POLITECNICO MILANO 1863

Nifty50 forecasting and trading strategies

Barbieri Claudio, Delera Giacomo, Huser Cyril, Howe Alessandro

Econometrics project 2024/2025

Lecturer: Azzone Michele

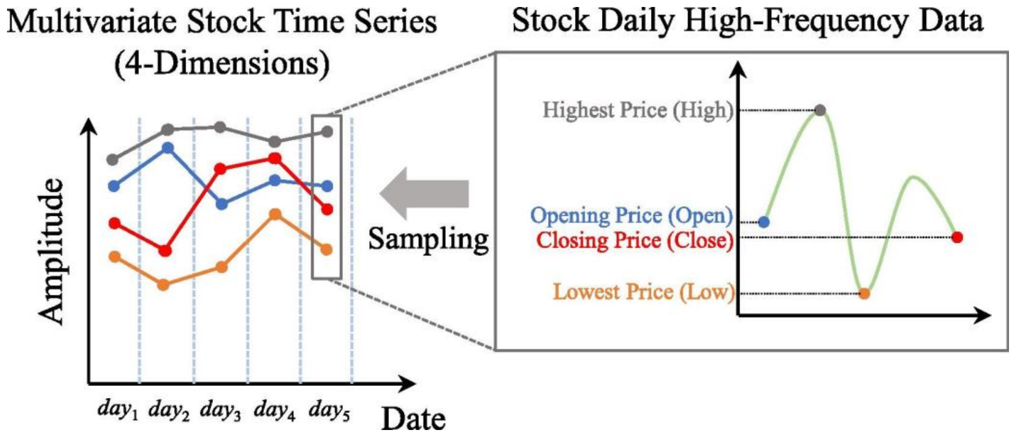
TA: Sala Edoardo

Data

High-frequency (minute-level) data of the Indian stock market from 2017 to 2021

- Indexes
- Stocks

- Select **Nifty50** index data from 2017,2018, and 2019 to avoid the COVID-19 pandemic
- Aggregate into *daily* **O**pen **H**igh **L**ow **C**lose prices to reduce noise level



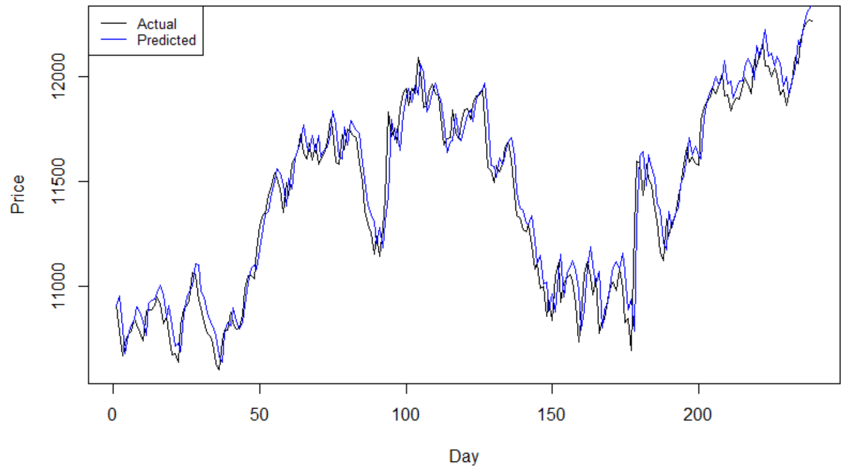
Category	Indicators
Trend-following	MACD (Moving Average Convergence/Divergence) DEMA (Double Exponential Moving Average) CP (Close Price)
Momentum	ROC (Rate of Change) RSI (Relative Strength Index) ULTOSC (Ultimate Oscillator) WR (Williams %R) CCI (Commodity Channel Index)
Volatility	ATR (Average True Range) VOL (Volatility)
Volume-based	OBV (On Balance Volume)

- Extract a proxy for the index **V**olume from the constituent underlying stocks
- Compute some **technical indicators** to use as features

Linear Model

Predict the **close price** of Nifty50 using technical indicators with a linear regression model

- Variable selection: backward/forward minimizing AIC
- Training set: 2017, 2018
- Test set: 2019



$$CLOSE_{t+1} = \beta_0 + \beta_1 CLOSE_t + \beta_2 ULTISC_t + \beta_3 ATR_t + \beta_4 OBV_t + \beta_5 WR_t$$

