Macro-Linked Energy ETF: Performance Evaluation, Risk Analysis, and Business Viability

Course: MSDS 451 - Financial and Algorithmic Trading
Student: Vrishani Shah
Term Project GitHub: `https://github.com/vrishanishah20/macro-linked-energy-etf`

1. Introduction

The energy sector is one of the most cyclical and volatile segments of global financial markets. Its performance is heavily influenced by macroeconomic variables such as crude oil prices, interest rates, inflation, volatility indices, and real economic activity. Over the past two decades, the sector has experienced multiple boom bust cycles, including the 2008 financial crisis, the 2014-2016 oil price collapse, and the 2020 COVID-19 shock. These episodes have created both opportunities and substantial risks for investors.

This research investigates whether a **macro-linked, rules-based energy ETF** could deliver superior **risk-adjusted returns** relative to a passive benchmark such as the S&P 500 (SPY). Specifically, the study evaluates whether systematic allocation strategies informed by macroeconomic signals—derived from crude oil returns, Federal Funds Rate changes, and VIX volatility—can add value for investors **net of management and performance fees**. The study uses over 22 years of daily data (2002-2024) and applies a full quantitative pipeline, including portfolio construction, backtesting, Monte Carlo simulation, walk-forward validation, and fee impact analysis.

1.1 Potential Users and Applications

The knowledge and tools developed in this project are potentially valuable to:

- **Institutional investors** seeking targeted exposure to the energy sector within diversified portfolios.
- **Hedge funds and quantitative asset managers** evaluating sector-based trading strategies.
- **Individual investors** pursuing tactical or sector rotation approaches using ETFs.
- **Financial advisors** assessing the role of energy allocations relative to broad equity benchmarks.

- **Academic researchers and students** studying the interaction between macroeconomic variables, sector returns, and portfolio construction methods.

1.2 Objective and Deliverables

The objective is to define and evaluate a **macro-linked energy ETF strategy** built on:

1. A **rules-based allocation framework** using multiple portfolio strategies (equal-weight, volatility parity, risk parity, mean-variance, and macro-adaptive allocations).
2. A **backtesting engine** over ~22.4 years of historical data (2002-2024) with realistic assumptions about transaction costs.
3. **Monte Carlo simulations** (multivariate normal and bootstrap) to test robustness and quantify uncertainty.
4. **Walk-forward backtesting** to evaluate out-of-sample performance and detect overfitting.
5. A **fee structure analysis** exploring the impact of different management and performance fees on investor returns.
6. A **business viability assessment**: given the empirical evidence, is this ETF a product that should be launched?

In line with the course objectives and checkpoint guidelines, this report summarizes the research conducted so far, evaluates the performance and fee structure of the proposed ETF, and answers whether this represents a viable business opportunity.

2. Literature Review

This section reviews the literature in four thematic areas relevant to the project: (1) energy markets and macro linkages, (2) portfolio optimization, (3) systematic trading strategies, and (4) backtesting and performance evaluation. The goal is to situate the proposed fund within the existing quantitative finance literature and to identify the research gap this project seeks to address.

2.1 Energy Markets and Macro Linkages

The relationship between oil prices and macroeconomic conditions has been extensively studied. Hamilton (2009) documents the effects of oil price shocks on economic growth and business cycles, showing that oil shocks can trigger recessions and significant market dislocations. Kilian and Murphy (2014) decompose oil price movements into demand and supply shocks, highlighting the complexity of interpreting oil market dynamics.

Downey (2009) provides a practitioner's view of crude oil fundamentals, emphasizing the structural drivers of supply and demand, storage dynamics, and term structure behavior. Edwards (2017) and Bouchouev (2023) further explore quantitative and trading-oriented perspectives on energy markets, including the use of derivatives, risk management, and the integration of macroeconomic variables into trading systems.

Overall, these works show that the energy sector is heavily influenced by macro factors, but they stop short of fully integrating these macro signals into a systematic, fee-adjusted ETF-like portfolio strategy in a comprehensive, long-horizon backtest.

2.2 Portfolio Optimization

Modern portfolio theory originates with Markowitz (1952, 1956), who introduced **mean-variance optimization**, linking portfolio expected return and variance into an efficient frontier. Sharpe (1963) extends this framework into the Capital Asset Pricing Model (CAPM), formalizing the concepts of **systematic risk (beta)** and **alpha** as excess return over what is explained by market exposure.

Qian (2005) and subsequent practitioners popularize **risk parity** and **volatility parity** approaches, emphasizing allocation based on risk contributions rather than nominal weights. Asness et al. (2012) discuss leverage aversion and risk parity, providing empirical justification for risk-balanced portfolios.

