Policy Research Working Paper 10611

Forecasting Industrial Commodity Prices

Literature Review and a Model Suite

Francisco Arroyo-Marioli
Jeetendra Khadan
Franziska Ohnsorge
Takefumi Yamazaki

WORLD BANK GROUP
Development Economics
Development Prospects Group
November 2023

Policy Research Working Paper 10611

Abstract

Almost two-thirds of emerging market and developing economies rely heavily on resource sectors for economic activity, fiscal and export revenues. In these economies, economic planning requires sound baseline projections for the global prices of the commodities they rely on and a sense of the risks around such baseline projections. This paper presents a model suite to prepare well-founded forecasts for the global prices for oil and six industrial metals (aluminum, copper, lead, nickel, tin, and zinc). The model suite adapts six approaches used in the literature and tests their forecast performance. Broadly speaking, futures prices or bivariate correlations performed well at short horizons, and consensus forecasts and a large-scale macroeconometric model performed well at long horizons. The strength of Bayesian vector autoregression models lies in generating forecast scenarios. The sizable forecast error bands generated by the model suite highlight the need for policy makers to engage in careful contingency planning for higher or lower prices.

This paper is a product of the Development Prospects Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at jkhadan@worldbank.org.

Forecasting Industrial Commodity Prices: Literature Review and a Model Suite

Francisco Arroyo-Marioli, Jeetendra Khadan, Franziska Ohnsorge, and Takefumi Yamazaki*

Keywords: Oil price; metal price; commodity prices; forecasting.
JEL Codes: F10; F41; Q02; Q40; Q41; Q47.

* Arroyo-Marioli: World Bank, Prospects Group; farroyomarioli@worldbank.org; Khadan: World Bank, Prospects Group; jkhadan@worldbank.org; Ohnsorge: World Bank, Prospects Group; and CAMA; fohnsorge@worldbank.org. Yamazaki: World Bank, Prospects Group; tyamazaki@worldbank.org. We thank John Baffes, Christiane Baumeister, Marek Kwas, Peter Nagle, Pablo Pincheira-Brown and many colleagues at the World Bank Group for helpful suggestions and comments on our background work for this paper. Kaltrina Temaj and Muneeb Naseem provided excellent research assistance. The findings, interpretations, and conclusions expressed in this paper are those of the authors. They do not necessarily represent the views of the institutions they are affiliated with.

1. 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 (Alquist and Kilian 2010).

This paper offers a framework for commodity price forecasts, at least for the subset of seven industrial commodities (aluminum, copper, lead, nickel, oil, tin, and zinc) that account for 8.5 percent of global exports and 31.6 percent of global commodity trade.[1] It presents a suite of models adapted from the literature that forecasts commodity prices. The development of a model suite, rather than the attempt to identify a single "best" model, is in the spirit of Baumeister and Kilian (2014, 2015). For oil prices, their findings show that forecast performance can be significantly improved if several forecasting approaches are combined.

Specifically, the paper addresses the following questions. First, which models are included in the model suite? Second, how does the forecasting performance of these models compare? Third, what are the implications for policy makers in EMDEs?

This paper contributes to the literature in several ways. The previous literature is summarized in the next section. For each of the commodities included here, there are several studies arguing that they have identified the best-performing forecasting model for their sample horizon, frequency, and commodity.

First, this paper differs from the existing literature in opting for *consistency*: it applies a consistent set of models to all commodities under consideration, using data at the quarterly frequency, for the same quarterly forecasting horizon from 2015-2022.

Second, this paper selects the approaches in the model suite as those that, according to the existing literature, have arguably performed best for their selected commodity. For example, Bayesian vector autoregression models (BVARs) have been explored extensively in the literature on oil prices. Other approaches, especially machine learning techniques, have been heavily used in the literature on metals prices. We choose particularly well-established candidate models from all strands of the literature. We evaluate their relative forecast performance testing for statistically significant differences in bias (mean error) and root mean squared forecast errors as well as the

[1] The analysis is restricted to 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; World Bank 2019). Oil prices are treated as representative of energy prices more broadly because, until recently, they have generally correlated closely with non-oil energy prices.

test for forecast accuracy developed by Diebold and Mariano (1995) and the test for directional forecast accuracy developed by Pesaran and Timmermann (2009).[2]

The paper documents three main findings. First, all approaches other than the BVAR (and, for copper and lead, consensus forecasts) did well in terms of directional accuracy, at least at horizons of less than one year. Forecast biases typically did not differ significantly across models for most forecast horizons and commodities. However, forecast precision, as captured in root mean squared errors, differed significantly.

Second, futures prices or bivariate correlations performed well at short horizons for most commodities. Consensus forecasts and a macroeconometric model (*Oxford Economic Model*) were preferred at long horizons.

Third, the reduced form BVAR model had significantly poorer forecast performance than all other approaches, both in terms of bias and precision, for all commodities and all horizons.[3] Also, the strength of BVAR models lies not only in forecast performance, but in scenario analysis: Bayesian vector autoregressive (BVAR) models are an important approach that allows a straightforward translation of global output growth forecasts into industrial commodity price forecasts.

Section 2 reviews the extensive literature on commodity price forecasting. Section 3 presents the six models selected for forecasting the seven industrial commodity prices. Section 4 discusses these models' forecast performance. Section 5 summarizes the main findings and highlights the uncertainty in commodity price forecasts.

2. Literature review

The empirical literature on commodity price forecasting distinguishes between quantitative methods and qualitative methods. The most common quantitative methods used in the forecasting literature include financial models, Bayesian time series models, univariate time series models, and non-standard methods such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM). Qualitative approaches include methods such as belief networks, the Delphi procedure, fuzzy logic and expert systems, and text mining techniques (Behmiri and Manso 2013; Drachal 2016; Frey et al. 2009). This review of the commodity price forecasting literature focuses on quantitative methods for crude oil and metal commodity prices (**Table 1**).

2.1. Crude oil price forecasting

The review of the literature on crude oil price forecasting draws on 40 studies in peer-reviewed journals, most of which examine West Texas Intermediate (WTI) prices and five of which examine

---

[2] Diebold (2015), in a review of the use of the test of Diebold and Mariano (1995) argue that sometimes simpler approaches remain appropriate. Here, therefore, we have used both the comparisons of bias and root mean squared errors and the formal tests of Diebold and Mariano (1995) and Pesaran and Timmerman (2009).

3 There are caveats. First, the sample size of the BVAR is much smaller than the other approaches due to data availability. The BVAR in this paper uses the quarterly average of monthly data while the other methods use monthly data. Second, the BVAR predicts real prices, then inflates real values to nominal prices by actual inflation rates.

Brent prices.[4] An about equal number of studies examine the performance of (univariate or multivariate) time series models and machine learning techniques (**Figure 1**). The vast majority of studies benchmark their models under examination against a no-change forecast 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 above one year were almost equally common.

2.1.1 Futures prices

Futures prices are widely used for forecasting purposes, including by international organizations such as the Asian Development Bank, the Inter-American Development Bank, the International Monetary Fund, and central banks such as the Bank of England and the ECB (Inter-American Development Bank 2022; International Monetary Fund 2014; Nixon and Smith 2012; Svensson 2006). They provide a useful forecasting tool for policy institutions since they are based on market expectations of future spot prices and are easy to communicate (Baumeister 2022; Manescu and Van Robays 2017).

In principle, futures prices incorporate the collective judgment of market participants, but they differ systematically from expected future spot prices. The deviation of futures prices from the expected spot price for storable commodities such as crude oil is related to a risk premium (the compensation to speculators), a convenience yield (the benefit of holding physical inventories), and costs related to storage and interest. Futures prices usually underpredict future spot prices because the cost component tends to be smaller than the risk and convenience yield components (Manescu and Van Robays 2017; Reeve and Vigfusson 2011).[5]

In practice, several studies have found that futures prices tend to be unbiased predictors of future spot oil prices but they are not always efficient predictors.[6] Futures prices have underperformed forecasts from a no-change benchmark (Abosedra and Baghestani 2004; Alquist and Kilian 2010; Alquist, Kilian and Vigfusson 2013; Chernenko, Schwarz and Wright 2004; Chu et al. 2022; Drachal 2016), VAR models (Baumeister and Kilian 2012; Baumeister and Kilian 2014), machine learning techniques (Moshiri and Foroutan 2006) and univariate time series models (Jin 2017; Miao et al. 2017; Naser 2016; Yousefi, Weinreich and Reinarz 2005). This finding holds for both WTI and Brent crude oil prices with forecast horizons up to one year. The forecasting performance

[4] The WTI price has increasingly reflected U.S.-specific rather than global oil market dynamics since 2010 (Berk 2016; Manescu and Robays 2017).
[5] The futures oil market is characterized by backwardation more than two-thirds of the time (Pakko 2005). This occurs when the spot price of oil is higher than prices trading in futures market (Alquist and Arbatli 2010; Alquist and Kilian 2010; Emmons and Yeager 2002; Hamilton and Wu 2014; Reeve and Vigfusson 2011; Singleton 2014).
[6] See, for example, Abosedra (2006); Abosedra and Baghestani (2004); Bopp and Lady (1991); Chinn, LeBlanc and Coibion (2005); Fritz and Weber (2012); Jiang, Xie and Zhou (2014); Moosa and Aloughani (1994); Shambora and Rossiter (2007).

of futures prices has tended to deteriorate with the forecast horizon (Alquist, Kilian and Vigfusson 2013; Reichsfeld and Roache 2011).[7]

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. Moreover, in an evaluation of alternative approaches to the random walk benchmark for commodity price forecasts, Kwas and Rubaszek (2021) found that alternative benchmarks to random walk can produce similar and, in some cases, superior forecast accuracy. The authors noted that for nominal commodity prices, the random walk benchmark should be supplemented by futures-based forecasts, while a local projections approach should serve as an alternative benchmark for real commodity prices.

There is some evidence that information contained in futures prices can improve forecasts when combined with other information or when examined over long periods. For example, models including both futures and spot WTI prices have yielded more accurate forecasts than raw oil futures prices (Wu and McCallum 2005). Vector error correction models (VECM) and vector autoregression (VAR) models suggest that the long-run relationship between spot and future WTI price fluctuations explains a sizable portion of in-sample oil price movements (Coppola 2008). However, in-sample relationships between spot and future prices do not necessarily translate into improved out-of-sample forecasts (Rubaszek et al. 2020).

2.1.2 Univariate time series models

Several studies have shown that univariate time series models perform poorly against other approaches. Univariate models produce less accurate forecasts than futures prices (Abosedra 2006; Chinn and Coibion 2014), BVAR models (Baumeister and Kilian 2012), and machine learning techniques (Fernandez 2007; Godarzi et al. 2004; Li et al. 2018; Mostafa and El-Masry 2016; Xie et al. 2006; Yu, Wang and Lai 2008) although they outperform the no-change benchmark (Alquist, Kilian and Vigfusson 2013; Chen 2014; Coppola 2008; Cortazar, Ortega and Valencia 2021; Jin 2017).

For horizons up to 12 months, an autoregressive–moving-average (ARMA) model—a widely used univariate time series model in the forecasting literature—lacks directional accuracy and produces larger forecast errors than VAR models (Baumeister and Kilian 2012). Machine learning methods have outperformed ARMA models in both level and directional forecasting accuracy (Lu et al. 2021). Autoregressive integrated moving average (ARIMA) models have also had poorer out-of-sample forecasting power than non-standard methods such as nonlinear ANN and SVM (Fernandez 2007; Mostafa and El-Masry 2016; Xie et al. 2006).

[7] An exception is Chu (2022), who found that while futures prices are inferior to no-change forecasts for WTI up to one year, they outperform no-change forecasts for horizons from one to five years.

While other time series models have been used to forecast oil price *levels*, the generalized autoregressive conditional heteroskedasticity (GARCH) family of models is the main approach used to model and forecast oil price *volatility*. Markov switching stochastic volatility models have outperformed both GARCH and historical volatility models for forecasting WTI and Brent oil price volatility for both in-sample and out-of-sample forecasts (Vo 2009; Wang, Wu and Yang 2016).[8] The predictive accuracy of oil price realized volatility has been enhanced by accounting for implied volatility and additional information on stocks, exchange rates, macroeconomic variables, and other market variables (Degiannakis and Filis 2017; Haugom et al. 2014; Pincheira-Brown et al. 2022; Wen, Gong and Cai 2016).

2.1.3 Multivariate time series models

Models that include the behavior of economic agents and economic variables can improve forecasts of crude oil prices. Baumeister, Korobilis and Lee (2022) showed that the most accurate models for Brent oil prices forecast included their global economic conditions indicator, which covers several dimensions of the global economy. OPEC (Organization of the Petroleum Exporting Countries) decisions relating to production quotas, overproduction, spare capacity, and capacity utilization have had statistically significant effects on crude oil price forecasts in the short term (Dées et al. 2007; Kaufmann et al. 2004; Tang and Hammoudeh 2002; Zamani 2004). The world output gap and the U.S. dollar real effective exchange rate have played a statistically significant role in explaining oil and copper price dynamics (Lalonde, Zhu and Demers 2003, Arroyo-Marioli and Letelier 2021). Petroleum inventories, including nonlinear inventory variables such as low and high inventory states, have improved short-run oil price predictions for both in-sample and out-of-sample forecasting (Ye, Zyren and Shore 2002; Ye, Zyren and Shore 2005; Ye, Zyren and Shore 2006).