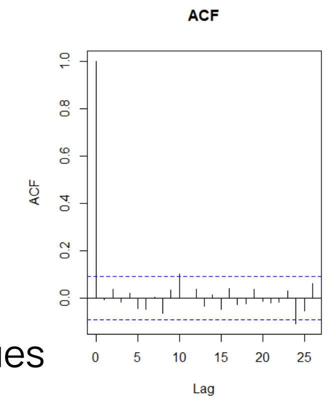
Adj R-squared	DW test	BP test	JB test	RMSE
0.9881	p-value = 0.4736	p-value $<$ 0.01	p-value $<$ 0.01	106.6839

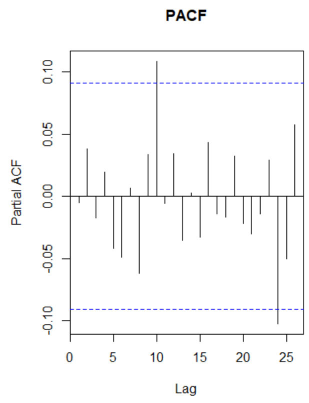
No interest in the statistical reliability here, just focus on the forecast power

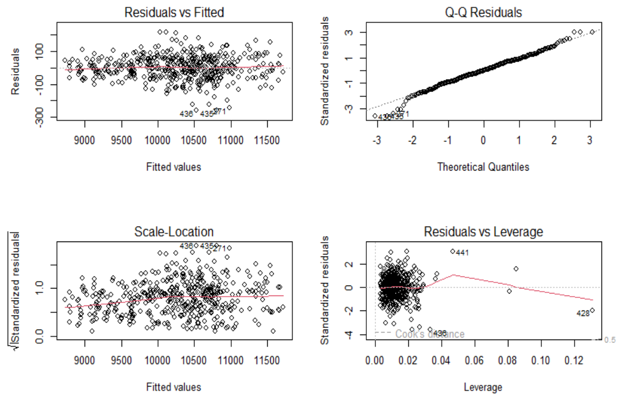
- Spurious high R^2 from non-stationary features
- Heteroskedastic residuals
- Non-normal error distribution

Diagnostics for guiding future modeling

- No autocorrelation detected after including lagged values







Trading strategies

Absolute Change Threshold Strategy

Objective: Take a position only when the predicted change in log returns is significant.

Threshold-Based Switching:

- Compute change in predictions: $\Delta pred_t = \left|\frac{\hat{r}_t - \widehat{r_{t-1}}}{\widehat{r_{t-1}}}\right|$
- Enter or switch position only if $\Delta pred_t \geq$ threshold and the sign of the return is the opposite.

Position Logic:
Long if prediction > 0, Short if < 0, otherwise Neutral.
Hold previous position if change is small.

Transaction Cost Handling:
When a position change occurs.

Volatility-Scaled Strategy

Objective: Adjust position size dynamically based on predicted log return and market volatility, aiming to maintain consistent risk exposure.

Volatility-Based Sizing:
- Estimate rolling volatility from past true returns: $vol_t = std([r_{t-n}, \dots, r_{t-1}])$
- Compute position size: $position_t = scale \cdot \frac{\hat{r}_t}{vol_t + \epsilon}$
- Clip positions within $[-max\ leverage, +max\ leverage]$

Position Logic:
- Take proportionally larger positions when predicted return is high and volatility is low.
- Reduce exposure when volatility rises or signal weakens.
- No fixed thresholds — position is continuous and dynamically adjusted.

Transaction Cost Handling:
When a position change occurs.

Our choice: **Fixed Threshold Sign Strategy**

Objective: Daily trading strategy, taking long or short positions based on the predicted log return. While not the most advanced approach, it best allowed us to evaluate the index forecast — our primary objective.

Signal-Based Positioning:
Long if prediction > threshold, Short if < -threshold, Neutral otherwise.

2 Variants:

1. **Log(Close / Previous Close)** measures the **daily return**, encompassing both overnight and intra-day price changes between two consecutive closes.
2. **Log(Close / Open)** captures the **intra-day return**, reflecting the price movement from the market open to close within the same trading day.

Even if the 2nd is the "classical" definition of log return, strategies based on **Log(Close/Open)** often provide **more consistent and reliable signals** for short-term trading.