DeMiguel, Garlappi, and Uppal (2009) famously show that "naive" 1/N diversification often competes surprisingly well with optimized portfolios, especially when estimation error is high. This is particularly relevant for energy, where parameter instability and regime shifts are common.

2.3 Systematic Trading Strategies

Systematic trading and factor-based investing have a long literature. Grinold and Kahn (2000) formalize the concept of information ratio and active portfolio management, emphasizing the rarity of persistent alpha. Jegadeesh and Titman (1993) demonstrate the profitability of **momentum strategies**, buying past winners and selling losers. Antonacci (2014, 2017) extends momentum concepts to global tactical asset allocation, emphasizing dual momentum strategies.

Greyserman and Kaminski (2014) explore long-run trend-following and managed futures, while D'Souza et al. (2016) and Gray (2020) discuss practical implementation of momentum and trend strategies in equity and ETF portfolios. Velissaris (2020) shows how mean reversion and momentum tactics can be combined.

Although much of this work focuses on broad asset classes or multi-asset portfolios, the conceptual tools of momentum, trend-following, and regime-aware strategies inspire the macro-adaptive and risk-based approaches used in this project.

2.4 Backtesting and Performance Evaluation

Accurate performance evaluation and robust backtesting are central to algorithmic trading. López de Prado (2018) highlights pitfalls in traditional backtesting, including multiple testing bias, overfitting, and the need for walk-forward analysis and Monte Carlo resampling. Trivedi and Kyal (2021) provide Python-based examples of backtesting techniques and highlight the importance of realistic assumptions and robust validation.

Chan (2020) and others stress the need to adjust for transaction costs, slippage, and fees when evaluating strategy performance. Gray (2023) and Quant Radio (2024) emphasize the importance of **out-of-sample testing**, cross-validation, and scenario-based analysis for systematic portfolios.

A key takeaway from this literature is that **negative results are as important as positive ones**: a strategy that fails robust testing can still provide valuable insight and prevent costly real-world mistakes.

2.5 Research Gap

While prior research has examined:

- The macroeconomic sensitivity of energy markets;
- Portfolio optimization methods (mean-variance, volatility parity, risk parity);
- Systematic trend and momentum strategies; and
- Backtesting and risk evaluation techniques,

there is relatively little work integrating all of the following into a **single, unified framework**:

1. Energy-focused portfolio strategies (WTI, XLE, XOM, CVX, IEF),
2. Macro-based regime classification and tilting,
3. Multi-strategy portfolio construction,
4. Full-fee, long-horizon backtesting over two decades,
5. Monte Carlo simulation and walk-forward out-of-sample testing, and
6. An explicit business viability assessment from an investor perspective.

This project aims to fill that gap by building a **macro-linked energy ETF prototype**, rigorously testing its performance, and answering whether it constitutes a viable commercial product.

3. Methods

This section describes the data sources, feature engineering, portfolio strategies, backtesting framework, robustness testing, fee structure analysis, and performance metrics used to evaluate the proposed fund.

3.1 Data Sources and Universe

The investment universe consists of five assets:

- **WTI Crude Oil** (proxied via a liquid futures or ETF series),
- **XLE** - Energy Select Sector SPDR ETF,
- **XOM** - Exxon Mobil Corporation,
- **CVX** - Chevron Corporation,
- **IEF** - iShares 7-10 Year Treasury Bond ETF (diversifying bond exposure).

The benchmark for market performance is:

- **SPY** - SPDR S&P 500 ETF Trust (proxy for S&P 500 index).

Data are sourced from:

- **Yahoo Finance** for daily adjusted closing prices of all ETFs and equities.
- **FRED (Federal Reserve Economic Data)** for macroeconomic series, including:
    - Federal Funds Rate,
    - Inflation (e.g., CPI),
    - VIX index (or equivalent volatility proxy),
    - Industrial Production,
    - Treasury yield spreads.

The analysis period covers **July 2002 through December 2024**, yielding approximately 5,634 trading days (~22.4 years). This falls slightly short of the ideal 1999-2024 period due to data limitations for some instruments, but is sufficient for meaningful long-horizon conclusions.

Daily prices are converted to daily log-returns and then aggregated into rolling statistics and features as described below.

3.2 Feature Engineering

A large set of features is constructed to capture return dynamics, risk, macro sensitivity, and inter-asset relationships. Key feature categories include:

3.2.1 Drawdowns

For each asset, drawdown metrics quantify the severity of losses from prior peaks:

- Current drawdown from all-time peak,
- Maximum drawdown over 63, 126, and 252 days.

