A support vector regression model for time series forecasting of the COMEX copper spot price

Esperanza García-Gonzalo[1][0000-0002-3194-4448], Paulino José García Nieto[1][0000-0001-8880-6348], Javier Gracia Rodríguez[2][0000-0001-9548-7322], Fernando Sánchez Lasheras[1][0000-0002-7052-2811] and Gregorio Fidalgo Valverde[2][0000-0002-7106-5747]

[1] University of Oviedo, Faculty of Sciences, c/ Federico García Lorca 18, 33007 Oviedo, Spain
espe@uniovi.es
pjgarcia@uniovi.es
sanchezfernando@uniovi.es
[2] University of Oviedo, School of Mining, Energy and Materials Engineering c/ Independencia 13, 33004 Oviedo, Spain
graciajavier@uniovi.es
gfidalgo@uniovi.es

Abstract. The price of copper is unstable but it is considered an important indicator of the global economy. Changes in the price of copper point to higher global growth or an impending recession. In this work, the forecasting of the spot prices of copper from the New York Commodities Exchange (COMEX) is studied using a machine learning method, support vector regression (SVR) coupled with different model schemas (recursive, direct and hybrid multi-step). Using these techniques, three different time series analysis are built and its performance compared. The numerical results show that the hybrid direct-recursive obtains the best results.

Keywords: New York commodity exchange (COMEX), Support vector machine (SVM), Time series analysis, Commodities price forecasting

1 Introduction

Nonferrous metals are essential raw materials that are crucial for measuring the global economy. However, these materials, such as fossil fuels, are a limited resource. The production of nonferrous metals is strongly affected by several factors: supply, demand and share prices of non-ferrous metal companies. Copper is one of the main metal commodities and a nonferrous metal that is traded in the physical futures trading exchanges [1-3]: the New York Commodity Exchange (COMEX), the London Metal Exchange (LME) and the Shanghai Futures Exchange (SHFE). Prices are related to the demand and supply of this metal around the world, but they are also affected by the price of currencies and movements in investments that are linked to temporary price fluctuations that may be affected by variations in the economic cycle. [4-6].

The chemical symbol for copper is Cu and its atomic number is 29. From the physical point of view it is a ductile, malleable and soft metal that stands out for its high electrical and thermal conductivity. The color of pure copper is orange with a pink tint. Given its high conductivity, copper is used directly as such in construction. It is also used as part of alloys. For example, with silver as a material for jewelry pieces, with nickel to make coins and as cupronickel does not corrode in seawater, for various elements of marine use. It is frequently used to manufacture mechanical, electrical and medical equipment. Copper is third in the ranking of most used metals. The first is iron and the second, aluminum, and is used mostly for electrical applications. Porphyry copper is the main source of copper in the world. It is extracted in open-pit mines as copper sulfides. Chile is the main copper producer in the world, followed by Peru, China, the Democratic Republic of the Congo, and the USA. Copper needs are increasing in developing countries, but reserves do not seem able to meet this growing demand [4,5].

Copper has been in use for 10,000 years, but massive copper mining begun around 1900, and half of the total has been mined in the last 30 years. Although the amount of copper on Earth is very large with approximately 1000 tons in the surface layer of the Earth's crust, only a small part is accessible with current techniques. Currently, it is estimated that there are copper reserves for 30 years according to some estimates while others grant it up to 60 years, at the present rate of consumption growth. Currently, copper from recycling is a substantial part of the produced copper. This makes it difficult to estimate the role it will play in copper production and what part of the needs it will be able to cover. (see Fig. 1).

