



VIP-VXU

Commodity Volatility Prediction: Final Report

Shyam Sai Bethina, Ronak Agarwal, Arnav Patidar,
Rohit Prasanna, Alex Yu-Hsin Chen, Sidharth
Subbarao, Harold Huang

Table of Contents

01 General Overview
- Flow Chart

02 Methodology

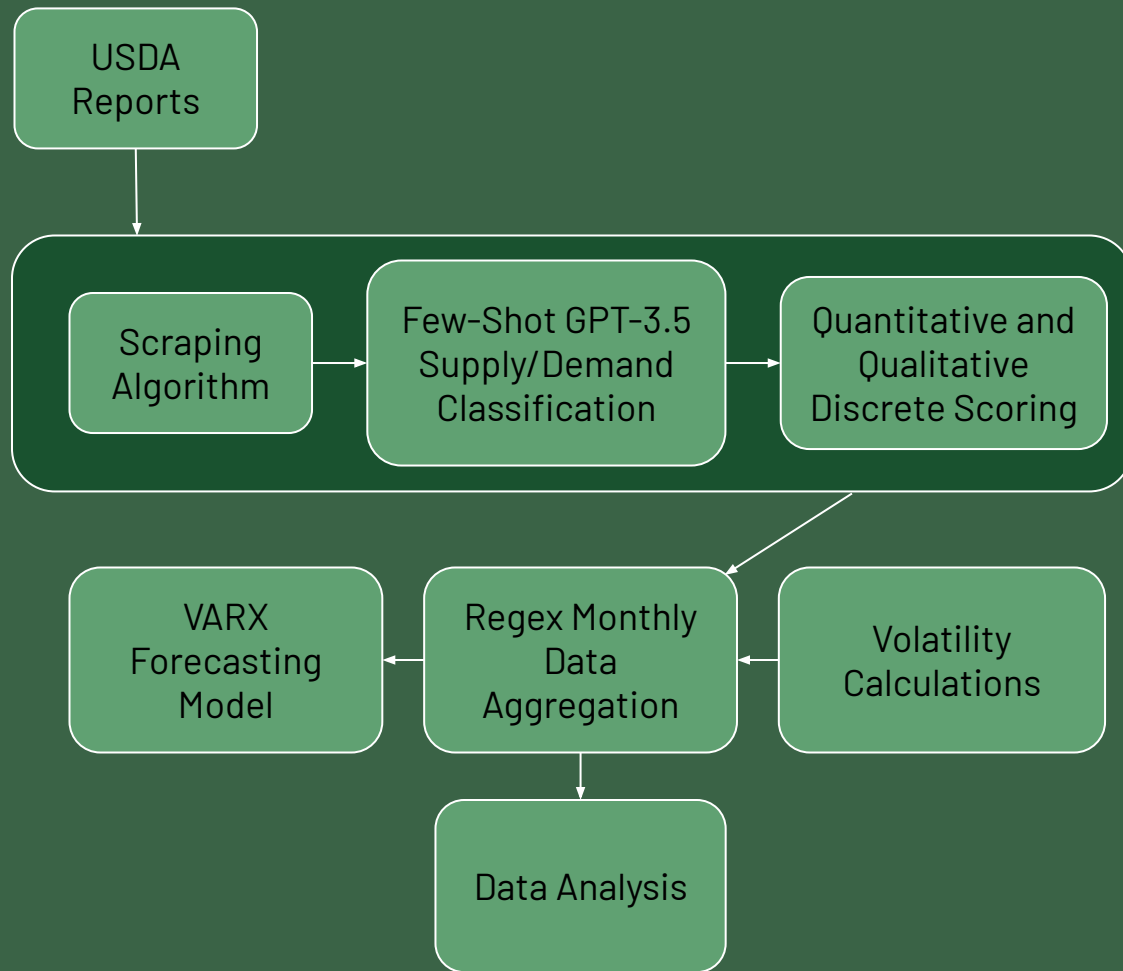
03 Results and
Data Analysis

04 Future Ideas

General Overview

- The agriculture sector is the most crucial sector for food production in the world.
- Numerous factors impact supply and demand, which in turn significantly shapes the prices of agricultural commodities. Analyzing the volatility of the most common crops is essential to understanding the commodity market.
- On monthly basis, USDA releases WASDE reports, which offer critical projections of supply and demand in the agricultural sector, focusing on both short-term (the upcoming month) and long-term (the forthcoming year) outlooks.
- Can a few-shot based GPT-3.5 approach be used to analyze supply & demand in USDA WASDE reports to forecast the future volatility of wheat futures?