The drawdown at time

t is defined as: $DD_t = (\max(P_{0:t}) P_t - 1) \times 100$
where P_t is the asset price at time t.

3.2.2 Beta (Market Sensitivity)

Rolling **beta vs SPY** is computed over 63, 126, and 252-day windows:
$\beta = Var(R_{SPY}) Cov(R_{asset}, R_{SPY})$

This measures each asset's sensitivity to broad market movements.

3.2.3 Correlation

Rolling correlations between asset pairs (e.g., WTI-XLE, XLE-IEF) are computed over 63, 126, and 252-day windows: $\rho = Corr(R_i, R_j)$

These capture the degree of co-movement and inform diversification potential.

3.2.4 Sharpe Ratios

Rolling Sharpe ratios for each asset are computed over 63, 126, and 252-day windows:

$Sharpe = \sigma(R) E[R] - R_f \times 252$
Where R is daily return and R_f is the risk-free rate.

3.2.5 Macro Sensitivities

To quantify how assets respond to macro changes, rolling betas to macro factors are computed:

- Rolling beta to changes in Federal Funds Rate,
- Rolling beta to changes in inflation,
- Rolling beta to changes in VIX,

- Rolling beta to changes in industrial production.

For a generic macro factor $\Delta Macro_t$: $Sensitivity = \frac{Cov(R_{asset}, \Delta Macro)}{Var(\Delta Macro)}$

3.2.6 Composite Macro Score

A **composite macro score C_t integrates energy-specific and macro indicators to classify regimes: $C_t = 0.4 \times z(r_t^{WTI}) - 0.3 \times z(\Delta FEDFUNDS_t) - 0.3 \times z(\Delta VIX_t)$**

where $z(\cdot)$ is a z-score computed over a 126-day rolling window: $z(x) = \frac{x - \mu}{\sigma}$

This score is designed so that higher values correspond to more favorable conditions for energy (strong oil returns, stable/lower rates, lower volatility).

3.2.7 Regime Classification

Based on C_t, the macro environment is classified into three regimes:

- **BULLISH:** $C_t > 0.75$ (214 days, ~3.8%)
- **NEUTRAL:** $-0.75 \le C_t \le 0.75$ (4,870 days, ~86.4%)
- **BEARISH:** $C_t < -0.75$ (277 days, ~4.9%)

These regimes drive the tilts in the macro-adaptive allocation strategy.

3.3 Portfolio Strategies

Six portfolio strategies are tested:

1. **SPY Benchmark (Buy-and-Hold)**
   - 100% SPY, no rebalancing, no transaction costs.
   - Serves as baseline for performance, alpha, and beta.
2. **Equal-Weight Strategy**
   - Weights: $w_i = 1/N = 20\%$ for each of the five assets (WTI, XLE, XOM, CVX, IEF).
   - Rebalanced monthly.
   - Provides a simple, naive diversification benchmark within the energy + bond universe.
3. **Volatility Parity**
   - Weights are inversely proportional to recent volatility: $w_i = \frac{1/\sigma_i}{\sum_j 1/\sigma_j}$

- σ_i is annualized volatility over the past 126 days.
- Lower-volatility assets receive higher weights; aims to equalize volatility contributions.

4. **Risk Parity**

   - Objective: equalize **risk contributions** across assets.
   - Optimization problem solved via SLSQP:
     - Constraints: $\sum_i w_i = 1, w_i \geq 0$.
     - Uses 126-day covariance matrix.
   - If optimization fails, falls back to equal-weight allocations.

5. **Mean-Variance Optimization (Markowitz)**

   - Objective: maximize Sharpe ratio: $\max_w \frac{w'\mu - r_f}{\sqrt{w'\Sigma w}}$
   - Constraints: $\sum_i w_i = 1, w_i \geq 0$.
   - Parameters estimated over a 126-day window:
     - μ: annualized expected returns (mean daily returns × 252),
     - Σ: annualized covariance matrix (daily covariance × 252).
   - Risk-free rate: $r_f = 2\%$ annually.

6. **Macro-Adaptive Strategy**

   - Starts from equal-weight base (20% per asset).
   - Tilt logic:
     - If **BULLISH** ($C_t > 0.75$): increase energy weights by +20% total, reduce bond (IEF).
     - If **BEARISH** ($C_t < -0.75$): reduce energy weights by −20% total, increase bond.
     - If **NEUTRAL**: maintain equal weights.
   - Energy assets: WTI, XLE, XOM, CVX (indices 0-3); bond asset: IEF (index 4).
   - Weights are rescaled to remain non-negative and sum to 1 after tilting.

