SPECIAL FOCUS

Forecasting Industrial Commodity Prices
Introduction

Almost two-thirds of emerging market and developing economies (EMDEs) depend heavily on commodities for export or fiscal revenue and economic activity. Among commodity-exporting EMDEs, resource sectors accounted for an average of 39 percent of exports of goods and non-factor services, 31 percent of goods exports, and 10 percent of value added in 2019. In some commodity-importing EMDEs, in turn, commodities account for a large share of imports and, in the presence of subsidies, fiscal spending.

For both public and private sectors, the ability to engage in sound economic and financial planning, therefore, depends heavily on the quality of commodity price forecasts. Yet, many institutions rely on futures prices for commodity price forecasts, despite their well-known shortcomings in providing accurate predictions (Alquist and Kilian 2010).

This Special Focus reviews the large literature of price forecasting approaches for a subset of seven industrial commodities: oil as well as six industrial metals commodities (aluminum, copper, lead, nickel, tin, and zinc). Together, these commodities account for 10 percent of global goods exports and 32 percent of global commodity trade.

The analysis focuses on these commodities because their demand is primarily driven by economic growth, in contrast to demand for agricultural commodities, which is mainly driven by population growth (Baffes and Nagle 2022). Oil prices are treated as representative of energy prices more broadly because, until recently, they have generally correlated closely with non-oil energy prices.

This Special Focus is a comprehensive review of the broadest range of studies and for the widest range of commodities thus far. Earlier studies have compiled reviews of forecasting models, but only for individual commodity prices.

This Special Focus reviews 60 studies in peer-reviewed journals. For crude oil price forecasts, it draws on 40 studies, most of which examine West Texas Intermediate (WTI) prices and five of which examine Brent prices (figure 18). For industrial metal prices, it draws on 20 price forecast studies covering aluminum (11 studies), copper (14), lead (8), nickel (8), tin (5), and zinc (8).

Most of these studies analyze monthly forecasts, focusing on forecast horizons of less than one year for oil prices and on forecast horizons longer than one year for metal prices. The most common forecasting approaches used in this literature have included futures prices, time series models (univariate and multivariate), and machine learning techniques (Behmiri and Manso 2013).

Note: This Special Focus was prepared by Jeetendra Khadan and Franziska Ohnsorge. The discussion is drawn from Arroyo-Marioli et al. (forthcoming). Helpful comments were provided by Paolo Agnolucci, John Baffes, Christiane Baumeister, Valerie Mercer Blackman, Pablo Pincheira Brown, Marek Kwas, Dawit Mekonnen, and Peter Nagle.

1 The WTI price has increasingly reflected U.S.-specific rather than global oil market dynamics since 2010 (Berk 2016; Manescu and Van Robays 2017).
FIGURE 18 Summary of studies of crude oil price forecast performance

The review of the literature on crude oil price forecasting draws on 40 studies. The vast majority of studies examine the performance of time series models and machine learning techniques—mostly benchmarked against no-change forecasts or futures prices. Most studies examined time periods that ended before the collapse in oil prices in mid-2014 and relied on monthly data frequencies. Forecast horizons between 3 and 12 months and of more than one year were almost equally common.

Most studies evaluate forecast performance based on directional accuracy, precision, and unbiasedness. Directional accuracy assesses whether the forecast and actual prices move in the same direction. Precision, usually measured by the root mean square forecast error, measures the degree of forecast accuracy. Unbiasedness evaluates whether forecasts systematically over- or under-predict their actual values. The remainder of this Special Focus examines the most common forecasting approaches.

Futures prices

Futures prices are based on the collective judgment of market participants’ expectations of future spot prices. They provide insights into perceptions of factors that influence future prices, such as demand and supply dynamics. Futures prices are widely used for forecasting purposes.