Process Flow Chart



Methodology - WASDE Reports

Data Collection:

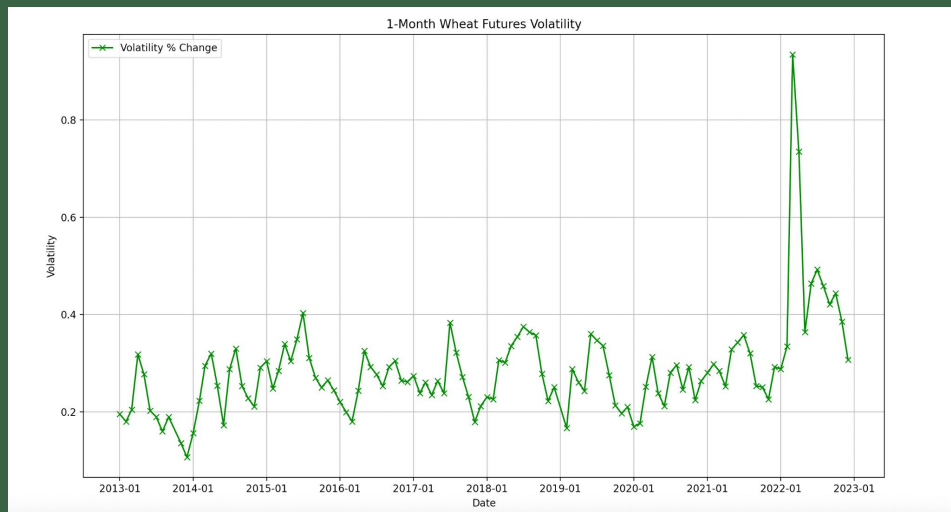
- USDA World Agricultural Supply and Demand Estimates Reports (January 2013 - December 2022; 118 total reports)
- Used USDA ESMIS API to download reports in .XLS format
- String manipulation to get monthly highlights wheat in a JSON format

WHEAT: The outlook for 2023/24 U.S. wheat this month is for larger supplies, decreased domestic use, unchanged exports, and higher ending stocks. Supplies are raised on increased imports, up 10 million bushels to 145 million, on a strong pace to date and expectations for the rest of the marketing year. Total domestic use is projected 4 million bushels lower to 1,155 million, all on a reduction in food use following the release of the latest NASS *Flour Milling Products* report. July-September wheat used in milling is the smallest for this quarter since at least 2014 when NASS began reporting this series. With no other changes to the U.S. balance sheet, projected ending stocks are raised 14 million bushels to 684 million. The projected 2023/24 season-average farm price is lowered \$0.10 per bushel to \$7.20 on lower expected prices for the remainder of the marketing year.

The global wheat outlook for 2023/24 is for increased supplies, fractionally lower consumption, less trade, and larger ending stocks. Supplies are projected up 0.6 million tons to 1,051.5 million as increased beginning stocks more than offset a decline in global production. World production is lowered 1.5 million tons to 782.0 million on decreases to many countries including India, Argentina, Kazakhstan, the United Kingdom, and Brazil. The decrease for production in India is based on revised government estimates. In Argentina, production is forecast 1.5 million tons lower to 15.0 million as rains in October were too late to benefit the crop in Cordoba and Santa Fe. These production declines are partially offset by a 5.0-million-ton increase in the forecast for Russia, up to 90.0 million, based on near-final harvest data from the Ministry of Agriculture that indicates more harvested area and higher yields. The global forecast for trade is lowered 1.3 million tons to 205.0 million, primarily on lower exports from Argentina, India, and Egypt that are only partly offset by an increase for Ukraine. Projected global ending stocks are raised 0.6 million tons to 258.7 million, with larger forecasts for Russia, China, and Argentina more than offsetting declines for India, Ukraine, and Brazil.