3.4 Backtesting Framework

The backtesting engine simulates real-world implementation under consistent assumptions.

3.4.1 Rebalancing

- Frequency: **Monthly**, on the last trading day of each month.
- Rebalance is calendar-based, not trigger-based.
- All trades are assumed to execute at the closing price on the rebalance day (with transaction cost applied).

3.4.2 Transaction Costs

- Transaction cost: **10 basis points (0.10%)** per rebalance.
- Applied as a one-time reduction to portfolio value at each rebalance: $\text{Portfolio\_value}_t = \text{Portfolio\_value}_{t-1} \times (1-0.001)$
- Intended as a conservative approximation capturing trading costs and bid-ask spreads.

3.4.3 Lookback Windows

- Parameter estimation uses a **126-trading-day (~6-month) lookback** for volatility, covariance, and expected returns.
- Strategies are only active after the initial 126-day burn-in period.

3.4.4 Execution Assumptions

- Perfect execution at end-of-day prices.
- No slippage, no market impact, and fractional shares allowed.
- No explicit modeling of exchange fees beyond the 10 bps trading cost.

3.5 Monte Carlo Simulation

To assess robustness and quantify uncertainty, two Monte Carlo approaches are applied:

3.5.1 Multivariate Normal Scenarios (1,000 Runs)

- Parameters: μ and estimated from historical returns.
- Scenarios: returns simulated from a multivariate normal distribution $N(\mu,\Sigma)$.
- Horizon: 252 trading days (1 year).

- Advantages:
    - Preserves the estimated correlation structure.
- Limitations:
    - Assumes normality, which underestimates fat tails and extreme events.

3.5.2 Bootstrap Resampling (1,000 Runs)

- Method: sample daily return vectors with replacement from historical data.
- Horizon: 252 days per scenario.
- Advantages:
    - Preserves empirical return distribution, including skewness and kurtosis.
    - Non-parametric; fewer assumptions about distributional form.

For both methods, the **equal-weight or macro-adaptive strategy** is applied to each scenario, computing:

- Annualized return,
- Sharpe ratio,
- Maximum drawdown.

Percentiles (5th, 25th, 50th, 75th, 95th) and distributional statistics are reported.

3.6 Walk-Forward Backtesting

To test for overfitting and measure true out-of-sample performance:

- **Training period:** 5 years (1,260 trading days).
- **Testing period:** 1 year (252 trading days).
- **Step size:** 1 year.
- **Total windows:** 17, covering roughly 2007-2024.

For each window:

1. Train parameters and optimize strategies on the 5-year in-sample period.
2. Apply the resulting rules and weights to the next 1-year out-of-sample period.

3. Record annual return, Sharpe ratio, max drawdown, and other metrics.
4. Slide the window forward by one year and repeat.

This approach simulates realistic deployment where only past data are available at decision time.

3.7 Fee Structure Analysis

The impact of fees on investor outcomes is analyzed through:

3.7.1 Management Fees

- Annual management fee levels: 0%, 1%, 2%, 3%, 4%.
- Deducted daily: Daily fee=252Annual fee
- Net return: rnet =rgross −252fee%

3.7.2 Performance Fees

- Performance fee levels: 0%, 5%, 10%, 15%, 20%, 25%.
- Benchmark: SPY returns or risk-free rate.
- High-water mark methodology:
  - Performance fees charged only on returns above the previous peak net asset value.
  - Prevents charging fees on recovery from prior losses.

3.7.3 Combined Fee Structures

- Test 30 combinations: 5 management levels × 6 performance levels.
- Compare net returns to **SPY**, which charges only ~0.03% management fee and no performance fee.
- Summarize fee drag and investor vs manager wealth outcomes.

3.8 Performance Metrics

The following metrics are calculated for each strategy:

- **Returns**
  - Total return,

- Annualized return,
- Cumulative returns series.

- **Risk**
  - Annualized volatility,
  - Downside deviation (returns < 0),
  - Maximum drawdown,
  - Value at Risk (VaR, e.g., 5th percentile).
- **Risk-Adjusted**
  - Sharpe ratio,
  - Sortino ratio,
  - Calmar ratio (annual return / |max drawdown|).
- **CAPM-Based**
  - Beta vs SPY,
  - Alpha (Jensen's alpha): $\alpha = R_p - [R_f + \beta(R_m - R_f)]$
  - Tracking error: standard deviation of $R_p - R_m$.
- **Consistency**
  - Win rate (% positive periods),
  - Percentage of positive months,
  - Qualitative consistency measures (e.g., Sharpe × √(win rate)).

