Advanced Programming 2025

The Predictive Content of Armed Conflict for Commodity Price Volatility

Final Project Report

De Martino Lorenzo Davide
lorenzo.demartino@unil.ch
Student ID: 21612817

11 January 2026

Abstract

This project studies whether information on armed conflict contains predictive content for commodity price volatility. Standard volatility models rely almost exclusively on past price dynamics, while geopolitical events are often associated with heightened uncertainty, particularly for energy commodities and safe-haven assets. The analysis focuses on three major commodities: WTI crude oil, natural gas, and gold.

Conflict intensity is measured using daily event-level data from the UCDP Georeferenced Event Dataset. Organized violence is proxied by the number of reported fatalities. To reduce skewness and capture persistence in conflict dynamics, fatalities are transformed using a logarithmic transformation and smoothed through an exponentially weighted moving average. Both global and region-specific conflict indices are constructed to reflect the economic exposure of each commodity.

Methodologically, the project compares a benchmark Heterogeneous Autoregressive (HAR) model with an augmented HAR-X specification that incorporates conflict information. Model performance is first assessed in-sample to evaluate explanatory power, and then evaluated out of sample using a strict walk-forward forecasting procedure. In addition, a Random Forest model is included as a nonlinear benchmark using the same information set.

The results show limited predictive gains overall. Conflict information slightly improves volatility forecasts for natural gas when region-specific indices are used, while no robust improvements are found for WTI crude oil and gold. These findings highlight that the predictive role of armed conflict for commodity volatility is heterogeneous and strongly commodity-specific.

Keywords: Python, Commodity Volatility, Armed Conflict, HAR Models, Machine Learning

Contents

1 Introduction

Commodity price volatility plays a central role in financial markets, risk management, and economic decision-making. It directly affects investors' portfolio allocation and hedging strategies and is central to the implementation of asset pricing models (Andersen Bollerslev, 1998, p. 885). In commodity markets, volatility also influences producers and consumers by increasing uncertainty and complicating planning along global supply chains (Pindyck, 2004, p. 1030). Accurately forecasting commodity price volatility is therefore a key issue in empirical finance.

Geopolitical tensions and armed conflicts represent an important source of uncertainty for financial markets. Armed conflicts generate complex economic consequences by destroying infrastructure, disrupting production, and interrupting trade and supply chains (Cappelen et al., 1984, p. 371; Schneider Troeger, 2006, p. 642). These effects are particularly relevant for commodities, whose production and transportation networks are often geographically concentrated and strategically important.

The objective of this project is to examine whether conflict-related information improves volatility forecasts for three major commodities: WTI crude oil, natural gas, and gold. Conflict intensity is measured using data from the UCDP Georeferenced Event Dataset. Methodologically, the analysis builds on the HAR model and evaluates the incremental forecasting value of conflict-related variables in an in-sample and out-of-sample setting. The aim is not merely to enhance predictive accuracy, but to assess whether conflict information provides economically meaningful value beyond standard volatility dynamics.

The remainder of the report is organized as follows. Section 2 reviews the related literature. Section 3 describes the data and methodology. Section 4 presents the empirical results. Section 5 discusses the findings and their limitations. Section 6 concludes.

2 Literature Review

2.1 Volatility and volatility forecasting in commodity markets

Volatility measures the variability of asset returns over time and is widely interpreted as a proxy for risk. A key stylized fact is that commodity price volatility is time-varying and persistent, which complicates its measurement and forecasting. A large empirical literature models volatility dynamics using econometric frameworks designed to capture persistence and clustering. GARCH-type models formalize conditional heteroskedasticity and volatility clustering in financial time series (Engle, 1982; Bollerslev, 1986, p. 308). More recently, the Heterogeneous Autoregressive (HAR) model has emerged as a parsimonious alternative that approximates long-memory behavior in realized volatility. By combining daily, weekly, and monthly volatility components, the HAR framework provides an intuitive and empirically successful approach to volatility forecasting (Corsi, 2009, pp. 174–196). Owing to its simplicity and strong performance, HAR has become a standard benchmark in the realized volatility literature, including applications to commodity markets.