Methodology - Volatility

- Scraped volatility data (1 month timeframe) for wheat futures (ZW=F)
- Extracted 2013-2022 futures price data from Yahoo Finance 2013-2022
- Calculated Historical Volatility from this extracted data
- Aim: Predict volatility patterns using NLP analysis



Methodology - NLP:

- Tested Models: GPT-3.5 Turbo, Davinci-002, ETC.
 - Consider the differences in capabilities, computational requirements, and performance
- Compiled a comprehensive dataset of USDA reports, covering various time-frames (months) for temporal analysis.
 - Preprocessed the text data to remove noise, irrelevant information, and ensure consistency
- Implemented varying training approaches
 - 0-shot: Model predicts sentiment without specific training examples
 - 1-shot: Model exposed to one example per category during training
 - Few-shot: Model trained on a few examples per category to enhance contextual understanding
- Tested with different scoring parameters to determine the most effective metric for measuring sentiment accuracy and volatility.

Methodology - LLM:

Langchain + GPT 3.5 Turbo

- Labeling: Conditioning GPT to provide a numeric score for
 - qualitative sentences (-0.5, 0, 0.5)
 - quantitative sentences (-1, -0.5, 0, 0.5, 1)

Supply Qualitative:

Wheat: Supply outlook is stable, due to unchanged U.S. wheat outlook, Metric Score is 0.

Demand Qualitative:

Wheat: Global consumption is projected down 24.6 million tons year to year, Metric Score is -0.5.

Supply Quantitative:

Wheat: Global exports are 41 million tons, decreased 2.5 million tons, Metric Score is -1.

Demand Quantitative:

Wheat: Global consumption is 673.4 million tons, virtually unchanged, Metric Score is 0.

```
{'supply': [-0.25], 'demand': [-0.5]}
```

Prompt

Using the provided USDA report, categorize the information into four lists: Supply Qualitative, Supply Quantitative, Demand Qualitative, and Demand Quantitative. Each sentence should be classified based on its content.

For qualitative sentences, label them as positive, negative, or neutral, corresponding to a score of 0.5, -0.5, and 0, respectively.

For quantitative sentences, label them (bad, slightly bad, neutral, slightly good, good) according to the examples below on the score (-1, -0.5, 0, 0.5, 1).

For example the following sentences should have these Metric Scores:

"Quantitative: Wheat: Feed and residual use in Russia is 21.0 million tons, increased 1.0 million tons." (1)

"Quantitative: Wheat: Global consumption is 791.0 million tons, increased 2.4 million tons." (0.5)

"Qualitative: Wheat: Supply outlook is stable, due to unchanged U.S. wheat outlook." (0)

"Quantitative: Wheat: Ending stocks are 862 million bushels, decreased 15 million bushels." (-0.5)

"Quantitative: Wheat: Global exports are 41 million tons, decreased 2.5 million tons" (-1)

Give me only the lists and the sentences in the following format:

Qualitative: [Commodity]: [Qualitative Aspect] is [Trend/Outlook], due to [Cause/Factor], [Metric] score is [Metric Score].

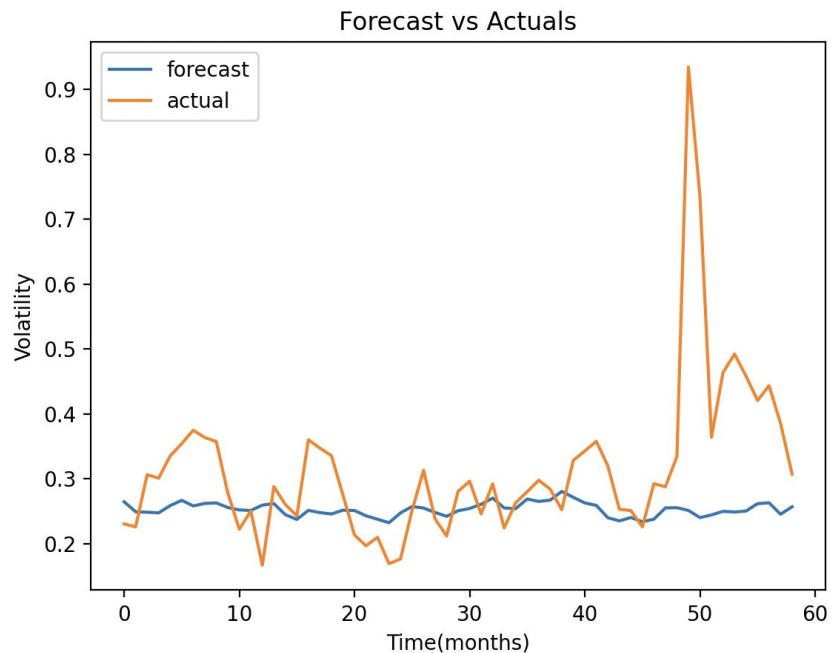
Quantitative: [Commodity]: [Metric] is [Value][Unit], [Change Direction][Change Amount] [Unit], [Metric] score is [Metric Score].