4. Results

This section summarizes the empirical results from the full-period backtest, alpha and beta analysis, Monte Carlo simulations, walk-forward validation, fee impact analysis, and overall business viability.

4.1 Full-Period Performance (2002-2024)

Table 1. Full-Period Performance Metrics

Strategy	Total Return	Annual Return	Volatility	Sharpe	Sortino	Max DD	Calmar
SPY Benchmark	559.91%	8.81%	18.92%	0.44	0.43	-59.58%	0.15
Volatility Parity	164.39%	4.44%	13.52%	0.31	0.23	-42.82%	0.10
Macro-Adaptive	155.50%	4.28%	21.38%	0.25	0.14	-62.93%	0.07
Equal Weight	156.50%	4.30%	20.99%	0.25	0.14	-63.98%	0.07
Mean-Variance	69.02%	2.38%	6.89%	0.21	0.08	-26.72%	0.09

Key findings:

- **SPY dominates all energy strategies.**
  Over ~22.4 years, SPY delivers **559.91% total return** (8.81% annual), while energy-focused strategies deliver only 69-164% total return (2.38-4.44% annually). The annual performance gap is **4.37-6.43 percentage points**.

- **Best energy strategy: Volatility Parity.**
  Among energy allocations, volatility parity offers the highest return (4.44% annually), lowest volatility (13.52%), best Sharpe ratio (0.31), and the smallest max drawdown (-42.82%).

- **Mean-variance optimization is highly defensive.**
  Mean-variance achieves the lowest volatility (6.89%) and smallest drawdown (-26.72%) but also the lowest return (2.38% annually). It behaves more like a bond-heavy, defensive allocation than an energy growth strategy.

- **Macro-Adaptive adds little value.**
  The strategy using macro-regime tilts (4.28%) performs essentially the same as equal weight (4.30%) with higher volatility, indicating that the macro signal provides little incremental benefit.

Dollar perspective (Initial $100,000 in 2002):

- SPY: ≈ **$659,910**
- Volatility Parity: ≈ **$264,390**
- Equal Weight: ≈ **$256,500**
- Macro-Adaptive: ≈ **$255,500**
- Mean-Variance: ≈ **$169,020**

The **lost opportunity cost** versus SPY ranges from roughly $395,520 to $490,890.

4.2 Alpha and Beta Analysis

Table 2. CAPM-Based Risk-Adjusted Performance

Strategy	Alpha (annual)	Beta	Correlation	Portfolio Vol	Benchmark Vol	Tracking Error
Equal Weight	-0.38%	0.69	0.62	20.99%	18.92%	17.41%
Volatility Parity	0.49%	0.45	0.62	13.52%	18.92%	14.88%
Mean-Variance	2.40%	-0.11	-0.31	6.89%	18.92%	22.07%
Macro-Adaptive	-0.42%	0.71	0.62	21.38%	18.92%	17.60%

Key findings:

- **Alpha:**
    - Positive alpha strategies: Volatility parity (+0.49%) and mean-variance (+2.40%).
    - Negative alpha: Equal weight (-0.38%), macro-adaptive (-0.42%).
      Even where alpha is positive (mean-variance), the **absolute return is still low**, and the strategy behaves more like a hedge than a growth vehicle.
- **Beta:**
    - Equal weight and macro-adaptive: $\beta \approx 0.7$ (moderately sensitive to market).
    - Volatility parity: $\beta = 0.45$ (more defensive).
    - Mean-variance: $\beta = -0.11$ (slightly negatively correlated to market).
- **Tracking Error:**
  Tracking error is high (15-22%) for all energy strategies, indicating significant deviation from SPY, **without corresponding excess return**, which is unattractive to most institutional investors.
- **Correlation:**
    - Energy strategies have moderate positive correlation with SPY (≈0.62).
    - Mean-variance shows modest negative correlation (-0.31), making it a potential hedge but with low returns.

Overall, most energy strategies **destroy value relative to SPY** once risk is accounted for, with only limited positive alpha from defensive positioning.