Transaction Cost Handling:

1. When a position change occurs.
2. Exiting and entering the market costs.

Result on Linear Model:

Intraday trading

Annualized Return: -31.34%
Sharpe Ratio: -3.00
Max Drawdown: 30.79%
Sortino Ratio: -4.98

Whole day trading

Annualized Return: 17.79%
Sharpe Ratio: 1.24
Max Drawdown: 11.38%
Sortino Ratio: 2.40

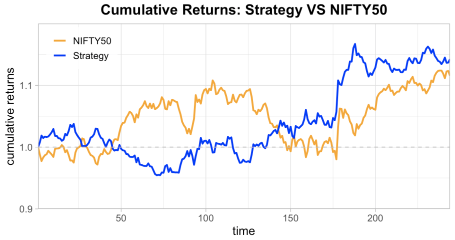
Remark: strategies with frequent trades suffer cost accumulation.

Zero transaction cost

Annualized Return: 44.22%
Sharpe Ratio: 3.13
Max Drawdown: 4.79%
Sortino Ratio: 6.33

Transaction costs

5 bps transaction cost

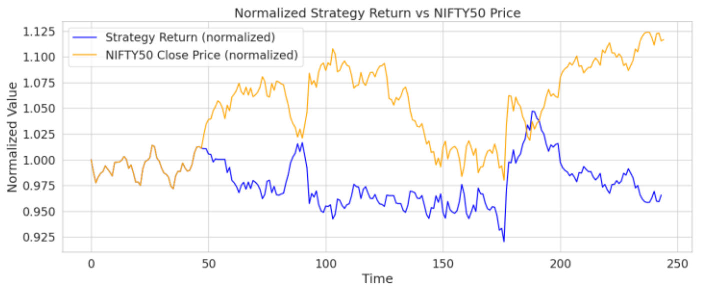


Annualized Return: 14.73%
Sharpe Ratio: 1.21
Max Drawdown: 8.00%
Sortino Ratio: 2.28

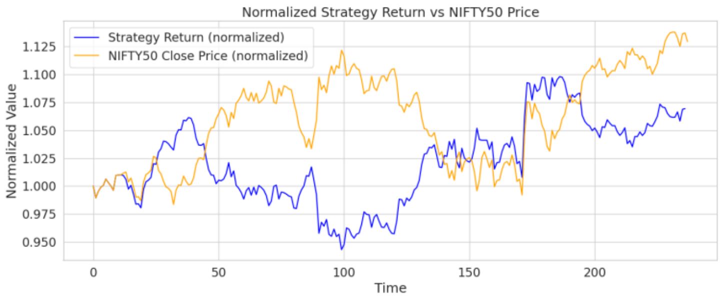
(Conv-)LSTM

- Leverage LSTM's sequence "memory" plus local feature extraction via 1D convolutions, common in financial forecasting literature
- Input tensor: N x W x F
  N: total samples (sliding windows)
  W: window size (days of history)
  F: features per day
- Sliding-window encoding
  stride 1 (one-day step)
  many-to-one (single output)
- Feature scaling: Min-Max scaling
- Option 1: NIFTY50 (prior-day log-return + prior-day volume + technical indicators)
  Option 2: NIFTY50 + SP500 (prior-day NIFTY50 and SP500 log-return + prior-day NIFTY50 and SP500 volume + NIFTY50 technical indicators

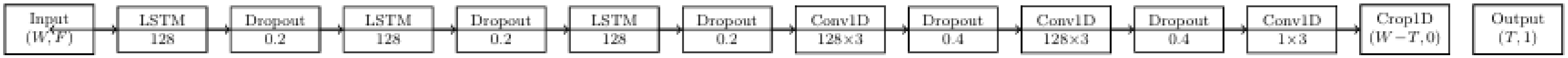
Whole-day trading



Annualized Return: -3.12% **Max Drawdown**: 9.49%
Sharpe Ratio: -0.16 **Sortino Ratio**: -0.28