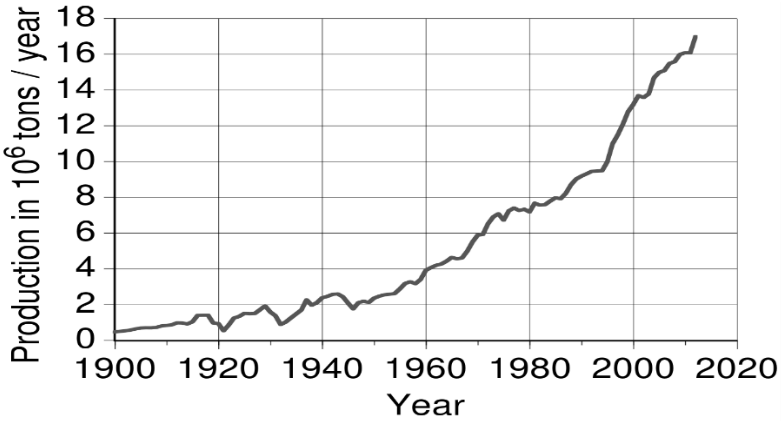


Fig. 1. World observed production trend of Copper

Copper appears in sulfides, in particular bornite (Cu_5FeS_4) and chalcopyrite ($CuFeS_2$), but also in chalcocite (Cu_2S) and covellite (CuS). But on average, the copper concentration does not exceed 0.6%. To increase their concentration, these minerals are crushed and subjected to *bioleaching* or *froth flotation* processes, which increases the concentration of Cu up to 15%. Heating the resulting material in flash smelting with silica removes much of the iron in the slag. This procedure converts iron sulfides into oxides, which, when reacting with silica, form a silicate that is part of the slag. The result is a *copper matte* that contains Cu_2S and that is roasted, obtaining oxides from the sulfides according to the chemical reaction [4,5]:

$$2\ Cu_2S + 3\ O_2 \rightarrow 2\ Cu_2O + 2\ SO_2$$

along with the following chemical reaction:

$$2\ Cu_2O \rightarrow 4\ Cu + O_2$$

blistering copper is obtained from copper oxide. The so-called Sudbury matting process only converts half the sulfur to oxide, but this oxide is used to remove the sulfur that was left as oxide. In a later step, the copper is electrolytically refined and gold and platinum are obtained from the anode mud because reducing the copper oxide to metal is relatively easy. This is done by blowing natural gas to eliminate the remaining oxygen and the product of this process is electro refined to obtain pure copper [4,5]:

$$Cu^{2+} + 2\ e^{-} \rightarrow Cu$$

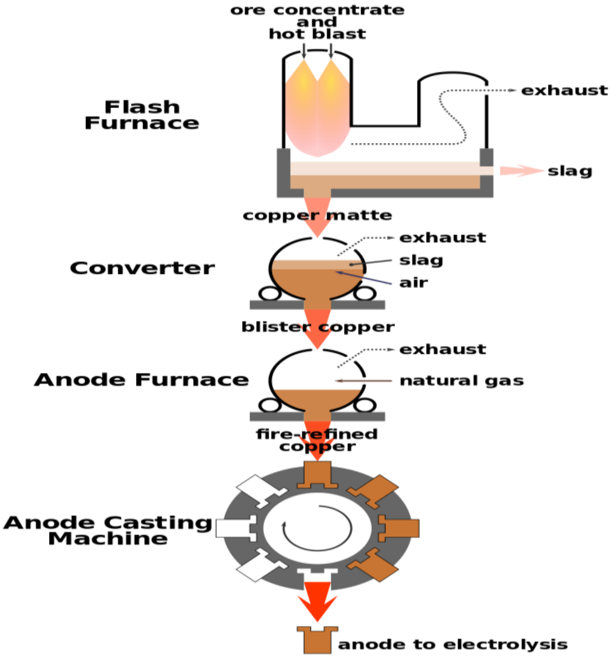


Fig. 2. Scheme of flash smelting process

The price of copper is unstable but it is considered an important indicator of the global economy. The price of basic resources is directly affected downward when demand expectations fall in times of crisis. Changes in the price of copper point to higher global growth or an impending recession. Several methodologies have been used for metal price forecasting. Dooley and Lenihan [7] used two time-series forecasting techniques to conclude that the forecast obtained with ARIMA method gives better results than lagged forward price modelling. Cortazar and Eterovic [8] proposed multicommodity models that forecasted long term prices for silver and copper.

On the other hand, Khashei et al. [9] prefer artificial neuronal networks for time series forecasting. Ma et al. [10] proposed a grey model, optimized by particle swarm algorithm, to forecast iron ore import and consumption in China. Kriechbaumer et al. [11] decompose time series into its frequency and time domain to capture this cyclic behaviour dominant in the metal market. Finally, Sánchez Lasheras et al. [12] examine the forecasting performance of ARIMA model and two different neuronal networks to forecast the COMEX copper spot price.

