Imperial College London

Department of Earth Science and Engineering

MSc in Applied Computational Science and Engineering

Independent Research Project
Project Plan

Enhancing Commodity Price Predictions: Advanced Machine Learning Models for Oil and Precious Metals Forecasting

by

Peifeng Tan

Email: pt623@imperial.ac.uk
GitHub username: acse-pt623
Repository: https://github.com/acse-pt623/irp-pt623-external.git

Supervisors:

Antony Sommerfeld
Yves Plancherel

June 2024

Abstract

Copper is an indispensable raw material in modern industry and is widely used in a variety of sectors, including power electronics, household appliances, transportation and energy, machinery manufacturing and construction. However, as copper is both a commodity and a financial product, its price forecasting becomes exceptionally complex. This study aims to improve copper price forecasting by considering a range of influencing factors, including global supply and demand, energy costs, financial indicators and related commodity prices. Traditional econometric models such as ARIMA and ARMA have limitations in capturing nonlinear features in price changes. Therefore, we explore advanced machine learning models such as LSTM and the attention mechanism transformer. inspired by the success of the attention mechanism transformer in the field of natural language processing and image processing, we propose to use a transformer-like model incorporating a multi-head attention mechanism to capture feature-level interdependencies. In addition, we will incorporate sentiment analysis information from social media to improve prediction accuracy. This approach aims to provide a more robust and interpretable model for copper price forecasting. In the future, it is hoped that this model can be applied to other commodity forecasts.

1 Introduction

Oil and metals are crucial raw material and strategic resource for national economic development and energy industry. As a kind of vital raw materials, metals find wide applications in power electronics, household appliances, transportation energy, machinery manufacturing, and construction. [27, 22, 14] Especially, copper which is a key industrial raw material, the global demand for it has continued to increase worldwide. Its electrical conductivity in electrical equipment, corrosion resistance in construction and versatility in manufacturing make it an indispensable resource.

However, copper is more than just an industrial raw material. Its trading in the commodity market also gives it the attributes of a financial product. This has made the forcasting of copper prices more complicated and difficult to control. Various factors such as supply and demand in the market, political factors, economic policies, and speculative behaviors can affect the price fluctuation of copper, which increases the difficulty of price forecasting.

Based on the complexity of copper prices, this study will consider a variety of influencing factors to forecast market changes in copper prices. First, as a commodity, global supply and demand, as well as the cost of energy for smelting and production, will influence the medium- to long-term trend of non-ferrous metal prices. Specifically, we consider coal prices, gasoline prices, crude oil prices, and natural gas prices [3]. Second, as a financial product, copper's price fluctuations are also influenced by a variety of financial factors. We will consider factors such as the U.S. dollar exchange rate, consumer indices, and price levels. These include the GDP and EURUSD exchange rates, price conditions in the LME and COMEX markets, the U.S. Dow Jones Industrial Average, the price of Brent crude oil, gold, silver and iron ore futures, as well as the U.S. Dollar Index, and the U.S. Dollar to Renminbi (USD/CNY) and Australian Dollar (USD/AUD) exchange rates [12, 8]. Finally, copper prices are also influenced by other related products such as aluminum prices, gold prices, lead prices, iron ore prices, zinc prices, tin prices, silver prices and nickel prices. Price fluctuations in these products tend to have a strong correlation with the price of copper. Specifically, we take into account price movements of other major commodities in the international markets as well as relevant financial indicators [20, 6].

In the past, many researchers used econometric models such as ARIMA [2] and ARMA [25] to forecast copper prices. However, these linear models cannot effectively capture the nonlinear features in price changes. With the development of machine learning, researchers began to try to use models such as NARX [18, 10], but these models usually require predefined model structures and have certain limitations. Subsequently, machine learning methods such as ANN [26], KNN [1, 7], SVM [9], Gaussian processes [13], decision trees and random forests [5, 11] have been applied to copper price prediction with good results. However, these methods also face the problem of not being able to effectively capture the pre- and post-relationships of time series.

In order to better solve the above problems, RNN was introduced into time series prediction —citerumelhart1986learning, but RNN suffers from the problem of gradient vanishing when dealing with long sequences [4].LSTM network, as an improvement of RNN, can better capture long time dependencies and has achieved remarkable results in time series prediction [15]. With further reference to models such as CNN-LSTM [23, 16] and DA-LSTM [21], our first goal is to construct an LSTM network with a view to obtaining better prediction results.

Recently, attention mechanisms have achieved great success in the fields of natural language processing and image processing. Especially in the field of natural language processing, the emergence of Transformer [17, 24] has contributed to the epochal achievement of large language models such as GPT. Inspired by this, we plan to try to use a model similar to Transformer to handle time series forecasting of copper prices. We will innovatively try to use a multi-attention mechanism to extract feature-level interdependencies instead of focusing only on time-level dependencies refering on DA-LSTM [21].This will help to explain and understand the impact of individual factors on price.

In addition, this study also plans to use information from social media to assist in predicting copper prices. Referring to studies such as “Sentiment Analysis with LSTM - NLTK” [19], we introduce the element of sentiment analysis into the model to analyze the tendency of sentiment on social media in order to further improve the accuracy of the prediction.

