



Monsoon & Macro Signals to Predict Stock Break-Out

Student: Abhilash Panicker

Mentor: Abhijit Chakraborty



Introduction

- **Problem Statement**
India's monsoon drives GDP, inflation and investor sentiment, yet most equity models ignore weather completely.
Can rainfall & macro surprises improve **next-quarter** forecasts for mid-cap returns?
- **Gap in Literature**
Existing Indian studies focus on index re-balancing or daily weather quirks—none combine **rainfall anomaly** + **GDP/CPI/PMI** + **repo-rate** in a predictive model.
- **Goal**
Build a lightweight, reproducible pipeline—collect public data, test a baseline vs. enriched model, and stress-check results over COVID and rate-hike regimes

Appendix

Term	Meaning
Excess Return	A security's return minus the return of a benchmark (here, NIFTY 50 Total-Return Index). Shows <i>out-performance</i> or <i>under-performance</i> .
Break-Out Quarter	A three-month period in which returns jump far above the benchmark, signaling the start of “blue-chip-like” behaviour.
Sharpe Ratio	Reward-to-risk metric: average excess return ÷ volatility. Higher = better risk-adjusted performance.
Max Drawdown	Largest peak-to-trough fall during the test period; measures worst-case loss.
R² (Coefficient of Determination)	% of the ups-and-downs in returns that the model explains. 0 = no skill; 1 = perfect.
MAE (Mean Absolute Error)	Average size of prediction mistakes, in percentage-points. Lower = more accurate.
F1-Score	Harmonic mean of Precision and Recall in classification; balances false alarms vs. missed signals. Range 0–1.
GDP YoY	Year-over-year percentage change in Gross Domestic Product; high → faster economic growth.
CPI YoY	Year-over-year inflation rate (Consumer Price Index).
PMI (Purchasing Managers' Index)	Survey index (> 50 = expansion, < 50 = contraction) that captures business sentiment.
Repo-Rate Δ	Quarterly change in the RBI's policy lending rate; rate cuts usually stimulate growth.
Monsoon Rainfall Anomaly	% deviation of June-Sep rainfall from the 50-year average; +ve = wetter than normal.
ElasticNet	Linear regression with built-in feature selection; balances Lasso and Ridge penalties.
LightGBM	Fast gradient-boosted decision-tree algorithm—great for small tabular datasets.
Walk-Forward CV	Time-series cross-validation: train on early quarters, test on the next slice, then roll forward.
TreeSHAP	Explainer that shows how much each feature pushes a tree model's prediction up or down.



Scope and objectives- Research Questions

- ☐ Predictive Boost: Does adding rainfall anomaly and macro surprises (GDP YoY, CPI YoY, PMI, repo-rate Δ) let a model beat a simple “lag-return” baseline when forecasting **1-quarter-ahead excess returns** of the NIFTY Midcap 100?
- ☐ Monsoon Effect on Break-Out Quarters: Do quarterly excess returns differ meaningfully between seasons with “good rain” (rainfall anomaly $\geq +4\%$) and “poor rain” ($\leq -4\%$)?
- ☐ Robustness: Does the enriched model from RQ-A keep most of its predictive power during market shocks—the COVID crash (2020 Q1-Q2) and the rate-hike cycle (2022 Q2-2023 Q4)?
- ☐ Lead–Lag Channel: Does monsoon rainfall anomaly **lead** next-quarter GDP growth, and does using that “rain \rightarrow GDP” link further improve equity-return forecasts?



Scope and objectives (Sample size calculation)

- **Observations:** 2010 Q1 \rightarrow 2025 Q2 \Rightarrow 62 quarters.
- **Regression power check** ($\alpha = 0.05$, desired power = 0.80, expected $f^2 \approx 0.15$): needs 55 points \rightarrow *we have 62 \Rightarrow sufficient.*
- **Good-vs-poor monsoon t-test:** rule-of-thumb ≥ 30 observations per group; we have 34 “good” & 28 “poor”.

*Time-series methods rely more on **rolling splits** than sheer N , but we comfortably exceed minimums.*



Data Description

Layer	Source	Key Fields
Index price	NSE bhavcopy (daily) → quarterly TR	Close, splits
Rainfall anomaly	IMD Southwest-Monsoon report (annual XLS)	% vs 10-yr mean
GDP YoY	MOSPI / RBI Handbook CSV	q/q annualised %
CPI YoY	RBI time-series	monthly % (rolled to Q)
PMI	IHS-Markit press releases	index (≥ 50 = expansion)
Repo-rate Δ	RBI DB	change in bps during Q

All series merged to **quarter-end**, stored in a single CSV ($\approx 62 \times 7$).