In this article, a new methodology to foretell the COMEX copper spot price has been built and implemented. This paper introduces a novel methodology to estimate, the copper price by means of support vector machine regression (SVR) used for time series analysis [13,14] coupled with three different schemas: recursive multi-step, direct multi-step, and direct-recursive hybrid. The proposed method uses a kernel-penalized optimization of all hyperparameters in SVR identifying nonlinear input features with success.

2 Materials and methods

2.1 The dataset

Data from this study are the COMEX price of copper [15] in the period of time that starts in January 1960 and finishes in October 2019.

2.2 Support vector regression (SVR) for time series analysis

This section presents $\varepsilon -$ SVR [16,17] for time series analysis. Given a set of time series data, a training set consisting of a continuous dependent variable $y_i \in \Re, \forall i = 1, 2, ..., m$ and covariates $\mathrm{x}_i \in \Re^p, \forall i = 1, 2, ..., m$ can be constructed by taking p lags of y_i. The method $\varepsilon -$ SVR constructs a function $f\left(\mathrm{x}\right) = \mathrm{w}^T \mathrm{x} + b$, $\mathrm{w} \in \Re^n, b \in \Re$ that has at most a deviation of ε from y_i for all x_i, and is as flat as possible [16-19]. Flatness is encouraged by minimising the Euclidean norm of **w**, while model fit is achieved by penalising the sum of the deviations higher than ε. The $\varepsilon -$ SVR method aims at solving the following optimisation problem [16-19]:

$$\min_{\mathrm{w},b,\xi,\xi^*} \frac{1}{2}\left\|\mathrm{w}\right\|^2 + C\sum_{i=1}^{m}\left(\xi_i + \xi_i^*\right) \tag{1}$$

such that

$$\begin{cases} y_i - \left(\mathrm{w}^T \mathrm{x}_i + b\right) \geq \varepsilon + \xi_i, & i = 1, 2, ..., m \\ \left(\mathrm{w}^T \mathrm{x}_i + b\right) - y_i \geq \varepsilon + \xi_i^*, & i = 1, 2, ..., m \\ \xi_i, \xi_i^* \geq 0, & i = 1, 2, ..., m \end{cases} \tag{2}$$

where the slack variables $\xi, \xi^* \in \Re^m$ are introduced for each training vector in order to allow deviations higher than ε , but penalising this deviations in the objective function. The parameter C controls the trade-off between model fit and complexity reduction [16-19].

Nonlinear model estimation can be obtained by mapping the data to a higher dimensional feature space H. Instead of using a projection function, expressions (1) and (2) can be represented in its dual form, where the data points appear only in the form of dot products. The mapping is performed by a kernel function $K\left(\mathrm{x}_i, \mathrm{x}_j\right)$ which defines an inner product in H [19]. Once applying the Karush–Kuhn–Tucker (KKT) conditions, the following dual formulation is obtained:

$$\max_{\alpha,\alpha^*} \sum_{i=1}^{m} y_i \left(\alpha_i - \alpha_i^*\right) - \varepsilon \sum_{i=1}^{m} \left(\alpha_i + \alpha_i^*\right) - \frac{1}{2} \sum_{i,s=1}^{m} \left(\alpha_i - \alpha_i^*\right)\left(\alpha_s - \alpha_s^*\right) K\left(\mathrm{x}_i, \mathrm{x}_j\right) \tag{3}$$

such that

$$\begin{cases} \sum_{i=1}^{m} \left(\alpha_i - \alpha_i^*\right) = 0 & \\ 0 \leq \alpha_i \leq C, & i = 1, 2, ..., m \\ 0 \leq \alpha_i^* \leq C, & i = 1, 2, ..., m \end{cases} \tag{4}$$

The decision rule $f\left(\mathrm{x}\right)$ for a new sample **x** is thus:

$$f\left(\mathrm{x}\right) = \sum_{i=1}^{m} \left(\alpha_i - \alpha_i^*\right) K\left(\mathrm{x}, \mathrm{x}_i\right) + b \tag{5}$$

Several usual functions used as kernels [16–19] are:

- Polynomial:

$$K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left(\sigma \mathbf{x}_i \cdot \mathbf{x}_j + a\right)^b$$

- Radial basis function (RBF):

$$K\left(\mathbf{x}_i,\mathbf{x}_j\right)=e^{-\sigma\left\|\mathbf{x}_i-\mathbf{x}_j\right\|^2}$$

- Sigmoid:

$$K\left(\mathbf{x}_i,\mathbf{x}_j\right)=\tanh\left(\sigma\mathbf{x}_i\cdot\mathbf{x}_j+a\right)$$

Among a variety of kernel functions available, the radial basis function (RBF) kernel is chosen in many applications and in this research due to its superior performance [17-19].

