Forecasting Stock Volatility

TEAM
Cameron Cinel
George Mitchell
Palak Arora
Thomas Polstra
Waleed Ahmed

The Erdos Institute Data Science Boot Camp 2023

What is volatility and why is it important?

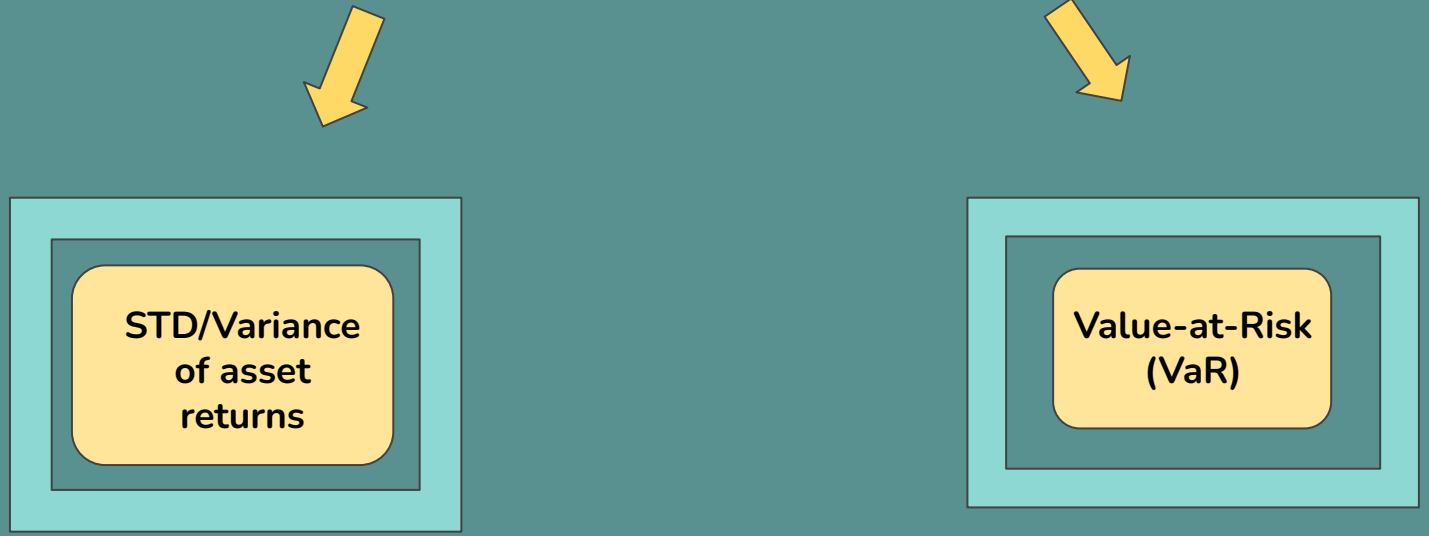
Volatility refers to the degree of variation or fluctuation in the price or value of a financial instrument, such as stocks, bonds, or commodities, over a specific period of time.

Risk Assessment:
Volatility is a key indicator of investment risk

Investment Decisions:
Depending on volatility investors can seek opportunities for quick profits or stable investments

Market Analysis:
By studying historical volatility patterns, financial analysts can identify trends to make predictions about future price movements.

Common Risk Measures



Data Collection and Pre-processing

Examined daily closing prices of stocks spread among the biotech, healthcare, industrial, and tech industries.

The closing price data was collected over all trade days from Jan. 5, 2010 to Dec. 30, 2022. A total of 3271 trading days.

Pre-processing:
- calculate log returns
- checked for stationarity of time-series

Date	JNJ	LLY	MRK	NVO	RHHBY	...
2010-01-05	43.12716674804688	23.49225807189941	22.52070617675781	9.98923397064209	13.39768409729004	...
2010-01-06	43.47796249389648	23.66582489013672	22.82372665405273	10.106612205505373	13.515424728393556	...
2010-01-07	43.16765213012695	23.545663833618164	22.86008071899414	10.020121574401855	13.351204872131348	...
...	...	...	...	...	...	...
2022-12-28	174.08499145507812	363.09771728515625	110.31306457519533	132.48841857910156	37.76811218261719	...
2022-12-29	174.9718780517578	364.8872375488281	110.05485534667967	134.12530517578125	38.1730842590332	...
2022-12-30	174.0751190185547	363.7141418457031	110.1839599609375	134.26419067382812	37.74883270263672	...