VAR models, the most commonly used Bayesian models, have produced lower 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). VAR models have produced more accurate real-time short-run forecasts than futures prices, no-change forecasts and regression models, while Bayesian vector autoregression (BVAR) models have offered the best combination of low forecast error and high directional accuracy (Baumeister and Kilian 2012; Baumeister and Kilian 2014).[9]

However, at very short horizons of one to five days, error-correction models have performed better than unrestricted VARs or random walk models (Zeng and Swanson 1998). ANN techniques such

---

[8] Even among the GARCH family of models, there have been differences in forecast performance. For *Brent* price volatility, a standard GARCH model forecasts performed better than other GARCH variants at short horizons, but an Asymmetric Power ARCH (APARCH)-normal model performed better at longer forecast horizons (Cheong 2009). For *WTI* price volatility, the Fractionally Integrated Asymmetric Power ARCH (FIAPARCH) model had the highest accuracy at both short and long horizons (Mohammadi and Su 2010).

[9] VAR forecasting models using monthly data outperform those based on quarterly data (Baumeister and Kilian 2014).

as back-propagation networks-genetic algorithms have also outperformed VAR models in the forecast direction of movement (Mirmirani and Li 2004).

Combinations of several forecasting models have tended to generate more accurate out-of-sample forecasts. For horizons up to two years, a combination of four models yields WTI price forecasts with statistically significantly better directional accuracy than no-change forecasts (Baumeister and Kilian 2015). Over horizons up to 11 quarters, Manescu and Van Robays (2017) found similarly improved directional accuracy and unbiasedness over both futures prices and no-change forecasts for Brent oil price forecasts with a combination forecasting model built from futures, risk-adjusted futures, a SVAR, and a dynamic stochastic general equilibrium model.

2.1.4. Machine learning techniques

Numerous studies have applied nonstandard or machine learning techniques to forecast crude oil prices. These models have the advantage of accounting for nonlinearity and being well-suited for noisy data series such as crude oil prices. The application of machine learning techniques to crude oil prices has included approaches such as ANN, SVM, empirical mode decomposition (EMD) models, and gene expression programming (Cheng et al. 2019; Gabralla, Jammazi and Abraham 2013; Mostafa and El Masry 2016; Zhang, Zhang and Zhang 2015).

Despite their reported success, these machine learning-based techniques have several shortfalls. They tend to lack theoretical foundations, their forecast accuracy notwithstanding. In addition, Yu, Wang and Lai (2008) and Yu, Zhao and Tang (2014) pointed to local minima and overfitting in ANN models, the requirement of long time series in SVM models, and sensitivity parameter selection in ANN, SVM, and genetic programming models. Zheng, Cheng and Yang (2014) and Lei et al. (2013) point to the sensitivity of EMD models to statistical problems in irregular, noisy data. Peng et al. (2014) cite issues with convergence and efficiency in gene expression programming.

Machine learning techniques have generally shown better forecast performance 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. For example, ARIMA models have outperformed ANN models at very shortest forecast horizons but, at longer horizons, ANN and SVR have outperformed ARIMA models (Fernandez 2007). Similar results were found by Cheng et al. (2019) in the comparison of a hybrid vector error correction and nonlinear autoregressive neural network (VEC-NAR) model with linear time-series models such as VAR, VEC, and GARCH models.

A gene expression programming algorithm has outperformed traditional statistical techniques such as ARIMA models, and even ANN models in predicting oil prices (Mostafa and El-Masry 2016).

Also, Chiroma, Abdulkareem and Herawan (2015) found that genetic algorithm and neural network (GA–NN) methods have superior forecasting performances for monthly WTI prices than other benchmark algorithms.

An EMD-based neural network ensemble learning model has performed better than an ARIMA model and other nonlinear methods, for both WTI and Brent prices, in terms of forecast prediction and directional accuracy (Yu, Wang and Lai 2008). Similar results for EMD-based machine learning approaches were found by Xiong et al. (2013), Zhang, Zhang and Zhang (2015) and Ahmad et al. (2021).

Wavelet-based machine learning models have performed better than conventional back propagation neural network models for both level and directional predictive accuracy in WTI oil markets (Jammazi and Aloui 2012).

2.1.5 Conclusion from the literature on oil price forecasting

The following general conclusions can be drawn from the above literature review. *First*, many studies have empirically established that forecasts of WTI and Brent oil prices based on futures contracts are inferior to several model-based approaches. *Second*, model-based approaches have generally outperformed other methods. *Third*, several studies have found that incorporating relevant external regressors and controlling for time series properties embedded in oil prices can improve forecast accuracy. *Fourth*, machine learning techniques have tended to yield better forecasting performance than traditional benchmarks and univariate methods, but they have been sensitive to the choice of specifications. Their comparisons with model-based approaches have been limited. The few available studies show that, in at least two cases, machine learning techniques have outperformed reduced-form VAR models, but only at the very shortest forecast horizons (Cheng et al. 2019) or up to one year (Mirmirani and Li 2004).

2.2 Metal commodity price forecasting

This literature draws on 20 studies of price forecasts for aluminum (11 studies), copper (14), lead (8), nickel (8), tin (5), and zinc (8). The most evaluated methods are univariate and multivariate time series models—mostly benchmarked against no-change forecasts—although the number of studies based on machine learning techniques is also growing rapidly (**Figure 2**). As for oil prices, most studies use sample periods that end before the commodity price collapse of mid-2014 and most examine monthly data. In contrast to oil prices, where studies examining forecast horizons up to one year are common, the most examined forecast horizon for metals prices have exceeded one year.

2.2.1 Futures prices

Like for oil prices, futures prices of metals have underperformed the no-change benchmark, but they have predictive content that can improve model forecasts. In a study of several metals (aluminum, copper, lead, nickel and tin), Chinn and Coibion (2014) showed that the no-change benchmark has modestly outperformed futures prices at horizons 3, 6 and 12 months. On the other hand, Bowman and Husain (2004) showed for several metals that models that incorporate futures prices in an error correction model performed better than models based entirely on historical data or judgment, in terms of directional accuracy and precision, particularly at longer forecast horizons. Complementing futures price data with other information—such as industrial production, exchange rate dynamics, commodity currencies, international metals stock index, structural breaks, and short-run common-cycle restrictions—has further improved forecast performance (Gong and Lin 2018; Issler, Rodrigues and Burjack 2014; Pincheira-Brown and Hardy 2019; Pincheira-Brown and Hardy 2021; Pincheira-Brown et al. 2021).

2.2.2 Univariate time series models

Univariate time series models have performed better than no-change forecasts but underperformed other quantitative methods (Rubaszek, Karolak and Kwas 2020; Alipour, Khodaiari and Jafari 2019). For aluminum, copper, nickel, and zinc, univariate autoregressive models delivered significantly better forecasts than the no-change approach (Rubaszek, Karolak and Kwas 2020).

However, univariate time series models have generally underperformed other quantitative methods. For copper, the forecast performance of ARIMA and no-change forecasts was inferior to that of neural networks, dynamic averaging and selection models and stochastic differential equations (Alipour, Khodaiari and Jafari 2019; Buncic and Moretto 2015; Lasheras et al. 2015). For aluminum, VECMs have had better out-of-sample forecast accuracy than ARIMA and VAR models (Castro, Araujo and Montini 2013). 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 or ARMA methods (Chen, He and Zhang 2016). For aluminum, copper, lead, and zinc, wavelet-autoregressive integrated moving average (ARIMA)-based models have outperformed traditional ARIMA models in terms of forecast accuracy (Kriechbaumer et al. 2014). For lead and zinc, He et al. (2015) found that a curvelet-based multiscale forecasting approach was superior to traditional benchmark models such as ARMA and random walk. For lead, ARIMA models have generated slightly better forecasts than models based on lagged forward prices (Dooley and Lenihan 2005).

2.2.3. Multivariate time series models

Multivariate time series models such as VARs have generally outperformed the no-change benchmark and, in many cases, univariate models. In a comprehensive analysis of price forecasts for copper, lead, nickel, tin and zinc, Issler, Rodrigues and Burjack (2014) using annual and monthly data from 1900 to 2010 found that model performance differs by data frequency and commodity. For annual data, univariate autoregressive models were the best for aluminum and copper prices, while VARs produced the best results for lead and zinc, and VECMs had the best results 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. In contrast, Rubaszek, Karolak and Kwas (2020) found that taking into account nonlinearities, such as the inclusion of a threshold structure to autoregressive models, did not materially improve the forecast performance for aluminum, copper, nickel, and zinc prices.

2.2.4 Machine learning techniques

Machine learning techniques have shown superior forecasting performance over several other approaches (Lasheras et al. 2015). For copper prices, ANN and support vector regression (SVR) models have produced a better forecast performance 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). For copper prices, forecasts from hybrid neural network models have outperformed other models (Du et al. 2021) and traditional ANN techniques in terms of level and directional predictions (Wang et al. 2019).

2.2.5 Conclusion from the literature on metal price forecasting

The following general conclusions can be drawn from the literature on metal price forecasting. *First*, similar to crude oil, futures prices have generally had inferior forecast performance to the no-change benchmark, but they have had some predictive information when combined with other modeling approaches. *Second*, multivariate time series models have performed better than univariate time series methods and no-change forecasts. *Third*, machine learning techniques have outperformed both univariate time series methods and no-change. However, to the best of our knowledge, there is no comparison between forecasts from multivariate time series models and machine learning techniques. *Fourth*, as in the case of oil, several studies found that the addition of other economic variables and controlling for properties of the metal price series, such as structural breaks, can improve forecast accuracy for metal prices.

3. A model suite

This paper draws on these existing studies to inform the choice of models for a broad range of model-based metal price forecasts. The model suite includes six approaches: futures prices and consensus forecasts; bivariate correlations; a Bayesian vector autoregression estimation (BVAR); a model-based approach (the *Oxford Economics Model*); and a machine learning-based approach (EMD-based support vector regression and GARCH). Although the literature review highlights the promise of forecast combination approaches, we do not explore this approach since the aim of our model suite is to develop a range of commodity price forecasts rather than a single "best" forecast.

These approaches were used to establish quarterly forecasts one to eight quarters ahead. The estimated parameters of the models were updated using an expanding scheme that allows for the inclusion of new observations as data availability increases. The forecasts derived from the six approaches were compared with realized prices for the period 2015Q1-2022Q1.

This period is one of substantial commodity price volatility. It included the plunge from mid-2014 to its trough in early 2016, the subsequent rebound, the collapse during the pandemic, in spring 2020, and again the subsequent rebound—for many commodities to historic highs. Since many of these swings were unusually strong by historical standards, model performance is likely to be generally poorer than in previous studies.[10]

3.1 Futures prices and consensus forecasts

Consensus forecasts were drawn from Consensus Forecast Reports accessed through the IMF-World Bank library for 2000-2022. Data is available monthly, with quarterly forecasts until the end of the subsequent year.

Generic futures prices were drawn from Bloomberg for all seven commodities for 2000-2022: Oil (tickers CL3-CL6-CL9-CL12); copper (tickers HG3-HG6-HG9-HG12); aluminum (tickers AA3-AA6-AA9-AA12); lead (tickers LL3-LL6-LL9-LL12); nickel (tickers LN3-LN6-LN9-LN12); tin (tickers LT3-LT6-LT9-LT12); and zinc (tickers LX3-LX6-LX9-LX12). Monthly average data is used, with 3-, 6-, 9, and 12-month-ahead futures prices.

3.2 Bivariate correlations

Alquist, Kilian, and Vigfusson (2013), in their review of oil price forecasting models, identify several variables that are strongly correlated with oil prices over the subsequent 6-12 months. These include the Commodities Research Bureau (CRB) Raw Industrial Commodity Index, U.S. M1 growth, futures prices, and consensus (or other professional) forecasts. These variables are augmented with indicators for China, which plays a large role in global metal markets, and purchasing managers indicators. These include China's Manufacturing Purchasing Managers' Index (PMI), the Global Manufacturing PMI, and the Global Composite PMI.

Based on Alquist, Kilian, and Vigfusson (2013), the forecast is obtained by regressing the future change of a commodity price against past independent variables:

$$\boldsymbol{\Delta_h Y_t = \beta M_t + \varepsilon_{t,t+h}}$$

$$\boldsymbol{\Delta_h Y_t = Y_{t+h} - Y_t}$$

where $\boldsymbol{\Delta_h Y_t}$ indicates the percent change of a commodity price within the next *h* months, and $\boldsymbol{M_t}$ is one input.

Six inputs are used separately: U.S. base money (M1 from the FRED database), 10-year Treasury bond yields (from the FRED database), and the CRB Raw Materials Index (from Bloomberg) in past percent differences equivalent to the horizon.[11] For example, the percent change over the last

[10] Several studies employed different approaches to address the breaks introduced by the COVID-19 pandemic in forecasting (Schorfheide and Song 2021; Ng 2021; Lenza and Primiceri 2022). There was no need to make assumptions about the treatment of the pandemic in the econometric exercise here as the COVID-19 period formed part of the out-of-sample evaluation period in our comparison.

[11] The CRB Raw Industrial Commodity Index includes 22 commodities. The 22 commodities are combined into an "All Commodities" grouping, with two major subdivisions: Raw Industrials, and Foodstuffs. Raw Industrials include burlap, copper scrap, cotton, hides, lead scrap, print cloth, rosin, rubber, steel scrap, tallow, tin, wool tops, and zinc.

three months of the U.S. M1 is used to forecast the percent change in oil prices over the next three months. Global PMIs (composite and manufacturing), and China's manufacturing PMI are used in levels and lagged *h* months. In total, each commodity has six forecasted values for each horizon. The final forecast is an average of all statistically significant forecasts. Equations with nonsignificant results are discarded. The exercise is conducted with monthly data.