Moreover, representative parameters of the SVM approach can be summarized as [16,19]:

- Regularization constant (*C*): also term *cost function*. This factor defines balance between the margin, that defines the model flatness and the importance of the slack variables and is linked to the training error. Furthermore, this constant *C* must be chosen a priori, being a parameter of the machine learning.
- ε parameter: this factor controls the width of the error margin allowed. The second term of the objective function (see Eqs. (1) and (2)) is known as empirical error determined by means of the $\varepsilon -$ insensitive loss function, which indicates that it does not disregard errors below ε (i.e. to a distance ε of the true value).
- *a*, *b* and σ : these factors determine the expression of the different kernels in the subsequent model.

2.3 Computational procedure and numerical schemes

The training dataset comprises the data from January 1960 to August 2018 while the forecasted monthly prices start in September 2018 and end in August 2019. Thus, in this particular case, we must forecast twelve steps ahead. Thus, we will be performing multi-step forecasting. Three different strategies for the building of the training data will be used:

1. Direct multi-step forecast;
2. Recursive multi-step forecast; and
3. Direct-recursive hybrid forecast.

We have started using only one variable. The obvious variable is the copper price in previous years. Once this model is constructed, we have tried to improve the best model adding new variables from the dataset but no significant improvement was observed and thus, we have not included these other models in this study. Next, we are going to describe below the three different strategies for this problem of multi-step forecast.

Direct multi-step forecast

In this scheme, we construct different models for the different ahead forecasting:

$$\begin{aligned}
pred\left(t+1\right) &= model1\left(obs\left(t\right), obs\left(t-1\right), \ldots, \ obs\left(t-s\right)\right)\\
pred\left(t+2\right) &= model2\left(obs\left(t\right), obs\left(t-1\right), \ldots, \ obs\left(t-s\right)\right)\\
&\ldots\\
pred\left(t+12\right) &= model12\left(obs\left(t\right), obs\left(t-1\right), \ldots, \ obs\left(t-s\right)\right)
\end{aligned} \tag{6}$$

As we can see, the training set $\left(obs\left(t\right), obs\left(t-1\right), \ldots, \ obs\left(t-s\right)\right)$ is the same for all the models but twelve different models have been constructed, one for each prediction. These models depend on five parameters: the first one is the lag, that is, the time period of observations used for each sample in the training set. In this case, we use s+1 observations per model. The observations in a given time can comprise one or more variables. We have started with only one variable, the copper price. The second parameter is the number of samples used. It depends on how much we go back in time taking samples into account to construct our model. Sometimes, the behavior of a variable changes with time and the model benefits from dropping samples during the first years. Finally, the last three parameters are those related with the method used, in this case, SVR technique with RBF kernel.

Recursive multi-step forecast

In this case, we construct only a model that could be the same as *model1* of the previous method. Then, at each step, we forecast only the next value. Then, we incorporate the predicted value, drop the oldest value and predict the next value. Thus, once the model has been constructed, the prediction process will be as follows:

$$\begin{aligned}
pred\left(t+1\right) &= model\left(obs\left(t\right), obs\left(t-1\right), \ldots, \ obs\left(t-s\right)\right)\\
pred\left(t+2\right) &= model\left(pred\left(t+1\right), obs\left(t\right), obs\left(t-1\right), \ldots, \ obs\left(t-s+1\right)\right)\\
pred\left(t+3\right) &= model\left(pred\left(t+2\right), \ pred\left(t+1\right), obs\left(t\right), \ldots, \ obs\left(t-s+2\right)\right)\\
&\ldots\\
pred\left(t+12\right) &= model\left(pred\left(t+11\right), \ pred\left(t+10\right), \ldots, \ obs\left(t-s+12\right)\right)
\end{aligned} \tag{7}$$

As we can see, we have a unique model. When we are predicting we move forward one step, incorporate the last prediction and drop the oldest observation. We have the same parameters as in the previous case.

