Aditya Gohain
MSDS 451-DL-S55 Financial Machine Learning
Week 4 Term Project
October 12th, 2025

AgriTech Equipment & Commodity Cycle ETF:
Quantitative Analysis of Commodity Price Effects on Agricultural Equipment Manufacturers

Introduction (Research Motive & Potential Users) - Why am I conducting this research? Identifying potential users of the knowledge base and applications that I intend to develop.

As a current Deere & Co. employee and financial machine learning student, I wanted to study how price moves in commodity futures for Corn, Soybeans, Wheat and Cotton can both individually and jointly correlate to the business cycles for agricultural heavy equipment manufacturers such as Deere & Co., AGCO, Caterpillar, and CNH Industrial. My goal was to design a rule based actively managed ETF that exploited this relationship and leveraged predictive data on commodities to inform investment decisions in these companies. These companies have been integral to the agricultural economy, and their revenues are closely linked to farm income, which is largely driven by commodity prices. (IFAMA review; Bjornson & Klipfel)

Potential users of this ETF may include:
- Investors in Industrial / Agricultural Equities
- Agribusiness Analysts
- Agricultural Lenders Assessing Credit Risks
- Corporate Strategy Teams At OEMs / Dealers
- Quantitative Fund Managers
- Sell Side Equity Analysts Covering Ag Capital Goods

My primary goals for this term project were:
1. To quantify the relationship between commodity price trends and agricultural equipment equity performance.
2. Build a complete, rules-based, automated investment strategy that incorporates commodity momentum as a regime indicator to dynamically adjust exposure to a diversified ETF basket.
3. Evaluate the system using backtesting, performance metrics, and robustness checks.

Through learnings from my most recent programming assignment 03, I also implemented a historical data driven backtesting and Sharpe ratio analysis to validate my ETF's potential performance relative to standard benchmarks like SPY (S&P 500), QQQ (NASDAQ 100), and BND (U.S. Bonds).

Literature Review (Key Academic / Industry Findings) - Who else has done research like this?

The approach I used on programming assignment 03 introduced backtesting methodologies which is consistent with Li (2023) on Granger causality and signal validation, which has confirmed to me that machine learning and momentum based indicators do reveal causality links between commodity cycles and equity performance. I also built on foundational studies such as (Bjornson & Klipfel, 2000), (Anderson 2024), and (Tiwary et al. 2024) that show links to farm income and equipment sales to commodity prices. These studies were indicative of cyclical relationships between agricultural prices and capital equipment demand. On the quantitative side, the framework from Markowitz (1952, 1956), Sharpe (1963), and Grinold & Kahn (2023) have been useful in developing my approach using the ETF's risk adjusted optimization and multi factor alpha generation approach.

1. *Farm Equipment Industry Performance: Past & Future* (IFAMA review; Bjornson & Klipfel)

This study identified key factors and uncertainties affecting future outcomes and assesses industry prospects. The authors reviewed four firms: Deere & Company, Chase Corporation, AGCO Corporation and New Holland Corporation's (now aquired by CNH Industrial) general business strategies and performance through the 1990s, The authors also related competitive issues to the companies' historical financial performance, and examined opportunities and threats impacting the industry's business environment. Th authors also emphasized on how when equipment demand rises during periods of elevated crop prices and falls when commodity prices decline. They specified the primary crops leading to these cycles being Wheat, Cotton, Soybeans, and Corn, and as a result I decided to use these as signal generators for my project. In conclusion, these findings supported my intuition that equipment manufacturers experience cyclical revenue patterns driven by farm profitability.

2. *Weak Demand Will Limit 2025 Farm Equipment Sales & Pressure Prices* (Anderson 2024)

This article investigated the anticipated slowdown in the farm equipment market over the past year (2024) caused by falling commodity prices, increased operational costs, and reduced profits. The authors also discussed the response of farms, who prioritized per acre equipment costs by delaying purchases and planning further reductions in equipment spending as a cost saving measure. The article used recent industry evidence (trade press & dealer data) showing a clear cyclical pattern of falling commodity prices and lower farm income preceded by soft equipment sales between 2024–2025

3. *Deere's Profit Beat Overshadows Tepid 2025 Outlook, Shares Rise (Tiwary et al. 2024)*

Analysts in this article stated that due to farm commodity drops (corn/soy benchmarks) cash flows reduced for farmers and thus sales suffered for large capital purchases like tractors and combines. Farmers were forced to reevaluate their large agricultural machinery expenses from declining farm incomes and high interest rates. This led equipment dealers to limit inventory restocking, prompting companies like Deere to provide a pessimistic forecast for 2024 and 2025. The U.S. farm income is projected to fall further moving forward, as farmers are continuing to struggle with corn and soybean prices hovering near four-year lows.