For oil prices, several studies have found that futures prices tend to be unbiased predictors of future spot oil prices, meaning they do not systematically over- or under-predict actual prices. However, they are not always efficient predictors and can generate large forecast errors in either direction (Abosedra and Baghestani 2004; Chinn, LeBlanc, and Coibion 2005; Jiang, Xie, and Zhou 2014). Futures prices have underperformed forecasts from a no-change benchmark (Alquist and Kilian 2010; Alquist, Kilian, and Vigfusson 2013), vector autoregression (VAR) models (Baumeister and Kilian 2012, 2014), machine learning techniques (Moshiri and Foroutan 2006), and univariate time series models (Jin 2017).

However, the predictive content of futures prices appears to have improved since the mid-2000s, possibly due to increased financialization of commodity markets (Ellwanger and Snudden 2023). Using weekly data, Rubaszek et al. (2020) found that futures prices outperformed the random walk benchmark.

Futures prices of metals have also underperformed the no-change benchmark. Chinn and Coibion (2014) showed this for aluminum, copper, lead, nickel, and tin at horizons of 3, 6, and 12 months. On the other hand, Bowman and Husain (2004) showed that incorporating futures prices in an
error correction model improved the directional accuracy and precision of forecasts for several metal prices, particularly at longer forecast horizons, compared to models that relied solely on historical data or judgment.

Univariate time series models

Univariate time series forecasting involves modeling and predicting future values of prices by correlating the price with its own lagged values. The most common univariate time series approach used in commodity price forecasting is the family of autoregressive integrated moving average (ARIMA) models.

Several studies have shown that oil price forecasts based on univariate time series models perform poorly against other approaches, although they outperform the no-change benchmark (Alquist, Kilian, and Vigfusson 2013; Jin 2017). Univariate models produce less accurate forecasts than futures oil prices (Abosedra 2006), Bayesian vector autoregression (BVAR) models (Baumeister and Kilian 2012), and machine learning techniques (Mostafa and El-Masry 2016). ARIMA models have also had poorer out-of-sample forecasting power than non-standard methods, such as nonlinear artificial neural network models and support vector machines (Mostafa and El-Masry 2016; Xie et al. 2006). Autoregressive moving average (ARMA) models have been found to lack directional accuracy and precision compared with VAR models (Baumeister and Kilian 2012).

Similarly, for metals prices, univariate time series models have performed better than no-change forecasts but underperformed other quantitative methods (figure 19; Alipour, Khodaiari, and Jafari 2019; Buncic and Moretto 2015; Rubaszek, Karolak, and Kwas 2020). For aluminum, copper, nickel, and zinc, univariate autoregressive models delivered significantly better forecasts than the no-change benchmark (Rubaszek, Karolak, and Kwas 2020). For lead, ARIMA models have generated slightly better forecasts than models based on lagged forward prices (Dooley and Lenihan 2005). For copper, the forecast performance of ARIMA and no-change forecasts was inferior to that of neural networks, dynamic averaging and selection

FIGURE 19 Summary of studies of metal price forecast performance

The review of the literature on forecasting metals prices draws on 20 studies. The most commonly evaluated methods are time series models, mostly benchmarked against no-change forecasts. Most studies use sample periods that end before the commodity price collapse of mid-2014 and most examine monthly data. The most commonly examined forecast horizon for metals prices is above one year.

A. Metals being evaluated

<table>
<thead>
<tr>
<th>Metal</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminum</td>
<td>11</td>
</tr>
<tr>
<td>Copper</td>
<td>14</td>
</tr>
<tr>
<td>Lead</td>
<td>8</td>
</tr>
<tr>
<td>Nickel</td>
<td>8</td>
</tr>
<tr>
<td>Tin</td>
<td>6</td>
</tr>
<tr>
<td>Zinc</td>
<td>8</td>
</tr>
</tbody>
</table>

