# Optimal Stopping Time in Commodity Trading: A Machine Learning Approach

Vittorio Balestrieri

**Abstract**

This paper explores the application of machine learning techniques to predict the optimal stopping time in commodity trading, with a focus on the oil market. We collect historical oil price data, preprocess it, train a machine learning model to predict future prices, simulate trading based on the predictions, and analyze risk management strategies. The goal is to develop a robust trading strategy that maximizes profits and minimizes risks.

## 1 Introduction

Commodity trading, particularly in the oil market, presents unique challenges and opportunities for investors. Determining the optimal time to buy or sell commodities can significantly impact profitability. Traditional approaches rely on technical analysis, fundamental analysis, and market sentiment to make trading decisions. However, these methods may be subjective and prone to human biases.

In recent years, machine learning techniques have gained popularity in financial markets for their ability to analyze large datasets, identify patterns, and make data-driven predictions. In this project, we leverage machine learning algorithms to predict the optimal stopping time for trading oil futures. By accurately forecasting price movements and implementing effective risk management strategies, we aim to develop a profitable trading strategy.

## 2 Data Collection

The first step in our project is to collect historical oil price data. We obtain this data from Yahoo Finance using the `yfinance` library in Python. The dataset includes daily price information such as Open, High, Low, Close, Adjusted Close prices, and trading volume. The data spans a specified time period, typically several years, to capture a wide range of market conditions.

# 3 Data Preprocessing

Before training our machine learning model, we preprocess the raw data to ensure its quality and suitability for analysis. The preprocessing steps include:

## 3.1 Handling Missing Values

We check for missing values in the dataset and handle them appropriately. Missing values can occur due to various reasons, such as data collection errors or trading holidays. We use techniques such as interpolation or imputation to fill missing values or remove rows with missing data.

## 3.2 Calculating Returns

To train our machine learning model, we need to calculate the returns on the asset. Returns represent the percentage change in price from one day to the next and provide valuable information about the asset's volatility and trend. We calculate returns using the formula:

$$\text{Return}_t = \frac{\text{Price}_t - \text{Price}_{t-1}}{\text{Price}_{t-1}}$$

where $\text{Price}_t$ is the closing price on day $t$.

## 3.3 Feature Engineering

In addition to returns, we may generate additional features that could be predictive of future price movements. These features could include technical indicators such as moving averages, relative strength index (RSI), or Bollinger Bands. Fundamental factors such as supply-demand dynamics, geopolitical events, or macroeconomic indicators may also be considered.

# 4 Machine Learning - Predictive Modeling

With the preprocessed data, we train a machine learning model to predict the optimal stopping time for trading oil futures. Our choice of algorithm depends on various factors such as the dataset size, feature space, and desired predictive performance. Common machine learning algorithms used in financial forecasting include:

## 4.1 Random Forest

Random Forest is an ensemble learning technique that combines multiple decision trees to improve predictive accuracy and reduce overfitting. It is particularly effective for regression tasks, such as predicting asset prices. Random Forest models are robust to noisy data and can handle both numerical and categorical features.

### 4.2 Gradient Boosting

Gradient Boosting is another ensemble learning method that builds a sequence of weak learners (usually decision trees) to minimize the loss function iteratively. It is known for its high predictive accuracy and ability to capture complex relationships in the data. Gradient Boosting models are widely used in financial forecasting and have been shown to outperform traditional statistical methods.

### 4.3 Long Short-Term Memory (LSTM)

LSTM is a type of recurrent neural network (RNN) designed to capture temporal dependencies in sequential data. It is well-suited for time series forecasting tasks, such as predicting stock prices or commodity prices. LSTM models can learn from past price movements and adapt their predictions based on recent market conditions.

### 4.4 ARIMA

Autoregressive Integrated Moving Average (ARIMA) is a classical time series forecasting method that models the linear relationship between past observations and future values. ARIMA models are widely used for forecasting stationary time series data and can capture both short-term fluctuations and long-term trends.

## 5 Simulation and Optimal Stopping Strategy

Once we have trained our machine learning model, we simulate trading based on the predicted prices. The trading strategy aims to maximize profits while minimizing risks. A simple approach is to buy when the predicted price is lower than the current price and sell otherwise. However, more sophisticated strategies may incorporate additional factors such as transaction costs, market liquidity, and risk appetite.

## 6 Risk Management

Risk management is a crucial aspect of successful trading strategies. It involves identifying, assessing, and mitigating risks to protect capital and preserve profits. Common risk management techniques include:

### 6.1 Stop-Loss Orders

Stop-loss orders automatically sell a security when its price falls below a predefined threshold. They help limit losses and prevent emotional decision-making during periods of market volatility.

### 6.2 Position Sizing

Position sizing determines the amount of capital allocated to each trade based on risk tolerance and account size. It ensures that no single trade significantly impacts overall portfolio performance.

### 6.3 Diversification

Diversification involves spreading investments across different assets, sectors, or regions to reduce the impact of individual security risk. It helps mitigate losses from adverse market movements and improve portfolio stability.

### 6.4 Risk-adjusted Returns

Risk-adjusted returns measure the performance of an investment relative to its risk level. Common risk-adjusted metrics include the Sharpe ratio, which calculates the excess return per unit of risk, and the Sortino ratio, which focuses on downside risk.

## 7 Conclusion

In conclusion, this project demonstrates the application of machine learning techniques in predicting the optimal stopping time for trading oil futures. By collecting and preprocessing historical data, training predictive models, and simulating trading strategies, we aim to develop robust and profitable trading strategies. Risk management techniques play a crucial role in protecting capital and maximizing returns in commodity trading. Future research may explore more advanced machine learning algorithms, incorporate alternative data sources, or optimize trading strategies for specific market conditions.