2.2 Armed conflict, uncertainty, and financial markets

Armed conflicts generate persistent economic disruptions through infrastructure damage, supply-chain interruptions, and heightened uncertainty, with heterogeneous effects across countries, assets, and time periods (Cappelen et al., 1984, p. 371; Schneider Troeger, 2006, p. 643; Guidolin La Ferrara, 2010, p. 682). Financial market reactions depend critically on the severity and unexpectedness of conflict events. Sudden escalations tend to trigger sharp increases in uncertainty and volatility, whereas anticipated developments may already be partially priced in (Schneider Troeger, 2006, p. 642; Barro, 2006, pp. 825, 864–865).

Commodity markets are particularly exposed to geopolitical shocks because they rely on geographically concentrated production and complex global trade networks. Recent conflicts, most notably the Russia–Ukraine war, illustrate how geopolitical disruptions can trigger sharp increases in commodity prices and volatility, especially in energy and agricultural markets (Alam et al., 2022, p. 1; Liadze et al., 2022, p. 4; Izzeldin et al., 2023, p. 10). These observations motivate a focus on volatility rather than average returns as a key outcome variable when studying the financial impact of armed conflicts.

2.3 Measuring geopolitical risk and conflict intensity

A central challenge in this literature concerns the measurement of geopolitical risk. Two main approaches dominate. Event-based measures rely on observed violent events and allow the construction of quantitative indicators of conflict intensity based on fatalities and locations. These measures capture realized violence and are closely linked to physical disruptions affecting commodity supply.

In contrast, text-based approaches rely on news coverage to capture perceived geopolitical risk and investor attention. While these indicators reflect expectations and sentiment, they do not directly measure realized violence and may therefore capture a different dimension of geopolitical risk.

Far fewer examine whether conflict-related indicators improve ex-ante forecasts of financial risk. In particular, the predictive value of high-frequency, event-based measures of physical violence remains largely unexplored.

This work assesses whether georeferenced conflict intensity improves out-of-sample volatility forecasts, contributing new insights to the literature. The focus remains strictly on predictive performance rather than correlations, offering fresh evidence on the utility of armed-conflict data for commodity forecasting.

3 Methodology

3.1 Financial Data and Volatility Construction

The analysis relies on daily futures price data for WTI crude oil, natural gas, and gold obtained from Investing.com (1990-2024). Futures prices are preferred to spot prices as they are highly liquid, continuously traded, and play a central role in price discovery in commodity markets. Raw price files are downloaded as CSV files and preserved unchanged to ensure full reproducibility of the analysis.

All price series are processed using a unified Python pipeline that performs data cleaning, date normalization, and chronological alignment in a consistent manner across commodities. Daily log-returns are computed from settlement prices according to

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right),$$

where P_t denotes the futures price on day t.

Volatility is measured using realized volatility. Daily realized volatility is defined as the squared log-return,

$$RV_t = r_t^2.$$

To capture volatility persistence at different horizons, weekly and monthly components are constructed as rolling averages:

$$RV_t^{(w)} = \frac{1}{5}\sum_{i=1}^{5} RV_{t-i}, \qquad RV_t^{(m)} = \frac{1}{22}\sum_{i=1}^{22} RV_{t-i}.$$

These components form the core inputs of the Heterogeneous Autoregressive (HAR) framework.

The forecasting target is next-day realized volatility. At each date t, models predict RV_{t+1} using exclusively information available up to date t, ensuring a strictly causal forecasting setup without look-ahead bias.

3.2 Conflict Data and Index Construction

Conflict-related information is drawn from the UCDP Georeferenced Event Dataset (GED), which provides daily, georeferenced records of organized violent events worldwide. Each event includes the date, location, and an estimated number of fatalities. Conflict intensity is proxied by the number of reported fatalities.

