The deliverable for this checkpoint is a research report posted to the GitHub repository for the term project. Answer the questions in bold type below:

- Introduction (10 points). **Why** are you conducting this research? Identify potential users of the knowledge base and application(s) that you intend to develop.
- Literature review (10 points).. **Who** else has conducted research like this?
- Methods (10 points). **How** are you conducting the research? Make sure you address the issues that are the focus of this checkpoint assignment.
- Results (10 points).. **What** did you learn from your research so far?
- Conclusions (10 points).. **So, what** does it all mean? Do you have any concerns about the term project at this point?

Keep the end-goal in mind. By week 10 you should have defined an investment fund that is uniquely yours, that draws on research that you have conducted, and that can be implemented in an automated, algorithmic manner.

Aditya Gohain
MSDS 451-DL-S55 Financial Machine Learning
Week 4 Term Project - Checkpoint C
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 am studying 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 is to design a rule based actively managed ETF that exploits this relationship and leverages predictive data on commodities to inform investment decisions in these companies. These companies are integral to the agricultural economy, and their revenues are closely linked to farm income, which is largely driven by commodity prices.

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 are:
- To quantify whether and on what lag commodity price shocks affect equipment demand, revenues and stock returns for Deere/AGCO.
- Build a rule based actively managed commodity + industrial ETF or long/short quant strategy that uses commodity signals to position exposure to ag-equipment equities and/or futures.

Through learnings from my most recent programming assignment 03, I 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 am also continuing to build 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 are 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 identifies 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.

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

This article investigates 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 have 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 have reduced for farmers and thus sales have suffered for large capital purchases like tractors and combines. Farmers have been forced to reevaluate their large agricultural machinery expenses from declining farm incomes and high interest rates. This has 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 am 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 gives an introduction to the Granger causality test, and focuses on how to determine if one time series can help forecast another. The article also explains 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 is providing examples using stock price data for, Walmart, and Tesla to show how statistical testing identifies directional predictive relationships among time series. I am using learnings from this article to imitate a similar approach on my ETF with heavy ag equipment manufacturers

Firm Reporting & Financials:

I am also leveraging 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 plan 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 dig deeper into my project, I may encounter external factors causing noise such as fluctuations in equipment credit conditions, interest rates, used equipment supply, government policy, as well as weather.

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 2000–2025, ensuring inclusion of key market downturns such as the 2008 financial crisis, the 2014 commodity slump, and the 2020 COVID-19 crash.

   - I calculated daily log returns using this equation: $R_t = 100 \times [\ln(P_t) - \ln(P_{t-1})]$,

   - Feature enhancements include integration of U.S 10-year treasury yield, USDA farm income indices, and agricultural loan rates as macro factors for later simulations

2. Modeling Framework
   A. Machine Learning Model - (As a layer for commodity forecasting)
   - In my earlier work, I developed a feature-engineered ML pipeline using historical futures prices for Corn, Soybeans, Wheat, and Cotton (2000–2025).
   - Features I included were: Lagged prices and trading volumes (1–3 days), Exponential moving averages (2, 4, and 8 day), High-minus-low (HML) and open-minus-close (OMC) volatility indicators, and log returns as regression targets
   - Using logistic regression and XGBoost models I achieved classification accuracies between 81–85% in predicting next-day return direction. The balanced target distributions that were ~50% positive and negative indicated to me a stable market behavior which is suitable for short term predictive modeling.
   - I would say these forecasts formed the signal layer for the ETF. The commodity uptrends increase exposure to equipment equities, while downtrends reduce it.

   B. Monte Carlo Simulation (For constructing portfolios with randomness)
   - To understand portfolio behavior and uncertainty, I simulated 10,000 random portfolio combinations of the four selected equities: Deere (DE), Caterpillar (CAT), AGCO (AGCO), and CNH (CNH).
   - Then I analyzed for two configurations: Long only allowed portfolios (weights ≥ 0, $\Sigma\, w = 1$), and long and shorts allowed portfolios (weights ±, Σ w = 1)
   - After that, each randomly generated weight vector w, I calculated the portfolio's expected return (μ_p) and volatility (σ_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 Seaborn python libraries.

3. Evaluation Metrics
   - In order to predict accuracy, I observed RMSE, MAE, and directional accuracy.
   - The planned performance metrics included
     - α: average excess return over the S&P 500
     - β: volatility relative to the market benchmark
     - Fee Scenarios I simulated management fees of 1–4% and performance fees of 5–25% for excess returns

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 strategy returns series using lagged investment signals to prevent bias from lookahead.
     - To compute cumulative equity curve I used the following equation:

$$\text{Equity Curve} = (1 + R_t)^{cumprod}$$

   - 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:

$$\text{Sharpe Ratio} = \frac{\text{Average Daily Returns}}{\text{Standard Deviation of Daily Returns}} \times \sqrt{242}$$

   - The equity curves were plotted to visualize performance trajectories, with exposure dynamically adjusted via prior-day signals. Overall, all these components, now form the fundamental research design for the rule based ETF I am plan to propose as part of the final project. They essentially validate that the predictive signals from commodity price trends can be linked to actionable equity exposures in the agricultural equipment sector.