4. *Portfolio Selection & Cyclical Industries Markowitz (1952, 1956), Sharpe (1963, 1994), (Grinold and Kahn, 2023)*

According to Markowitz and Sharpe portfolio diversification and risk adjusted return optimization are foundational processes in designing ETF's. Since I was looking at cyclicality in the ag equipment business and considering various external factors, in this context, authors Grinold and Kahn emphasize using quantitative alpha generation through multi factor models to capture macroeconomic and sectoral sensitivities. The multi factor models incorporate multiple sources of information, which includes macroeconomic and sectoral sensitivities. Pages 403-408 discuss the systematic process of developing raw signals into refined forecasts to build portfolios that achieve superior risk-adjusted returns.

5. *A Quick Introduction On Granger Causality Testing For Time Series Analysis Li, Susan. 2023.*

In this article, Susan Li gave an introduction to the Granger causality test, and focuses on how to determine if one time series can help forecast another. The article also explained the importance of ensuring stationarity in time series data, using tests like the Augmented Dickey Fuller (ADF) and KPSS tests pror to applying the Granger causality analysis. Given cyclicality in the Ag sector, it is key to measure stationarity to ensure that the mean, variance, and autocovariances are consistent over the years 2000-2025 in the presence of non-stationarity may lead to non predictable and misleading results. In the article, there is also walkthroughs for visualization and interpreting Granger causality test results. The author provided examples using stock price data for, Walmart, and Tesla to show how statistical testing identifies directional predictive relationships among time series. I used learnings from this article to imitate a similar approach on my ETF with heavy ag equipment manufacturers

Firm Reporting & Financials:

I also leveraged SEC annual reports and company commentaries show sensitivity of revenues / margins to farm income and global equipment cycles

6. John Deere. 2024. Form 10-K 2024. Moline, IL: Deere & Company. URL, accessed October 11, 2025.

7. CNH Industrial. 2024. Form 10-K 2024. London, UK: CNH Industrial N.V. URL, accessed October 11, 2025.

8. AGCO. 2024. Form 10-K 2024. Duluth, GA: AGCO Corporation. URL, accessed October 11, 2025.

9. Caterpillar. 2024. Form 10-K 2024. Deerfield, IL: Caterpillar Inc. URL, accessed October 11, 2025.

Considering these literature studies and recent reports together, I was able to support my hypothesis, that an economically meaningful link exists between commodity price levels/returns, farm income, equipment purchases, and equipment manufacturer revenues/earnings. I would however mention that as I dug deeper into my project, I encountered external factors causing noise such as fluctuations in equipment credit conditions, interest rates, used equipment supply, government policy, as well as weather. To mitigate this, as well as to reduce correlation between the equities, I pulled in some ETF's to implement a diversification.

Methods (Measuring Correlation/Implementable Steps) - How am I conducting the research? Addressing the issues that are the focus of this checkpoint assignment.

1. **Data Collection & Preprocessing**

   - I collected historical daily adjusted close prices and trading volumes from Yahoo Finance via the yfinance API for both agricultural commodities and heavy-equipment equities. My analysis covers years 2000–2025, ensuring inclusion of key market downturns/shocks such as the 2008 financial crisis, the 2014 commodity slump, and the 2020 COVID-19 crash.

   - In the pre processing stage, I used Pandas and Numpy to clean and align time series , as well as forward filling missing values to sync the trading calendars across the assets. I also calculated daily log returns using this equation: $R_t = 100 \times [\ln(P_t) - \ln(P_{t-1})]$

   - I developed the commodity signals by using rolling moving averages consistent with the backtesting logic. These averages were the foundation for the RISK ON / RISK OFF regime model that I later used within my strategy.