Monthly data are available for 2004-2022, constrained by the availability of China's Manufacturing PMI. All data are from Haver Analytics.

3.3 Bayesian vector autoregression model

A reduced-form Bayesian VAR model is used to forecast oil prices. In the estimation step, to ensure compatibility between prediction and scenario analysis, we impose sign-restrictions on the VAR model. The BVAR employs quarterly data for commodity production, commodity prices, and global real GDP growth. Forecasts are conditioned on realized data or the June 2022, *Global Economic Prospects* report (GEP).

The BVAR model of the global oil or metal market is written as:

$$\boldsymbol{Y_t} = \sum_{\boldsymbol{i}=\mathbf{1}}^{\mathbf{4}} \boldsymbol{A_i Y_{t-i}} + \boldsymbol{u_t}$$

where $\boldsymbol{Y_t}$ is the vector of endogenous variables, $\boldsymbol{u_t}$ is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma_u}$.. $\boldsymbol{A_i}, \boldsymbol{i} = \mathbf{1,2,3,4}$ denotes the coefficient matrices. Endogenous variables are $\boldsymbol{Y_t} = [\boldsymbol{\Delta q_t},\ \boldsymbol{\Delta y_t},\ \boldsymbol{p_t}]'$, where $\boldsymbol{\Delta q_t},\ \boldsymbol{\Delta y_t},\ \boldsymbol{p_t}$ denote the log-differences of metal or oil productions, GDP growth rates and metal or oil prices.[12] Since GDP growth is only available on a quarterly basis, the exercise is conducted with quarterly averages of monthly data. Other variables that have been used in forecasting commodity prices include indicators of global economic conditions, world output gap, capacity utilization, industrial production and exchange rates (Baumeister, Korobilis and Lee 2022; Dées et al. 2007; Kaufmann et al. 2004; Tang and Hammoudeh 2002; Zamani 2004; Lalonde, Zhu and Demers 2003; Ye, Zyren and Shore 2006).

In the prediction step, 1-step-ahead prediction at the time origin $\boldsymbol{h}$, $\boldsymbol{\widehat{Y}_h}(\mathbf{1})$, and the associated forecast error, $\boldsymbol{e_h}(\mathbf{1})$ are:

$$\boldsymbol{\widehat{Y}_h}(\mathbf{1}) = \sum_{\boldsymbol{i}=\mathbf{1}}^{\mathbf{4}} \boldsymbol{A_i Y_{h+1-i}}, \qquad \boldsymbol{e_h}(\mathbf{1}) = \boldsymbol{u_{h+1}}.$$

Foodstuffs include butter, cocoa beans, corn, cottonseed oil, hogs, lard, steers, sugar, and wheat. It is developed by the Commodities Research Bureau.

[12] We adopt a log-level commodity price model following Kilian and Murphy (2014). As they suggested, it is not clear whether commodity prices should be modeled in log-levels or log-differences. The advantage of the level specification is that impulse responses of log-level price models are consistent with sign-restrictions while, in many cases, log-difference price models create meaningless impulse responses that would not be extended to scenario analysis in practice. The disadvantage of not imposing unit roots in estimation is a loss of asymptotic efficiency.

For 2-step-ahead forecasts:

$$\widehat{\boldsymbol{Y}}_{\boldsymbol{h}}(\mathbf{2}) = \boldsymbol{A}_{\mathbf{1}}\widehat{\boldsymbol{Y}}_{\boldsymbol{h}}(\mathbf{1}) + \sum_{i=2}^{4} \boldsymbol{A}_{i}\boldsymbol{Y}_{\boldsymbol{h}+2-i}, \quad \boldsymbol{e}_{\boldsymbol{h}}(\mathbf{2}) = \boldsymbol{u}_{\boldsymbol{h}+\mathbf{2}} + \boldsymbol{A}_{\mathbf{1}}\left[\boldsymbol{Y}_{\mathbf{1}} - \widehat{\boldsymbol{Y}}_{\boldsymbol{h}}(\mathbf{1})\right].$$

We extend the 3 to 8-step ahead forecasts in a similar fashion. Forecasted commodity prices $\widehat{\boldsymbol{p}}_{\boldsymbol{h}}$ (one of $\widehat{\boldsymbol{Y}}_{\boldsymbol{h}}$) are in real prices, thus we re-inflate $\widehat{\boldsymbol{p}}_{\boldsymbol{h}}$ using actual (realized) U.S. CPI.

In the estimation step, we estimate the following VAR representation:

$$\boldsymbol{B}_{\mathbf{0}}\boldsymbol{Y}_{\boldsymbol{t}} = \sum_{i=1}^{4} \boldsymbol{B}_{i}\boldsymbol{Y}_{\boldsymbol{t}-i} + \boldsymbol{\varepsilon}_{\boldsymbol{t}}$$

where $\boldsymbol{\varepsilon}_{\boldsymbol{t}}$ is the structural shocks which follow standard normal distribution. $\boldsymbol{B}_{\boldsymbol{i}}$ denotes the coefficient matrices. The structural shocks comprise of supply shocks, demand shocks, and residual shocks.

The identification problem consists of finding a mapping from the errors in the reduced-form representation to its structural counterpart:

$$\boldsymbol{u}_{\boldsymbol{t}} = \boldsymbol{B}_{\mathbf{0}}^{-\mathbf{1}}\boldsymbol{\varepsilon}_{\boldsymbol{t}}.$$

We exploit the following relation:

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{u}} = \boldsymbol{E}[\boldsymbol{u}_{\boldsymbol{t}}\boldsymbol{u}_{\boldsymbol{t}}'] &= \mathbb{E}\left[\boldsymbol{B}_{\mathbf{0}}^{-\mathbf{1}}\boldsymbol{\varepsilon}_{\boldsymbol{t}}\left(\boldsymbol{B}_{\mathbf{0}}^{-\mathbf{1}}\boldsymbol{\varepsilon}_{\boldsymbol{t}}\right)'\right] = \boldsymbol{B}_{\mathbf{0}}^{-\mathbf{1}}\mathbb{E}[\boldsymbol{\varepsilon}_{\boldsymbol{t}}\boldsymbol{\varepsilon}_{\boldsymbol{t}}']\left(\boldsymbol{B}_{\mathbf{0}}^{-\mathbf{1}}\right)' = \boldsymbol{B}_{\mathbf{0}}^{-\mathbf{1}}\boldsymbol{\Sigma}_{\varepsilon}\left(\boldsymbol{B}_{\mathbf{0}}^{-\mathbf{1}}\right)' \\ &= \boldsymbol{B}_{\mathbf{0}}^{-\mathbf{1}}\left(\boldsymbol{B}_{\mathbf{0}}^{-\mathbf{1}}\right)'.\end{aligned}$$

To explore $\widetilde{\boldsymbol{B}}$, the estimate of $\boldsymbol{B}_{\mathbf{0}}^{-\mathbf{1}}$, we generate the random orthogonal matrix $\boldsymbol{QQ}' = \boldsymbol{I}$ and consider Cholesky factor $\boldsymbol{\Sigma}_{\boldsymbol{u}} = \boldsymbol{PP}'$ as follows:

$$\boldsymbol{\Sigma}_{\boldsymbol{u}} = \boldsymbol{PQQ}'\boldsymbol{P}' = (\boldsymbol{PQ})(\boldsymbol{PQ})'.$$

Relating the above equations, we consider the matrix $\widetilde{\boldsymbol{B}} = \boldsymbol{PQ}$ as a valid candidate. The structural shocks are identified using the sign restriction of Kilian and Murphy (2014) as shown in **Table 2**.[13] A positive demand shock on impact is assumed to raise the real price of oil or metals and stimulate oil or metal production, as well as raise GDP growth. A negative supply shock is assumed to lower oil or metal production on impact. It will also lower global economic activity while increasing the real price of oil or metals.

We simulate impulse responses based on a candidate $\widetilde{\boldsymbol{B}}$. The candidate $\widetilde{\boldsymbol{B}}$ is retained if the resulting impulse responses meets the sign restrictions, otherwise discarded.

[13] The quantitative restrictions on supply elasticities used by Kilian and Murphy (2014) are avoided since such elasticities cannot easily be benchmarked for metals prices.

The integrated estimation steps are as follows:

1. Run an unrestricted VAR and find $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{u}}$. Implement Cholesky decomposition to extract $\boldsymbol{P}$.[14]
2. Draw a random orthogonal matrix $\boldsymbol{Q}$ and compute $\widetilde{\boldsymbol{B}} = \boldsymbol{PQ}$.
3. Compute impulse responses using $\widetilde{\boldsymbol{B}}$ calculated in the step 2. If all implied impulse response functions satisfy the sign restrictions (**Table 2**), retain $\widetilde{\boldsymbol{B}}$. Otherwise discard $\widetilde{\boldsymbol{B}}$.
4. Repeat the first two steps 50,000 times, recording each $\widetilde{\boldsymbol{B}}$ that satisfies the restrictions and record the corresponding impulse response functions. About one-fifth of the draws are discarded.

The commodity price data for aluminum, copper, lead, nickel, oil, tin, and zinc are drawn from the World Bank's Commodities Price Data (*Pink Sheet*). The price for oil is the unweighted average of the Brent, West Texas Intermediate, and Dubai oil prices. These commodity prices are deflated by the U.S. CPI from the Federal Reserve Economic Data (FRED) database. Quarterly GDP growth rates are from Haver Analytics, with forecasts based on the World Bank's June 2022, *Global Economic Prospects* report. Commodity production for aluminum, copper, zinc, lead, nickel, oil, tin, and zinc is drawn from the World Bureau of Metal Statistics. Quarterly averages of monthly data are available for 2000—2022Q1.

3.4 Oxford Economics Model (OEM)

The OEM is a macroeconometric model that is routinely used for growth forecasting in many international institutions and central banks (Guenette and Yamazaki 2021). It includes price series for world food, world beverages, world agricultural raw materials, aluminum, copper, iron, lead, nickel, tin, zinc, coal, oil, and natural gas.[15] These are extracted from the latest OEM forecasts.

The OEM is a macroeconometric model with 46 countries, 6 regional blocs and the Eurozone (Oxford Economics 2019). Most components are specified as error correction models. In the short run, shocks to demand generate economic cycles that can be influenced by fiscal and monetary policy. Over the long-run, output is determined by supply side factors: investment, demographics, labor participation and productivity. The resulting dynamics of short-run fluctuations and long-run trend are integrated in terms of cointegration. We use biannual (Q1 and Q3) forecast data from 2015Q1–2022Q1 due to data availability.[16]

3.5 Machine learning approach

The EMD-based support vector regression and GARCH model of Zhang, Zhang and Zhang (2015) is selected as a machine learning approach, in line with its generally successful forecast

[14] We adopt the Gibbs-sampler, an estimation approach that allows for structural identifications such as elasticity and sign restrictions and prior beliefs about future economic events, to estimate a restricted VAR. Although the identification procedure is not required for forecasting, we impose sign-restrictions to ensure compatibility between prediction and scenario analyses.

[15] Natural gas is a composite of Henry Hub, Japan, and European natural gas prices.

[16] Since the OEM uses biannual forecasts while other models use quarterly forecasts, its performance may be underestimated by Diebold-Mariano statistic in **Figure 3** because the number of forecasts for OEM is smaller than other models.

performance documented in the literature on oil prices. First, the EMD approach decomposes the seven industrial commodity price series into multiple nonlinear components and time-varying components. Then, the price forecasts are compiled from the predicted values from a support vector regression for the nonlinear components and a GARCH (1,1) model for the time-varying components. Specifically, the EMD approach is conducted in the following five steps.

First, each of the seven industrial commodity price series ($\boldsymbol{P_t}$) is represented as being subject to multiple white noise processes:

$$\boldsymbol{P_t^i = P_t + \varepsilon_t^i, (i \ = \ 1, 2, \dots, M)}$$

where $\boldsymbol{\varepsilon_t^i}$ denotes the i[th] white noise series, and $\boldsymbol{P_t^i}$ represents the industrial commodity price in the i[th] trial. The standard deviation of $\boldsymbol{\varepsilon_t^i}$ is assumed to be 0.005 times the standard deviation of the original series.

Second, an EMD model decomposes $\boldsymbol{P_t^i}$ into finite intrinsic mode functions (IMFs), $\boldsymbol{c_{ij}}$ ($\boldsymbol{j = 1, 2, \dots, J}$), and residual series. The ensemble mean of $\boldsymbol{M}$ trials for each IMFs is $\boldsymbol{C_t^j = \frac{1}{M}\sum_{i=1}^{M} c_t^{ij}}$ and the ensemble mean of the residuals is $\boldsymbol{R_t^M = \frac{1}{M}\sum_{i=1}^{M} r_t^i}$. The relationship between the price series ($\boldsymbol{P_t}$), the ensemble mean IMFs ($\boldsymbol{C_t^i}$), and the residuals ($\boldsymbol{R_t^M}$) is represented as:

$$\boldsymbol{P_t = \sum_{j=1}^{J} C_t^j + R_t^M}.$$

Third, $\boldsymbol{C_t^j}$ and $\boldsymbol{R_t^M}$ are rearranged into a time varying component ($\boldsymbol{S_t^l}$) and nonlinear component ($\boldsymbol{N_t^k}$), such that the equation above is reshaped into the following: [17]

$$\boldsymbol{P_t = \sum_{k=1}^{m} N_t^k + \sum_{l=m+1}^{n} S_t^l + R_t^M}.$$

Fourth, GARCH (1,1) is applied to $\boldsymbol{S_t^l}$ to forecast the time varying component ($\boldsymbol{\widehat{S}_t^l}$) and a support vector machine is applied to $\boldsymbol{N_t^k}$ and $\boldsymbol{R_t^M}$ to forecast the nonlinear component ($\boldsymbol{\widehat{N}_t^k}$) and residuals ($\boldsymbol{\widehat{R}_t^M}$).