Please provide the lists and sentences in the above formats based on the report content. The report is {value}.

Methodology - VARX:

- Extract sentiment scores from the NLP model for each USDA report
 - Aggregate these scores on a monthly basis to align with the temporal granularity of the supply and demand data
- Collect both qualitative and quantitative supply and demand data for each corresponding month
- Align NLP sentiment scores with qualitative and quantitative supply and demand data on a monthly timeline
 - Compute monthly averages of NLP sentiment scores, qualitative and quantitative values for both supply and demand
- Use VARX to forecast volatility by using monthly volatility as an endogenous variable, and Supply Sentiment and Demand Sentiment as the exogenous variables
- See how accurate the forecasted values are compared to the actual volatility values

Results (VARX forecast)

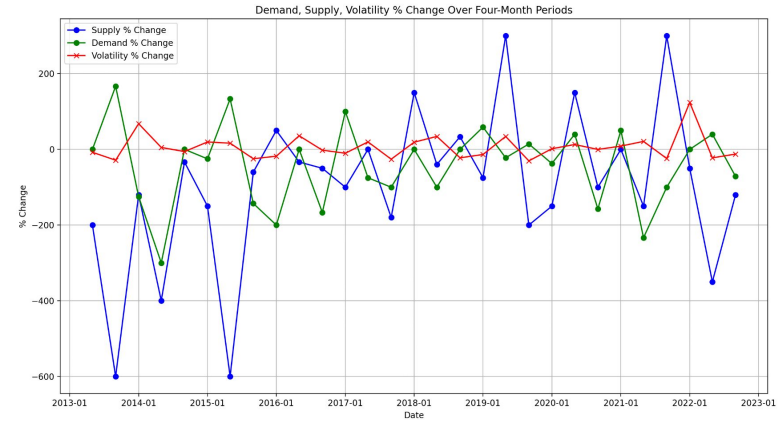
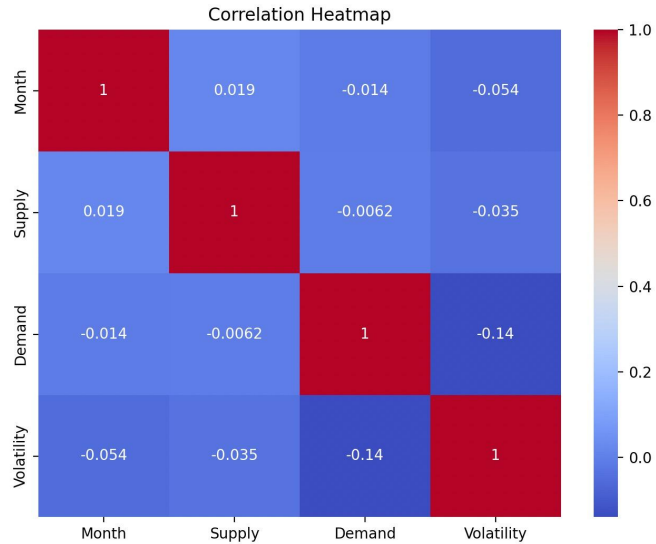


