Commodity Market LLM-based Explainable Forecasting

Dana Golden * Dikshya Mohanty† Khushboo Singh †

June 2024

1 Motivation

Commodity prices play a significant role in international trade and inflation dynamics. Although commodities constitute a relatively minor component of the overall cost of most goods, commodity prices are markedly more volatile compared to other economic sectors. This volatility underscores the importance of understanding commodity price fluctuations to gain insights into inflationary trends. Furthermore, movements in commodity markets are more heavily driven by tangible supply and demand factors compared to other financial markets. The agricultural and energy sectors are extremely sensitive to changes in weather patterns, while minor geopolitical shifts can dramatically affect metal prices. We create a model, which takes in news data, futures prices, and commodity market reports, and returns a prediction of price movements alongside a text-based explanation of the reasons for the prediction.

Given the pivotal role of commodity markets and the influence of physical supply and demand dynamics, numerous governments issue "outlook" reports. These reports create an authoritative second draft of history, provide the underlying causes of price fluctuations, and detail the prevailing factors impacting market supply and demand. Outlook reports are instrumental in shaping the market's understanding of commodity supply, demand, and market functioning. These reports go hand-in-hand with forecasting. The forecasting techniques have evolved alongside the literature. Machine learning techniques have been developed for commodity price forecasting. However, they often struggle to provide explanations for their predictions, posing a challenge in contexts where interpretability is crucial.

We create a system utilizing large-language models alongside other deep learning models to digest commodity outlook reports, news, and futures prices as training data to forecast futures prices and provide rationales for price movements. Our approach integrates numerical and textual data, offering predictions of price movements alongside text-based explanations for anticipated shifts. The model combines commodity reports, which focus on long-term factors, continuously updating our model with the real-time news data that help provide improved predictions of price movements. We report on the model's performance in both explanatory and forecasting capacities.

The research builds on work done by Sarah Mouabb, n.d. to explain based on news reports the economic reasons for futures price movements based on news data. Her work classified price movements into discrete bins, limiting the breadth of the explanations provided. We build upon this by incorporating expert analysis from government officials along with news reports. Forecasting commodity markets is one of the oldest challenges in economics, yet the methodology has traditionally lagged behind more contemporary forecasting techniques. A recent branch of the literature has attempted to add machine learning techniques (Bernard et al. 2008; Cortez et al. 2018; Gifuni 2023; Gu et al. 2022; Hegde, Hulipalled, and Simha 2021), but these techniques tend to be poorly explained and lack validity in the eyes of economists. Adding an explainable component to the forecasts improves the value of the forecasts and improves the ability to provide answers to interesting questions about why markets move. The question of how government reports impact commodity markets and contain unique information has been of perennial interest to economics going back to at least 1970s with most papers finding information content in the reports but some work finding declining value

*Stony Brook University, Department of Economics, dana.golden@stonybrook.edu
†Stony Brook University, Department of Computer Science, dikshya.mohanty@stonybrook.edu, khusingh@cs.stonybrook.edu

for market prediction as markets become more efficient (Adjemian 2012; Bunek et al. 2015; Ederington et al. 2019; Falk and Orazem 1985; Huang, Serra, and Garcia 2021; Karali et al. 2019; Li, Shang, and Wang 2019; Linn and Zhu 2004; Ye and Karali 2016; Ying, Chen, and Dorfman 2019; Yun 2006). A budding literature also exists combining NLP-analysis with economic analysis, but this literature would benefit from the implementation of recent techniques developed in deep learning Baker et al. 2021; Chen 2021; Tang and Lei 2023.

On the more computational side, we build two broad categories of models - a forecasting model and a large language model - that generate numerical predictions and corresponding rationales. We explore state-of-the-art models for numerical and text data individually. Using the vanilla transformer architecture for time-series forecasting (Rasul et al., n.d.; Vaswani et al. 2017) as motivation, we create a custom n-layered encoder-decoder structure to fit our commodities' time-series data. For the reason model, we draw inspiration from the success in reasoning with fine-tuned Large Language Models (LLMs) using Instruction Tuning of (Zhang et al. 2023). We utilize a domain-adapted version of Llama-2 (Heitz et al. 2008) to create a model specific to commodities markets but general enough to provide text-based explanations of complex market trends.

Our research combines these separate strands of literature to incorporate the information in government reports and news sources to explain forecasts made by a state-of-the-art transformer-based system. The combination of text and numerical data for forecasting commodity markets with explanations presents a major advancement of the commodity forecasting literature. Furthermore, we introduce a comprehensive model that predicts trends across various commodity sectors concurrently, revealing novel correlations between seemingly unrelated commodities such as Germanium and wool.

