Term Project: Checkpoint A

Ivy Liu

MSDS 451: Financial Machine Learning

July 25, 2025

Introduction

The goal of this project is to design an actively managed exchange-traded fund (ETF) that employs automated trading based on data science insights. The prospective investors for this fund are clients who seek a more professionally managed multiple asset portfolio than index-tracking ETFs to achieve a stated investment objective and take on a flexible trading approach without regard to an index. By integrating machine learning models, the fund aims to deliver risk-adjusted returns while maintaining high liquidity and regulatory compliance.

Literature Review

The U.S. Securities and Exchange Commission (SEC) regulates ETFs differently for index-based ETFs, which passively follow a benchmark, and actively managed ETFs, which are not based on an index and may adjust their holding daily ((U.S. Securities and Exchange Commission n.d.). Its active management strategy acts similarly to hedge funds' to potentially outperform a benchmark, but still differs significantly in terms of investment strategies, fees, and accessibility. Advances in data science, machine learning and low-cost computing have allowed fund sponsors to incorporate complex models into ETF portfolios. These trends align with the rise of quantitative hedge funds such as AQR Capital and Citadel, both of which blend technology and behavioral finance to exploit market inefficiencies ((AQR n.d.; Citadel n.d.). Citadel's Global Quantitative Strategies group, founded in 2012, develops algorithmic strategies across equities, futures, fixed income, and currencies (Citadel n.d.).

Methods

This project utilizes a comparative analysis of quantitative investment practices to inform the design of an actively managed ETF. To ensure compliance with U.S. securities laws, the SEC’s website will be referenced for creation/redemption mechanisms and regulatory requirements. The

current gaps in integrating hedging strategies with the ETF's structure for risk management will be addressed in future empirical research.

An initial pipeline for designing the fund is established below.

1. Define the investable universe. Select asset classes and identify liquid securities within each class.
2. Data collection and preprocessing. Gather historical prices, earnings, book value, and leverage data. Clean and align data across frequencies. Handling missing values and outliers. Maintain a secure and auditable data pipeline.
3. Feature engineering. Compute indicators such as moving averages, and relative strength indices. Build trend-following signals by comparing recent returns to long-term averages. Derive risk measures such as volatility forecasts.
4. Modeling. Combine signals using machine learning models (e.g. regression and gradient boosting) trained on historical data with cross-validation. Integrate domain knowledge to avoid overfitting and ensure interpretability.
5. Portfolio optimization. Formulate an optimization that maximizes expected return subjects to defined constraints. Solve using mathematical programming (to be determined after stage 2) to obtain target weights for each asset.
6. Backtesting, simulation, and evaluation. Run historical backtests to estimate performance. Perform Monte-Carlo simulation to assess robustness under different market regimes. Compare the strategy to relevant benchmarks and passive portfolios.

Results

The fund will be registered as an actively managed ETF, enabling daily trading and dynamic allocation across assets. A transparent disclosure of methodology and holdings will satisfy

regulatory requirements. Because the active management strategy incurs higher trading and research costs, the fund must justify its fee by delivering higher risk-adjusted returns. Following MPT, the fund will allocate capital across multiple asset classes with low correlations to form a diversified base (Investopedia 2025). Capital will be partitioned into a core allocation and a tactical overlay for signal-driven trades. Screening rules will eliminate companies with weak fundamentals. Recognizing the success of the buy-and-hold strategy, the fund will still hold passive exposures to market indices and high-quality bonds. These positions provide long-term growth, but they will be complemented with tactical overlays including trend following and hedges to reduce drawdowns. This hybrid approach aims to retain the benefits of passive investing while addressing its common limitations during market crashes given the current uncertainties of the market.

Conclusion

This report records the initial stage of building an actively managed ETF through not only academic researches but also technical studies of the market. The fund's philosophy blends fundamental screening and systematic trend following, which diversifies sources of return and improves market-timing decision. A transparent modeling process, informed by data science principles, ensures that model outputs interpretable and actionable recommendations rather than decisions to keep human oversight central. Next steps include selecting the initial asset classes, building and validating predictive models. Through iterative testing and refinement, the proposed ETF can evolve from a research idea into a viable product offering data-driven insights and robust downside protection.

References

AQR. n.d. *AQR Homepage*. https://www.aqr.com.

Citadel. n.d. *Global Quantitative Strategies*. https://www.citadel.com/what-we-do/global-quantitative-strategies.

Investopedia. *"Modern Portfolio Theory: What MPT Is and How Investors Use It."* Investopedia. https://www.investopedia.com/terms/m/modernportfoliotheory.asp.

U.S. Securities and Exchange Commission. *"Exchange-Traded Funds (Etfs)."* Exchange-Traded Funds (ETFs) | Investor.gov. https://www.investor.gov/introduction-investing/investing-basics/investment-products/mutual-funds-and-exchange-traded-2#:~:text=,market%20or%20its%20chosen%20benchmark.