Direct-recursive hybrid forecast

This numerical scheme is a mixture of the two previous ones. We create a different model for each prediction but, in the predicting stage, the models are able to incorporate the predicted values one by one. In this case, the lag for each model increases as we advance in the prediction. That is, if we start with s+1 observations for the first model, the second model will use one observation more, as it incorporates (in the forecasting stage) the newly predicted value.

$$\begin{aligned}
pred(t+1) &= model1\left(obs(t), obs(t-1), \ldots, obs(t-s)\right)\\
pred(t+2) &= model2\left(pred(t+1), obs(t), obs(t-1), \ldots, obs(t-s)\right)\\
pred(t+3) &= model3\left(pred(t+2), pred(t+1), obs(t), obs(t-1), \ldots, obs(t-s)\right) \quad (8)\\
&\ldots\\
pred(t+12) &= model12\left(pred(t+11), pred(t+10), \ldots, obs(t-s)\right)
\end{aligned}$$

In this case, we incorporate the predictions but we do not drop old observations as we advance in the prediction.

3 Results and discussion

For the three numerical schemes, only a variable (copper price) has been used. All the available data has been used as training data. The available data set for training consist in the monthly copper prices between January 1960 and August 2017. The data between September 2017 and August 2018 has been used as validation set to optimize the hyperparameters with the grid-search method. Different models where created with the training data and the optimal hyperparameters were obtained with the grid-search method, using the validation set. The number of training samples varies with the lag. The shorter the lag, the greater the number of available samples, as a sample uses less observations and they span for a shortest period of time, allowing more samples with the same data. As the aim is to forecast monthly prices from September 2018 till August 2019, all the data related with this period of time (and the following one) have not been used during the training phase.

Evaluating the forecast accuracy

It is crucial to be sure that we can rely on forecasting. The choosing, construction, and interpretation of forecast evaluation statistics are just as important as making forecasts. For using the forecasts, is the accuracy of the "future" forecast that is most important [20].

The evaluation of forecast accuracy is based on the measurement of the errors, considering by an “error” the difference between the predicted value and real value. The fundamental forecast evaluation statistics that we can use to test our predictive model, and to evaluate the forecast accuracy, are: the mean absolute error (MAE), the root mean square error (RMSE), the mean percentage error (MPE), the mean absolute percentage error (MAPE) [21,22]. The MAE and RMSE statistics deals with measures of accuracy whose size depends on the scale of the data. Thus, they use absolute error measures and do not facilitate comparison across different time series and time intervals. To make a comparison like these, we need to work with relative or percentage error measures, as MPE and MAPE [23].

Table 1 indicates the accuracy forecast statistics (absolute, MAE and RMSE, and relative, MAPE and MPE, error measures) for the three different numerical schemes.

Table 1. Accuracy forecast statistics for the three different numerical schemes.

Numerical scheme	MAE	RMSE	MPE(%)	MAPE(%)
Direct multi-step	569.76	660.16	-9.3513	9.4621
Recursive multi-step	343.33	400.26	-0.9502	5.7013
Direct-recursive hybrid	144.21	170.15	-0.7971	2.3647

The greater accuracy of the forecast, the lower the values of these statistics. Therefore, we can see that the direct-recursive hybrid scheme has the greatest forecast accuracy.