This metric acts as a filter for economic impact. Violent events happen daily, but most are too small to affect global markets. By tracking the number of fatalities, the model separates minor local clashes from major conflicts. High casualty counts signal serious instability and these are the specific events that disrupt supply chains and drive volatility.

Fatality counts are highly skewed and characterized by episodic spikes. To mitigate the influence of extreme values and ensure numerical stability, daily fatalities are transformed using a logarithmic transformation,

$$d_t = \log(1 + \text{fatalities}_t).$$

To account for the persistence of geopolitical risk, the transformed series is smoothed using an exponentially weighted moving average (EWMA),

$$C_t = \lambda C_{t-1} + (1 - \lambda)d_t,$$

with a decay parameter fixed at $\lambda = 0.94$. This transformation is backward-looking by construction and relies solely on observed past and contemporaneous events.

Conflict indices are constructed at the daily frequency and aligned with the financial data. In addition to a global conflict index, region-specific indices are created to reflect the economic exposure of each commodity. For WTI crude oil, the regional index focuses on the Middle East; for natural gas, it focuses on Europe; and for gold, the global index is retained given its role as a global safe-haven asset.

All conflict variables enter the models in lagged form. When forecasting RV_{t+1}, only conflict information up to time t is used. This strict temporal alignment guarantees that the forecasting exercise remains free from information leakage.

3.3 Econometric and Machine Learning Models

The baseline specification is the standard Heterogeneous Autoregressive (HAR) model:

$$RV_{t+1} = \beta_0 + \beta_d RV_t + \beta_w RV_t^{(w)} + \beta_m RV_t^{(m)} + \varepsilon_{t+1}.$$

To test whether conflict information improves volatility forecasts, the HAR model is augmented with conflict indices (HAR-X). First, a global specification is considered:

$$RV_{t+1} = \beta_0 + \beta_d RV_t + \beta_w RV_t^{(w)} + \beta_m RV_t^{(m)} + \gamma C_t^{glob} + \varepsilon_{t+1}.$$

For commodities with strong geographic exposure (WTI and natural gas), a regional component is added:

$$RV_{t+1} = \beta_0 + \beta_d RV_t + \beta_w RV_t^{(w)} + \beta_m RV_t^{(m)} + \gamma_1 C_t^{glob} + \gamma_2 C_t^{reg} + \varepsilon_{t+1}.$$

This specification allows testing whether regional conflict intensity contains incremental predictive information beyond global geopolitical risk.

In addition to linear models, a non-linear machine learning benchmark is implemented using a Random Forest regressor. The Random Forest is trained on the same information set as the HAR models and evaluated within the same rolling out-of-sample forecasting framework. While HAR and HAR-X models are implemented explicitly by constructing regressors and estimating parameters via ordinary least squares, the Random Forest model is imported as a standard machine learning benchmark.

3.4 Forecast Evaluation

All models are evaluated using a rolling walk-forward forecasting procedure. At each iteration, models are estimated on a fixed-length training window and used to produce one-day-ahead forecasts of realized volatility. The estimation window is then advanced forward in time, ensuring that all forecasts are genuinely out-of-sample.

Let t denote the forecast origin. At each date t, models generate a forecast for RVt+1 using only information available up to date t. This strict temporal alignment guarantees the absence of look-ahead bias and ensures the causal interpretation of the forecasting exercise.

To balance statistical rigor and computational feasibility, forecasts are generated every five trading days rather than daily. This choice affects only the frequency of forecast production and does not alter the definition of the variables, the lag structure, or the underlying daily data. Model parameters are re-estimated at each forecast origin using the rolling window, allowing coefficients to adapt over time to changing volatility regimes.

Forecast accuracy is assessed using standard error-based loss functions commonly employed in volatility forecasting. Specifically, we report the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE), defined as

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}\left|RV_t - \widehat{RV}_t\right|, \qquad \text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(RV_t - \widehat{RV}_t\right)^2}.$$

In addition, the RMSE computed on the logarithm of realized volatility is reported to place greater emphasis on relative errors and to reduce the influence of extreme volatility spikes.