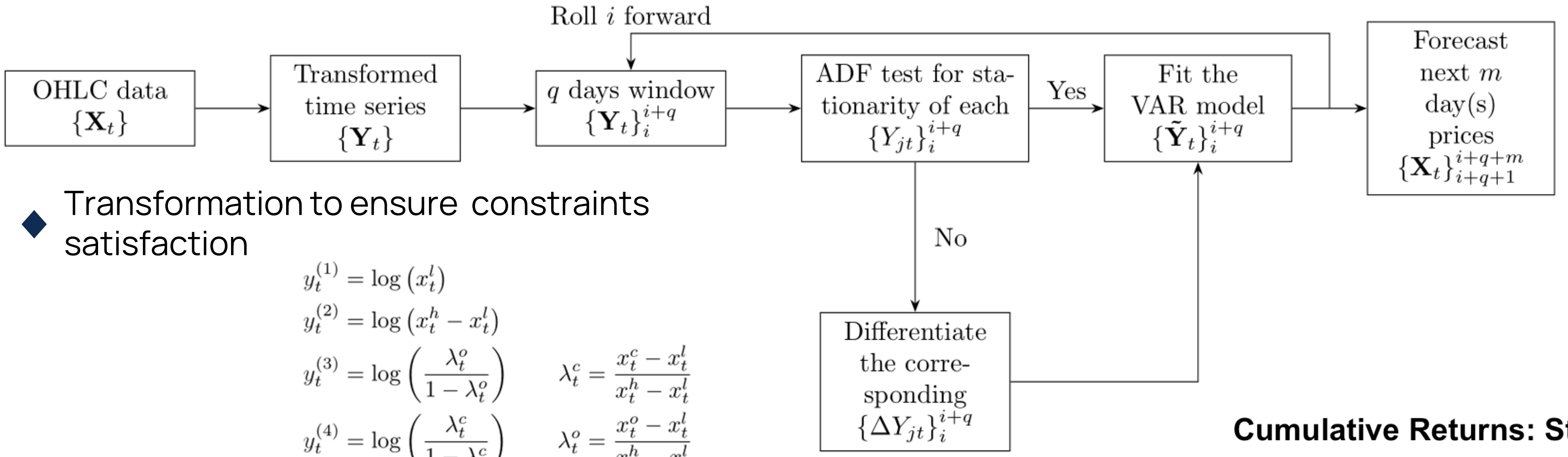


Annualized Return: 5.95% **Max Drawdown**: 11.13%
Sharpe Ratio: 0.49 **Sortino Ratio**: 0.78



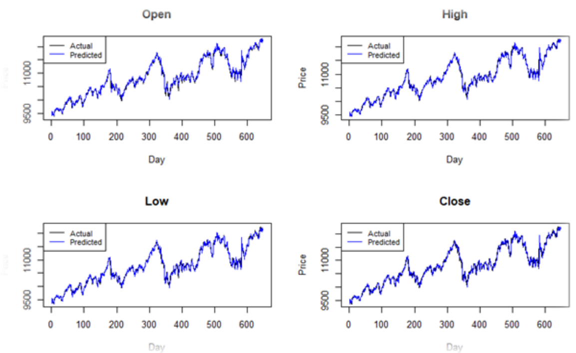
VAR on OHLC

Enable mutual influence among OHLC variables via dynamic multivariate model: **VAR**



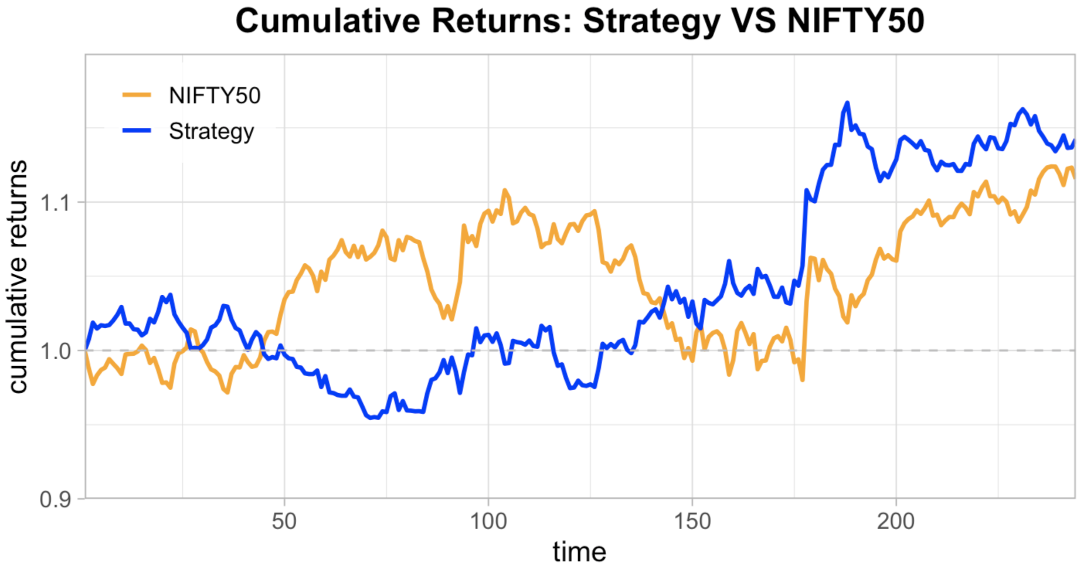
- Transformation to ensure constraints satisfaction