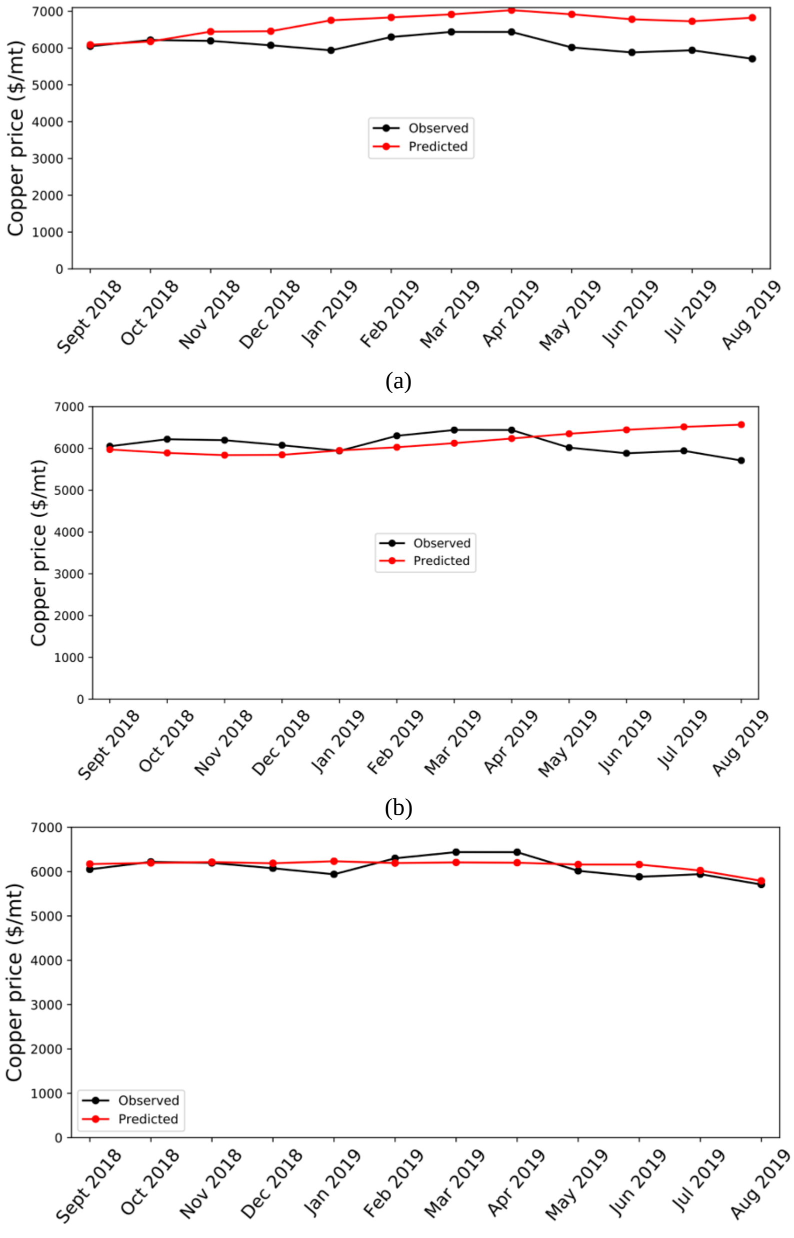
These relative measures (MPE and MAPE) give equal weight to all errors in contrast to the RMSE, which squares the errors and thereby emphasizes large errors. It would be helpful to have a measure that considers both the disproportionate cost of large errors and provides a relative basis for comparison with naïve methods. Measures that have these characteristics are the U-statistics developed by Theil [20]. Table 2 indicates the Theil's U-statistics for the three different numerical schemes.

Table 2. Theil's U-statistics for the three different numerical schemes.

Numerical scheme	U_1	U_2
Direct multi-step	0.0517	3.3909
Recursive multi-step	0.0333	2.1172
Direct-recursive hybrid	0.0139	0.8448

The greater accuracy of the forecast, the lower the values of the U_1 and U_2 statistics. Therefore, the values obtained for each of the methods used allow us to conclude that the direct-recursive hybrid scheme is the method with the best performance and greatest forecast accuracy.

Finally, Fig. 3 indicates observed and predicted COMEX copper spot price values using as predictor the SVR technique with a RBF kernel for the three different schemes



(a)

(b)

(c)

Fig. 3. Observed and predicted COMEX copper spot price values using as predictor the SVR technique with a RBF kernel for the following multi-step schemas: (a) Direct; (b) Recursive; and (c) Direct-recursive hybrid.

4 Conclusions

According to the numerical results of the present research obtained with public data of copper in the COMEX market, it can be stated using as predictor the SVR technique that the performance level of the direct-recursive hybrid scheme is higher than those achieved by the recursive multi-step and direct multi-step schemes when analyzed in terms of statistics such as the mean absolute error (MAE). In this case, the direct multi-step method is the one that performs worst.