These complementary metrics allow for a nuanced evaluation of forecasting performance. While MAE reflects average forecast accuracy, RMSE penalizes large errors more heavily and is therefore particularly informative during periods of market stress. The comparison across models focuses on whether the inclusion of conflict-related information or the use of non-linear machine learning methods leads to systematic improvements in out-of-sample predictive accuracy relative to the baseline HAR model.

3.5 Approach

The empirical analysis is centered on the Heterogeneous Autoregressive (HAR) model, which provides a parsimonious and economically intuitive representation of volatility persistence through the combination of daily, weekly, and monthly realized volatility components. Due to its strong empirical performance and interpretability, the HAR model serves as the primary econometric benchmark.

To examine whether armed conflict contains incremental predictive information, the baseline HAR specification is extended to a HAR-X model by including a lagged conflict intensity variable. The conflict variable enters the model with a one-day lag, ensuring that forecasts rely exclusively on information available at the time of prediction. This parsimonious extension preserves interpretability and limits the risk of overfitting.

In addition to linear models, a Random Forest regressor is included as a nonlinear benchmark. The Random Forest is trained on the same information set as the HAR-X model and allows for nonlinear interactions between volatility components and conflict variables. Its role is not to replace the econometric models, but to assess whether additional predictive flexibility yields superior out-of-sample performance.

3.6 Implementation

The project is implemented entirely in Python using a modular and fully reproducible pipeline. Core libraries include `pandas` and `NumPy` for data manipulation, `statsmodels` for econometric estimation, and `scikit-learn` for the Random Forest benchmark. All figures are generated using `matplotlib`.

The system architecture follows a clear separation of tasks. Data loading, feature construction, model estimation, and forecast evaluation are handled in separate modules, which reduces code complexity and limits the risk of information leakage. Commodity price data and conflict-event data are processed independently and merged only after all transformations and lag constructions are completed.

Key code components include functions for realized volatility feature construction, conflict index building, rolling out-of-sample forecasting, and model evaluation. In-sample estimation, walk-forward forecasting, and performance assessment are implemented in dedicated scripts. The entire workflow can be executed from a single main script, which reproduces all results from the raw input data to the final tables and figures, ensuring transparency and full reproducibility.

Key code components

```python
def build_features_df(df, price_col="Price", date_col="Date"):
    # Ensure proper typing and ordering
    df[date_col] = pd.to_datetime(df[date_col], errors="coerce")
    df[price_col] = pd.to_numeric(df[price_col], errors="coerce")
    df = df.dropna(subset=[date_col, price_col]).sort_values(date_col)

    # Log-returns
    df["Log_Ret"] = np.log(df[price_col] / df[price_col].shift(1))

    # Realized volatility components
    df["RV_Daily"] = df["Log_Ret"] ** 2
    df["RV_Weekly"] = df["RV_Daily"].rolling(5).mean()
    df["RV_Monthly"] = df["RV_Daily"].rolling(22).mean()

    # Remove initial observations with undefined long-horizon volatility
    df = df.dropna(subset=["RV_Monthly"])

    return df
```

Listing 1: Construction of realized volatility features used in the HAR models

This function processes raw price series into the required HAR regressors. It computes daily realized volatility as squared log-returns, followed by the aggregation of weekly and monthly volatility components using rolling averages of 5 and 22 trading days, respectively.

```python
for t in range(window_size, len(data), step_size):

    train = data.iloc[t - window_size : t]
    test = data.iloc[[t]]

    # HAR baseline
    har_model = sm.OLS(
        train["Target_RV"],
        sm.add_constant(train[har_features])).fit()

    har_pred = har_model.predict(
        sm.add_constant(test[har_features])).iloc[0]

    # HAR-X model
```

```
    harx_model = sm.OLS(
        train["Target_RV"],
        sm.add_constant(train[harx_features])).fit()

    harx_pred = harx_model.predict(
        sm.add_constant(test[harx_features])).iloc[0]
```

Listing 2: Walk-forward out-of-sample forecasting procedure

The code uses a rolling out-of-sample framework to avoid look-ahead bias. At each time step, models are re-estimated using past data only, and a one-day-ahead forecast is produced to assess predictive performance

4 Results

This chapter presents the empirical findings of the study, organized into two main evaluations. First, we assess the in-sample explanatory power of armed conflict variables to determine if geopolitical instability provides statistically significant information for describing historical volatility dynamics. The results indicate that the direct explanatory power of conflict intensity is generally limited within a linear framework, with the exception of a marginal improvement observed for Natural Gas.

