Technical and behavioral analysis of ETF groups

Angad Bhatti

August 11, 2025

Introduction and Motivation

Exchange-traded funds (ETFs) are baskets of securities that track a particular market index or sector. Sector ETFs, such as those tracking technology, finance or real estate, allow investors to gain focused exposure to industries without buying individual stocks. Because the composition and behaviour of each sector can differ dramatically, trading strategies that work in one sector may fail in another. The aim of this internship project was to evaluate whether widely used technical analysis indicators can be combined into a systematic strategy for multiple sector ETFs and to understand the behavioural characteristics of each sector. Technical indicators provide computed measures from historical price and volume data that purport to capture momentum, trend strength and price extremes. Examples include the relative strength index (RSI), moving-average convergence divergence (MACD), Bollinger bands and rolling volatility. Traders frequently use these indicators to identify overbought or oversold conditions and to time entry and exit points.

Each indicator has its own interpretation. RSI measures the magnitude of recent price changes; readings above 70 suggest overbought conditions while readings below 30 suggest oversold. MACD is a trend-following momentum indicator computed as the difference between a short– and a long–term exponential moving average (EMA); a nine-period EMA of this difference forms the signal line and crossovers between the MACD line and the signal line are used to generate buy or sell signals. Bollinger bands consist of a moving average with upper and lower bands set at a fixed number of standard deviations; when price moves near the upper band the asset is considered overbought, whereas movement near the lower band signals an oversold condition. Volatility, measured as the standard deviation of returns over a rolling window, provides a sense of market turbulence and is frequently used to adjust position sizing or to detect risk regimes.

Although technical indicators are simple to compute, applying them in a predictive time–series model is challenging. Indicators are functions of past prices and therefore exhibit strong autocorrelation. This leads to late identification of changes in market trends and heavy drawdowns if not handled carefully. Moreover, raw price series are typically non-stationary, and many statistical methods require stationary inputs. Thus, a careful workflow is required: compute indicators in a rolling manner using only past data, transform features to achieve stationarity when necessary, and evaluate strategies on out-of-sample data via walk-forward testing.

Data and Preprocessing

Data Collection

The study used daily closing prices from January 2011 through mid-2023 for nine widely traded ETFs: SPY (broad U.S. equity market), QQQ (technology), SOXX (semiconductor index), SMH (semiconductor manufacturers), FDN (internet/technology), XLF (financials), XLV (healthcare), XLRE (real estate) and GDX (gold miners). These tickers span a range of market sectors and volatility regimes. Prices were downloaded from freely available sources and aligned to a common calendar. Because technical indicators require previous data points for their computation, the first few observations were dropped after indicator calculation.

Indicator Computation and Stationarity

For each ETF the following indicators were computed using rolling windows to avoid lookahead bias:

- **Relative Strength Index (RSI).** RSI is computed from the differences between consecutive closing prices. Average gains and average losses over a specified window are calculated using exponential weighting, and the ratio of these averages is converted to a 0–100 scale. Low RSI values indicate oversold conditions and high values indicate overbought conditions.

- **Moving-Average Convergence Divergence (MACD).** Two EMAs of the close are computed (fast and slow). The MACD line is the difference between the fast and slow EMAs, and a second EMA of the MACD line serves as the signal line. Crossovers of these lines are used as momentum signals.

- **Bollinger Band Z-Score.** A simple moving average (SMA) and standard deviation are computed over a window. The distance of the current price from its moving average, normalised by the standard deviation, produces a z-score. Values beyond ± 1 standard deviation flag potential extreme prices.

- **Rolling Volatility.** The standard deviation of percentage returns over a recent window provides a measure of volatility. High values correspond to turbulent regimes.

All indicator calculations were implemented in a modular fashion and housed in the code under `src/tech_ts/indicators/indicators.py`. To ensure stationarity of the inputs, stationarity tests (Augmented Dickey–Fuller and KPSS) were applied to the indicators. Some indicators, such as volatility, required log-differencing to achieve stationarity; these transformations were applied by the function `force_stationary`.

Avoiding Lookahead Bias

It is easy to introduce lookahead bias when technical indicators are used in models. In this project indicators were computed using only historical data: for each date t the indicator value is based on prices up to t. Signal generation functions operate on these indicator series and then produce positions for the next day. Walk-forward validation further ensures that each model training window uses only past data and that performance is assessed on unseen data.