4.3 Monte Carlo Simulation

Table 3. Monte Carlo Results - Multivariate Normal

Metric	Mean	Median	Std Dev	5th %ile	25th %ile	75th %ile	95th %ile	Min	Max
Annual Return (%)	7.46	4.40	22.37	-24.97	-8.29	20.48	50.78	-47.52	91.56
Sharpe Ratio	0.34	0.31	0.99	-1.27	-0.30	1.01	2.01	-3.23	3.16
Max Drawdown (%)	-19.86	-18.97	7.23	-33.34	-24.19	-14.46	-10.11	-49.76	-4.79

Table 4. Monte Carlo Results - Bootstrap Resampling

Metric	Mean	Median	Std Dev	5th %ile	25th %ile	75th %ile	95th %ile	Min	Max
Annual Return (%)	7.61	6.18	23.32	-26.54	-9.00	21.02	50.15	-53.33	107.78
Sharpe Ratio	0.37	0.40	1.03	-1.30	-0.31	1.04	2.05	-2.96	3.68
Max Drawdown (%)	-19.77	-17.83	8.26	-35.45	-24.39	-13.76	-9.28	-55.07	-6.53

Key findings:

- **Extreme outcome uncertainty:**
  Annual returns range from roughly -**47% to +90% (normal)** and up to +108% (bootstrap). The 5th percentile returns (≈ –25-27%) reflect **substantial downside risk**.

- **Overoptimistic simulations vs history:**
  Mean simulated returns (~7.5%) exceed the realized historical returns (~4.3-4.4%). This suggests that **simple return models (normal or bootstrap)** may understate structural regime changes and sector-specific headwinds that actually occurred.

- **Sharpe ratio dispersion:**
  Sharpe ratios range from deeply negative (≈ –3) to very high (≈ +3), with roughly a quarter of scenarios exhibiting negative Sharpe. This underscores the **unreliability and regime dependence** of the strategy.

Overall, Monte Carlo analysis confirms that outcomes are highly uncertain, with significant tail risk and no guarantee of robust excess returns.

4.4 Walk-Forward Validation

Table 5. Walk-Forward Out-of-Sample Summary

Metric	Value
Number of Windows	17
Average Annual Return	0.09%
Std Dev of Returns	13.25%
Worst Window	-29.36%
Best Window	16.92%
Positive Windows	10 / 17 (58.8%)
Average Sharpe	0.02

Selected window results:

Window	Train Period	Test Period	Annual Return	Sharpe	Max DD
1	2002–2007	2007–2008	10.95%	0.63	-14.49%
2	2003–2008	2008–2009	-26.06%	-0.50	-41.69%
5	2006–2011	2011–2012	16.92%	0.94	-12.46%
8	2009–2014	2014–2015	-29.36%	-1.85	-30.68%
17	2018–2023	2023–2024	2.14%	0.15	-11.22%

Key findings:

- **Out-of-sample failure:**
  The average out-of-sample annual return is **0.09%**, compared to ~4.28% in-sample. This represents roughly a **98% degradation** in performance.

- **High variability:**
  Returns range from −29.36% to +16.92%, with standard deviation ~13%. Performance is highly unstable across market regimes.

- **Crash periods are devastating:**
  The strategy performs especially poorly during crisis periods (e.g., 2008-2009, 2014-2015), failing to avoid large drawdowns despite the presence of macro signals and treasuries.

- **Marginally better than random:**
  The win rate is 58.8% (10 of 17 windows positive), only slightly better than a coin flip. The average Sharpe ratio (~0.02) is essentially zero.

Walk-forward validation strongly indicates that the strategy is **overfit to historical data** and **does not generalize** to new periods.

4.5 Fee Impact Analysis

Table 6. Fee Structure Sensitivity (Macro-Adaptive Strategy)

Fee Structure	Mgmt Fee	Perf Fee	Gross Return	Net Return	Fee Drag	vs SPY Gap
No Fees	0%	0%	5.25%	5.25%	0.00%	-3.53%
1% Mgmt	1%	0%	5.25%	4.21%	1.04%	-4.57%
2% Mgmt	2%	0%	5.25%	3.17%	2.08%	-5.61%
3% Mgmt	3%	0%	5.25%	2.14%	3.11%	-6.64%
4% Mgmt	4%	0%	5.25%	1.12%	4.13%	-7.66%
2% + 20% Perf	2%	20%	5.25%	1.92%	3.33%	-6.87%
2% + 25% Perf	2%	25%	5.25%	1.61%	3.64%	-7.17%

SPY's net annual return over the same period is approximately **8.78%** after its ~0.03% fee.

Key findings:

- **Fees devastate returns.**
  Under a typical hedge fund fee structure (2% management + 20% performance), net return for the macro-adaptive strategy drops from 5.25% to **1.92%**, a **63% reduction**.
- **Manager vs investor wealth:**
  Over 22 years on a $100,000 investment, an illustrative analysis shows:
    - Investor final value under 2% + 20%: ~$152,000 (gain of $52,000),
    - Manager collects ~$161,000 in fees, roughly **3× the investor's gain**.
- **Even at 0% fees, the strategy is inferior to SPY.**
  With no fees, the strategy yields 5.25% vs SPY's 8.78%. At 1-4% management fees, the gap widens to **4.6-7.7 percentage points**.