Mean Absolute Error: 0.0803

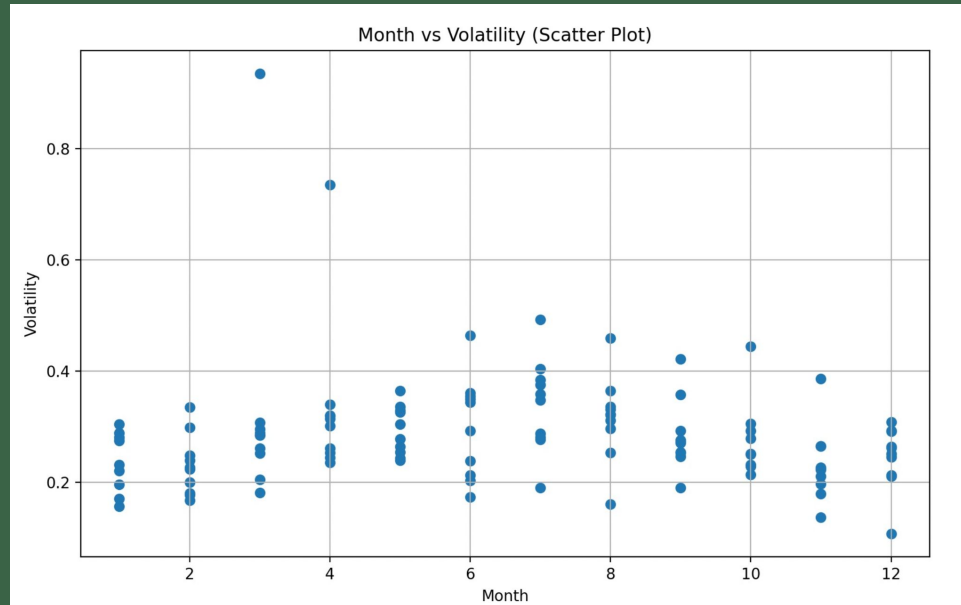
Mean Squared Error: 0.0187

Root Mean Squared Error: 0.1368

Analysis of Results



Analysis of Results



Analysis of Results (Granger's Causality)

Causality Test	Lag	F-Test P-Value	Chi-Square P-Value	Likelihood Ratio P-Value	Parameter F-Test P-Value
Demand → Volatility	1	0.3942	0.3852	0.386	0.3942
Demand → Volatility	2	0.6783	0.6638	0.6649	0.6783
Demand → Volatility	3	0.2936	0.2558	0.2643	0.2936
Demand → Volatility	4	0.3127	0.2568	0.2697	0.3127

Causality Test	Lag	F-Test P-Value	Chi-Square P-Value	Likelihood Ratio P-Value	Parameter F-Test P-Value
Supply → Volatility	1	0.3675	0.3583	0.3592	0.3675
Supply → Volatility	2	0.1301	0.1119	0.1171	0.1301
Supply → Volatility	3	0.2152	0.18	0.189	0.2152
Supply → Volatility	4	0.1492	0.1064	0.1187	0.1492

Limitations

- Inability for GPT-3.5 turbo to give supply and demand scores
 - Interpretation of highly specific domain content is not as accurate
 - Cannot give continuous score
 - Even discrete scores (.5 intervals) varied
- Prompt engineering
 - The prompt given to GPT may have been flawed and not specific enough
 - A bad prompt can lead to variable responses from GPT, even with the few-shot learning methodology used
- Supply and demand not correlated with wheat futures volatility
 - Supply and demand simply may not be correlated with volatility in wheat futures
 - There are many factors that impact wheat prices such as weather, war, politics, etc

Conclusion

- ChatGPT is not very good at labeling supply & demand scores
 - Large variation between responses
- ChatGPT is especially bad at giving continuous scores, even with a few-shot approach
- USDA WASDE reports on supply and demand may not be a good indicator of the volatility in commodity futures
 - This may be because the data in these reports are already priced in
- VARX output was unable to give a statistically significant output

Continued Work:

- Aim to refine NLP model precision, tailoring it more closely to the agricultural market sentiment.
- Plan to integrate diverse data sources, including weather trends/patterns and economic indicators, for a holistic and more comprehensive market analysis.
- Focus on a more granular temporal analysis to uncover shorter-term volatility trends within monthly cycles.
- Expand the research scope to compare volatility trends across various agricultural markets or regions.