Analytic approach (Overview)

- ETL – read CSV/XLS, resample to Q, forward-fill rainfall.
- **Feature set**
 - Lagged index return (baseline)
 - Lagged GDP, CPI, PMI, repo-rate, rainfall anomaly (enriched)
- **Models**
 - **Baseline:** ElasticNet on lag return only
 - **Enriched:** LightGBM on all features
- **Validation:** 5-split walk-forward (train : test windows $\approx 10 : 2$ Q).
- **Tests**
 - Question 1: R^2 , MAE comparison + SHAP bar chart
 - Question 2: two-sample t & KS on good vs poor monsoon returns
 - Question 3: run fixed model through crisis quarters; check $\Delta R^2 / \Delta \text{Sharpe} < 15\%$.



Analytic approach (Question 1)

Research question 1: Do rain + macro lags beat a lag-return baseline for predicting a break-out quarter (excess return surge vs. Nifty-50)?

- **Target:** next-quarter excess return (NIFTY Midcap 100 TR – NIFTY 50 TR).
- **Baseline vs Enriched:** ElasticNet vs LightGBM.
- **Success Rule:** Enriched model raises out-of-sample R^2 by ≥ 0.10 *and* cuts MAE ≥ 10 %.



Analytic approach (Question 2)

Research question 2: Do returns differ between “good rain” and “poor rain” seasons?

- **Define seasons**
 - *Good rain*: rainfall anomaly $\geq +4$ % (IMD definition “Above Normal”)
 - *Poor rain*: ≤ -4 % (“Below Normal”)
- **Test**: two-sample t and KS on the distribution of same-quarter excess returns.
- **Significance**: $|\Delta \text{ mean}| > 0.5 \sigma$ and $p < 0.05 \rightarrow$ rainfall matters.



Analytic approach (Question 3)

Research question 3: Does the enriched model keep most of its skill during COVID & rate-hike shocks?

- **Shock windows**
 - COVID crash: 2020 Q1–Q2
 - Rate hikes: 2022 Q2–2023 Q4
- **Metrics:** ΔR^2 , ΔSharpe , max drawdown vs full sample.
- **Pass mark:** degradation < 15 %.



Analytic approach (Question 4)

Research question 4: Does rain lead GDP, and does that extra feature help forecasts?

- **Step 1 Econometrics**
 - OLS: $GDP_{t+1} = \alpha + \beta \cdot Rain_t$ and same for $t+2$.
 - Significance: β $p < 0.05$, $R^2 \geq 0.08$.
- **Step 2 Pipeline Boost**
 - Add `gdp_pred_from_rain` to feature list.
 - Re-run walk-forward; compare ΔR^2 & ΔMAE .
- **Pass mark** – Additional $R^2 \geq 0.03$ **or** $MAE \downarrow \geq 5\%$ over RQ-A enriched model.



Evaluation Matrix

Metric	Formula (text)	Used in
R ²	$1 - \Sigma(\hat{y}-y)^2 / \Sigma(y-\bar{y})^2$	A, C
MAE	Σ	$\hat{y}-y$
Sharpe	$(\mu_p - r_f) / \sigma_p$	C
KS Statistic	max	$F_{\text{good}}(x) - F_{\text{poor}}(x)$
t-statistic	$(\mu_{\text{good}} - \mu_{\text{poor}}) / SE$	B



Recommendation and applications

- **Fund-house mid-cap desks** – use rainfall & macro watch-list as an overlay on existing factor models.
- **Sell-side research** – quarterly “Monsoon Tracker” note becomes a value-add service.
- **Policy analysts** – quantify how repo-rate actions mediate weather shocks on equity risk premium and other capital market components.
- **Open notebook** – fast, transparent code for quants to extend.



References and bibliography

- *IMD. End-of-Season Southwest Monsoon Report (2001-2024).*
- *RBI. Handbook of Statistics on Indian Economy, 2024.*
- *MOSPI. National Accounts, Quarterly GDP press releases, 2010-2025.*
- *IHS-Markit. India Manufacturing PMI Methodology, 2024.*
- *Fama & French. "Common Risk Factors," J. Finance, 1993.*
- *Tsay, R. Analysis of Financial Time Series, 2023.*
- *Wright Research Blog. "Does Monsoon Impact Nifty?" 2024.*
- *StockEdge Blog. "Monsoon & Market Returns," 2025.*
- *Bansal et al. "Predicting Nifty Upgrades," IIM W.P., 2022.*
- *Lalwani & Meshram. "Machine-Learning Indian Cross-Section," SSRN 4511996, 2023.*