Fee analysis shows that **no reasonable fee structure** can make the strategy competitive with SPY.

4.6 Business Viability Summary

Table 7. Strategy Comparison vs SPY (Net Returns)

Fund	Gross Return	Fee Structure	Net Return	vs SPY Gap	Viable?
SPY (Passive)	8.81%	0.03%	8.78%	Baseline	Yes
Volatility Parity	4.44%	~0.75% assumed	3.69%	-5.09%	No
Macro-Adaptive	5.25%	2% + 20%	1.92%	-6.87%	No
Equal Weight	5.27%	2% + 20%	~1.94%	-6.84%	No

The evidence is clear: **this energy-only macro-linked fund is not commercially viable as a standalone product**.

5. Conclusions

5.1 Main Findings

Research question:
Can macroeconomic signals and advanced allocation methods improve risk-adjusted returns in an energy-focused ETF relative to a low-cost passive benchmark such as SPY?

Answer: No.

Key results:

1. **Performance deficit:** All energy strategies underperformed SPY by **4.37-6.43% per year** over ~22.4 years.

2. **Best-in-class still inferior:** Volatility parity, the best energy strategy, returned **4.44% per year** versus SPY's **8.81%**.

3. **Macro-adaptive strategy failed to add value:** Macro-regime tilting produced almost identical returns to equal weight, with higher volatility.

4. **Negative or modest alpha:** Most strategies exhibited **negative alpha**; positive alpha (mean-variance, volatility parity) came with low or defensive returns and did not surpass SPY.

5. **Out-of-sample breakdown:** Walk-forward testing yielded **0.09% average annual return**, demonstrating extreme overfitting.

6. **Monte Carlo uncertainty:** Simulations showed wide dispersion of returns and substantial downside risk, with a non-trivial chance of large annual losses.

7. **Fees amplify underperformance:** Typical institutional fee structures (e.g., 2% + 20%) reduced net returns by **~63%** and made the strategy dramatically worse than simply holding SPY.

From an investor's point of view, there is no rational justification for choosing this energy-only macro-linked fund over a low-cost, diversified index ETF.

5.2 Why the Energy Strategies Failed

Several structural and methodological factors contributed to the failure of the proposed ETF:

1. **Sector concentration risk:**
   The portfolio is heavily concentrated in a single, cyclical sector. Over the study period, the energy sector suffered repeated large drawdowns (2008, 2014-2016, 2020), and these crashes were severe enough to overwhelm any tactical allocation or diversification within the sector.

2. **Structural headwinds to energy:**
   The last two decades have seen the rise of technology and the declining relative weight of traditional energy in major indices, along with the emergence of renewables and ESG constraints. These trends structurally disadvantaged fossil energy equities relative to broad markets.

3. **Limited effectiveness of macro signals:**
   The composite macro score and simple tilting mechanism were insufficient to predict or sidestep major energy crashes. Many adverse events (e.g., 2008 financial crisis, 2020 COVID shock) were either too sudden or too complex to be captured reliably by a small set of macro indicators.

4. **Fee and cost drag:**
   Even if the gross returns had been slightly better, realistic management and performance fees would have eroded most, if not all, of the value proposition. Given negative or modest alpha, fees are effectively a transfer from investors to managers.

5. **Overfitting and instability:**
   The dramatic drop from 4.28% in-sample returns to 0.09% out-of-sample in walk-forward testing is strong evidence of overfitting. The strategy appears to capture noise and regime-specific patterns rather than robust, persistent edges.

5.3 What Worked (Relatively)

Despite the overall failure as an investable product, the study yields useful insights:

- **Volatility parity** performed the best among energy strategies, offering a more balanced risk profile and slight positive alpha relative to SPY, albeit with much lower returns.

- **Treasury diversification (IEF)** provided partial downside protection and reduced volatility.
- **Mean-variance optimization** created a low-volatility, negative-beta portfolio that could serve as a partial hedge—but it looked more like a bond-heavy risk-minimizing allocation than an energy growth product.

These findings reinforce the role of **diversification** and **risk balancing**, even when they cannot fully compensate for structural sector underperformance.

5.4 Business Opportunity Assessment

The project's central business question is:

Given the performance evaluation and fee structure, is this ETF a business opportunity worth pursuing?