Fifth, the final price forecasts ($\boldsymbol{\widehat{P}_t}$) is constructed from these component forecasts using the following equation:

$$\boldsymbol{\widehat{P}_t = \sum_{k=1}^{m} \widehat{N}_t^k + \sum_{l=m+1}^{n} \widehat{S}_t^l + \widehat{R}_t^M}.$$

The approach is applied to monthly commodity price data for oil, aluminum, copper, lead, nickel, tin, and zinc and then converted into quarterly averages. The data are drawn from the World Bank's

[17] In this paper, IMF1, IMF2 and IMF3 are classified as time-varying components and the other IMFs are classified as nonlinear components based on Zhang, Zhang and Zhang (2015).

Commodities Price Data (Pink Sheet). The price for oil is the unweighted average of the Brent, West Texas Intermediate, and Dubai oil prices. Monthly data from January 1995 to December 2014 are used for the training sample, the model is updated based on expanding scheme, and data from 2015Q1 to 2022Q1 are used as the testing period.

4. Forecast performance

The six models are compared in terms of their bias and precision.[18] The forecast of each model is compared against all other approaches in the model suite, rather than a single benchmark (**Figure 3**).[19] The mean error (bias) is defined as the difference between the actual price and the price forecast one to eight quarters ahead for each of the seven industrial commodities. Precision is captured by the root mean squared error (RMSE). Model comparisons are tested for statistical significance in a t-test (for absolute bias) or an F-test (for RMSE) or in the significance of the Diebold-Mariano statistic. Futures prices are only available up to one year ahead with reasonable liquidity, hence longer forecast horizons are not considered here for futures prices.[20] Bivariate correlations are only significant for horizons up to one year ahead; hence, forecasts from bivariate correlations are dropped for horizons beyond one year. OEM forecasts are only available on a semi-annual basis and the statistical tests are adjusted for the fewer degrees of freedom.

4.1 Oil prices

None of the oil price forecasts generated by any of the six methodologies had a statistically significant bias (**Table 3.A**). However, the RMSEs were sizable for several approaches and have tended to be larger at longer horizons (except for OEM-based forecasts, **Table 3.B**). At horizons up to one year, all approaches except the BVAR predicted the direction of forecast changes accurately; at horizons over one year, only consensus forecasts and the OEM did so (**Tables 3.C-D**).

At the one- to four-quarter horizons, futures prices had particularly small RMSEs that were either significantly smaller than those of most other approaches or no larger (**Tables 4.A-B**). Similarly, the Diebold-Mariano test suggests significantly better or no worse forecast accuracy of futures prices than all other approaches (**Figure 3**). At longer horizons, both OEM-based forecasts had smaller biases, lower RMSEs and better forecast accuracy than other forecasting approaches. At all horizons, BVAR-based forecasts and bivariate correlations generally performed more poorly than one or more of the other approaches, despite being unbiased.

[18] The sample period used to compute forecast and forecast errors was restricted to 2015-2022 to ensure sufficient observations for estimation.

[19] The evaluation period for our model suite is 2015-2022. Other studies have compared forecast performance for longer periods for several commodities, particularly for futures prices, for example Reeve and Vigfusson (2011) and Bowman and Husain (2004).

[20] A comparison of the forecast accuracy of futures prices against other methods using the Diebold-Mariano statistic and RMSEs for time horizons beyond one year showed that futures performed better than other methods in some cases. However, we opted to exclude forecast comparisons with futures prices beyond one year from the forecast evaluation exercise over concerns about their overall robustness. Futures contracts for distant delivery months tend to have low trading volumes and liquidity. This lack of activity can reduce price discovery, leading to less reliable forecasts and a potentially biased comparison of futures prices with other models.

4.2 Aluminum prices

Like for oil prices, none of the approaches generated significantly biased forecasts of aluminum prices and none of the approaches generated statistically significantly inaccurate forecast directions. That said, biases for futures prices were significantly larger than those of all other approaches (**Table 5.A**).

Bivariate correlations (at horizons up to two quarters) and machine learning approaches (at horizons of three to five quarters) had significantly smaller RMSEs and significantly better forecast accuracy than other approaches (**Table 5.B**). At horizons above five quarters, the OEM had significantly better forecast accuracy than other approaches (**Figure 3**). The BVAR underperformed on both metrics.

4.3 Copper prices

Again, none of the approaches had statistically significant forecast biases at any forecast horizon. However, only futures prices, bivariate correlations, and machine learning approaches generated accurate forecast directions at horizons below one year, and only the OEM at horizons above one year, as suggested by the Pesaran-Timmerman test (**Table 3.D**). The BVAR, bivariate correlations, and machine learning techniques produced very particularly imprecise forecasts with large RMSEs.

Biases, even if statistically indistinguishable from zero, were significantly larger for bivariate correlations than for most other approaches at horizons above one quarter. At the one-quarter horizon, bivariate correlations had significantly smaller RMSEs than most other approaches (**Tables 6.A-B**). But at the two- to four-quarter horizon, this advantage switched to futures prices. Based on the Diebold-Mariano test, futures prices generated statistically significantly more accurate forecasts than other approaches at horizons up to one year (**Figure 3**). At horizons beyond one year, consensus forecasts and OEM-based forecasts had similar or smaller errors than other approaches. BVAR-based forecasts had significantly higher RMSEs and poorer forecast accuracy, but no larger biases, than all other approaches at almost all forecast horizons.

4.4 Lead prices

With the exception of some few instances, all approaches underpredicted prices at virtually all horizons during the forecast period. However, none of these biases were statistically significant at any forecast horizon. Like for copper prices, only futures, bivariate correlations, the OEM and the machine learning approach produced forecast that were directionally accurate at horizons up to one year, and only the OEM at horizons above one year. RMSEs were particularly large for the BVAR, the machine learning approach (at longer horizons) and bivariate correlations (at shorter horizons).

At horizons up to one year, futures prices were more accurate than almost all other approaches, with significantly smaller RMSEs, although at the one-quarter horizon, bivariate correlations performed similarly to futures prices (**Tables 7.A-B**). At horizons above one year, none of the approaches differed in their biases but consensus forecasts were the most accurate forecasts, with the smallest RMSEs. The BVAR approach was again outperformed by all other approaches.

4.5 Nickel prices

Again, none of the forecast approaches generated statistically significantly biased forecasts. The only approach that did not produce forecast that were directionally accurate at horizons up to one year was the BVAR but, at longer horizons, only the OEM produced directionally accurate forecasts.

At horizons up to one year, futures prices were significantly more accurate than other approaches, with significantly smaller RMSEs(**Tables 8.A-B**). At horizons beyond one year, OEM-based forecasts outperformed all other approaches. Both the BVAR and, to a lesser extent, the machine learning approach performed worse than other approaches.

4.6 Tin prices

Except for bivariate correlations and machine learning methods, all approaches underpredicted prices at all forecast horizons, but not statistically significantly. All approaches, except for the machine learning approach at horizons above one-year, generated forecasts that were directionally accurate. RMSEs were particularly large for the BVAR. RMSEs were particularly large for the machine learning and BVAR approaches.

At horizons up to one year, futures prices were no less accurate or significantly more accurate than all other approaches; at horizons up to half a year, the forecast accurate of bivariate correlations was on par with that of futures prices (**Tables 9.A-B**). At horizons above one year, the OEM produced the most accurate forecasts. At all forecast horizons, the BVAR performed the same or worse than all other approaches.

4.7 Zinc prices

Again, with the exception of bivariate correlations and machine learning-based models, all approaches underpredicted prices at all forecast horizons, but not statistically significantly. All approaches other than the BVAR generated directionally accurate forecasts. RMSEs were particularly large for the BVAR and machine learning approaches.

At horizons up to one year, futures prices were more accurate than all other approaches, except for bivariate correlation which were equally accurate at the one-quarter horizon (**Tables 10.A-B**). At horizons beyond one year, consensus and OEM-based forecasts were more accurate than the BVAR and machine learning approaches.

4.8 Comparison across industrial commodities

A few patterns hold across all commodities. First, forecast biases typically did not differ significantly across models for most forecast horizons and commodities. However, forecast precision, as captured in root mean squared errors, and forecast accuracy, as captured by the Diebold-Mariano test, differed significantly across models (**Tables 11.A-G; Table 12**).

Second, futures prices or bivariate correlations performed well at short horizons for most commodities, but especially for metals commodities. Consensus forecasts or the macroeconometric OEM were preferred at long horizons.

Third, BVAR models had significantly poorer forecast performance than all other approaches, in terms of bias, precision, and direction of change, for all commodities and all horizons. The strength of BVAR models lies not so much in forecast performance, but in scenario analysis: among the approaches considered here, BVAR models allow the most straightforward translation of global output growth forecasts into industrial commodity price forecasts (**Table 13**).

There are some caveats on the comparison. Broadly speaking, these caveats bias the comparison against the BVAR approach. First, future prices, consensus forecasts, and the machine learning approach directly predict nominal price levels while bivariate correlations forecast percent changes of nominal prices. BVAR models predict real prices, then inflate nominal prices using actual inflation rates. Second, all the models except for the BVAR approach generate monthly price forecasts that are aggregated into quarterly forecasts for comparison; in contrast, the BVAR uses the quarterly average of monthly data as input to generate quarterly price forecasts since GDP data is only available at the quarterly frequency. Third and most important, forecasts were done without any judgement. None of the forecasting approaches here have pre-designed scenarios or priors. However, in reality, it is often of interest to condition the forecasts on different scenarios. The OEM is particularly useful to conduct scenario exercises that take into account changes in policy variables, global growth, inflation, and structural variables. Another purpose of conditional forecasts is to incorporate information from higher frequency data or judgment into the model (Karlsson 2013). This underscores the main advantage of the BVAR model, which allows the forecaster to simulate scenarios or test priors while maintaining the statistical properties of model. The forecaster can then make inferences about the posterior forecast conditional on the prior. The version of the BVAR model used in this exercise leads to a larger forecast bias than other models for oil prices. However, in practice, the forecaster could subsequently adjust the parameters to reflect changes in priors to arrive at more informed forecasts.

5. Conclusion

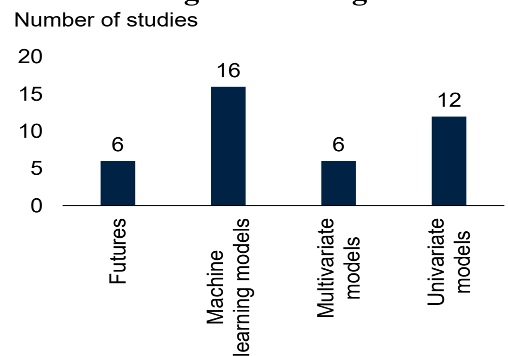
This paper presents a model suite for forecasting prices of seven industrial commodities: oil, aluminum, copper, lead, nickel, tin, and zinc. It includes six approaches, based on the review of a rich literature of commodity price forecasting: consensus forecasts, futures prices, univariate correlation, a Bayesian VAR, a large-scale macroeconometric model (*Oxford Economic Model*), and a machine learning-based approach (EMD-based SVR and GARCH).

It finds that no single approach is the best but rather that model performance depends on the commodity and the forecast horizon. As a rule, futures prices or bivariate correlations performed well at short horizons; consensus forecasts and a macroeconometric model at long horizons. The strength of a BVAR model lies in its ready applicability to forecast scenarios and the incorporation of external knowledge.

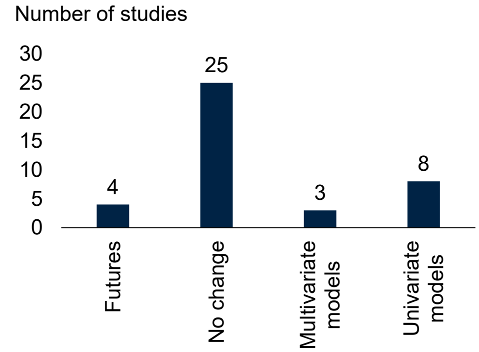
While these approaches are useful to anchor forecasts, they will always need to be supplemented by judgment. For policy makers, the wide range of forecasts and forecast errors is a reminder of the uncertainty around commodity price forecasts and the need to develop contingency plans for alternative outcomes.

