Copper price forecasting

Data Description

This paper adopts the copper prices in RMB per metric ton from October 31, 2011, to January 1, 2021, excluding public holiday and closed days, i.e., both Saturday and Sunday, with a total of 480 weekly observations. The other variables used to predict the copper price are crude oil price, inflation rates of US, inflation rates of China, coffee price, USD/CLP, USD/PEN, USD/CNY, USD/EURO, copper spot price from LME and Yangtze River market, gold price, silver price, nickel price, aluminum price, zinc price, and iron price, respectively.

Data preprocessing

At this stage, we do time lag establishment, feature scaling, and train/test split. Firstly, according to previous literature on forecasting time series, time lag has a significant impact on forecasting results (Bouktif et al., 2018; Rubaszek et al., 2020). In this research, we consider the previous one to four weeks' historical copper prices as input factors respectively. Furthermore, the feature with the larger scale will dominate the other when the features are on different scales. To cope with this problem, we used min-max method to scale the features within a specific range to truly present the impact of each variable change on copper prices.

Finally, the predictive model is also very important to maintain a certain degree of effectiveness for out-of-sample. The convenient ratios of splitting samples such as 70:30%, 80:20%, and 90:10% (Alameer et al., 2019). From the perspective of data volume, this research divided the training and test data into the ratio of 80:20, in which the first 324 observations (from October 31, 2011 to February 13, 2019) were used for training the model, whereas the final 36 observations (from February 20, 2019 to January 1, 2021) were employed to test the performance and accuracy of the proposed model.

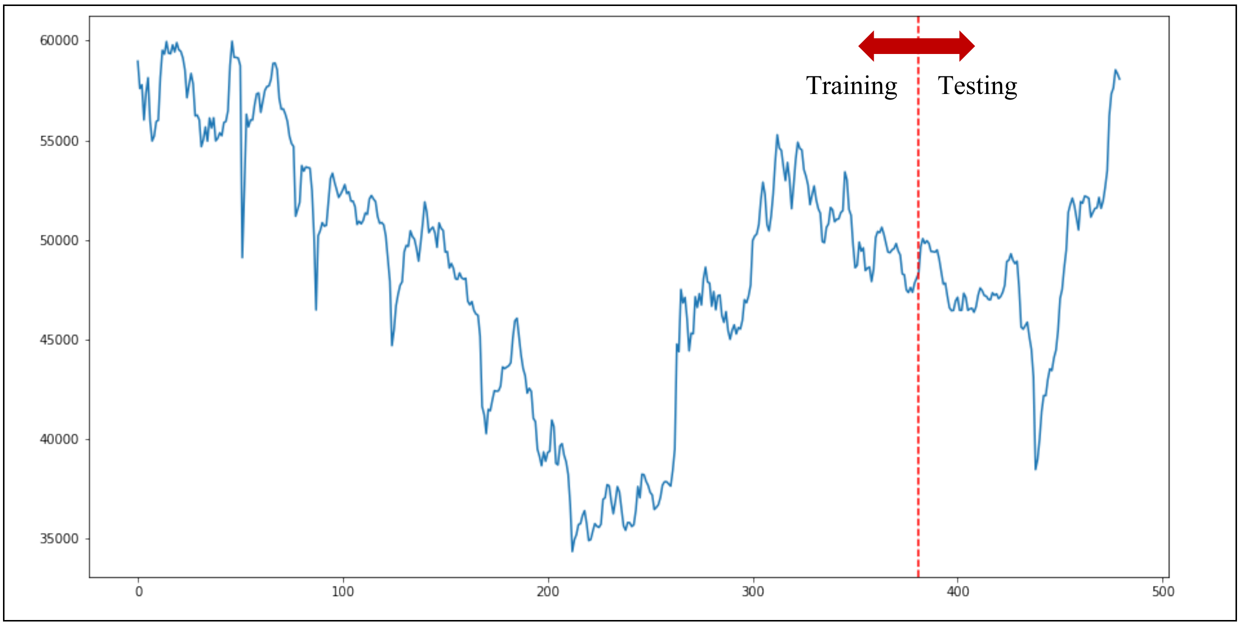

Figure 1. Historical copper prices from October 31, 2011 to January 1, 2021

Experiment and Insights

Currently, we applied three types of models like Autoregressive Integrated Moving Average model (ARIMA), Support Vector Regression (SVR), and Single-hidden Layer Feed-forward Neural Networks (SLFN) to forecast the price of copper in the coming week. Furthermore, the root mean square error (RMSE) is utilized to verify the effectiveness and accuracy between these models for forecasting copper price fluctuations. Each model considers the copper price data from the previous one to four weeks as the input factor and utilizes out-of-sample for testing. Table 1 summarizes the out-of-sample performance measure of the models used in this study. The results indicated that SLFN can get the best results among these prediction models.