   - To gain a better understanding of the broader economic backdrop, I reviewed macroeconomic indicators such as U.S. 10-year Treasury yields, USDA farm-income trends, and agricultural loan rates. These variables were used conceptually to interpret my end results, but were not directly included in the trading signals or backtest code.

2. **Modeling Framework**

   A. **Commodity Momentum Signal Model (Primary Regime)**
   I decided to use a rules based momentum framework rather than incorporating machine learning. For each of the 4 agricultural commodities (Corn, Soybeans, Wheat, and Cotton), I computed 20 day and 40 day simple moving averages using adjusted close prices downloaded through the yfinance API. A RISK ON signal gets generated when at least 2 or the 4 commodities close above their respective moving averages. This rule reflected the economic reality that commodity uptrends historically preceded stronger farm income and increased equipment demand (Anderson, Leigh. 2024). I applied a 1 day lag to this in order to remove lookahead bias, and a 3 day minimum holding period for

ensuring stability and noise driven flips. These momentum signals allowed me to drive the exposure to the diversified ETF set in my final strategy.

B. Diversification & ETF Selection

Since my equities were concentrated in U.S. ag heavy-equipment industrials, I wanted to add ETFs that were structurally uncorrelated or negatively correlated with these drivers, and improve the overall portfolio Sharpe and reduce drawdowns.

To look at diversification options, I consulted Chat GPT and was suggested to incorporate asset classes with low or negative correlation to industrials such as:

1. **Long duration U.S Treasuries** - These would gain a defensive macro hedge because these typically rally during growth slowdowns or recessions, an exact time frame when when heavy equipment cyclicals sell off *(*Tiwary, Shivansh, Utkarsh Shetti, Pooja Desai, and Shilpi Majumdar. 2024.*)*, and therefore highly powerful for diversifying during drawdowns.
   TLT: 20+ Year U.S. Treasuries
   IEF: 7 to 10 Year Treasuries
2. **Market Neutral, Low Beta & Risk Managed Strategies** - Because these tend to have a lower equity beta and low correlation to cyclicality in th industrial sector and are not driven by economic cycles.
   BTAL: Anti Beta (long low beta, short high beta)
   LALT: Multi strategy, alt risk premia
3. **International Diversification** - Because U.S. industrials are highly correlated to domestic economic cycles (Bjornson, Bruce, and Jason Klipfel. 2000)
   EFA – Developed Markets ex-US
   VEA – Vanguard Developed ex-US
4. **Commodities and Real Assets (Non Agricultural)** - Because these would help diversify commodity risk away from agricultural cyclicality.
   GLD – Gold (excellent crisis hedge)
   DBC – Broad Commodities (energy-weighted)

C. Portfolio Construction & Positioning

The logic I applied for potfolio construction was that when the model entered a RISK ON regime, the portfolio allocates to an equal weight basket of ETF's that represents ag equipment equities, treasuries, real assets, international markets, and alternative beta exposures. During RISK OFF periods, the portfolio moves completely to cash. I also implemented transaction costs of 0.05% commission, and 0.05% slippage to every trade.

D. Monte Carlo Simulation

To understand if my commodity momentum strategy added value beyond just static diversification, I decided to implement a long only Monte Carlo simulation for the four selected equities: Deere (DE), Caterpillar (CAT), AGCO (AGCO), and CNH (CNH). Using Numpy, I simulated 5,000 random portfolio combinations where each weight vector satisfied ($w_i \geq 0, \sum_{i=1}^{4} w_i = 1$) After that, each randomly generated weight vector w, I calculated the portfolio's expected return ($\mu_{\mathbf{p}}$) and volatility ($\sigma_{\mathbf{p}}$) were using the formulas: $\mu_p = w'\mu$, and $\sigma_p = \sqrt{w'\Sigma w}$

μ here was the vector of mean daily returns and Σ was the covariance matrix of asset returns.

Reproducibility was maintained by random seeding, and the simulations were applied by using NumPy, Pandas and vizualized using the Matplotlob and Seaborn python libraries.