Figure 1. Summary of studies of crude oil price forecast performance

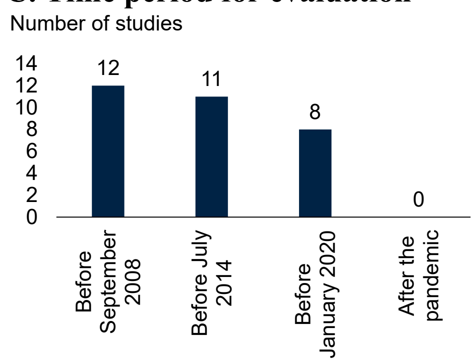
A. Forecasting methodologies evaluated



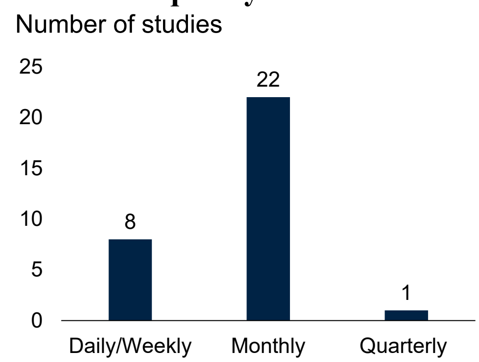
B. Benchmarks used for evaluation



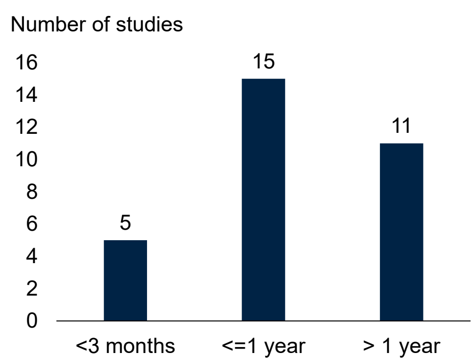
C. Time period for evaluation



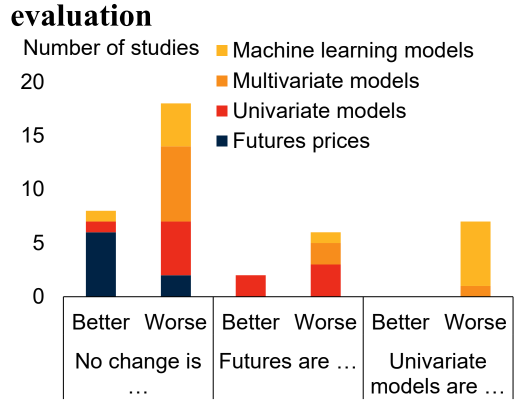
D. Data frequency



E. Forecast horizon



F. Outcomes of forecast performance evaluation



Source: World Bank.

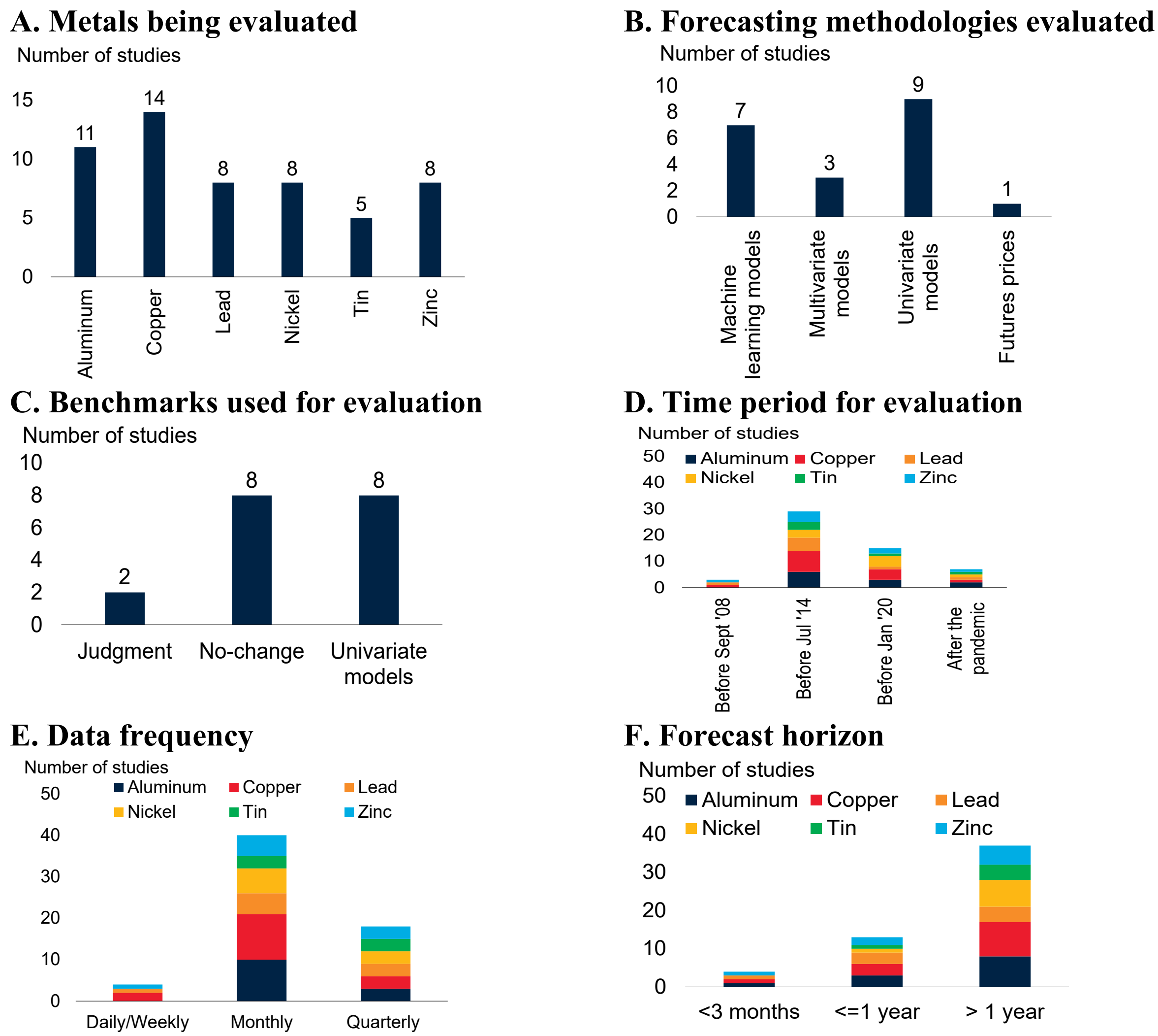
A. Number of studies that examine the forecast performance of futures prices, machine learning techniques, multivariate models (including structural VARs and Bayesian VARs), and univariate time series models against a benchmark. The two studies that examine both multivariate and univariate models are shown in the category for multivariate models.

B. Number of studies that benchmark forecast performance against latest spot prices ("no change"), futures prices, multivariate models (including structural VARs and Bayesian VARs), and univariate time series models. Studies that benchmark against both no-change forecasts and futures prices (9) and against both no-change forecasts and univariate models (2) are shown in the category for no-change benchmarks.

C.-E. Number of studies by end date of sample period (C), data frequency (D), and forecast horizon (E).

F. Number of studies in which benchmark approaches on the x-axis (no-change forecasts, futures, univariate models) had better or worse forecast performance than the approaches listed in the legend (futures prices, univariate models, multivariate models, and machine learning techniques).

Figure 2. Summary of studies of metal price forecast performance



Source: World Bank.

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.

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 examine both machine learning techniques and univariate models are 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 latest spot prices ("no change"), futures prices, multivariate models (including structural VARs and Bayesian VARs), and univariate time series models. Studies that benchmark against both no-change forecasts and futures prices (1) and against both no-change forecasts and univariate models (1) are shown in the category for no-change benchmarks. The one study that conducts a qualitative analysis is included in the category for judgment-based forecasts.

D. Number of studies that evaluate forecast performance for each metal price.

Table 1.A. Literature review of forecasting methods for crude oil prices

Authors and year	Crude oil prices	Methods	Main findings	Real or nominal prices	In-sample or out-of-sample forecasts
Abosedra (2006)	WTI	Futures, univariate models	Futures and univariate models are unbiased and weakly inefficient.	Nominal prices	In-sample forecasts
Abosedra and Baghestani (2004)	WTI	Futures, no-change	Futures prices underperform no-change forecasts.	Nominal prices	Out-of-sample forecasts
Alquist and Kilian (2010)	WTI	Futures, no-change	Futures are not the most accurate predictor of the spot oil price.	Nominal prices	Out-of-sample forecasts
Alquist, Kilian and Vigfusson (2013)	WTI	Futures, no-change, VAR	VAR models have lower out-of-sample forecasting errors than no-change. Long-term futures prices are less accurate than the no-change.	Both prices	Out-of-sample forecasts
Baumeister and Kilian (2014)	WTI	Futures, SVAR, no-change	VAR models generate more accurate forecasts than futures prices and no-change forecasts.	Real prices	Out-of-sample forecasts
Baumeister and Kilian (2015)	WTI	VAR model, futures prices, no-change, forecast combination models	Forecast combinations generate more accurate out-of-sample forecasts than no-change forecasts.	Real prices	Out-of-sample forecasts
Baumeister and Kilian (2012)	WTI	AR, ARMA, BVAR, VAR, futures prices, no-change	Recursive VAR forecasts have lower forecast errors at short horizons than other models.	Real prices	Out-of-sample forecasts
Chen (2014)	Brent; WTI	Predictive regression models, no-change	Oil-sensitive stock prices contain substantial information for predicting nominal and real crude oil prices at short horizons.	Both prices	Both types of forecasts

Table 1.A. Literature review of forecasting methods for crude oil prices (continued)

Authors and year	Crude oil prices	Methods	Main findings	Real or nominal prices	In-sample or out-of-sample forecasts
Cheng et al. (2019)	Brent; WTI	VEC-NAR, VAR, VECM, and GARCH models	The hybrid VEC-NAR model outperforms other models for longer forecast horizons.	Nominal prices	Out-of-sample forecasts
Chernenko, Schwarz and Wright (2004)	WTI	Futures, no-change	Forward or futures rates are not rational expectations of actual future prices.	Nominal prices	Both types of forecasts
Chinn and Coibion (2014)	WTI	Futures, GARCH, linear regressions, no-change	Futures prices outperform forecasts from reduced-form empirical models and no-change.	Nominal prices	Out-of-sample forecasts
Chu et al. (2022)	Brent	Futures, no-change	No-change forecast performs better than futures prices in the short term.	Nominal prices	Out-of-sample forecasts
Coimbra and Esteves (2004)	Crude oil price	Futures, random walk (carry over assumption)	No difference between opting for futures prices or using the carry-over assumption for short-term forecast horizons.	Nominal prices	In-sample forecasts
Coppola (2008)	WTI	VEC model, no-change	Forecasts from VEC models outperform no-change.	Nominal prices	Both types of forecasts
Cortazar, Ortega and Valencia (2021)	WTI	Multifactor stochastic pricing model; no-change; Bloomberg's consensus expected price model	Multifactor stochastic pricing models perform better than no-change and Bloomberg's consensus expected price models.	Both prices	Both types of forecasts
Drachal (2016)	WTI	ARIMA, futures prices, model averaging methods, no-change	Forecasts based on futures contracts produce larger errors than no-change.	Nominal prices	Out-of-sample forecasts

Table 1.A. Literature review of forecasting methods for crude oil prices (continued)

Authors and year	Crude oil prices	Methods	Main findings	Real or nominal prices	In-sample or out-of-sample forecasts
Fernandez (2007)	Arab Gulf Dubai	ARIMA, artificial neural networks, support vector regression	ARIMA model forecasts outperform artificial neural networks and support vector regression approaches only in the short term.	Nominal prices	Out-of-sample forecasts
Godarzi et al. (2014)	Crude oil price	Time series models, dynamic Nonlinear Auto Regressive model with eXogenous input (NARX)	NARX model is more accurate than the time series models in predicting oil prices.	Nominal prices	Out-of-sample forecasts
He (2018)	WTI	ARIMA, simple exponential smoothing, moving average, support vector regression	Support vector regression and ARIMA models have similar forecasting accuracy.	Nominal prices	Out-of-sample forecasts
Jin (2017)	WTI	Futures, no-change, unobserved components model	A futures-based unobserved components model outperforms no-change and futures prices.	Real prices	Out-of-sample forecasts
Kaboudan (2001)	Crude oil price	Genetic programming, no-change	No-change is inferior to forecasts derived from genetic programming models but outperforms those from artificial neural networks.	Nominal prices	Out-of-sample forecasts
Lalonde, Zhu and Demers (2003)	WTI	SVAR, VAR, AR (1) model, no-change	SVAR models outperform other models in out-of-sample oil price forecasting.	Real prices	Out-of-sample forecasts
Li et al. (2018)	Brent; WTI	Time series, ensemble empirical mode decomposition, sparse Bayesian learning	Ensemble empirical mode decomposition with sparse Bayesian learning and addition outperforms other forecasting methodologies.	Nominal prices	Out-of-sample forecasts

Table 1.A. Literature review of forecasting methods for crude oil prices (continued)

Authors and year	Crude oil prices	Methods	Main findings	Real or nominal prices	In-sample or out-of-sample forecasts
Lin and Sun (2020)	WTI	ARIMA, random walk, ensemble empirical mode decomposition	Ensemble empirical mode decomposition performs better than other models forecasting oil prices.	Nominal prices	Out-of-sample forecasts
Lu et al. (2021)	WTI	ARIMA, artificial neural networks, random walk, machine learning methods	The long short-term memory network method outperforms benchmark methods in both level and directional forecasting accuracy.	Both prices	Out-of-sample forecasts
Miao et al. (2017)	WTI	No-change, futures-based forecast, factor-based model, LASSO method, stepwise regression method	LASSO regression improves the accuracy of price forecasts compared to no-change and futures-based models.	Real prices	Out-of-sample forecasts
Mirmirani and Li (2004)	Crude oil price	Artificial neural networks, VAR	Artificial neural networks outperform VAR models.	Nominal prices	Out-of-sample forecasts
Moosa and Al-Loughani (1994)	WTI	Error-correction models, Futures prices, GARCH	Futures prices are neither unbiased nor efficient forecasters of spot prices.	Nominal prices	In-sample forecasts
Moshiri and Foroutan (2006)	Crude oil price	ARIMA, GARCH models, artificial neural networks	Artificial neural networks yield better forecasts than ARIMA and GARCH models.	Nominal prices	Out-of-sample forecasts
Mostafa and El-Masry (2016)	Crude oil price	ARIMA, artificial neural networks, gene expression programming (GEP)	The GEP model outperforms artificial neural networks and ARIMA models in predicting oil prices.	Nominal prices	Out-of-sample forecasts
Ramyar and Kianfar (2017)	Brent, WTI, Dubai Fateh	Artificial neural networks, VARs	Multi-layer perceptron neural networks can more accurately predict crude oil prices than a VAR model.	Nominal prices	Both types of forecasts