The movements of commodity markets are crucial for policymakers given commodities' significant impact on economic growth and consumer welfare. Our research offers a toolkit to aid policymakers in understanding how commodity markets fluctuate and provides forecasters with a method to interpret current events and the influence of these events on markets. It is rare for text data to include narrative explanations that pinpoint the exact reasons behind a variable's movement. This unique feature enhances the value of the dataset and methods we create.

Future research could extend these techniques to the prediction of central bank interest rate movements and to analyze explanations provided by central bankers for their decisions. Employing these methods can facilitate a deeper understanding of how to contextualize news and determine which narratives are most likely to influence market prices. This approach promises to enhance the accuracy and relevance of real-time economic forecasts by integrating narrative data with quantitative analysis.

2 Data Sources

Taking advantage of APIs and webscraping, we create a novel dataset containing tens of thousands of documents from various governmental agencies from around the world including futures prices, outlook reports, news data, and documents for pre-training. Contextual pre-training data are taken from the CFTC, IAEA, IEA, OPEC, IMF, IRENA, OECD, USDA, USGS, and academic literature from ScienceDirect. These documents help the model learn the distinctive features of the commodities market. Commodity "outlook" report data are taken from EIA, USDA, USGS, OPEC, and the EU Quarterly Gas Report. Reports cover a wide range of commodity areas including seventy commodities from energy, livestock, grains, softs, precious metals, and industrial metals. The reports are manually annotated with reasons for price movements to create the training set. Futures data are taken from the Energy Information Administration (EIA) and Trading Economics, encompassing price data for these commodities dating back to 1985. This extensive historical dataset allows for in-depth analysis and modeling of commodity price trends over several decades. News data are taken from Factiva, S&P capital IQ Global, and News API. Together, the dataset provides high quality data on commodities markets, which improves the value of the state-of-the-art final model.

3 Methodology

Utilizing historical data from commodity reports, news, and futures prices, we predict futures price movements, along with text-based reasons for price fluctuations. We develop a forecasting model that leverages historical prices to predict monthly commodity price. To predict reasons for price movements, we create a domain-adapted LLAMA-2 model, and use chunks of relevant information to train our model to generate the explanations. To sieve through all the data available and generate relevant chunks, we use a retriever system.

Continual pretraining of Large Language models is required to adapt an off-the-shelf LLM/foundation model to a specific domain. In our project, we use the workhorse LLAMA-2 model developed by Meta and train the model further on large quantities of unstructured reports from the commodity. By doing this, we have develop a customized LLM (CommodityLLM) with increased domain knowledge, and enhanced capabilities to perform downstream tasks.

Because most LLMs have an input token limit, we restrict the amount of data we feed into the model for training. We use a retrieval system that helps us select relevant chunks of information from the reports and news data. We use a dense retriever as it is able to take into account semantic similarity, rather than just doing a surface level, word-to-word matching.

We provide two types of models: Time Series/Forecasting and reason prediction model. For forecasting prices, we implement two models - a) Time Series Transformer model - which utilizes the vanilla transformer architecture, and b)Supervised Fine-Tuned Model using a RoBERTa-base model with a regression objective. As inputs to our model, we use historical price trends and text to generate future prices.

For our reason model, we use our domain adapted CommodityLLM, and fine-tuned it further on relevant reports and news data chunks to generate reasons for price fluctuations. We implemented a) Parameter-Efficient Fine-Tuning (PEFT) using Low-Rank Adaptation(LoRA)(Hu et al. 2021) on Llama-2-7b model, and b) full fine-tuning on GPT-2-M model, for predicting the reason for the future price movements.

For evaluating the forecasting models, we used mean-squared error, mean-absolute error, and mean-absolute percentage error. For evaluating the reason models, since it's a text generation task, we used Rouge Scores (Rouge1, RougeL & ROUGE-Lsum) (Lin 2004). The current models perform well in terms of forecasting error and provide economically valid predictions of market movements. In addition to standard metrics, we manually review the validity of forecasted reasons.

