Commodity Investing By Students

Cross-Metal Statistical Arbitrage Engine with LSTM Model Integration

Rishon Reddy Nimmakayala

04/2025

I. Overview

This project proposes the development of a quantitative framework and strategy that signals trading opportunities between cointegrated metals (Gold, Silver, Copper) through a statistical arbitrage lens. The goal is to identify historically co-moving metal pairs, measure their deviations in price behavior, and quantify the opportunity for mean-reverting relative value strategies.

II. Motivation

Throughout this semester, COINS has placed a significant interest and importance on returning to spread-based trades within the certain commodity divisions and it is expected that the growth of these types of trades will increase in the foreseeable semesters. The normal arbitrage spread models provide indicators based on fixed values and assume that a linear relationship exists between a commodities' corresponding macroeconomic variables such as geopolitics and weather, however we understand that these inputs are not linear which reduces the efficiency of these linear based models for commodity spread trades. In order to create a tool and indicator for the metals division that can sufficiently take into account macroeconomic data with a higher accuracy than linear-based models, I propose researching and creating Long-Short Term Memory Reconcurrent Networks which are known for effectively capturing temporal dependencies in time-series data.

III. LSTM Foundations & Technicals

A. Prologue: Recurrent Neural Network (RNN)

This project aims to create a version of a LSTM (Long-Short Term Memory) model, which is a formulated structure founded on the principles of a recurrent neural network (RNN). While normal neural networks take into consideration one input at a time and generate output, RNN's are specifically designed to tackle multiple inputs which make it extremely efficient for handling time-series data, such as forecasting spread data between two different metals. Normal neural networks are restricted to one input value, while RNNs have feedback loops which makes it possible to use input values of futures prices over time to make predictions. RNNs process inputs one at a time, while passing through hidden states and computing through fixed weights and biases which allows the network to remember information from previous events. However, there is a major issue with RNNs which is the exploding/vanishing gradient problem. When a RNN has multiple chains (suppose 50-day data from a commodity market), the gradient gets constantly multiplied by the weights, which makes it either too large if the weights are larger than 1, or extremely small if the weights are less than 1. As a result, RNNs struggle to handle long-term time-series data (such as a 1-year data for a commodity which is very important). To work around the vanishing/exploding gradient problem, LSTMs create three different gates to contain long term and short term memory.

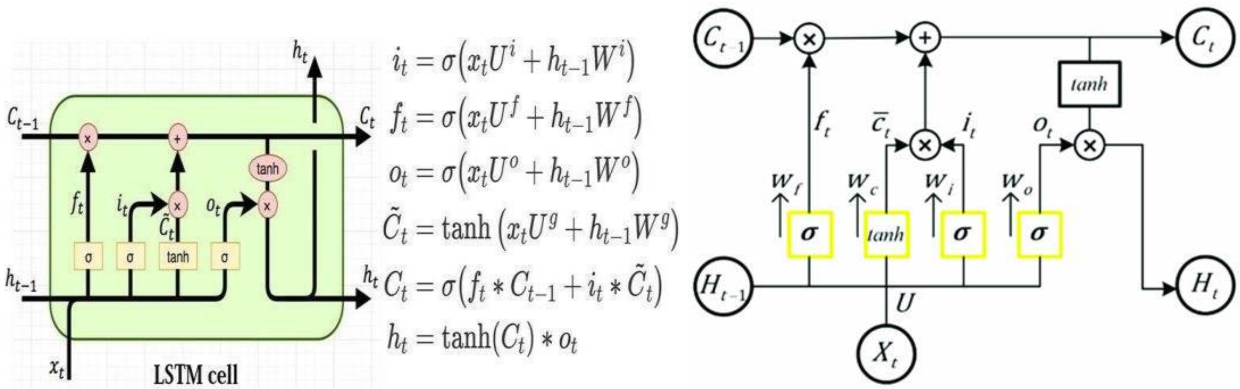
B. Long-Short Term Memory Technicals

Basic Overview

An LSTM is a specially designed recurrent neural network that is built to overcome the challenges of the vanishing/exploding gradient of a normal recurrent neural network. It uses a mathematical structure with three "gates": the forget gate, input gate, and output gate. These gates are built to control the flow of time-series data. These three gates allow the system to retain important information for long periods of time whilst also discarding irrelevant information. The goal of this project is to design an LSTM that will learn the various movements of price patterns of different metals and identify when prices between these metals will return to their normal relationships. A successful LSTM model will be a useful trading signal to support fundamental trades in the future.

Overview for Technical Individuals

Each LSTM contains a cell state which passes through long-term memory. Within the architecture, there exists different gates that process short-term memory and regulate the flow of it in and out of the cell state. Each of the three gates within each cell use nonlinear mathematical activation functions to process the data. The first gate that information passes through is the short-gate. This 'gate' determines which of the temporary short-term data gets discarded. It combines data from the previous short-term memory with the current input (start of the corresponding chain), and passes it through a sigmoid activation function and multiplies it to the long-term memory stored in the cell state. The sigmoid function is a function that takes any x-axis coordinate and outputs a value between zero and one, basically representing the strength value of the memory. The next stage of the chain is the input gate which determines the percentage of new information that should be added to the existing information. A candidate memory is passed through the Tanh function (the Tanh/hyperbolic tangent function takes in x coordinate and assigns value between -1 and 1). The output from the function is then sent to the sigmoid function which is then applied to the long-term memory contained in the cell state. These gates allow for the LSTM model to understand when and how much information gets stored to the long-term memory. The last stage of the process is through the output gate which updates the hidden state/short-term memory. The new-long term memory gets passed through the Tanh function which then produces an output of potential short-term memory. It then gets passed through a sigmoid function which determines how much of the potential short-term memory gets added to the long-term memory. This short-term memory then becomes the output at the current step and the input for the next step.



IV. Metals Division

To demonstrate the effectiveness of this project, it will be tested on multiple spreads with more focus on the gold-silver spread. Gold and silver adopt an extremely interesting relationship as they are both tied to the same macroeconomic data such as U.S. interest rate expectations and inflation data. Despite this, their prices tend to deviate from each other due to their corresponding industrial use and data concerning. These factors make the gold-silver spread an excellent candidate for a macro-influenced mean-reversion indicator to quantify and supplement a fundamentals based trade. The model will first define the spread as the difference of gold futures price minus a scaled value of the silver futures price: $Spread = P_{gold(t)} + \beta * P_{silver(t)}$ .The hedge ratio (β) will be determined through cointegration testing, which aligns the metals based on their historical futures prices and behavior. Time series of the spread (amount of data yet to be determined) will be imputed into the LSTM model including variables such as real interest rates, the U.S. dollar index, and Fed data. The goal of the LSTM model is to learn the temporal patterns visualized in the spread and create a bearish/bullish indicator useful for trades forecasting the potential direction of price. Using this LSTM model for gold-silver spread observations will help the metals division quantify and identify the timing of mean reversion more accurately and adjust the potential strategy based on the non-linear relationships between macroeconomic and political information.