Table 1.A. Literature review of forecasting methods for crude oil prices (continued)

Authors and year	Crude oil prices	Methods	Main findings	Real or nominal prices	In-sample or out-of-sample forecasts
Reichsfeld and Roache (2011)	WTI	ARMA, futures prices, exponential smoother, error correction model, random walk	Futures prices and random walk models outperform other models over the short horizon.	Nominal prices	Both types of forecasts
Wu and McCallum (2005)	WTI	Hotelling's model, no-change, futures model, futures-spot spread model	Raw oil futures prices provide relatively less accurate forecasts than the futures-spot spread model.	Nominal prices	Both types of forecasts
Xie et al. (2006)	WTI	ARIMA, artificial neural networks, support vector machines	Support vector machines are better than other forecasting methods but sometimes underperform ARIMA and artificial neural network methods.	Nominal prices	Out-of-sample forecasts
Xiong, Bao and Hu (2013)	WTI	Empirical mode decomposition models, random walk	The empirical mode decomposition-slope-based method has the best prediction accuracy.	Nominal prices	Out-of-sample forecasts
Ye, Zyren and Shore (2005)	WTI	No-change, relative stock model, modified alternative model	The relative stock model produces the best out-of-sample forecast results and no-change has the worst.	Nominal prices	Both types of forecasts
Yousefi, Weinreich and Reinarz (2005)	Crude oil price	Futures prices, wavelets	Wavelet-based forecasts outperform futures prices in the short term.	Nominal prices	Out-of-sample forecasts
Yu, Wang and Lai (2008)	Brent; WTI	ARIMA, artificial neural networks	The neural network ensemble learning model performs better than other models.	Nominal prices	Out-of-sample forecasts

Table 1.A. Literature review of forecasting methods for crude oil prices (continued)

Authors and year	Crude oil prices	Methods	Main findings	Real or nominal prices	In-sample or out-of-sample forecasts
Zeng and Swanson (1998)	Crude oil price	Random walk, VAR, VECM	Error-correction models perform better in shorter forecast horizons.	Nominal prices	Both types of forecasts
Zhao, Li and Yu (2017)	WTI	Deep learning neural network model, multivariate forecasting models	The deep learning approach outperforms multivariate forecasting models.	Nominal prices	Out-of-sample forecasts

Note: AR (Autoregressive Model), ARMA (Autoregressive–Moving-Average Model), ARIMA (Autoregressive Integrated Moving Average Model); BVAR (Bayesian Vector Autoregressive Model); GARCH (Generalized Autoregressive Conditional Heteroskedasticity); Least Absolute Shrinkage and Selection Operator (LASSO); SVAR (Structural Vector Autoregressive Model); VECM (Vector Error Correction Model); VEC-NAR (Vector Error Correction Model and nonlinear autoregressive neural network); VAR (Vector Autoregressive Model); WTI (West Texas Intermediate).

Table 1.B. Literature review of forecasting methods for metals prices

Authors and year	Metals	Models	Main findings	Real, nominal or both prices	In-sample, out-of-sample or both types of forecasts
Alipour, Khodaiari and Jafari (2019)	Copper	ARIMA, TGARCH, stochastic differential equations	Stochastic differential equations are better at forecasting copper price movements than traditional linear or non-linear functional forms.	Nominal prices	Out-of-sample forecasts
Bowman and Husain (2004)	Aluminum, copper, lead, nickel, tin, zinc, others	ARMA, error-correction models, judgmental models	Futures-based models have better statistical- and directional- forecast accuracy than historical-data-based or judgment approaches.	Nominal prices	Out-of-sample forecasts
Buncic and Moretto (2015)	Copper	Model averaging methods, random walk	Model averaging methods outperform random walk.	Nominal prices	Out-of-sample forecasts
Castro, Araujo and de Avila Montini (2013)	Aluminum	ARIMA, VAR, VEC models	VEC yields better forecast accuracy than VAR models.	Nominal prices	Out-of-sample forecasts
Chen, He and Zhang (2016)	Aluminum and nickel	ARMA, grey wave prediction method, random walk	Grey wave prediction methods forecasts outperform those from univariate models.	Nominal prices	Out-of-sample forecasts
Chinn and Coibion (2014)	Aluminum, copper, lead, nickel, tin, others	Futures, GARCH, linear regressions, random walk	Random walk modestly outperforms futures prices.	Nominal prices	Out-of-sample forecasts
Dooley and Lenihan (2005)	Lead and zinc	ARIMA, lagged forward price model	ARIMA models provide superior forecasting results for lead.	Nominal prices	Out-of-sample forecasts
Du et al. (2021)	Copper	Hybrid machine learning model, individual prediction model	Hybrid method significantly outperforms comparison models in metal price prediction.	Nominal prices	Out-of-sample forecasts

Table 1.B. Literature review of forecasting methods for metals prices (continued)

Authors and year	Metals	Models	Main findings	Real, nominal or both prices	In-sample, out-of-sample or both types of forecasts
He et al. (2015)	Lead and zinc	ARMA, curvelet based multi-scale forecasting, random walk	Curvelet-based forecasting algorithms are superior to traditional benchmark models.	Nominal prices	Both types of forecasts
Issler, Rodrigues and Burjack (2014)	Aluminum, copper, lead, nickel, tin, zinc	AR, VAR, VECM, restricted VECM	AR models are best for aluminum and copper, VARs are best for lead and zinc, and VECMs are best for nickel and tin.	Real prices	Out-of-sample forecasts
Kahraman and Akay (2022)	Aluminum, copper, lead, iron, nickel, tin, and zinc	Exponential smoothing, mean, naive, and ARIMA methods	The damped trend model is best for aluminum, copper, lead, and iron prices; the Holt model is best for nickel and zinc prices; and the Brown model is best for tin prices.	Real prices	Out-of-sample forecasts
Khoshalan et al. (2021)	Copper	Gene expression programming, artificial neural network, Adaptive neuro-fuzzy inference system	Artificial neural network was found to be the best approach for predicting copper prices.	Nominal prices	Out-of-sample forecasts
Kriechbaumer et al. (2014)	Aluminum, copper, lead, zinc	Wavelet-autoregressive integrated moving average	ARIMA model forecasts improve substantially when combined with wavelet-based multi-resolution analysis.	Nominal prices	Out-of-sample forecasts
Lasheras et al. (2015)	Copper	ARIMA and artificial neural networks models	Artificial neural network models perform better than ARIMA models.	Nominal prices	Out-of-sample forecasts

Table 1.B. Literature review of forecasting methods for metals prices (continued)

Authors and year	Metals	Models	Main findings	Real, nominal or both prices	In-sample, out-of-sample or both types of forecasts
Pincheira-Brown and Hardy (2019)	Aluminum, copper, lead, nickel, tin, zinc	AR, Random Walk, linear specifications, threshold regressions, Markov switching models	Accounting for Chilean peso dynamics improves metal price forecasts.	Nominal prices	Both types of forecasts
Reichsfeld and Roache (2011)	Aluminum, copper, others	ARMA, ARIMA, futures, random walk, exponential smoother, error correction model	Futures prices perform better at short horizons; time series models underperform random walk.	Nominal prices	Both types of forecasts
Mysen and Thornton (2021)	Aluminum	VECMs and a machine learning model	Machine learning models produce the most reliable and accurate forecasts.	Nominal prices	Out-of-sample forecasts
Rubaszek, Karolak and Kwas (2020)	Aluminum, copper, nickel, zinc	AR, threshold Autoregressive model, VAR model, threshold vector autoregressive (TVAR), random walk	Mean-reverting models provide better forecasts than naive random walk model; allowing for non-linearity does not improve the quality of forecasts.	Real prices	Out-of-sample forecasts
Villegas (2021)	Nickel	ARIMA, artificial neural networks, GARCH models	Artificial neural network methods yield more accurate forecasts than ARIMA and GARCH techniques.	Real prices	Out-of-sample forecasts
Wang et al. (2019)	Copper	Artificial neural networks	Hybrid artificial neural network techniques have more favorable forecasts in both level and directional accuracy compared with those of traditional artificial neural network techniques.	Nominal prices	Out-of-sample forecasts

Note: AR (Autoregressive Model), ARMA (Autoregressive–Moving-Average Model), ARIMA (Autoregressive Integrated Moving Average Model); GARCH (Generalized Autoregressive Conditional Heteroskedasticity); TVAR (Threshold vector autoregressive model); TGARCH (Threshold Generalized Autoregressive Conditional Heteroskedasticity); VECM (Vector Error Correction Model); VAR (Vector Autoregressive Model).

Table 1.C. Literature review of forecasting methods crude oil and metal prices volatility

Authors and year	Metals	Models	Main findings	Real, nominal or both prices	In-sample, out-of-sample or both types of forecasts
Astudillo et al. (2020)	Copper	Support vector regressions	Support vector regressions provide good prediction accuracy for copper price volatilities over the short horizon.	Nominal prices	Out-of-sample forecasts
Cheong (2009)	WTI and Brent	GARCH type models	The intensity of long-persistence volatility in WTI is greater than in the Brent.	Nominal prices	Out-of-sample forecasts
Degiannakis and Filis (2017)	Brent	Heterogeneous autoregressive models	Information channels improves predictive accuracy of oil price volatility.	Nominal prices	Out-of-sample forecasts
Dehghani (2018)	Copper	GEP, multivariate regression methods	GEP yields better prediction accuracy than time series and multivariate regression methods.	Real prices	Out-of-sample forecasts
Gong and Lin (2018)	Copper	Heterogeneous autoregressive models	Accounting for structural breaks in heterogeneous autoregressive models improves forecasts.	Nominal prices	Both types of forecasts
Haugom et al. (2014)	WTI	Heterogeneous autoregressive models	Including implied volatility and market variables improves volatility forecasts.	Nominal prices	Out-of-sample forecasts
Mohammadi and Su (2010)	Various benchmark prices	GARCH, EGARCH and APARCH and FIGARCH	Forecasting accuracy of the APARCH model outperforms the other GARCH models.	Nominal prices	Out-of-sample forecasts
Tang (2010)	Aluminum and copper	Standard GARCH models, Regime Switching GARCH (MRS-GARCH)	MRS-GARCH models outperform standard GARCH models in predicting metals prices.	Nominal prices	Both types of forecasts

Table 1.C. Literature review of forecasting methods crude oil and metal prices volatility (continued)

Authors and year	Metals	Models	Main findings	Real, nominal or both prices	In-sample, out-of-sample or both types of forecasts
Vo (2009)	WTI	Markov switching stochastic volatility (MSSV) model; stochastic volatility (SV) model, GARCH model and Markov switching (MS) model	Out-of-sample forecasts suggest that the MSSV outperforms other models.	Nominal prices	Both types of forecasts
Wang, Wu and Yang (2016)	WTI and Brent	Markov switching multifractal (MSM) volatility model, GARCH models, historical volatility (HV) model	MSM models have greater forecasting abilities than the GARCH or HV models.	Nominal prices	Both types of forecasts
Wen, Gong and Cai (2016)	WTI	Heterogeneous autoregressive models	Different models exhibit different predictive power in forecasting the 1-day, 1-week and 1-month volatility of crude oil futures.	Nominal prices	Out-of-sample forecasts

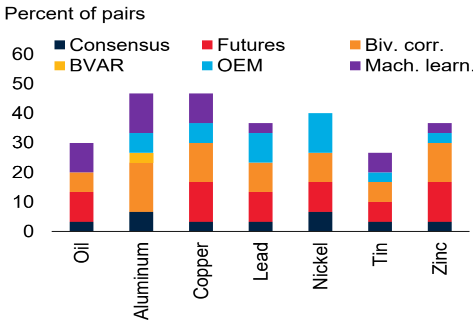
Note: APARCH (Asymmetric Power ARCH); EGARCH (Exponential GARCH); FIGARCH (Fractionally Integrated Generalized Autoregressive Conditionally Heteroskedasticity); GARCH (Generalized Autoregressive Conditional Heteroskedasticity); Gene expression programming (GEP); HAR (Heterogeneous Autoregressive-type volatility models); TGARCH (Threshold Generalized Autoregressive Conditional Heteroskedasticity).