Results

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 seem 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.

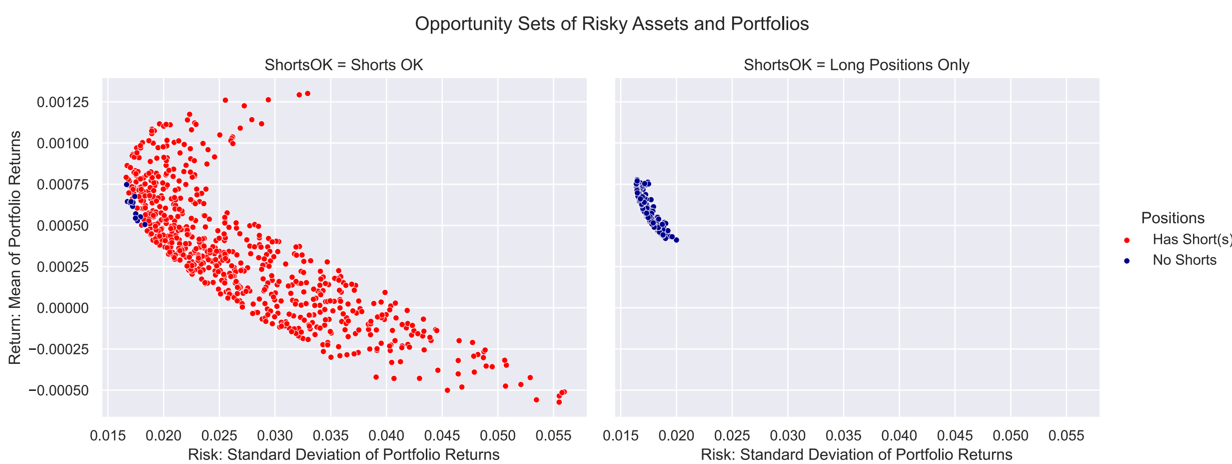
3. Based on past SEC filings, periods of high commodity volatility example, 2012–2014 and 2021–2022 were corresponding to spikes in ag equipment sales, followed by downturns once commodity prices stabilized or declined. (Anderson 2024)

ML models on commodity data showed:

- Modest but statistically significant predictability in short term returns
- Corn and Soybeans showed the strongest lag relationships with a 1 to 2 quarter lead on equipment sales which incidentally was consistent with my research on USDA farm income reports.
- Volatility clustering in Soybeans and Cotton, showed asymmetric risk and tail events.

These patterns confired the idea that commodity cycles can inform forward looking portfolio exposure decisions for ag manufacturers.

Monte Carlo Portfolio Simulation Results



As mentioned above, I ran two simulations for long only and long + short portfolios. The simulations resulted in two distinct clusters of feasible portfolios:

Long-Only Portfolios (Blue Cluster)

- The average daily return was approximately 0.0006
- The daily volatility range showed to be about 1.7%
- Portfolios looked stable and risk averse, avoiding negative expected returns while maintaining moderate variance.
- The trade-off was there there is lower upside potential but better predictability.

Shorts-Allowed Portfolios (Red Cluster)