Second, we examine the out-of-sample forecasting accuracy of the proposed models. This analysis compares the predictive performance of the baseline HAR model against the conflict-augmented HAR-X specification and a non-linear Random Forest benchmark. The findings reveal a stark heterogeneity across commodities: while WTI crude oil volatility is best predicted by non-linear machine learning approaches, Natural Gas favors parsimonious linear models, highlighting that the relationship between conflict and commodity volatility is highly asset-specific and method-dependent.

4.1 Experimental Setup

All experiments are conducted on a standard personal computer running a 64-bit operating system. No dedicated GPU acceleration is required. The analysis is designed to run on commodity hardware and does not rely on specialized computing infrastructure.

The project is implemented in Python. The main software environment relies on Python 3.11. Core libraries include NumPy and pandas for data manipulation, statsmodels for econometric estimation, and scikit-learn for the Random Forest benchmark. All visualizations are produced using matplotlib. Exact package versions are documented in the project repository to ensure full reproducibility.

Model hyperparameters are kept fixed throughout the analysis. For the HAR and HAR-X models, parameters are estimated via ordinary least squares without regularization. The rolling window length is set to 750 trading days, and forecasts are generated every five trading days. The forecasting horizon is one day ahead.

For the Random Forest model, the number of trees is fixed at 150, with a maximum tree depth of 10 and a minimum of 5 observations per leaf. All other hyperparameters are kept at their default values as implemented in scikit-learn. To limit computational cost in the rolling forecasting framework, the Random Forest is re-estimated periodically rather than at every forecast step.

4.2 Performance Evaluation

4.2.1 In-Sample Results

Table 1: In-sample fit: HAR vs HAR-X (best variant) with significance test

Commodity	N	Adj. R^2 (HAR)	Adj. R^2 (HAR-X)	Δ Adj. R^2	p-value
WTI	8900	0.236999	0.236937	-6.2×10^{-5}	0.358844
GAS	8834	0.044466	0.044737	$+2.70 \times 10^{-4}$	0.074542
GOLD	8826	0.060337	0.060265	-7.2×10^{-5}	0.426043

Note: The p-value tests whether HAR-X improves over HAR for the selected variant.

Table 1 reports the in-sample fit of the baseline HAR model and the best-performing HAR-X specification for each commodity. For WTI crude oil and gold, the inclusion of conflict variables does not improve in-sample fit. The change in R2 is negative and economically negligible, and the associated p-values indicate that the HAR-X model does not outperform the baseline HAR specification. For natural gas, the HAR-X model including European conflict intensity leads to a small increase in R2. The improvement is modest but statistically borderline, suggesting a limited in-sample explanatory role for regional conflict information in gas markets.

4.2.2 Out-of-Sample Results

Reminder : Out-of-sample evaluation is restricted to the 2015–2024 period to keep computation time manageable. This choice affects only the evaluation sample and does not change model specifications.

Table 2: WTI Out-of-Sample Forecast Metrics (HAR vs HAR-X vs Random Forest)

Model	**MAE**	**RMSE**	**RMSE_log**
HAR	0.002041	0.017263	2.857139
HAR-X	0.002132	0.017228	3.007395
RF	**0.001203**	**0.006539**	**2.714371**

Table 3: Natural Gas Out-of-Sample Forecast Metrics (HAR vs HAR-X vs Random Forest)

Model	**MAE**	**RMSE**	**RMSE_log**
HAR	**0.001589**	0.003928	**2.688271**
HAR-X	0.001607	**0.003923**	2.882217
RF	0.001773	0.004204	2.721167