B. Forecasting methodologies evaluated

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine learning models</td>
<td>7</td>
</tr>
<tr>
<td>Multivariate models</td>
<td>3</td>
</tr>
<tr>
<td>Univariate models</td>
<td>9</td>
</tr>
<tr>
<td>Futures prices</td>
<td>1</td>
</tr>
</tbody>
</table>

C. Benchmarks used for evaluation

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judgment</td>
<td>2</td>
</tr>
<tr>
<td>No-change</td>
<td>8</td>
</tr>
<tr>
<td>Univariate models</td>
<td>8</td>
</tr>
</tbody>
</table>

D. Time period for evaluation

<table>
<thead>
<tr>
<th>Period</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Sep 2008</td>
<td></td>
</tr>
<tr>
<td>Sep 2008–Jul 2014</td>
<td>11</td>
</tr>
<tr>
<td>Jul 2014–Dec 2019</td>
<td>8</td>
</tr>
<tr>
<td>After the pandemic</td>
<td>8</td>
</tr>
</tbody>
</table>

E. Data frequency

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily/Weekly</td>
<td></td>
</tr>
<tr>
<td>Monthly</td>
<td>50</td>
</tr>
<tr>
<td>Quarterly</td>
<td></td>
</tr>
</tbody>
</table>

F. Forecast horizon

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;3 months</td>
<td></td>
</tr>
<tr>
<td>&lt;=1 year</td>
<td>50</td>
</tr>
<tr>
<td>&gt;1 year</td>
<td></td>
</tr>
</tbody>
</table>

Note: Figures show the number of studies that included each commodity or applied different forecasting methods. Since several studies examine more than one metal price, the total can be larger than the number of studies.

A. Number of studies that evaluate forecast performance for each metal price.
B. Number of studies that examine the forecast performance of futures prices, machine learning techniques, multivariate models (including structural VARs and Bayesian VARs), and univariate time series models against a benchmark. The one study that examined both machine learning techniques and univariate models is shown in the category for machine learning techniques. The one study that examines both futures prices and univariate models is shown in the category for futures prices.
C. Number of studies that benchmark forecast performance against judgement-based models, latest spot prices (“no-change”), and univariate time series models.
D–F. Number of studies by end date of sample period (D), data frequency (E), and forecast horizon (F).
models, and stochastic differential equations (Alipour, Khodaiari, and Jafari 2019; Buncic and Moretto 2015; Lasheras et al. 2015). For aluminum and nickel prices, a modified grey wave forecasting technique—a univariate technique that explicitly accounts for irregular fluctuations in time series—performed better than no-change (and ARMA) methods (Chen, He, and Zhang 2016).

**Multivariate time series models**

Unlike univariate time series models, multivariate forecasting techniques account for multiple variables and their relationships when modeling and predicting the future values of a time series. VAR models are the most common multivariate time series models used in forecasting commodity prices.

For oil prices, VAR models have produced smaller out-of-sample forecast errors and more accurate directional accuracy at horizons up to 12 months than no-change forecasts and ARMA models (Alquist, Kilian, and Vigfusson 2013; Baumeister and Kilian 2012). They also have produced more accurate real-time short-run forecasts than futures prices, no-change forecasts, and regression models, while BVAR models have offered the best combination of low forecast error and high directional accuracy (Baumeister and Kilian 2012, 2014).

For metals prices, multivariate time series models have generally outperformed the no-change benchmark and, in many cases, univariate models. Issler, Rodrigues, and Burjack (2014) found that model performance differed by data frequency and commodity. For annual data, univariate autoregressive models performed best for aluminum and copper prices, while VARs produced the best forecasts for lead and zinc, and vector error correction models (VECMs) the best forecasts for nickel and tin. But for monthly data, VECMs of all metal prices and U.S. industrial production showed superior forecasting performance. Castro, Araujo, and Montini (2013) also showed that VECMs have had better out-of-sample forecast accuracy than VAR and ARIMA models for aluminum prices.