From Figure 2, these models are relatively incapable of accurately grasping the trend of copper prices. We suspect that there is concept drifting in the data, which means that the fluctuation pattern of the data will change over time. In other words, the model built on old data may be inconsistent with the new data. Thus, many methods are used to split data to facilitate dynamic training of the model, named instance weight (Klinkenberg 2004), detection concept change point (Kosina and Gama 2015), sliding window (Fornaciari and Grillenzoni 2017), and monitoring two different time window distribution (Gama and Kosina 2014). Regarding the incremental learning requirement, the sequence-based moving window (Babcock et al., 2001) is often used to handle the data expiration problem.

Subsequently, we will use a sliding window mechanism to adapt the model to changes in data (Gama et al., 2014; Kashani et al., 2012), i.e., the method is used to discard old data and update the model based on the new data .In addition to testing the hypothesis that copper price time-series data has concept drifting, the implemented model can also respond to copper price fluctuation patterns at different times.

Table 1. Out-of-sample performance comparison

Models	RMSE			
Lag time	1	2	3	4
ARIMA	0.267	0.263	0.257	0.257
SVR with linear kernal	0.157	0.109	0.094	0.113
SVR with polynomial kernal	0.336	0.760	0.767	0.590
SVR with RBF kernal	0.163	0.338	0.324	0.490
SLFN	0.110	0.092	0.111	**0.082**

Values in bold represent the best model

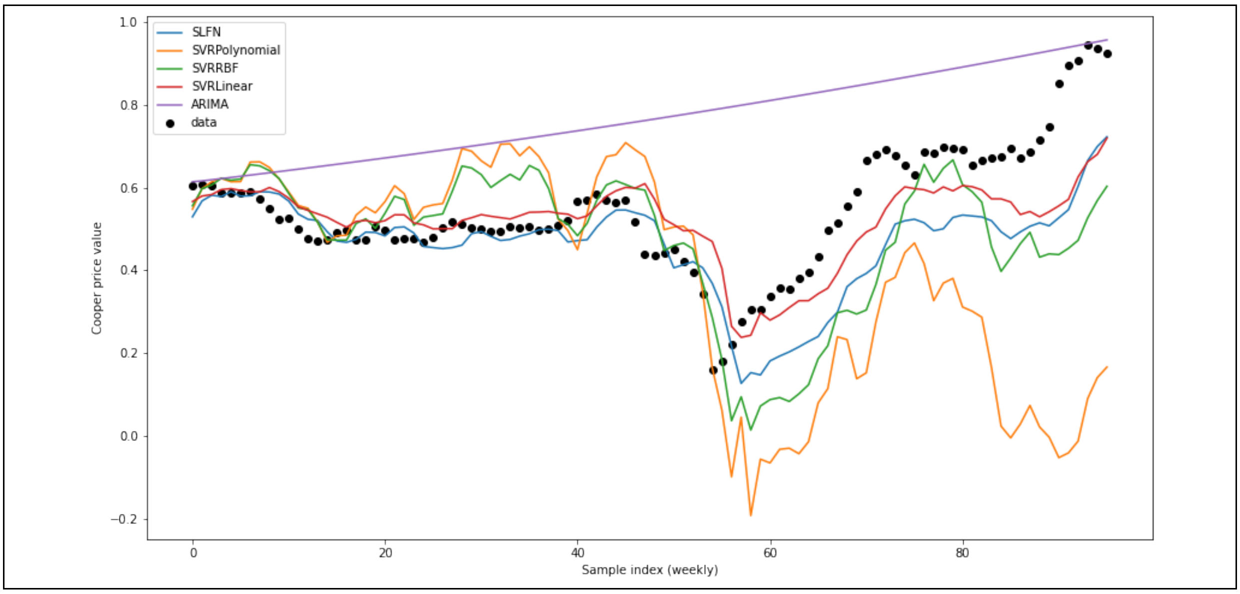

Figure 2. Actual and forecasted values during the test phase via ARIMA, SVR with linear kernal, SVR with polynomial kernal, SVR with RBF kernal, SLFN

Reference

Alameer, Z., Elaziz, M. A., Ewees, A. A., Ye, H., & Jianhua, Z. (2019). Forecasting Copper Prices Using Hybrid Adaptive Neuro-Fuzzy Inference System and Genetic Algorithms. *Natural Resources Research*, *28*(4), 1385–1401.

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, *11*(7), 1636.

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys (CSUR)*, *46*(4), 1–37.

Kashani, M. N., Aminian, J., Shahhosseini, S., & Farrokhi, M. (2012). Dynamic crude oil fouling prediction in industrial preheaters using optimized ANN based moving window technique. *Chemical Engineering Research and Design*, *90*(7), 938–949.

Rubaszek, M., Karolak, Z., & Kwas, M. (2020). Mean-reversion, non-linearities and the dynamics of industrial metal prices. A forecasting perspective. *Resources Policy*, *65*, 101538.