Project Scope and Objectives

Forecast the volatility of a portfolio of stocks using three different methods:

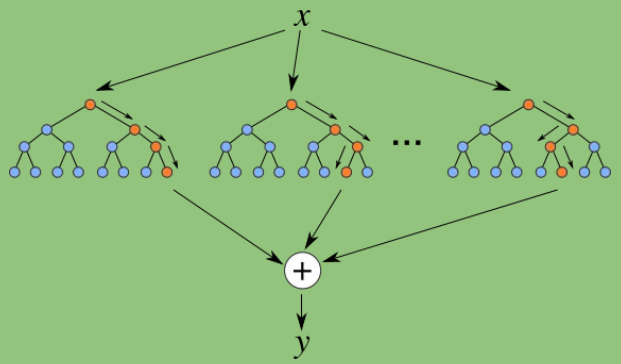
rolling window std deviation → benchmark

Random Forest	GARCH	LSTM
↓	↓	↓
Regression model	Autoregressive model	NN model

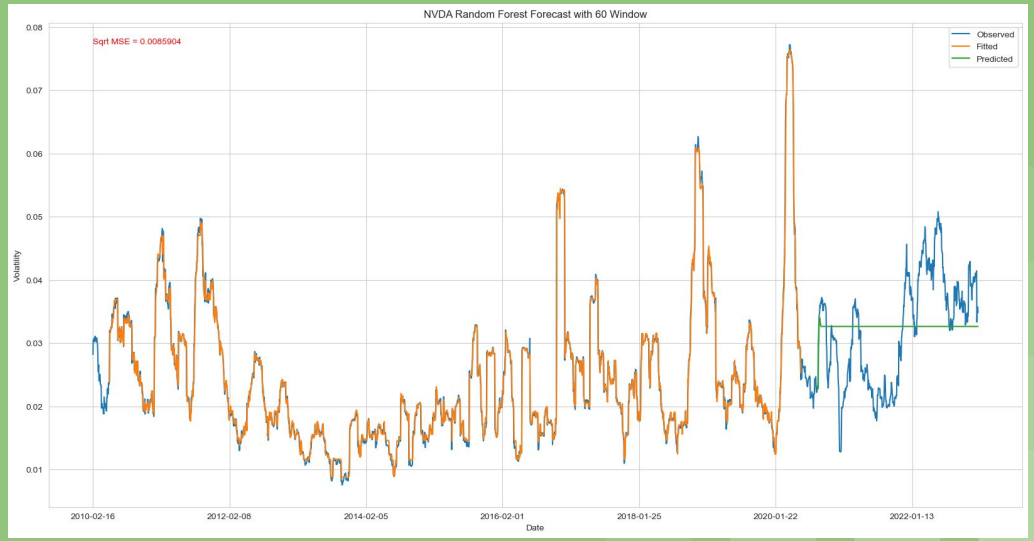
Backtest Random Forest and LSTM comparing mean squared error between projected variance and actual variance.

Backtest GARCH by analyzing modeling of Value at Risk.

Random Forest Results



- Convert time series to a supervised learning problem.
- Cross-validated for the best window size.
- Very strong fitted model.
- Predictions perform well on run-up, but stabilize to constant value long-term.

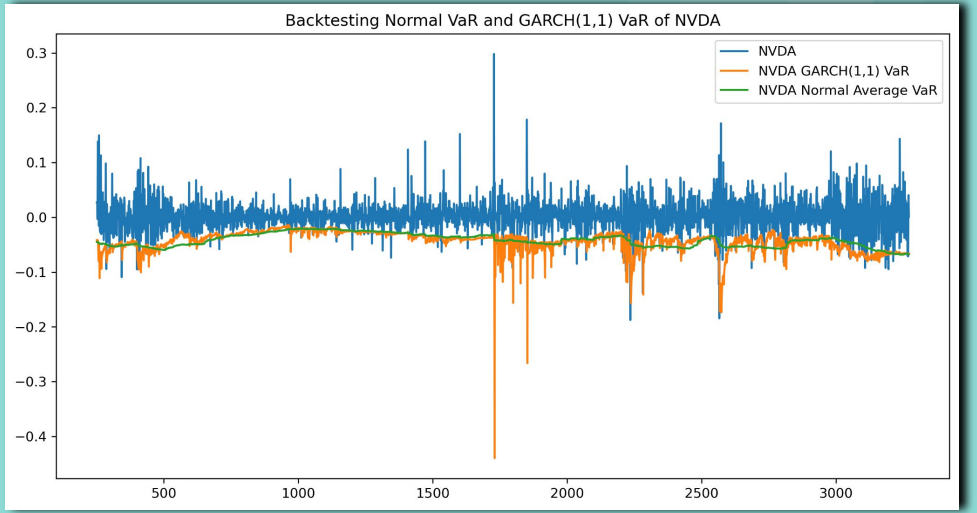


GARCH Results

GARCH exhibited similar failure ratios to baseline.