2 Methods

In this work, we wrote a simulation package for calculating the position of a particle in vertical motion as shown in Fig. 1. We compute the position $y(t)$ at time t using:

$$y(t) = v_0 t - \frac{1}{2} g t^2, \tag{1}$$

where v_0 is the initial velocity, and g is the acceleration due to the Earth’s gravity.

Python code we implemented for computing $y(t)$ is:

```
def position(t, v0=0, g=9.81):
    """Position of a particle in verticle motion."""
    return v0*t - 0.5*g*t**2
```

We made our computational workflows reproducible using Jupyter [26].

3 Results

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede.

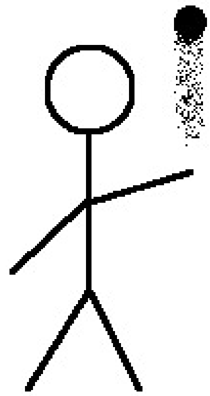

Figure 1: Particle in vertical motion.

Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4 Discussion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

5 Conclusion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

References

[1] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

[2] Dimitrios Asteriou and Stephen G. Hall. Arima models and the box-jenkins methodology. 2016.

[3] Niaz Bashiri Behmiri and Matteo Manera. The role of outliers and oil price shocks on volatility of metal prices. *Resources Policy*, 46:139–150, 2015.

[4] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.

[5] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[6] Daniel Buncic and Carlo Moretto. Forecasting copper prices with dynamic averaging and selection models. *The North American Journal of Economics and Finance*, 33:1–38, 2015.

[7] Sheng Chen, XX Wang, and Chris J Harris. Narx-based nonlinear system identification using orthogonal least squares basis hunting. *IEEE Transactions on Control Systems Technology*, 16(1):78–84, 2007.

[8] Cetin Ciner. Predicting white metal prices by a commodity sensitive exchange rate. *International Review of Financial Analysis*, 52:309–315, 2017.

[9] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.

[10] Eugen Diaconescu. The use of narx neural networks to predict chaotic time series. *Wseas Transactions on computer research*, 3(3):182–191, 2008.

[11] Juan D Díaz, Erwin Hansen, and Gabriel Cabrera. A random walk through the trees: Forecasting copper prices using decision learning methods. *Resources Policy*, 69:101859, 2020.

[12] Jeffrey A Frankel and Andrew K Rose. Determinants of agricultural and mineral commodity prices. *HKS Faculty Research Working Paper Series*, 2010.

[13] Roger Frigola and Carl Edward Rasmussen. Integrated pre-processing for bayesian nonlinear system identification with gaussian processes. In *52nd IEEE Conference on Decision and Control*, pages 5371–5376. IEEE, 2013.

[14] Xuyuan Han, Zhenya Liu, and Shixuan Wang. An r-vine copula analysis of non-ferrous metal futures with application in value-at-risk forecasting. *Journal of Commodity Markets*, 25:100188, 2022.

[15] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[16] Fei Li, Hanlu Zhou, Min Liu, and Leiming Ding. A medium to long-term multi-influencing factor copper price prediction method based on cnn-lstm. *IEEE Access*, 11:69458–69473, 2023.

[17] Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang, and Xifeng Yan. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in neural information processing systems*, 32, 2019.

[18] Tsungnan Lin, Bill G Horne, Peter Tino, and C Lee Giles. Learning long-term dependencies in narx recurrent neural networks. *IEEE transactions on neural networks*, 7(6):1329–1338, 1996.

[19] Saloni Mohan, Sahitya Mullapudi, Sudheer Sammeta, Parag Vijayvergia, and David C. Anastasiu. Stock price prediction using news sentiment analysis. In *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 205–208, 2019.

[20] Ali Can Ozdemir, Kurtuluş Buluş, and Kasım Zor. Medium-to long-term nickel price forecasting using lstm and gru networks. *Resources Policy*, 78:102906, 2022.

[21] Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison Cottrell. A dual-stage attention-based recurrent neural network for time series prediction, 2017.

[22] Qiushi Qu, Limao Wang, Zhi Cao, Shuai Zhong, Chufu Mou, Yanzhi Sun, and Chenran Xiong. Unfolding the price effects of non-ferrous industry chain on economic development: A case study of yunnan province. *Resources Policy*, 61:1–20, 2019.

[23] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28, 2015.

[24] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[25] Peter Whittle. Hypothesis testing in time series analysis. *(No Title)*, 1951.

[26] Hong Zhang, Hoang Nguyen, Diep-Anh Vu, Xuan-Nam Bui, and Biswajeet Pradhan. Forecasting monthly copper price: A comparative study of various machine learning-based methods. *Resources Policy*, 73:102189, 2021.

[27] Meirui Zhong, Ruifang He, Jinyu Chen, and Jianbai Huang. Time-varying effects of international nonferrous metal price shocks on china's industrial economy. *Physica A: Statistical Mechanics and its Applications*, 528:121299, 2019.