Methodology

Exploratory Analysis and Parameter Tuning

The first phase involved analysing each indicator in isolation. Rolling correlation analyses were performed to assess how well each indicator predicted next-day returns for different sectors. Parameter grids were defined for each indicator (for example, RSI window lengths of 10, 14 and 20; MACD fast EMAs of 10 or 12 and slow EMAs of 26 or 30). A rolling walk-forward tuning function measured the correlation between indicator signals and subsequent returns across multiple training and testing slices. The aim was to select parameters that maximised predictive correlation without overfitting. This analysis revealed that no single indicator consistently predicted returns across all sectors. Optimal parameters varied by ETF and over time. Consequently, combining multiple indicators was hypothesised to provide more robust signals.

Rule-Based Signal Generation

Simple deterministic rules were used to convert indicator values into trading signals:

- **RSI signals.** Go long when RSI falls below a lower threshold (e.g., 30) and go short when RSI rises above an upper threshold (e.g., 70). Positions are neutral when RSI lies between the thresholds.
- **MACD crossover signals.** Go long when the MACD line crosses above its signal line and go short when it crosses below.
- **Bollinger band signals.** Take a long position when the z-score is below a lower threshold (e.g., –1) and a short position when it is above an upper threshold (e.g., +1).
- **Volatility regime signals.** Classify the market into high- and low-volatility regimes by comparing rolling volatility to its median. In high volatility regimes the strategy may reduce position size or abstain from trades.

Individual indicator signals were then combined using a weighted voting scheme. Equal weights were used by default, but alternative weights (learned or regime-dependent) were explored. The sign of the weighted sum determined the final position (+1 for long, –1 for short, 0 for neutral).

Logistic Regression Meta-Indicator

Given the limited predictive power of single indicators, a data-driven approach was adopted to learn an optimal combination of features. For each ETF a feature vector was constructed comprising standardised technical indicators (RSI, MACD, Bollinger z-score and rolling volatilities) and external market indicators such as the CBOE Volatility Index (VIX) and the ratio of VIX to its three-month counterpart (VXV). A forward return over a horizon matching the median holding period was used as the target variable. The median holding period was estimated from how frequently the simple rule-based signals changed sign and was typically four to five trading days.

The predictive model chosen was logistic regression. Logistic regression is a fundamental classification technique that belongs to the family of linear classifiers. It models the log-odds (logit) of the probability of an outcome as a linear combination of the input variables and uses the sigmoid function to map this linear combination to a probability between 0 and 1. Although developed for binary classification, logistic regression can be extended to multiclass problems and is valued for its simplicity and interpretability. In this project the model estimated the probability that the forward return over the holding horizon was positive. Polynomial feature expansion of degree 2 captured interactions between indicators. A rolling walk-forward scheme retrained the model at each step using all available past data, allowing the coefficients to adapt to evolving market dynamics. At each date t the model output a probability P_t. A long position was taken when $P_t \geq 0.65$, a short position when $P_t \leq 0.35$, and no position otherwise. A modest transaction cost of 0.1% per trade was subtracted to simulate slippage and fees.

Regime Detection with Hidden Markov Models

Regime detection formed a core component of the final strategy. A hidden Markov model (HMM) is a probabilistic framework in which the system is assumed to switch among a finite number of unobservable (hidden) states, each associated with a distinct probability distribution for the observed data. HMMs are widely used to detect patterns or structures in sequential data. In this project an HMM with three states was fit to a multivariate feature set comprising daily percentage returns, short-term and long-term realised volatilities (e.g., 5-day and 21-day rolling standard deviations), changes in the CBOE VIX index and the VIX term-structure ratio. The fitting procedure uses multiple random initialisations and selects the model that maximises the log-likelihood.