Forecasting applications across industries observe an increase in time series frequency, from daily retail sales to hourly call centre volumes, half-hourly electricity demand, minute-by-minute internet traffic load, copper spot price and so on. Such high-frequency time series pose increasing challenges to forecasting methodologies, not only in the volume and velocity of each series, but in the need to model nonlinear interactions of exogenous variables not prevalent in monthly, quarterly, or yearly data. As a result, nonlinear algorithms from machine learning have seen an increasing popularity in forecasting, such as the support vector regression (SVR) used here based on kernel methods.

Finally, we believe there is a promising future for those lines of research combining hybrid models that are able to take full advantage of SVR models, creating models that combine machine learning techniques.

References

1. Streifel, S.: Impact of China and India on Global Commodity Markets Focus on Metals & Minerals and Petroleum. (2006).
2. Cuddington, J.T., Jerrett, D.: Super Cycles in Real Metals Prices? IMF Staff Pap. 55, 541–565 (2008).
3. Roache, S.K.: China's Impact on World Commodity Markets. (2012).
4. Lahart, J.: Ahead of the Tape: Dr. Copper, (2006).
5. Tilton, J.E., Lagos, G.: Assessing the long-run availability of copper. Resour. Policy. 32, 19–23 (2007).
6. Gordon, R.B., Bertram, M., Graedel, T.E.: Metal stocks and sustainability. Proc. Natl. Acad. Sci. 103, 1209–1214 (2006).
7. Dooley, G., Lenihan, H.: An assessment of time series methods in metal price forecasting. Resour. Policy. 30, 208–217 (2005).
8. Cortazar, G., Eterovic, F.: Can oil prices help estimate commodity futures prices? The cases of copper and silver. Resour. Policy. 35, 283–291 (2010).
9. Khashei, M., Bijari, M.: An artificial neural network (p,d,q) model for timeseries forecasting. Expert Syst. Appl. 37, 479–489 (2010).
10. Ma, W., Zhu, X., Wang, M.: Forecasting iron ore import and consumption of China using grey model optimized by particle swarm optimization algorithm. Resour. Policy. 38, 613–620 (2013).
11. Kriechbaumer, T., Angus, A., Parsons, D., Rivas Casado, M.: An improved wavelet–ARIMA approach for forecasting metal prices. Resour. Policy. 39, 32–41 (2014).
12. Sánchez Lasheras, F., de Cos Juez, F.J., Suárez Sánchez, A., Krzemień, A., Riesgo Fernández, P.: Forecasting the COMEX copper spot price by means of neural networks and ARIMA models. Resour. Policy. 45, 37–43 (2015).

13. Brockwell, P.J., Davis, R.A.: Introduction to Time Series and Forecasting, Springer. (2016).
14. Shumway, R.H., Stoffer, D.S.: Time Series Analysis and Its Applications: With R Examples, Springer. (2017).
15. World Bank Commodity Price Data (The Pink Sheet). Bloomberg; Engineering and Mining Journal; Platts Metals Week; and Thomson Reuters Datastream; World Bank. http://pubdocs.worldbank.org/en/561011486076393416/CMO-Historical-Data-Monthly.xlsx
16. Steinwart, I., Christmann, A.: Support Vector Machines, Springer. (2008).
17. Schölkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, The MIT Press. (2001).
18. Hamel, L.H.: Knowledge Discovery with Support Vector Machines, Wiley-Interscience. (2011).
19. James, G.,Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning: with Applications in R, Springer. (2017)
20. Makridakis, S.G., Wheelwright, S.C., Hyndman, R.J.: Forecasting: Methods and Application, 3rd Edition. Wiley. (1998).
21. Makridakis, S.G., Andersen, A., Carbone, R, Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R.L.: The accuracy of extrapolation (time series) methods: Results of a forecasting competition. J. of Forecasting. 1, 111-153 (1982).
22. De Gooijer, J.G., Hyndman, R.J.: 25 years of time series forecasting. Int. J. of Forecasting. 22, 443-473 (2006).
23. Hyndman, R.J., Athanasopoulos, G.: Forecasting: Principles and Practice, 3rd Edition. OTexts. (2021).