Table 4: Gold Out-of-Sample Forecast Metrics (HAR vs HAR-X vs Random Forest)

Model	**MAE**	**RMSE**	**RMSE_log**
HAR	0.000121	0.000284	**3.448249**
HAR-X	**0.000120**	0.000284	3.536235
RF	0.000129	**0.000277**	3.500213

Table 2 reports out-of-sample forecast errors for WTI volatility. The Random Forest (RF) clearly dominates the linear benchmarks, achieving substantially lower MAE, RMSE, and RMSElog than both HAR and HAR-X. By contrast, the difference between HAR and HAR-X is negligible: adding the conflict regressor does not produce a material improvement in predictive accuracy, and the ranking between HAR and HAR-X slightly varies across metrics.

Table 3 presents results for natural gas. Here, the baseline HAR model performs best overall, with the lowest MAE and RMSElog, while HAR-X does not improve upon the benchmark.

The Random Forest underperforms the linear models on MAE and RMSE, suggesting that additional nonlinear flexibility does not translate into better generalization for this commodity in the current information set and evaluation design.

Table 4 summarizes results for gold. Forecast errors are small in absolute terms and differences across models are limited. The HAR-X model yields a marginal improvement in MAE relative to HAR, while the Random Forest achieves the lowest RMSE. Overall, these results indicate that any incremental predictive contribution of conflict variables for gold is weak, and model comparisons depend on whether the focus is placed on average absolute errors (MAE) or on penalizing larger deviations (RMSE).

Taken together, the out-of-sample evidence highlights strong cross-commodity heterogeneity. Machine learning provides large gains for WTI, whereas natural gas volatility is best captured by the parsimonious linear HAR structure. For gold, performance differences remain modest and sensitive to the chosen loss metric.

4.3 Visualizations

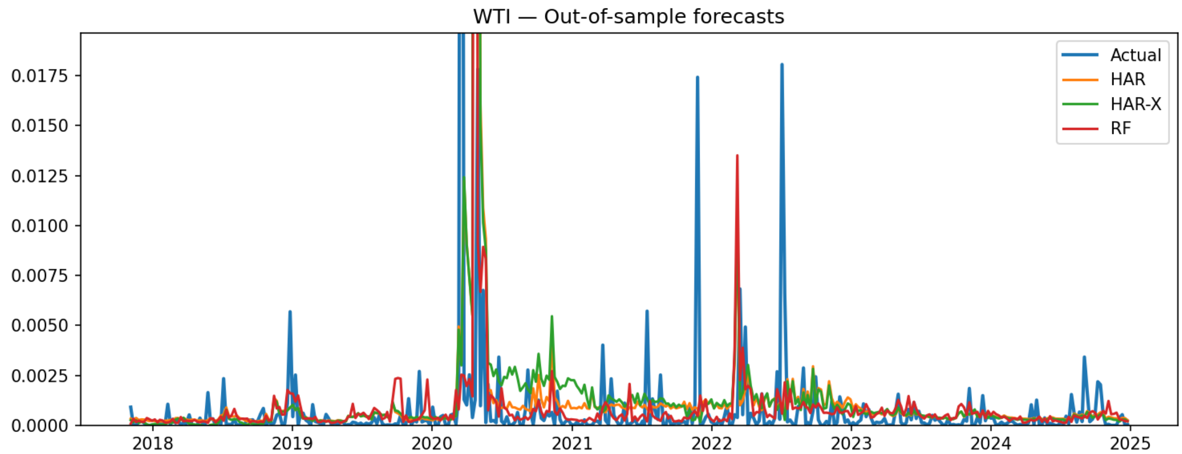

Figure 1: Out-of-sample volatility forecasts for WTI crude oil (2017-2024)

Figure 1 illustrates out-of-sample volatility forecasts for WTI crude oil. The plot highlights the ability of different models to track large volatility spikes, notably during the 2020 crisis. For brevity, analogous figures for natural gas and gold are omitted, as model comparisons are fully summarized in the out-of-sample performance tables.