Once trained on data up to the current window, the HMM produces two outputs for each date: (1) the most likely sequence of hidden regimes and (2) the posterior probabilities for each regime. To translate these regimes into trading signals, the historical performance of each indicator signal is analysed separately within each regime. Specifically, for each regime r and each indicator we compute the Spearman correlation (information coefficient) between the indicator and the forward return over the estimated holding horizon. A ridge-regularised regression then estimates a weight vector w_r that assigns higher weights to indicators with strong positive correlation and lower (or negative) weights to those with weak or opposite correlation. These weights are normalised to sum to one in absolute value. When the HMM identifies regime r at time t, the regime-weighted signal is computed as the dot product of the indicator signal vector with w_r and then thresholded to produce a position in {-1,0,1}.

Because the HMM-based position reflects only the historical efficacy of indicators under each regime, it is combined with the logistic regression signal described above to form a more balanced view. The base position at time t is defined as

$$\text{base\_pos}_t = 0.5\,\text{regime\_pos}_t + 0.5\,\text{logit\_pos}_t,$$

where regime_pos_t is the thresholded regime-weighted signal and logit_pos_t is the position implied by the logistic classifier. Averaging the two ensures that the model respects both macro-level regime dynamics and micro-level predictive features. This blended signal is then subjected to volatility-based position sizing as described in the next subsection.

Random Forest Regime Classifier (Unused Alternative)

As another alternative, a random forest classifier was trained to classify market regimes directly from the indicator features. Random forests are ensemble models that build multiple decision trees and average their predictions. The classifier attempted to predict whether the next day's return would be positive or negative. While the model achieved high accuracy on in-sample data, it generalised poorly out-of-sample, suggesting overfitting. Moreover, the random forest provided little interpretability and increased computational overhead. Consequently, this approach was not incorporated into the final strategy.

Volatility-Based Position Sizing

The blended base position described above still treats each trade identically regardless of market volatility. To adapt the strategy's exposure to changing risk conditions, a Carver-style volatility scaling overlay was applied. A volatility scaler maintains two exponentially weighted moving average (EWMA) estimates of realised variance: a fast variance σ^2_{fast} with decay parameter λ_{fast} and a slow variance σ^2_{slow} with decay λ_{slow}. The corresponding annualised volatilities σ_{fast} and σ_{slow} are combined into a target volatility $\sigma_{\text{target}} = w_{\text{fast}}\,\sigma_{\text{fast}} + w_{\text{slow}}\,\sigma_{\text{slow}}$, where w_{fast} and w_{slow} are user-defined weights. The leverage applied to the base position is then

$$L_t = \min\Big(\frac{\sigma_{\text{target},t}}{\max(\sigma_{\text{fast},t},\ \text{vol\_floor})},\ L_{\text{max}}\Big),$$

where vol_floor prevents division by very small volatilities and L_{max} caps the maximum leverage (e.g., 3×). The volatility-scaled position is simply L_t multiplied by the base position. Optionally, the leverage series can be smoothed with a small exponential moving average to avoid abrupt changes. When volatility is high, σ_{fast} rises and the leverage L_t decreases, resulting in smaller positions; conversely, in calm markets L_t increases to allow larger exposures. This overlay thus acts as a dynamic risk management tool, reducing drawdowns during turbulent periods and amplifying returns when volatility is subdued.

Backtesting Framework and Performance Metrics

Walk-Forward Backtesting

All strategies were evaluated using a walk-forward backtesting framework. The price series was divided into rolling training and testing windows. For example, a 252-day (one-year) training window and a seven-day testing window were rolled forward in weekly steps. In each iteration indicators were computed, parameters were tuned, models were trained on the training window and evaluated on the testing window. This process simulates how strategies would perform in real time and avoids contamination of training data with future information.

Performance Metrics

For each ETF the backtest produced daily strategy returns. From these returns several performance metrics were computed:

- **Compound Annual Growth Rate (CAGR).** The annualised growth rate of the equity curve.
- **Annualised Volatility.** Standard deviation of daily returns multiplied by the square root of 252.
- **Sharpe Ratio.** Ratio of the mean return to the standard deviation, annualised, assuming a zero risk-free rate.
- **Maximum Drawdown.** Worst peak-to-trough decline of the equity curve.
- **Total Trades.** Number of times the position changed sign (long to short or vice versa).
- **Hit Rate.** Proportion of trades that were profitable (positive return on a trade).
- **Average Holding Period.** Mean duration (in trading days) that positions were held.