Verdict: No.

Reasons:

- The fund **consistently underperforms SPY**, even before fees.
- After realistic fees, the performance gap is so large that no rational retail, advisory, or institutional investor would allocate capital to it.
- High tracking error and sector concentration would also raise risk and compliance concerns for fiduciaries.
- The fund's revenue model is weak: underperformance would quickly lead to redemptions, shrinking AUM, and unsustainable operations.

Launching this ETF would likely result in poor investor outcomes, reputational damage, and a high risk of business failure.

5.5 Concerns and Limitations

The study has limitations that are important to acknowledge:

- The data period (2002-2024) does not fully capture 1999-2002 due to availability constraints, although 22.4 years is still robust.
- Transaction costs and execution assumptions (10 bps, no slippage) may underestimate real-world frictions.

- The macro composite score and regime classification are relatively simple; more sophisticated machine learning models or alternative data might uncover more exploitable patterns.
- The study focuses on an energy-heavy universe; multi-sector designs might behave differently.

These limitations suggest directions for future refinement but do not overturn the core conclusion that this specific energy-only macro-linked ETF is not viable as a standalone product.

5.6 Lessons Learned

From a data science and financial engineering perspective, the project is highly successful despite (or because of) its negative investment conclusion:

1. **Diversification is non-negotiable.**
   Sector concentration, especially in cyclicals like energy, is extremely risky, and no amount of clever allocation can fully replace cross-sector diversification.
2. **Fees matter enormously.**
   Even a modest alpha can be fully offset by high management and performance fees. In the absence of strong, persistent alpha, low-cost indexing remains a formidable benchmark.
3. **Passive indexing is hard to beat.**
   SPY's diversification and low fees make it a tough hurdle. Most active strategies underperform over long horizons, and this project adds further evidence in that direction.
4. **Rigorous backtesting prevents costly mistakes.**
   By performing long-horizon backtests, Monte Carlo simulations, and walk-forward validation, the project shows how proper quantitative analysis can prevent launching a product that would likely fail investors.
5. **Negative results are valuable.**
   Admitting that a proposed fund should not be launched reflects scientific honesty and professional maturity. The project serves as a case study in **when not to proceed**.

5.7 Future Directions

Although the specific ETF studied here is not viable, the methodology and infrastructure developed can be repurposed:

- **Multi-sector macro-rotation strategies** that diversify across energy, technology, healthcare, financials, consumer sectors, and bonds.
- **Machine learning enhanced models** for regime detection, using richer macro and alternative data.

- **Defensive or income-oriented niche products** leveraging volatility parity and bond allocations.
- **Quantitative research tools or educational products** based on the backtesting framework and pipeline.

5.8 Final Reflection

This term project has achieved the course's core objectives:

- Building a full quantitative investment pipeline,
- Applying modern portfolio theory and risk measures,
- Implementing robust backtesting and simulation frameworks,
- Integrating fees and business considerations,
- Communicating complex results clearly and honestly.

The central takeaway is that rigorous analytical work is valuable even when it leads to a **"no-go"** decision. In this case, the most responsible conclusion for both investors and fund sponsors is simple:

Do not launch this energy-only macro-linked ETF.

References

- Antonacci, G. (2014, 2017). *Dual Momentum Investing* and related works.
- Asness, C., Frazzini, A., & Pedersen, L. (2012). Leverage aversion and risk parity.
- Bouchouev, I. (2023). *The Art of Alpha: Investing in Energy and Beyond*.
- Chan, E. (2020). *Algorithmic Trading: Winning Strategies and Their Rationale*.
- DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification.
- Downey, M. (2009). *Oil 101*.
- Edwards, R. (2017). *Energy Trading and Investing*.
- Fama, E., & French, K. (1992). The cross-section of expected stock returns.
- Grinold, R., & Kahn, R. (2000). *Active Portfolio Management*.
- Hamilton, J. (2009). Causes and consequences of the oil shock of 2007-08.

- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers.
- Kilian, L., & Murphy, D. (2014). The role of inventories and speculative trading in the global oil market.
- López de Prado, M. (2018). *Advances in Financial Machine Learning*.
- Markowitz, H. (1952, 1956). Portfolio selection.
- Qian, E. (2005). Risk parity and diversification.
- Sharpe, W. (1963). A simplified model for portfolio analysis.
- Trivedi, A., & Kyal, N. (2021). *Hands-On Algorithmic Trading with Python*.
- Velissaris, N. (2020). Combining mean reversion and momentum strategies.
- Additional course bibliography and references cited in the main text.