5 Discussion

This project demonstrates that a clean and well-structured data pipeline is essential for analyzing the relationship between armed conflicts and commodity price volatility. The pipeline performed reliably and allowed large volumes of heterogeneous data to be processed, merged, and analyzed in a fully reproducible way. Daily commodity prices and detailed conflict-event data were successfully aligned, enabling both econometric and machine learning models to be estimated within a consistent framework.

One of the main challenges concerned the size of the datasets and the associated computational burden, particularly for the Random Forest model. Training non-linear models in a rolling out-of-sample setting proved costly in terms of computation time, which motivated a restriction of the evaluation window to the period 2017–2024. This compromise preserved the realism of the forecasting exercise while ensuring that the analysis remained feasible within the project constraints. Another important difficulty arose in the evaluation stage. The QLIKE loss function, although standard in volatility forecasting, was numerically unstable in this setting because realized volatility and forecasts frequently take values close to zero. This led to numerical issues and required the adoption of alternative and more stable loss measures.

The empirical results are more limited than initially expected. Based on economic intuition, regional conflict indicators were expected to have a clear and statistically significant impact on volatility forecasts, especially for energy commodities. In practice, the additional predictive power of conflict variables is modest and varies across commodities. While machine learning models substantially improve forecasting performance for WTI crude oil, simpler linear models remain difficult to outperform for natural gas, and forecasting gold volatility remains particularly challenging. The limited predictive contribution of conflict variables suggests that much geopolitical information may already be incorporated into market expectations, or that its impact on volatility is highly episodic rather than persistent. As a result, conflict indicators may be more informative during discrete escalation episodes than in average forecasting performance.

Several limitations should be acknowledged. Conflict intensity is proxied solely by reported fatalities, which may not fully capture broader geopolitical risk. Moreover, the focus on one-day-ahead forecasting may limit the ability to detect slower-moving conflict effects that unfold over longer horizons. Finally, computational constraints restrict the extent to which more complex models and longer evaluation windows can be explored. Despite these limitations, the project provides a transparent and coherent framework for assessing the role of armed conflict in commodity volatility forecasting and highlights important differences across markets.

6 Conclusion and Future Work

6.1 Summary

This project examined whether armed conflict contains predictive information for commodity price volatility. Using a transparent and fully reproducible Python pipeline, the analysis compared a baseline HAR model, an augmented HAR-X specification incorporating conflict intensity, and a Random Forest benchmark.

The empirical results show that volatility persistence remains the dominant driver of short-term commodity volatility. Conflict-related variables provide, at best, limited and commodity-specific improvements. Small in-sample gains do not translate into systematic out-of-sample forecasting improvements.

Natural gas exhibits some sensitivity to regional conflict intensity, consistent with its geographic concentration and exposure to European supply disruptions. In contrast, WTI crude oil and gold show no robust predictive gains from conflict variables. These findings suggest that the predictive role of armed conflict for commodity volatility is limited, heterogeneous, and highly dependent on the structure of the underlying market.

The main contribution of this project lies in its rigorous evaluation framework. By combining high-frequency conflict data with a strict out-of-sample forecasting design, the study provides a clean assessment of the true predictive content of geopolitical information.

6.2 Future Directions

Several extensions could be considered. From a methodological perspective, alternative measures of geopolitical risk could be explored, such as news-based indices, sanction indicators, or forward-looking measures of political uncertainty.

Another promising extension would be to adopt an explicit event-study framework centered on the onset of armed conflicts. Rather than evaluating average forecasting performance over long samples, future work could analyze volatility dynamics in windows defined around the conflict start date. Specifically, an event window could be constructed around the first day of a major conflict (event day $t=0$), with a pre-event period capturing market conditions before the outbreak and a post-event period capturing the adjustment phase. This design would allow for a direct assessment of whether conflict onsets trigger an immediate volatility spike and whether volatility subsequently declines as markets incorporate the new information. Such an event-study approach would provide clearer insights into the short-run impact of geopolitical shocks and help distinguish transitory volatility responses from persistent changes in volatility dynamics.