Table 2. Sign restrictions on impulse responses

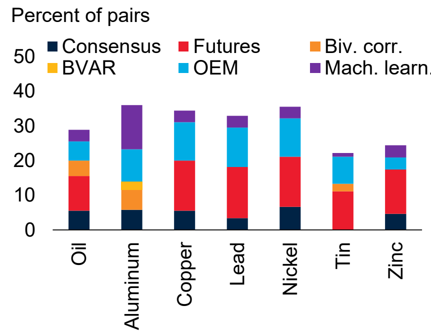
	Supply shocks	Demand shocks
Oil or metal production	−	+
Global economic activity	−	+
Real price of oil or metals	+	+

Figure 3. Approach with the best forecast performance

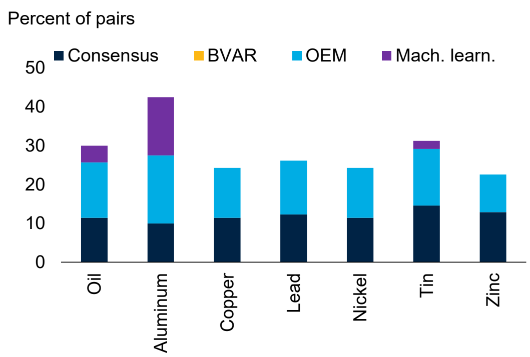
A. 3-month horizon, by Diebold-Mariano test



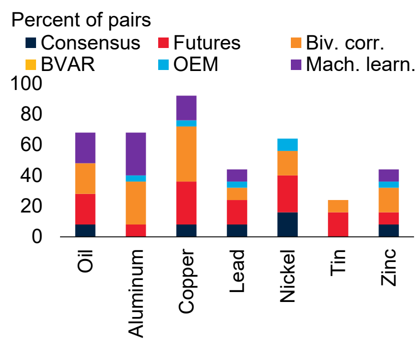
B. 6-12-month horizon, by Diebold-Mariano test



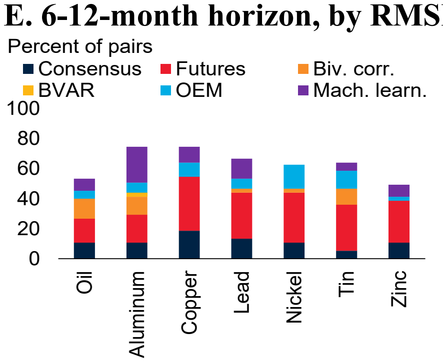
C. 15-24-month horizon, by Diebold-Mariano test



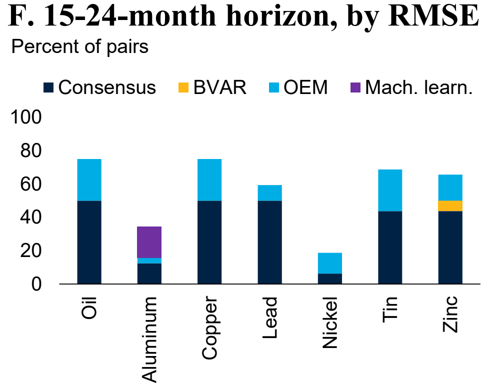
D. 3-month horizon, by RMSE



E. 6-12-month horizon, by RMSE



F. 15-24-month horizon, by RMSE



Source: World Bank.
Note: "Biv. corr." stands for bivariate correlations, "Mach. Learn." stands for machine learning approach. Charts show the percent of comparisons between all pairs of the six approaches in which the Diebold-Mariano statistic (A-C) or a lower RMSE (D-F) indicates that the approach shown on the x-axis has statistically significantly better forecast performance. Futures prices and bilateral correlations are not evaluated beyond the 12-month horizon, as discussed in the text.

Table 3.A. Bias of commodity price forecasts

(U.S. dollars per metric tonne unless otherwise specified)

Commodity	Horizon (months)	Consensus forecasts	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Oil	3	-3.1	-1.5	1.7	2.2	3.6	2.2
(U.S. dollars per barrel)	6	-2	-1.6	4.6	3.8	2.5	3.8
	9	-1	-2	5.8	5.4	1.5	3.6
	12	-0.1	-2.3	7.6	8	1.3	3.2
	15	0.7	…	…	10.7	1.3	2.6
	18	1.3	…	…	14.6	1.3	2.1
	21	1.8	…	…	17.3	1.5	2.3
	24	2.3	…	…	25.4	0	2.4
Aluminum	3	-6	354	6	-23	16	14
	6	-3	355	10	-28	0	22
	9	-3	363	114	-28	-9	22
	12	-2	371	235	-2	-7	14
	15	4	…	…	-49	-5	7
	18	12	…	…	-17	7	4
	21	19	…	…	11	17	-6
	24	30	…	…	23	-7	-16
Copper	3	-121	8	133	-69	-73	-49
	6	-112	29	527	-92	-92	-31
	9	-107	43	875	-138	-103	-28
	12	-93	52	1105	-95	-88	-30
	15	-71	…	…	-298	-64	3
	18	-36	…	…	-194	-12	43
	21	4	…	…	-70	43	78
	24	52	…	…	-119	117	104
Lead	3	-71	-39	-3	-49	-77	-115
	6	-75	-36	138	-68	-85	-124
	9	-76	-34	211	-82	-89	-125
	12	-76	-32	222	-50	-81	-119
	15	-78	…	…	-68	-88	-100
	18	-81	…	…	-35	-76	-78
	21	-79	…	…	69	-68	-88
	24	-78	…	…	103	-21	-87
Nickel	3	-546	-221	-125	-309	-222	-1279
	6	-344	-193	1123	-267	-127	-1090
	9	-167	-162	1057	-333	-128	-1143
	12	22	-132	1950	-155	23	-1276
	15	181	…	…	-393	253	-1371
	18	296	…	…	-371	442	-1407
	21	425	…	…	624	685	-1387
	24	598	…	…	1136	477	-1315
Tin	3	-910	-479	197	-1092	21	-38
	6	-1168	-675	691	-1636	-245	151
	9	-1357	-814	1542	-2614	-497	290
	12	-1402	-939	2622	-3149	-661	436
	15	-1422	…	…	-4092	-831	683
	18	-1430	…	…	-4287	-839	890
	21	-1381	…	…	-4028	-17103	1102
	24	-1380	…	…	-4202	-901	1347
Zinc	3	-98	-45	-20	-102	-55	-56
	6	-103	-52	130	-130	-64	-34
	9	-125	-62	227	-127	-73	-16
	12	-131	-75	331	-106	-84	2
	15	-135	…	…	-177	-86	27
	18	-141	…	…	-161	-75	57
	21	-147	…	…	-105	-69	68
	24	-157	…	…	-110	12	70

Source: World Bank.

Note: Bias is defined as the difference between the actual price and the predicted price. The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency.

Table 3.B. Root mean squared error of commodity price forecasts

(U.S. dollars per metric tonne unless otherwise specified)

Commodity	Horizon (months)	Consensus forecasts	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Oil	3	6	4	4	13	9	4
(U.S. dollars per barrel)	6	7	5	8	16	8	6
	9	8	6	7	21	7	10
	12	9	7	7	27	8	16
	15	10	…	…	31	8	21
	18	11	…	…	33	9	24
	21	12	…	…	34	10	25
	24	13	…	…	33	11	24
Aluminum	3	170	147	47	232	110	67
	6	176	133	37	266	109	75
	9	190	123	265	315	133	96
	12	203	117	567	306	175	133
	15	218	…	…	410	209	173
	18	233	…	…	373	221	206
	21	246	…	…	354	232	233
	24	249	…	…	358	276	258
Copper	3	441	145	74	790	334	277
	6	477	146	866	872	362	387
	9	551	149	1212	1166	407	684
	12	637	153	1190	1282	477	1062
	15	701	…	…	1791	528	1386
	18	769	…	…	1732	577	1637
	21	841	…	…	1630	604	1806
	24	859	…	…	1722	729	1925
Lead	3	154	104	106	253	95	134
	6	157	100	304	267	105	126
	9	165	97	300	337	131	150
	12	176	95	241	358	167	208
	15	188	…	…	361	204	278
	18	194	…	…	339	224	331
	21	195	…	…	353	236	364
	24	204	…	…	371	200	379
Nickel	3	1443	878	990	2763	959	2327
	6	1700	828	3134	2934	742	2380
	9	2044	794	2064	3307	763	2635
	12	2337	769	2520	3662	1221	2884
	15	2529	…	…	4033	1601	3053
	18	2756	…	…	4016	1784	3162
	21	2886	…	…	3196	2000	3208
	24	3011	…	…	3377	2233	3173
Tin	3	2171	1425	1576	2410	1959	1805
	6	2537	1280	1723	3242	1167	1969
	9	3021	1217	2377	4907	1395	2887
	12	3457	1181	3081	5482	1989	4114
	15	3743	…	…	6794	2672	5243
	18	3956	…	…	7066	3277	6103
	21	4077	…	…	6696	2369	6640
	24	4188	…	…	6632	4023	6880
Zinc	3	214	151	146	402	206	202
	6	213	146	313	389	190	200
	9	226	142	327	395	210	243
	12	255	142	343	419	260	341
	15	282	…	…	538	282	455
	18	309	…	…	553	298	559
	21	330	…	…	492	303	634
	24	357	…	…	426	290	686

Source: World Bank.

Table 3.C. Direction of commodity price forecasts errors

(Percent of forecast quarters)

Commodity	Horizon (months)	Percent of quarters with correctly forecast direction of change: Consensus forecasts	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Oil	3	69	97	97	61	86	83
(U.S. dollars per barrel)	6	76	97	90	50	79	83
	9	76	93	97	29	86	76
	12	76	93	93	39	86	62
	15	66	…	…	39	86	59
	18	79	…	…	54	79	55
	21	83	…	…	57	79	48
	24	79	…	…	57	46	55
Aluminum	3	86	72	86	79	100	97
	6	86	79	97	86	100	100
	9	83	83	90	79	100	100
	12	86	79	93	71	93	97
	15	83	…	…	79	93	93
	18	83	…	…	82	93	93
	21	86	…	…	86	93	93
	24	83	…	…	82	55	83
Copper	3	55	86	97	50	86	83
	6	55	86	93	61	93	79
	9	55	86	93	54	93	76
	12	52	86	93	43	86	72
	15	55	…	…	54	86	48
	18	52	…	…	57	86	41
	21	59	…	…	50	86	41
	24	66	…	…	61	55	41
Lead	3	59	83	93	54	79	72
	6	62	83	86	50	64	76
	9	59	83	83	46	79	76
	12	52	79	86	50	64	76
	15	48	…	…	50	64	62
	18	52	…	…	61	64	45
	21	66	…	…	54	64	41
	24	48	…	…	61	46	45
Nickel	3	72	93	97	50	93	86
	6	72	93	93	57	86	86
	9	72	93	93	61	86	72
	12	59	93	93	54	86	72
	15	59	…	…	54	79	59
	18	55	…	…	61	79	52
	21	59	…	…	57	79	52
	24	59	…	…	68	64	48
Tin	3	69	86	100	68	100	86
	6	69	90	100	68	93	72
	9	66	90	90	64	93	72
	12	69	93	93	57	93	62
	15	72	…	…	68	93	52
	18	69	…	…	68	86	31
	21	72	…	…	64	33	24
	24	66	…	…	71	64	31
Zinc	3	72	86	100	50	86	79
	6	69	86	93	54	93	76
	9	72	86	90	46	100	76
	12	72	83	90	50	100	69
	15	72	…	…	54	100	69
	18	76	…	…	61	100	72
	21	79	…	…	61	100	69
	24	59	…	…	68	64	69

Source: World Bank.

Table 3.D. Directional accuracy: Pesaran and Timmerman (2009) test

Commodity	Horizon (months)	Consensus forecasts	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Oil	3	1.7*	5.2***	5.2***	1.4†	3**	3.6***
(U.S. dollars per barrel)	6	2.6**	5.2***	4.3***	-0.1	2.3*	3.6***
	9	2.6**	4.7***	5.2***	-2.2	3**	2.9**
	12	2.7**	4.7***	4.8***	-1.1	3**	1.7*
	15	1.6†	…	…	-1.3	3**	1.5†
	18	3.1**	…	…	0.4	2.9**	1.2
	21	3.6***	…	…	0.7	2.9**	0.3
	24	3.2***	…	…	0.7	2.1*	0.9
Aluminum	3	4***	2.4**	4***	3.1***	4***	5.2***
	6	4***	3.3***	5.2***	3.9***	4***	5.6***
	9	3.6***	3.7***	4.4***	3.1***	4***	5.6***
	12	4***	3.3***	4.8***	2.3*	3.5***	5.2***
	15	3.6***	…	…	3.1***	3.5***	4.8***
	18	3.6***	…	…	3.5***	3.5***	4.8***
	21	4***	…	…	3.9***	3.5***	4.8***
	24	3.6***	…	…	3.4***	2.3*	3.6***
Copper	3	0.5	4.1***	5.2***	0	3**	3.6***
	6	0.5	4.1***	4.8***	1.1	3.5***	3.3***
	9	0.4	4.1***	4.8***	0.3	3.5***	3**
	12	0.1	4.1***	4.8***	-1	2.8**	2.5**
	15	0.4	…	…	0.4	2.8**	-0.1
	18	0.1	…	…	0.8	2.8**	-0.8
	21	0.8	…	…	0	2.8**	-0.9
	24	1.7*	…	…	1.2	2.1*	-0.9
Lead	3	0.9	3.7***	4.8***	0.4	2.3*	2.5**
	6	1.3†	3.7***	4***	0	1.3†	2.9**
	9	0.9	3.7***	3.6***	-0.4	2.3*	2.9**
	12	0.2	3.3***	4***	0	1.8*	2.9**
	15	-0.2	…	…	0	1.7*	1.5†
	18	0.2	…	…	1.2	1.7*	-0.5
	21	1.7*	…	…	0.4	1.7*	-1
	24	-0.1	…	…	1.2	0.7	-0.6
Nickel	3	2.5**	4.8***	5.2***	0	3.5***	4.1***
	6	2.5**	4.8***	4.8***	0.8	2.9**	4.1***
	9	2.5**	4.8***	4.8***	1.2	2.9**	2.5**
	12	0.9	4.8***	4.8***	0.4	2.9**	2.5**
	15	0.9	…	…	0.4	2.3*	0.9
	18	0.5	…	…	1.2	2.3*	0.1
	21	0.9	…	…	0.8	2.3*	0.1
	24	0.9	…	…	1.9*	3.5***	-0.3
Tin	3	2.1*	4.2***	5.6***	2*	4***	4.2***
	6	2.2*	4.5***	5.6***	2*	3.4***	2.8**
	9	1.8*	4.5***	4.4***	1.6†	3.4***	2.6**
	12	2.3*	4.8***	4.8***	0.8	3.4***	1.5†
	15	2.6**	…	…	2*	3.4***	0.3
	18	2.3*	…	…	2*	2.7**	-2.1
	21	2.6**	…	…	1.6†	-0.4	-2.9
	24	1.8*	…	…	2.4**	2.8**	-2.1
Zinc	3	2.4**	4.1***	5.6***	0.1	3**	3**
	6	2.1*	4.1***	4.7***	0	3.5***	2.7**
	9	2.4**	4.1***	4.3***	-0.8	4***	2.7**
	12	2.4**	3.8***	4.3***	-0.2	4***	2.1*
	15	2.4**	…	…	0	4***	2.1*
	18	2.5**	…	…	0.9	4***	2.7**
	21	3**	…	…	0.9	4***	2.5**
	24	1	…	…	1.8*	2*	2.1*

Source: World Bank.