Results

Table 1 summarises the performance of the logistic regression strategy for each ETF. The metrics were computed over the entire sample period with a transaction cost of 0.1% per trade. To comply with the report's formatting requirements, the metrics are divided into three narrower tables.

The results reveal varied performance across sectors. The strategy delivered positive CAGRs in most ETFs, with the strongest gains in GDX (7.01% CAGR and a Sharpe ratio of 0.53) and in the semiconductor fund SMH (4.44% CAGR, Sharpe 0.37). Financials (XLF) and the broader semiconductor index (SOXX) produced moderate returns around 2–3% per year with Sharpe ratios near 0.23–0.27. The broad market SPY and technology-heavy QQQ achieved modest positive returns of 2.24% and 0.43%, respectively, with low Sharpe

Table 1: CAGR and Sharpe ratio for each ETF.

ETF	CAGR	Sharpe Ratio
SPY	2.24%	0.25
QQQ	0.43%	0.10
SOXX	2.42%	0.23
SMH	4.44%	0.37
FDN	-3.39%	-0.16
XLF	2.69%	0.27
XLV	1.94%	0.27
XLRE	0.25%	0.08
GDX	7.01%	0.53

Table 2: Volatility and maximum drawdown for each ETF.

ETF	Volatility	Max Drawdown
SPY	11.14%	-28.93%
QQQ	11.87%	-37.75%
SOXX	16.38%	-36.69%
SMH	14.86%	-41.02%
FDN	14.74%	-67.98%
XLF	12.62%	-28.94%
XLV	8.37%	-19.19%
XLRE	10.25%	-25.54%
GDX	14.79%	-25.01%

ratios. Healthcare (XLV) and real estate (XLRE) generated small but positive gains (1.94% and 0.25%). Only FDN recorded a loss, with a –3.39% CAGR and negative Sharpe. Hit rates clustered around 50–52% across sectors, and average holding periods were about five trading days. Trade counts ranged from roughly 905 (XLRE) to 1,819 (SPY). Despite volatility scaling, some sectors still experienced substantial drawdowns, particularly FDN, underscoring the need for further risk management.

Discussion

Several insights emerge from the analysis:

- **Indicator effectiveness varies by sector.** Momentum-driven sectors like technology and financials benefited most from the strategy, while mean-reverting or commodity-driven sectors underperformed. This suggests that model parameters and feature sets should be tailored to each sector rather than applied uniformly.

- **Importance of regime awareness.** Large drawdowns in some ETFs highlight the need to recognise when market conditions are unfavourable for trend-following strate-

Table 3: Trading activity, hit rate and average holding period.

ETF	Total Trades	Hit Rate	Avg Holding Period
SPY	1,819	51.7%	5.40
QQQ	1,804	51.7%	5.34
SOXX	1,668	51.5%	5.28
SMH	1,534	52.2%	5.21
FDN	1,729	52.2%	5.35
XLF	1,534	52.1%	5.14
XLV	1,541	51.5%	5.13
XLRE	905	51.9%	5.20
GDX	1,255	50.9%	5.17

gies. Incorporating explicit regime detection or volatility filters may improve risk management.

- **Trade–off between complexity and robustness.** Simple rule-based signals were easy to interpret but lacked predictive power. The logistic regression meta-indicator improved performance by combining multiple features, yet it still struggled in certain regimes. More complex models (e.g., random forests) risk overfitting and reduced interpretability. A balance between model complexity and robustness is essential.

Conclusion and Future Work

This project developed and evaluated a technical-indicator-based trading strategy for multiple sector ETFs. After exploring individual indicators and simple rule-based signals, a logistic regression meta-indicator was implemented to combine RSI, MACD, Bollinger band z-scores and volatility measures. A walk-forward backtesting framework with transaction costs provided realistic performance estimates. The strategy achieved modest positive returns in some sectors but performed poorly in others. Future work should focus on refining regime detection (for example, by coupling HMMs with macro indicators), incorporating additional features such as fundamentals or cross-asset signals, and developing sector-specific models. Additionally, techniques to control drawdowns, such as stop-loss rules or volatility scaling, may enhance risk–adjusted returns. Finally, an in-depth analysis of the trial-and-error experiments (placeholders indicated above) should be added to capture personal learning and insights.