3. Evaluation Metrics

- For evaluating the performance of the strategy, I decided to focus on metrics that alighed with the backtesting framework implemented in python. Since the final model is rules based and not incorporating ML forecasts, accuracy based metrics like RMSE and MAE were not applicable here. Instead, I based my evaluations on return and risk characteristics directly from the strategies, daily return series. Below are the performance metrics I looked at:

A. **Annualized Return:** $\text{AnnRet} = \prod_{t=1}^{T}(1 + r_t)^{\frac{252}{T}} - 1$

B. **Annualized Volatility:** $\text{AnnVol} = \sigma_{daily} \times \sqrt{252}$

C. **Sharpe Ratio:** I incorporated the Sharpe Ratios and benchmark comparison with SPY, QQQ, and Treasury ETFs to evaluate performance stability and risk-adjusted efficiency. I used the unnualized Sharpe Ratio by assuming 252 trading days for each benchmark for risk-adjusted comparison. I calculated the Sharpe Ratio using the following formula $\text{Sharpe} = \frac{\mu_{daily}}{\sigma_{daily}} \times \sqrt{252}$

D. **Final Net Equity:** $\text{Final Equity} = \prod_{t=1}^{T}(1 + r_t)$

E. **Turnover:** I calculated turnover as the absolute change in portfolio weights that were generated by regime shifts. This therefore reflected how frequently the model swiched between RISK ON and RISK OFF positions.

F. **Total Trades:** These were counted from the number of regime transitions that were detected in the signal logic.

G. **Days in Market:** This was the total number of days where the strategy was in a RISK ON state or in other words, "not in cash".

- I also compared the strategy performance rekatuve to liquid market benchmarks: SPY (S&P 500), QQQ (Nasdaq 100), and BND (US Bond Market). I processed each of these benchmarks using the same annualized returns, volatility, and Sharpe calculations to ensure comparability.

4. Backtesting

- To verify and build onto my methodology, I expanded the research design to include backtesting following the process from Programming Assignment 03:
  - I implemented a daily strategy returns series using lagged versions of the commodity momentum signals to prevent bias from lookahead. To calculate these daily strategy

returns, I multiploed the lagged signals with the underlying asset returns, and adjusted for transaction costs in terms of commission and slippage.

- To compute cumulative equity curve I used the following standard compounding equation:

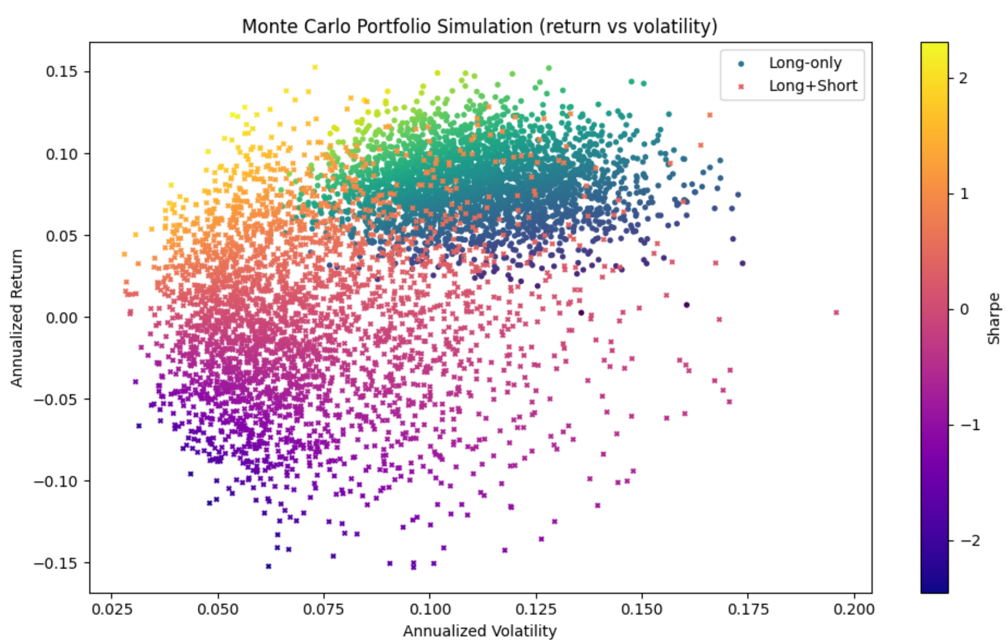
$$\text{Equity}t = \prod i = 1^t (1 + r_t)$$