**Machine learning techniques**

Machine learning forecasting techniques use algorithms and methods to learn patterns and relationships in data, and then make predictions based on these patterns in an atheoretical manner. The most common machine learning forecasting techniques used in commodity price forecasting are artificial neural network and support vector regressions.

Machine learning techniques have generally shown better forecast performance for oil prices than other approaches, such as univariate time series models (Moshiri and Foroutan 2006; Xie et al. 2006). A neural network ensemble learning model based on an empirical mode decomposition (EMD) has had better forecast prediction and directional accuracy than an ARIMA model and other nonlinear methods, for both WTI and Brent prices (Yu, Wang, and Lai 2008). But the comparison has been sensitive to the forecast horizon (figure 20). For example, ARIMA models have outperformed artificial neural network models at the very shortest forecast horizons. However, at longer horizons, artificial neural network models and support vector regressions have outperformed ARIMA models (Fernandez 2007).

For metals prices, machine learning techniques have shown superior forecasting performance over several other approaches (Lasheras et al. 2015). For copper prices, artificial neural network models and support vector regressions have produced better forecasts than a range of other models, and the gene expression programming method has generated more accurate predictions than time series and multivariate regression methods (Astudillo et al. 2020; Dehghani 2018; Khoshalan et al. 2021).

---

Approaches to enhancing forecast performance

Two approaches have been used to improve forecast performance: combining multiple models and adding more information.

Forecast combination approaches consist of weighted averages of forecasts from different models, such as futures, no-change forecasts, VARs, and general equilibrium models. These have generated more accurate out-of-sample forecasts with better directional accuracy than no-change forecasts (Baumeister and Kilian 2015; Issler, Rodrigues, and Burjack 2014). Manescu and Van Robays (2017) also found that forecast combination approaches have improved directional accuracy and unbiasedness over futures prices and no-change forecasts.

Additional information on the behavior of economic agents and relevant economic variables has been included in modelling approaches. Accounting for global economic conditions, petroleum inventories, world output gap, the U.S. dollar real effective exchange rate, and the possibility of speculation has been shown to improve oil price forecasts (Baumeister, Korobilis, and Lee 2022; Kaufmann et al. 2008; Lalonde, Zhu, and Demers 2003). Similarly, accounting for relevant external regressors—such as industrial production, exchange rate dynamics, commodity currencies, international metal stock indexes, structural breaks, and short-run common-cycle restrictions—has improved forecast performance of some metal prices (Gong and Lin 2018; Issler, Rodrigues, and Burjack 2014; Pincheira-Brown and Hardy 2019).

Conclusion

The following general conclusions can be drawn from the literature review for both oil and metal price forecasts. First, many studies have empirically established that forecasts of WTI and Brent oil prices and metal prices based on futures contracts are inferior to several model-based approaches. Yet, futures prices are still used extensively. Second, multivariate time series models have generally outperformed other methods covered in this literature review. Several studies have found that incorporating relevant external regressors and controlling for time series properties embedded in oil or metal prices can improve forecast accuracy.

Third, machine learning techniques have tended to yield better forecasts than traditional benchmarks (such as no-change forecasts) and univariate methods, but they have been sensitive to different specifications. Comparisons of machine learning techniques with multivariate time series models-based approaches have been limited. The few available studies show that, in at least two cases, machine learning techniques have outperformed unrestricted reduced-form VAR models of oil price forecasts, but only at very short forecast horizons (Cheng et al. 2019) or up to one year (Mirmirani and Li 2004). To our knowledge, no such comparison is available for industrial metal prices.
Future editions of the *Commodity Markets Outlook* will present detailed empirical results about the performance of forecasting models discussed here for seven industrial commodity prices (aluminum, copper, lead, nickel, oil, tin, and zinc).

**References**


Fernandez, V. 2007. “Forecasting Commodity Prices by Classification Methods: The cases of Crude Oil and Natural Gas Spot Prices.” Working Paper, Department of Industrial Engineering, University of Chile, Santiago.