4 Expected Research Findings

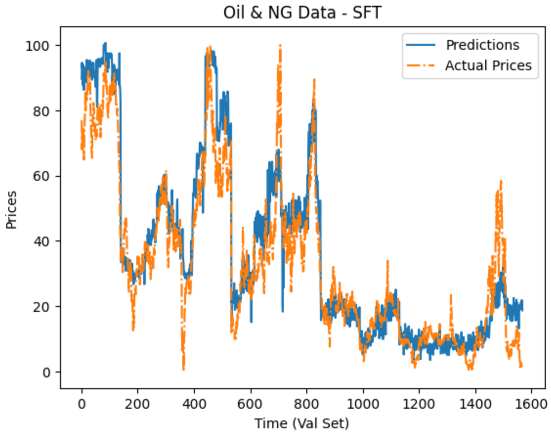
A version of our finalized model is currently operational on a subset of our documents, specifically those targeting energy market data. Preliminary results demonstrate that our model outperforms traditional time-series models. Additionally, our model provides predicted reasons for price movements, with evaluations based primarily on outlook reports. Our next steps include incorporating the news data into the models and expanding the range of commodities analyzed. We anticipate completing these tasks in time for the presentation of the paper.

Figure 1a shows price predictions made by the supervised fine tuned transformer trained only on text data. The figure will be updated with improved predictions and more commodities when the work is fully completed. Figure 1b shows evaluated Rouge scores measuring the quality of text output for the current iterations of the model. Table 1 provides a detailed analysis of these scores:

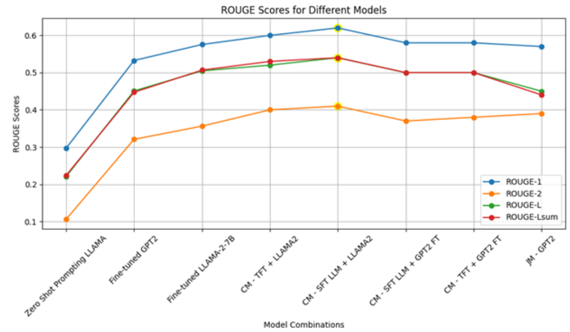
Predictive examples illustrating reasons for the movements in natural gas and oil prices provided by the model are presented in Figure 2. The model successfully identifies the reasons for movements in futures prices based on the commodity report immediately following the release of the report. Additionally, we provide an example of an inadequate explanation to illustrate the model's limitations. In this instance, the model fails to recognize actual trends and offers explanations that lack sound economic reasoning.

Reason for Movement

- Accurate example Natural Gas: "Today's natural gas price increased from last week due to volatile spot prices in the western United States driven by demand fluctuations in California and the Pacific Northwest, with SoCal Citygate and PG&E City"
- Inadequate example Petroleum: "Today's price decreased from last week's price due to a decrease in U.S. crude oil imports and refinery inputs, leading to lower supply and higher demand."



(a) SFT Predicted Vs. Actual.



(b) Rouge Score by Model.

Figure 1: Forecasting and Reason Prediction Model Results.

Table 1: Model Performance Explanation

Model &Commodity	**Evaluation Metrics**		
	Rouge1	***RougeL***	***Lsum***
Llama2 Zero-shot Oil	.268	.088	.201
Llama2 Zero-shot NG	.326	.244	.247
Llama2 Zero-shot STEO	.349	.214	.242
Llama2 PEFT	.57	.5	.5
TST+Llama2 FineTuned	.59	.519	.52
SFT+Llama2	.62	.53	.53
GPT2 Joint	.57	.45	.44
TST+GPT2-M FineTuned	.58	.50	.50
SFT+GPT2-M	.58	.49	.5

Figure 2: Example Reasons. Report on true reason.

Our work significantly advances the literature on both commodity market forecasting and explanability in machine-learning-based time series forecasts. We plan to extend the work by adding additional commodities and moving from predicting based on previous outlook reports to creating real-time predictions based on news data. Further, we will alter the information the model can access in order to understand the value of different types of information to forecasting. We will be able to provide insight into the predictive power of outlook reports as well as the information content of different news events.

References

Adjemian, Michael K. 2012. "Quantifying the WASDE announcement effect." *American Journal of Agricultural Economics* 94 (1): 238–256.

Baker, Scott, Nicholas Bloom, Steven J Davis, and Marco C Sammon. 2021. *What triggers stock market jumps?* Technical report. National Bureau of Economic Research Cambridge.

Bernard, Jean-Thomas, Lynda Khalaf, Maral Kichian, and Sebastien McMahon. 2008. "Forecasting commodity prices: GARCH, jumps, and mean reversion." *Journal of Forecasting* 27 (4): 279–291.

Bunek, Gabriel David, et al. 2015. "Characterizing the Effect of USDA Report Announcements in the Winter Wheat Futures Market Using Realized Volatility." PhD diss., Montana State University-Bozeman, College of Agriculture.