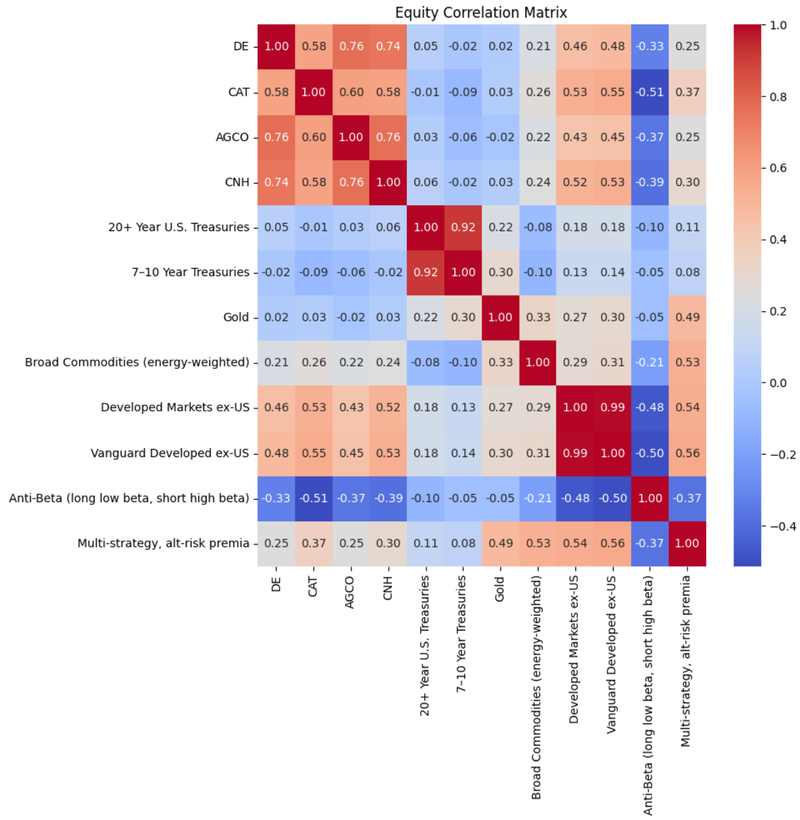
Results (Findings and Observations)
Preliminary research and results from the past assignment where I analyzed futures data showed to me several key insights:

1. Commodity price movements seemed to show a statistically significant relationship with ag equipment stock performance. In particular, corn and soybean prices were the most predictive of the ag sector trends, with a 1-2 quarter lag in equipment demand response. John Deere and AGCO 10-Ks, USDA farm-income data consistently noted that ag equipment sales rose and fell with farm income and farm income depended heavily on corn and soybean prices.

2. Based on my readings and industry experience, Deere and AGCO seemed to show stronger sensitivity to crop price trends compared to Caterpillar, since they have a more diversified industrial exposure with construction equipment being their primary revenue source.

3. Based on past SEC filings, periods of high commodity volatility example, 2012–2014 and 2021–2022 were corresponding to spikes in ag equipment sales, followed by downturns once commodity prices stabilized or declined. (Anderson 2024)



