic) to see whether cointegration or short-run drivers shift across regimes; (vii) granular linkage checks—pairwise cointegration alongside system tests and simple spillover diagnostics (lead–lag correlations, VAR impulse responses); and (viii) practical overlays and costs—lightweight momentum overlays for IT/FMCG and volatility-gated or mean-reversion filters for PHARMA, evaluated with turnover and transaction-cost sensitivity; as a robustness add-on, feature relevance will be stress-tested with alternative importance frameworks (e.g., corrected permutation, model-class-reliance) to confirm that the sector-specific signal hierarchy persists.

References

1. Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics, 26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134
2. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
3. Brock, W. A., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance, 47*(5), 1731–1764. https://doi.org/10.1111/j.1540-6261.1992.tb04681.x
4. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*(366), 427–431. https://doi.org/10.1080/01621459.1979.10482531
5. Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica, 55*(2), 251–276. https://www.jstor.org/stable/1913236
6. Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research, 20*, 1–81. http://jmlr.org/papers/v20/18-760.html
7. Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*(2–3), 231–254. https://doi.org/10.1016/0165-1889(88)90041-3
8. NSE Indices. (2023). *Methodology document of NIFTY sectoral index series*. https://archives.nseindia.com/content/indices/Method_Nifty_Sectoral.pdf
9. NSE Indices. (2025a). *NIFTY IT—Index factsheet*. https://www.niftyindices.com/Factsheet/ind_nifty_it.pdf
10. NSE Indices. (2025b). *NIFTY PHARMA—Index factsheet*. https://www.niftyindices.com/Factsheet/ind_nifty_pharma.pdf
11. Kumar, S. (2022). A cointegration analysis of Nifty index with sectoral indices of NSE. *Artha Vijnana*, *64*(3), 231–246. https://doi.org/10.1080/02522667.2022.2138203

12. Park, C.-H., & Irwin, S. H. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys*, *21*(4), 786–826. https://doi.org/10.1111/j.1467-6419.2007.00519.x
13. Phillips, P. C. B., & Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica, 58*(1), 165–193. https://www.jstor.org/stable/2938339 JSTOR
14. Shahani, R., & Sharma, A. (2020). Is there a cointegration amongst sector-specific indices of the Indian stock markets? An investigation using ARDL with single structural break. *The IUP Journal of Applied Finance, 26*(3), 28–46. SSRN abstract: https://ssrn.com/abstract=3810543
15. Sullivan, R., Timmermann, A., & White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. *The Journal of Finance, 54*(5), 1647–1691. https://doi.org/10.1111/0022-1082.00163
16. Wilder, J. W. (1978). *New concepts in technical trading systems*. Trend Research. (Original work published 1978).