Chen, Qinkai. 2021. "Stock movement prediction with financial news using contextualized embedding from bert." *arXiv preprint arXiv:2107.08721.*

Cortez, CA Tapia, S Saydam, J Coulton, and C Sammut. 2018. "Alternative techniques for forecasting mineral commodity prices." *International Journal of Mining Science and Technology* 28 (2): 309–322.

Ederington, Louis H, Fang Lin, Scott C Linn, and Lisa Yang. 2019. "EIA storage announcements, analyst storage forecasts, and energy prices." *The Energy Journal* 40 (5): 121–142.

Falk, Barry, and Peter F Orazem. 1985. "A Theory of Future's Market Responses to Government Crop Forecasts."

Gifuni, Luigi. 2023. "NLP for analysis and forecasting of crude oil prices." PhD diss., University of Glasgow.

Gu, Yeong Hyeon, Dong Jin, Helin Yin, Ri Zheng, Xianghua Piao, and Seong Joon Yoo. 2022. "Forecasting agricultural commodity prices using dual input attention LSTM." *Agriculture* 12 (2): 256.

Hegde, Girish, Vishwanath R Hulipalled, and JB Simha. 2021. "Price prediction of agriculture commodities using machine learning and NLP." In *2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE),* 1–6. IEEE.

Heitz, Geremy, Stephen Gould, Ashutosh Saxena, and Daphne Koller. 2008. "Cascaded classification models: Combining models for holistic scene understanding." *Advances in neural information processing systems* 21.

Hu, Edward J, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. "Lora: Low-rank adaptation of large language models." *arXiv preprint arXiv:2106.09685.*

Huang, Joshua, Teresa Serra, and Philip Garcia. 2021. "The Value of USDA Announcements in the Electronically Traded Corn Futures Market: A Modified Sufficient Test with Risk Adjustments." *Journal of Agricultural Economics* 72 (3): 712–734.

Karali, Berna, Olga Isengildina-Massa, Scott H Irwin, Michael K Adjemian, and Robert Johansson. 2019. "Are USDA reports still news to changing crop markets?" *Food Policy* 84:66–76.

Li, Xuerong, Wei Shang, and Shouyang Wang. 2019. "Text-based crude oil price forecasting: A deep learning approach." *International Journal of Forecasting* 35 (4): 1548–1560.

Lin, Chin-Yew. 2004. "ROUGE: A Package for Automatic Evaluation of Summaries." In *Text Summarization Branches Out,* 74–81. Barcelona, Spain: Association for Computational Linguistics, July. `https://aclanthology.org/W04-1013`.

Linn, Scott C, and Zhen Zhu. 2004. "Natural gas prices and the gas storage report: Public news and volatility in energy futures markets." *Journal of Futures Markets: Futures, Options, and Other Derivative Products* 24 (3): 283–313.

Rasul, Kashif, Arjun Ashok, Andrew Robert Williams, Hena Ghonia, Rishika Bhagwatkar, Arian Khorasani, Mohammad Javad Darvishi Bayazi, George Adamopoulos, Roland Riachi, Nadhir Hassen, et al. n.d. "Lag-Llama: Towards Foundation Models for Probabilistic Time Series Forecasting."

Sarah Mouabb, Adrien Rousset Planat, Evgenia Passari. n.d. "The Origins of Commodity Price Fluctuations."

Tang, Xiaobin, and Nuo Lei. 2023. "Research on CPI Prediction Based on Natural Language Processing." *arXiv preprint arXiv:2303.05666.*

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention is all you need." *Advances in neural information processing systems* 30.

Ye, Shiyu, and Berna Karali. 2016. "The informational content of inventory announcements: Intraday evidence from crude oil futures market." *Energy Economics* 59:349–364.

Ying, Jiahui, Yu Chen, and Jeffrey H Dorfman. 2019. "Flexible tests for USDA report announcement effects in futures markets." *American Journal of Agricultural Economics* 101 (4): 1228–1246.

Yun, W. 2006. "Predictability of wti futures prices relative to eia forecasts and econometric models." *Journal of Economic Research Seoul* 11 (1): 49.

Zhang, Shengyu, Linfeng Dong, Xiaoya Li, Sen Zhang, Xiaofei Sun, Shuhe Wang, Jiwei Li, Runyi Hu, Tianwei Zhang, Fei Wu, et al. 2023. "Instruction tuning for large language models: A survey." *arXiv preprint arXiv:2308.10792.*