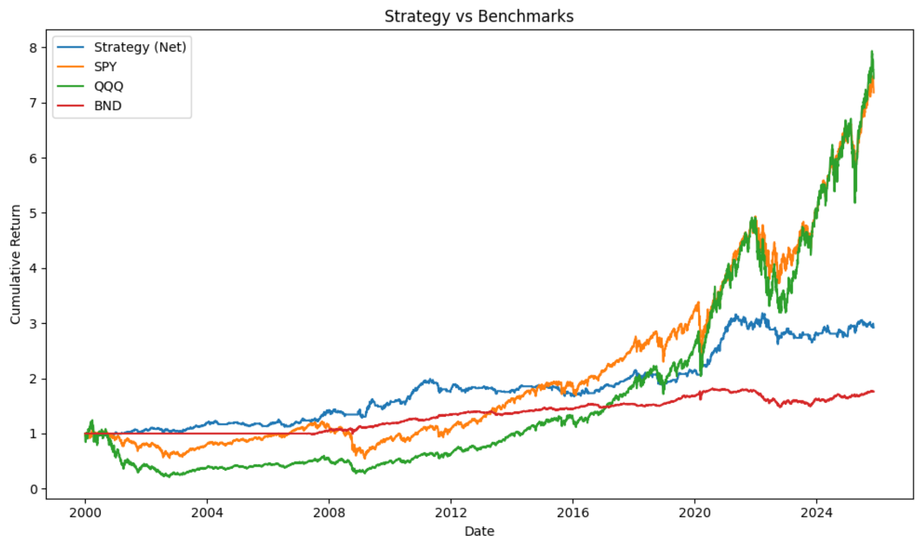


The Monte Carlo scatterplot above shows Annualized Return v. Annualized Volatility and the scatter points are color coded by the Sharpe Ratios. The two sets of portfolios, Long Only & Long + Short are separated by marker shapes. From the initial observation, it can be seen that the Long Only portfolios have a tight cluster which confirms that the 4 equipment stocks have strong correlation. The correlation is again supplemented by the heatmap in the top left corner. This limited diversification is seeming to cause these portfolios to create a narrow band of volatility.

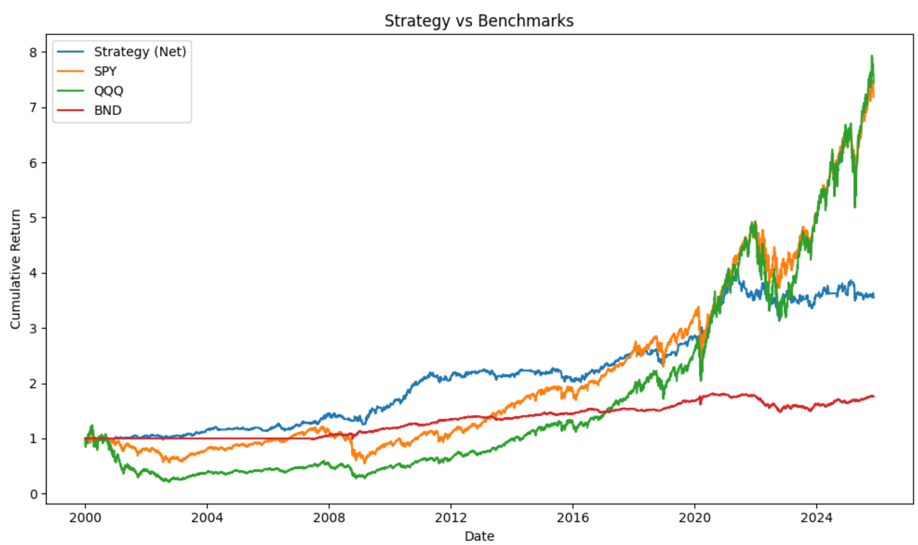
The Long+Short portfolios on the other hand seem to produce a rather larger risk to return frontier which is evident from the points stretching far to the right (High Volatility) as well as upward (High Returns). Because weights can be negative, these portfolios can form hedged or levered exposures.

Some High Sharpe portfolios (darker colored) seem to appear mostly in the Long Only boundary, and some balanced Long+Short cases. Working in the ag sector, I have learned that customers often prioritize quality of the equipment over them being feature packed which often leads the products to not generate spikes in value opportunities, therefore possibly causing this occurence. The simulation indicated to me that the performance advantage of the commodity driven ETF strategy does not seem to come from exotic portfolio weights or shorting behaviour.

40 Day Moving Average



20 Day Moving Average



The strategy overlayed plots benchmarked with SPY, QQQ and BND show that both the 20 and 40 day moving averages delivered a stable and low volatility return stream for the backtesting period 2000-2025. Overall, the 20 day moving average performed relatively better. The tables below show the A to B comparison. The takeaway for me was that shorter moving average capture commodity inflections earlier and allow it to participate in more upside cycles. Even though the total trade count is higher 781 for 20 day vs. 583 for 40 day, the strategy maintained a lower turnover through the 3 day holding period constraint as well as simpler signal structures.