Additional experiments could also include a broader set of commodities, or comparing multiple machine-learning models could further clarify whether nonlinear effects play a meaningful role in capturing short-lived geopolitical shocks.

From an applied perspective, the framework could be adapted for stress testing or scenario analysis. Conflict indices may be particularly informative during extreme geopolitical events, rather than for average volatility forecasting performance.

References

1. Alam, M. K., Tabash, M. I., Billah, M., Kumar, S., & Anagreh, S. (2022). The Impacts of the Russia Ukraine Invasion on Global Markets and Commodities: A Dynamic Connectedness among G7 and BRIC Markets. *Journal of Risk and Financial Management*, 15(8), 352. https://doi.org/10.3390/jrfm15080352

2. Andersen, T. G., & Bollerslev, T. (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review*, 39(4), 885. https://doi.org/10.2307/2527343

3. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal Of Econometrics*, 31(3), 307-327. https://doi.org/10.1016/0304-4076(86)90063-1

4. Bouri, E., Lei, X., Jalkh, N., Xu, Y., & Zhang, H. (2021). Spillovers in higher moments and jumps across US stock and strategic commodity markets.

5. Cappelen, A., Gleditsch, N.P., Bjerkholt, O., 1984. Military spending and economic growth in the OECD countries. *Journal of Peace Research.* 21 (4), 361–373

6. Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal Of Financial Econometrics*, 7(2), 174-196. https://doi.org/10.1093/jjfinec/nbp001

7. Davies, S., Pettersson, T., Sollenberg, M., Öberg, M. (2025). Organized violence 1989–2024, and the challenges of identifying civilian victims. *Journal of Peace Research*, 62(4). https://ucdp.uu.se/downloads/

8. Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* 50, 987-1008.

9. Guidolin, M., & La Ferrara, E. (2010). The economic effects of violent conflict: Evidence from asset market reactions. *Journal of Peace Research*, 47(6), 671–684. https://doi.org/10.1177/0022343310381853

10. Investing.com. (2025). *Commodities Historical Data (WTI Crude Oil, Natural Gas, Gold).* Available at: https://www.investing.com/commodities

11. Pindyck, R. S. (2004). Volatility and commodity price dynamics. *Journal of Futures Markets*, 24(11), 1029-1047. https://doi.org/10.1002/fut.20120

12. Schneider, G., & Troeger, V. E. (2006). War and the World Economy. *Journal of Conflict Resolution*, 50(5), 623–645. https://doi.org/10.1177/0022002706290430

13. Sieczka, P., & Hołyst, J. A. (2009). Correlations in commodity markets. *Physical And Statistical Mechanics And Its Applications*, 388(8), 1621-1.

A Code Repository

The full source code of the project is available on GitHub:

`https://github.com/LorenzoDavideDeMartino/https-github.com-project-DeMartino`

The full source code of the project is available on GitHub.

The repository is organized following a modular and reproducible structure. Raw commodity price data, processed datasets, feature construction, model estimation, and evaluation steps are clearly separated into dedicated folders and scripts. This structure facilitates readability, maintenance, and verification of the analysis pipeline.

Installation instructions and software dependencies are documented in the repository through a `requirements.txt` file. The entire project can be executed end-to-end by running the main script (`main.py`), which reproduces all results deterministically from the raw input data. Detailed instructions on how to reproduce the results are provided in the `README` file.

A AI Usage Declaration

During the development of this project, AI tools (ChatGPT, Gemini, DeepL) were utilized to assist with specific technical and editorial tasks:

- **Code Support:** Assistance with refactoring for modularity, debugging, and optimizing the computational efficiency.
- **Writing & Formatting:** Improving the clarity of academic English and assisting with overleaf typesetting for tables and document structure.

AI tools were **not** used to generate synthetic data, fabricate results, or determine final modeling choices. All experiments were conducted locally, and the author assumes full responsibility for the methodology, code execution, and interpretation of findings.