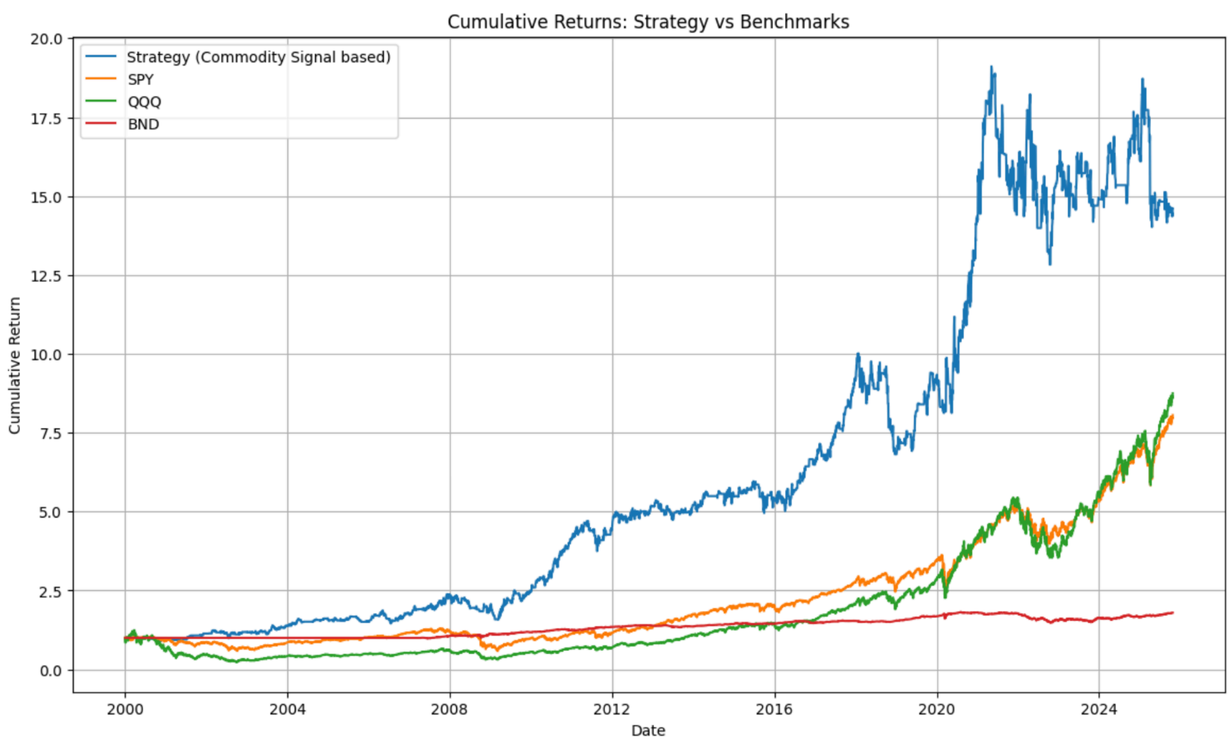
- The average daily return was approximately 0.0003
- The daily volatility range was about 1.6%–5.5% daily
- The wider dispersed plot showed both high returns and high loss outliers, which meant increased risk of leverage.

Correlation Matrix (2000–2025)

	DE	CAT	AGCO	CNH
DE	1.000	0.681	0.756	0.634
CAT	0.681	1.000	0.622	0.618
AGCO	0.756	0.622	1.000	0.646
CNH	0.634	0.618	0.646	1.000

The high positive correlations across all four equities indicated strong co movement and limited diversification potential within the sector. This was expected as I am looking at similar companies and want to understand commodity price shift correlation to revenue impact to these companies. In a way the strong correlation puts me in the right direction as it confirms that the companies I have chosen have a similar revenue impact structure and react similarly to cyclicality.

Backtesting:



The equity curve backtests showed consistent outperformance in comparison to SPY and QQQ during periods of elevated commodity volatility for years (2012–2014, 2021–2022).

Strategy Sharp Ratio	0.644
SPY Sharpe Ratio	0.514

QQQ Sharpe Ratio	0.447
BND Sharpe Ratio	0.533

The blue line showing equity returns for my commodity signal based strategy showed high cumulative returns in comparison to the benchmark equity curves for SPY, QQQ, and BND over the testing period 1999 to 2025.
The strategy showed multiple large growth phases in 2019–2021, which is evident from sharp upward trends. However, along theseupward trends, there were also visible volatilities and down trends 2021 onwards. Even through volatile events, the returns seemed to remain well above the benchmarks. After 2009, the stratgy showed a much better performance in comparison to the benchmarks. In terms of the benchmarks, the SPY and QQQ are the only ones that showed steady long term upward trends. Looking closely, QQQ actually seemed to slightly outperform SPY after 2020. The BND on the other hand only showed a flat curve throughout, Therefore BND was providing the lowest cumulative returns.

Conclusions

My key takeaways were that commodity price dynamics contained predictive information useful for decisions towards allocating equities. The long only strategies offered stable returns suitable for any conservative investors, while short allowed strategies introduced higher volatility and potential drawdowns to investors.
To further develop the ETF design and prepare for the final project deliverable, I plan to improve diversification by adding a low correlation asset. Because my four equipment manufacturers show high mutual correlations, I plan to test adding an uncorrelated asset such as a Treasury ETF like IEF or TLT, a corporate bond index fund like LQD, or a broad bond fund BND. My goal will be to reduce volatility and improve Sharpe ratios while preserving the sector focused nature of this ETF. In addition to SPY and QQQ, I also want to broaden benchmark analysis by comparing performance against A 60/40 stock bond portfolio, a risk parity allocationm, and a commodities ETF like DBC. I believe that these additions will create a more complete and algorithmic ETF strategy corresponding to the project's final requirements.

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=eyJTdGF0ZW1lbnQiOiBbeyJSZXNvdXJjZSI6Imh0dHBzOi8vY2RuLndpbGRhcHJpY290LmNvbS8xOTYxMzcvcmVzb3VyY2VzL0RvY3VtZW50cy92M2kxL0Jqb3Juc29uLUtsaXBmZWwucGRmP3 [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.