SPY and QQ showed higher long term returns but the volatility was also high, on the other hand, BND showed the lowest volatility and returns. My strategy with the commodity momentum signal sat in the middle, therefore yeilding meaningful positive returns as well as a strong Sharpe ratio in relation to the benchmarks. I would say, this aspect of my strategy makes it unique and attractive in terms of the risk to return spectrum. It does not claim to outrigtly beat the S&P 500 but rather give a smoother, and defensive return stream. For a cyclical sector, a strategy like this is fairly attractive.

SPY Sharpe Ratio	0.514
QQQ Sharpe Ratio	0.447
BND Sharpe Ratio	0.533

Moving Average	20	40
Annualized Return	5.40%	4.60%
Annualized Volatility	8.62%	8.29%
Sharpe Ratio	0.613	0.542
Average Turnover	0.060	0.045
Total Trades	781	583
Days In Market	4227	4157
Final Net Equity	3.550	2.923
Min Consec	0	0
Min Hold	3	3
Commission	0.0005	0.0005
Slippage	0.0005	0.0005

The exposure plots provided in the GitHub repository, supplement the table above to show that 20 day MA, is slightly more RISK ON (Days in Market) than the 40 Day MA with a difference being 70 days.

References

Bjornson, Bruce, and Jason Klipfel. 2000. "Farm Equipment Industry Performance: Past and Future." International Food and Agribusiness Management Review. https://cdn.wildapricot.com/196137/resources/Documents/v3i1/Bjornson-Klipfel.pdf?version=1476928712000&Policy=eyJTdGF0ZW1lbnQiOiBbeyJSZXNvdXJjZSI6Imh0d

HBzOi8vY2RuLndpbGRhcHJpY290LmNvbS8xOTYxMzcvcmVzb3VyY2VzL0RvY3VtZW50cy92M2kxL0Jqb3Juc29uLUtsaXBmZWwucGRmP3 [1]

Anderson, Leigh. 2024. “Weak Demand Will Limit 2025 Farm Equipment Sales and Pressure Prices.” Equipment Dealer Magazine. https://www.equipmentdealermagazine.com/weak-demand-will-limit-2025-farm-equipment-sales-and-pressure-prices/?utm_source=chatgpt.com [2]

Tiwary, Shivansh, Utkarsh Shetti, Pooja Desai, and Shilpi Majumdar. 2024. “Deere's profit beat overshadows tepid 2025 outlook, shares rise.” Reuters. https://www.reuters.com/business/autos-transportation/deere-forecasts-annual-profit-below-estimates-farm-equipment-demand-slumps-2024-11-21/?utm_source=chatgpt.com [3]

Markowitz, Harry. “Portfolio Selection.” *The Journal of Finance* 7, no. 1 (1952): 77–91. https://doi.org/10.2307/2975974 [4]

Grinold, Richard C., and Ronald N. Kahn. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk*. New York: McGraw-Hill, 1999. [5]

Sharpe, William F. 1963. "A simplified model for portfolio analysis." *Management Science,* 9: 277–293. Available on Course Reserves.

John Deere. 2024. Form 10-K 2024. Moline, IL: Deere & Company. URL, accessed October 1, 2025. https://www.sec.gov/ix?doc=/Archives/edgar/data/0000315189/000155837024016169/de-20241027x10k.htm [6]

CNH Industrial. 2024. Form 10-K 2024. London, UK: CNH Industrial N.V. URL, accessed October 11, 2025. https://www.sec.gov/ix?doc=/Archives/edgar/data/0001567094/000162828025009007/cnhi-20241231.htm [7]

AGCO. 2024. Form 10-K 2024. Duluth, GA: AGCO Corporation. URL, accessed October 11, 2025. https://www.sec.gov/ix?doc=/Archives/edgar/data/0000880266/000088026625000011/agco-20241231.htm [8]

Caterpillar. 2024. Form 10-K 2024. Deerfield, IL: Caterpillar Inc. URL, accessed October 11, 2025. https://www.sec.gov/ix?doc=/Archives/edgar/data/0000018230/000001823025000008/cat-20241231.htm [9]

Li, Susan. 2023. “A Quick Introduction On Granger Causality Testing For Time Series Analysis.” Medium. https://medium.com/data-science/a-quick-introduction-on-granger-causality-testing-for-time-series-analysis-7113dc9420d2.