$$y_t^{(1)} = \log\left(x_t^l\right)$$
$$y_t^{(2)} = \log\left(x_t^h - x_t^l\right)$$
$$y_t^{(3)} = \log\left(\frac{\lambda_t^o}{1-\lambda_t^o}\right) \qquad \lambda_t^c = \frac{x_t^c - x_t^l}{x_t^h - x_t^l}$$
$$y_t^{(4)} = \log\left(\frac{\lambda_t^c}{1-\lambda_t^c}\right) \qquad \lambda_t^o = \frac{x_t^o - x_t^l}{x_t^h - x_t^l}$$



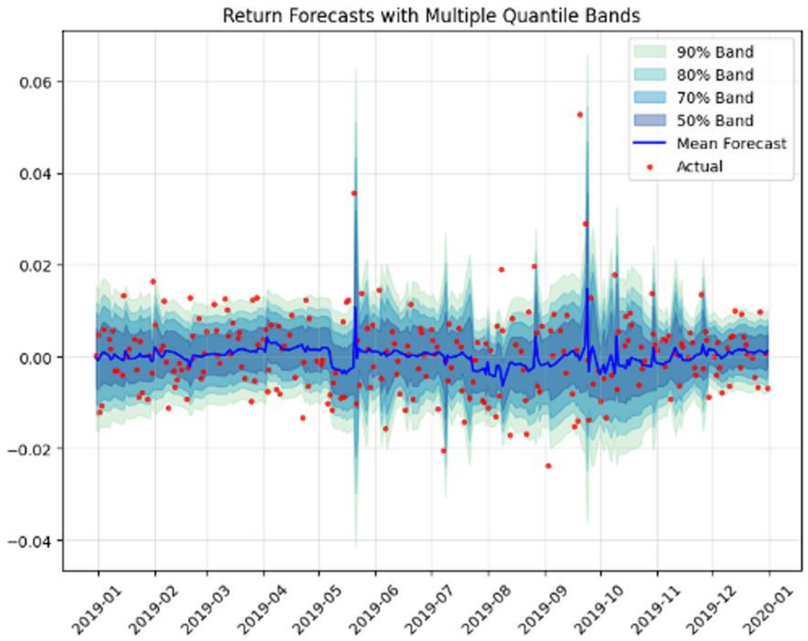
Best result on intraday with fixed threshold strategy

Ann. Return	Sharpe	Max DD	Sortino
14.35%	1.17	9.33%	2.28



GARCH

Forecast the **conditional quantiles** of Nifty50 closing prices by modeling volatility.



GARCH-M :

$$r_t = \mu + \lambda\sigma_t + \varepsilon_t,$$

$\varepsilon_t \sim \mathcal{N}(0, \sigma_t^2),$ **Strong assumption**: Normality of residuals

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2,$$

$$\mathrm{r}_t^{(q)} = E[r_t \mid \mathcal{F}_{t-1}] + z_q\sigma_t = \mu + \lambda\sigma_t + z_q\sigma_t, \qquad z_q = \Phi^{-1}(q),$$

Fit the model using a **50-day rolling window** to generate a 1-day-ahead forecast at each step

Results :

Backtesting the model reveals that the quantiles deviate by about 5%. The results become more consistent when applying shrinkage.

Confidence Band	Return Coverage	Return Winkler
90% (0.050–0.950)	86.99%	0.0395
80% (0.100–0.900)	75.20%	0.0324

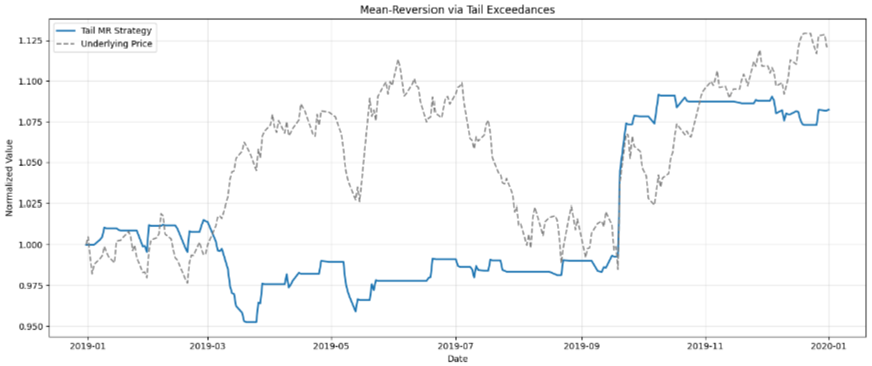
Probabilistic trading strategies

Tail Exceedance Mean Reversion Strategy

- Builds a signal from the frequency of returns breaching fixed quantiles over a 5-day rolling window.
- Uses the 25th and 75th percentiles as thresholds.
- Frequent downside breaches → signal a long position (overreaction).
- Frequent upside breaches → signal a short position (overextension).