GARCH did a better job of modeling Value at Risk around extreme disruptions.

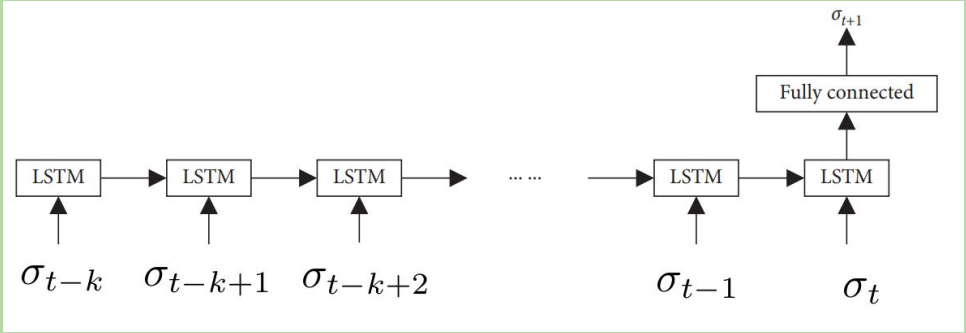
The baseline model performed better during consolidation phases.

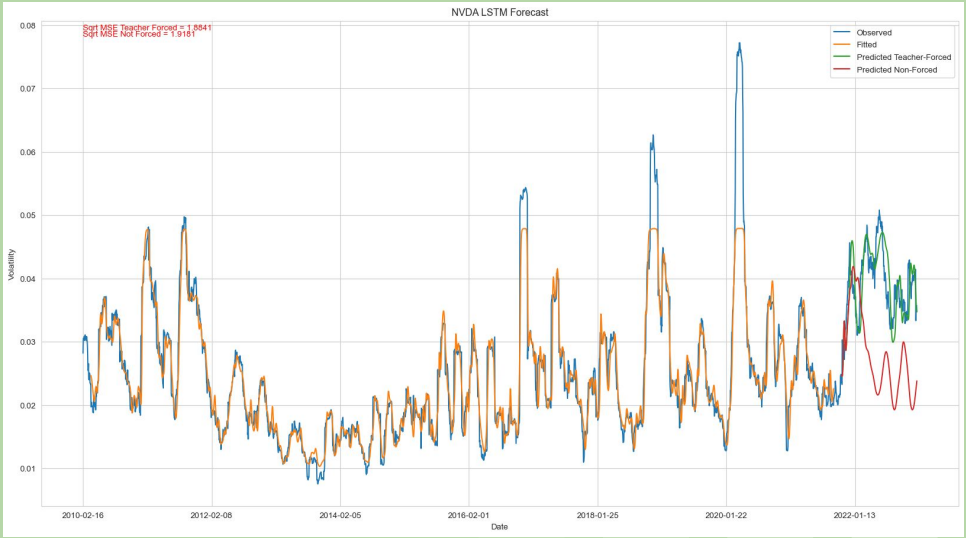


	NVDA (Normal VaR)	NVDA (GARCH VaR)
Failure Ratio	0.04899040052962595	0.042039059913935785
Bernoulli Test stat	0.06524	4.25133
Bernoulli Test p-value	0.79839	0.03922
Independent Test stat	7.1903	0.02441
Independent Test p-value	0.00733	0.87584

LSTM Results

- Trained 4 layer, 128 neuron LSTM with Adam optimizer
- Achieved a mean squared loss of 6.05 x 10^{-3} on the validation set and 5.86 x 10^{-2} on the test set
- Took significant amount of time to train, approximately 7.5 minutes per stock
- Able to autoregressively inference about 2-3 months with a 1 month warm-up
- Unable to fit to certain stocks (primarily industrial)





Summary

- The baseline model performed well for predicting volatility and value at risk over long periods of time.
- Random Forest was able to fit extremely well, but predictions became constant after a few horizons. Good for the early trend.
- GARCH(1,1) excelled in predicting value at risk following large movements in the asset price.
- LSTM was able to accurately predict volatility movements for some indexes for a two month time frame.