

ORIE 5741 Final Project: Prediction of Oil Prices

Harshavardhan Bapat (hsb57); Joel Dsouza (jnd74); Yash Ganatra (ybg3)

Problem Statement

Based on projected rising demand, the natural production decline from existing wells and decreases in drilling activity and industry investment – especially in the U.S. – the world’s oil needs could outpace production, with a possible shortfall in global oil supply as soon as next year! An undersupply potentially could put upward pressure on costs, impacting consumers, manufacturers and, generally, any process that utilizes oil. The aim of this project is to predict a oil prices and potentially an imminent oil crisis that may be caused by increased demand or reduced supply due to low production, oil spills or other external factors like weather.

Dataset

We will be using daily futures prices of WTI crude contracts with next month expiry, sourced from Bloomberg. This data is from January 1990 to December 2020, which includes several global and macroeconomic events such as the oil crises of 1990s and 2000s, global recession of 2008, as well as the recent COVID-19 pandemic and OPEC-Russia oil dispute of 2020. We will be compiling the data for our features from a variety of sources such as Bloomberg for financial data and volumes, Federal Reserve for production data estimates, NCDC for weather related features, and IATA for air traffic demand.

The features can be broadly classified as:

1. Pricing Data and possible lagged terms:
 - WTI Crude daily futures prices and volume traded
 - Gasoline pricing index in US
2. Supply Side:
 - OPEC, US, Russia oil production volumes
 - Global oil reserves
3. Demand Side:
 - GDP/GNP of world’s 10 largest economies
 - US and global industrial production
 - Air traffic volumes (ASKs)
4. Macroeconomic and other global features:
 - US Inflation index
 - Dummy variable for natural catastrophic events
 - Temperature at production sites

This is a preliminary feature list based on our economic intuition and the variables are subject to change on further data analysis.

Significance of Problem

The price of crude oil is primarily determined by demand and supply, but it is also influenced by a number of uncontrollable past, present and future events, such as climate change, stock levels, GDP growth, and political perspectives. These realities result in a market that is very variable and nonlinear, and the fundamental component of maintaining the complex dynamic is unknown. The extreme volatility in crude oil prices is difficult for humans to comprehend due to the nonlinear and complicated character of oil prices. We intend to use machine learning approaches to construct a model that can predict these oscillations with greater accuracy than a simple time series forecasting. Forecasting with high accuracy can help firms make more informed decisions for better planning and management.