Delivered solid performance with **8.42% annual return** and a Sharpe of 0.99.

Annual Return	Sharpe Ratio	Max Drawdown	Sortino Ratio
8.42%	0.99	6.11%	1.57



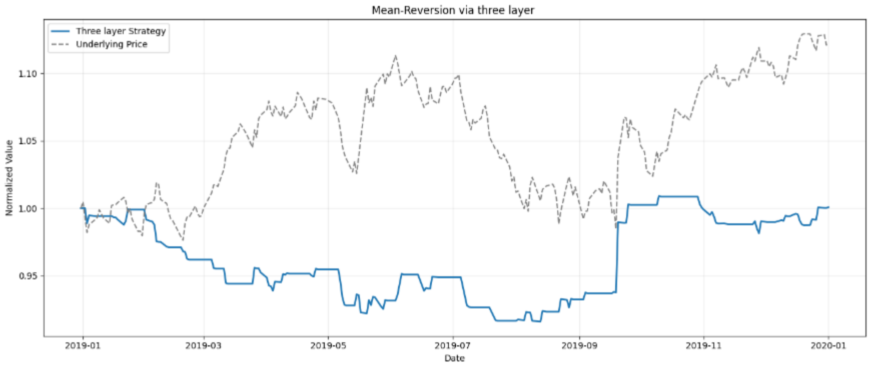
(a) Tail exceedance strategy.

Three-Layer Tail Mean Reversion Strategy

- Utilizes three quantile bands:
    - Outer: 10th–90th percentile
    - Middle: 15th–85th percentile
    - Inner: 25th–75th percentile
- Positions start when returns cross outer bands.
- Positions held while returns stay within middle or inner bands.
- Exit triggered by:
    - Crossing inner band favorably (profit)
    - Reversion within inner band (trend exhaustion)
    - Breaching opposite tail (regime shift)

Underperformed, showing near-zero return and weak risk-adjusted metrics.

Annual Return	Sharpe Ratio	Max Drawdown	Sortino Ratio
-0.24%	0.01	8.19%	0.14



(b) Three layer strategy.

Conclusion

Wrapping up:

- **Linear model** using technical indicators offers good prediction for *whole-day* trading
- **LSTM** with *log return* cannot generate profit with the used strategy (better with S&P500 information)
- **VAR** on the OHLC exhibits strong performance when targeting *intraday* returns
- **GARCH** leverages *probabilistic* strategies to achieve positive results

References

[1] Zahra Fathali, Zahra Kodia Aouina, and Lamjed Ben Said. "Stock Market Prediction of NIFTY 50 Index Applying Machine Learning Techniques". In: Applied Artificial Intelligence 36 (2022). doi: 10. 1080/08839514.2022.2111134. url: https://www.researchgate.net/publication/363632317_ Stock_ Market_ Prediction_ of_ NIFTY_ 50_ Index_ Applying_ Machine_ Learning_ Techniques.

[2] Gang Ji et al. "An adaptive feature selection schema using improved technical indicators for predicting stock price movements". In: Expert Systems with Applications 200 (2022), p. 116941. issn: 0957- 4174. doi: https://doi.org/10.1016/j.eswa.2022.116941. url: https://www.sciencedirect. com/science/article/pii/S0957417422003712.

[3] Fatemeh Moodi, Amir Jahangard-Rafsanjani, and Sajad Zarifzadeh. Feature selection and regression methods for stock price prediction using technical indicators. 2023. arXiv: 2310.09903 [q-fin.ST]. url: https://arxiv.org/abs/2310.09903.

[4] Huiwen Wang, Wenyang Huang, and Shanshan Wang. Forecasting open-high-low-close data contained in candlestick chart. 2021. arXiv: 2104.00581 [econ.EM]. url: https://arxiv.org/abs/ 2104.00581.

Thank you for the attention!