Note: Test statistics indicate statistically significantly accurate forecast directions for the model indicated in columns. *** indicates statistically significantly more accurate model in the row at the 0.1 percent significance level, ** at the 1 percent level, * at the 5 percent level, and † at the 10 percent level.

Table 4.A. Model comparison: Bias of oil price forecasts
(U.S. dollars per barrel)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	-3.1	…	…	…	…	…
Futures	3		-1.5	…	…	…	…
Bivariate correlations	3			1.7	…	…	…
BVAR	3				2.2	…	…
OEM	3					3.6	…
Machine learning	3						2.2
Consensus	6	-2	…	…	…	…	…
Futures	6		-1.6	…	…	…	…
Bivariate correlations	6			4.6	…	…	…
BVAR	6				3.8	…	…
OEM	6					2.5	…
Machine learning	6						3.8
Consensus	9	-1	…	…	…	…	…
Futures	9		-2	…	…	…	…
Bivariate correlations	9	*	*	5.8	…	…	…
BVAR	9				5.4	…	…
OEM	9					1.5	…
Machine learning	9						3.6
Consensus	12	-0.1	…	…	…	…	…
Futures	12		-2.3	…	…	…	…
Bivariate correlations	12	*	*	7.6	…	…	…
BVAR	12				8	…	…
OEM	12			*		1.3	…
Machine learning	12						3.2
Consensus	15	0.7	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15		…	…	10.7		…
OEM	15		…	…		1.3	…
Machine learning	15		…	…			2.6
Consensus	18	1.3	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18	*	…	…	14.6	…	…
OEM	18		…	…	*	1.3	…
Machine learning	18		…	…			2.1
Consensus	21	1.8	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21	*	…	…	17.3	…	…
OEM	21		…	…	*	1.5	…
Machine learning	21		…	…			2.3
Consensus	24	2.3	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24	*	…	…	25.4	…	…
OEM	24		…	…	*	0	…
Machine learning	24		…	…	*		2.4

Source: World Bank.
Note: Bias is defined as the difference between the actual price and the predicted price. The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate estimates of average bias for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to a t-test (bias).

Table 4.B. Model comparison: Root mean squared error of oil price forecasts

(U.S. dollars per barrel)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	6	…	…	…	…	…
Futures	3	*	3.5	…	…	…	…
Bivariate correlations	3	*		3.5	…	…	…
BVAR	3	*	*	*	13.1	…	…
OEM	3		*	*		8.5	…
Machine learning	3	*			*	*	4.1
Consensus	6	6.6	…	…	…	…	…
Futures	6		5	…	…	…	…
Bivariate correlations	6		*	7.5	…	…	…
BVAR	6	*	*	*	15.7	…	…
OEM	6				*	7.7	…
Machine learning	6				*		5.6
Consensus	9	7.9	…	…	…	…	…
Futures	9		6.2	…	…	…	…
Bivariate correlations	9			6.5	…	…	…
BVAR	9	*	*	*	21.1	…	…
OEM	9				*	7.4	…
Machine learning	9		*	*	*		10.1
Consensus	12	9.2	…	…	…	…	…
Futures	12		7.2	…	…	…	…
Bivariate correlations	12			6.6	…	…	…
BVAR	12	*	*	*	26.6	…	…
OEM	12				*	8.1	…
Machine learning	12	*	*	*	*	*	16.1
Consensus	15	10.3	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15	*	…	…	31		…
OEM	15		…	…	*	8.4	…
Machine learning	15	*	…	…		*	21.4
Consensus	18	11.1	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18	*	…	…	32.7	…	…
OEM	18		…	…	*	8.9	…
Machine learning	18	*	…	…		*	24.2
Consensus	21	11.9	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21	*	…	…	34.4	…	…
OEM	21		…	…	*	9.5	…
Machine learning	21	*	…	…		*	24.6
Consensus	24	12.8	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24	*	…	…	33.1	…	…
OEM	24		…	…	*	10.8	…
Machine learning	24	*	…	…		*	23.7

Source: World Bank.

Note: The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate average RMSE estimates for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to an F-test (RMSE).

Table 5.A. Model comparison: Bias of aluminum price forecasts

(U.S. dollars per metric tonne)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	-6.1	…	…	…	…	…
Futures	3	*	353.8	…	…	…	…
Bivariate correlations	3		*	5.8	…	…	…
BVAR	3		*		-23.3	…	…
OEM	3		*			15.9	…
Machine learning	3		*				14.2
Consensus	6	-2.7	…	…	…	…	…
Futures	6	*	354.7	…	…	…	…
Bivariate correlations	6		*	10.2	…	…	…
BVAR	6		*		-27.9	…	…
OEM	6		*			-0.3	…
Machine learning	6		*				22.4
Consensus	9	-2.6	…	…	…	…	…
Futures	9	*	362.7	…	…	…	…
Bivariate correlations	9		*	113.7	…	…	…
BVAR	9		*		-27.5	…	…
OEM	9		*			-9.4	…
Machine learning	9		*				21.8
Consensus	12	-1.9	…	…	…	…	…
Futures	12	*	370.8	…	…	…	…
Bivariate correlations	12	*		235.3	…	…	…
BVAR	12		*		-2.2	…	…
OEM	12		*			-7.3	…
Machine learning	12		*	*			13.5
Consensus	15	3.9	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15		…	…	-49		…
OEM	15		…	…		-4.6	…
Machine learning	15		…	…			7.4
Consensus	18	11.8	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18		…	…	-16.9	…	…
OEM	18		…	…		6.8	…
Machine learning	18		…	…			4.1
Consensus	21	18.9	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21		…	…	11.2	…	…
OEM	21		…	…		16.6	…
Machine learning	21		…	…			-5.7
Consensus	24	30.1	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24		…	…	22.8	…	…
OEM	24		…	…		-7.3	…
Machine learning	24		…	…			-15.7

Source: World Bank.

Note: Bias is defined as the difference between the actual price and the predicted price. The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate estimates of average bias for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to a t-test (bias).

Table 5.B. Model comparison: Root mean squared error of aluminum price forecasts

(U.S. dollar per metric tonne)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	170.3	…	…	…	…	…
Futures	3		146.6	…	…	…	…
Bivariate correlations	3	*	*	47.1	…	…	…
BVAR	3		*	*	232.4	…	…
OEM	3			*	*	109.5	…
Machine learning	3	*	*		*	*	67.3
Consensus	6	175.8	…	…	…	…	…
Futures	6		133.1	…	…	…	…
Bivariate correlations	6	*	*	37.3	…	…	…
BVAR	6	*	*	*	265.8	…	…
OEM	6			*	*	109.3	…
Machine learning	6	*	*	*	*		75.1
Consensus	9	190.4	…	…	…	…	…
Futures	9	*	123	…	…	…	…
Bivariate correlations	9		*	265.3	…	…	…
BVAR	9	*	*		314.5	…	…
OEM	9			*	*	133	…
Machine learning	9	*		*	*		96.3
Consensus	12	202.6	…	…	…	…	…
Futures	12	*	117.3	…	…	…	…
Bivariate correlations	12	*	*	567.4	…	…	…
BVAR	12	*	*	*	305.7	…	…
OEM	12			*	*	174.7	…
Machine learning	12	*		*	*		133.4
Consensus	15	218.1	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15	*	…	…	409.8		…
OEM	15		…	…	*	209.2	…
Machine learning	15		…	…	*		173
Consensus	18	232.9	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18	*	…	…	373.3	…	…
OEM	18		…	…		221	…
Machine learning	18		…	…	*		205.5
Consensus	21	246	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21		…	…	353.8	…	…
OEM	21		…	…		232	…
Machine learning	21		…	…	*		232.6
Consensus	24	249.3	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24		…	…	358.1	…	…
OEM	24		…	…		275.7	…
Machine learning	24		…	…			257.8

Source: World Bank.

Note: The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate average RMSE estimates for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to an F-test (RMSE).

Table 6.A. Model comparison: Bias of copper price forecasts

(U.S. dollars per metric tonne)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	-121	…	…	…	…	…
Futures	3		7.8	…	…	…	…
Bivariate correlations	3		*	132.6	…	…	…
BVAR	3				-69	…	…
OEM	3					-73.1	…
Machine learning	3						-49.1
Consensus	6	-111.6	…	…	…	…	…
Futures	6		29.1	…	…	…	…
Bivariate correlations	6	*	*	526.8	…	…	…
BVAR	6				-91.9	…	…
OEM	6			*		-92.1	…
Machine learning	6			*			-30.8
Consensus	9	-106.5	…	…	…	…	…
Futures	9		43.1	…	…	…	…
Bivariate correlations	9	*	*	875.4	…	…	…
BVAR	9			*	-137.8	…	…
OEM	9			*		-103.4	…
Machine learning	9			*			-27.8
Consensus	12	-93.4	…	…	…	…	…
Futures	12		52.1	…	…	…	…
Bivariate correlations	12	*	*	1105.1	…	…	…
BVAR	12			*	-95.2	…	…
OEM	12			*		-88.3	…
Machine learning	12			*			-29.5
Consensus	15	-70.9	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15		…	…	-297.6		…
OEM	15		…	…		-64.1	…
Machine learning	15		…	…			3.4
Consensus	18	-36.3	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18		…	…	-193.7	…	…
OEM	18		…	…		-12	…
Machine learning	18		…	…			43
Consensus	21	4.2	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21		…	…	-69.9	…	…
OEM	21		…	…		43.4	…
Machine learning	21		…	…			78.1
Consensus	24	51.6	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24		…	…	-119.4	…	…
OEM	24		…	…		116.6	…
Machine learning	24		…	…			103.6

Source: World Bank.

Note: Bias is defined as the difference between the actual price and the predicted price. The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate estimates of average bias for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to a t-test (bias).

Table 6.B. Model comparison: Root mean squared error of copper price forecasts

(U.S. dollars per metric tonne)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	440.6	…	…	…	…	…
Futures	3	*	145.2	…	…	…	…
Bivariate correlations	3	*	*	74	…	…	…
BVAR	3	*	*	*	789.8	…	…
OEM	3		*	*	*	334.2	…
Machine learning	3	*	*	*	*		277.1
Consensus	6	477.2	…	…	…	…	…
Futures	6	*	146.2	…	…	…	…
Bivariate correlations	6	*	*	865.8	…	…	…
BVAR	6	*	*		872	…	…
OEM	6		*	*	*	362.2	…
Machine learning	6		*	*	*		387.1
Consensus	9	551.4	…	…	…	…	…
Futures	9	*	148.7	…	…	…	…
Bivariate correlations	9	*	*	1212.2	…	…	…
BVAR	9	*	*		1165.7	…	…
OEM	9		*	*	*	406.7	…
Machine learning	9		*	*	*		684.4
Consensus	12	637.1	…	…	…	…	…
Futures	12	*	152.9	…	…	…	…
Bivariate correlations	12	*	*	1189.7	…	…	…
BVAR	12	*	*		1281.9	…	…
OEM	12		*	*	*	476.6	…
Machine learning	12	*	*			*	1062.3
Consensus	15	701.3	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15	*	…	…	1791		…
OEM	15		…	…	*	528.1	…
Machine learning	15	*	…	…		*	1385.7
Consensus	18	768.7	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18	*	…	…	1731.8	…	…
OEM	18		…	…	*	576.5	…
Machine learning	18	*	…	…		*	1637
Consensus	21	841.1	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21	*	…	…	1629.9	…	…
OEM	21		…	…	*	604	…
Machine learning	21	*	…	…		*	1805.7
Consensus	24	858.6	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24	*	…	…	1721.9	…	…
OEM	24		…	…	*	728.7	…
Machine learning	24	*	…	…		*	1924.5

Source: World Bank.

Note: The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate average RMSE estimates for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to an F-test (RMSE).

Table 7.A. Model comparison: Bias of lead price forecasts

(U.S. dollars per metric tonne)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	-70.6	…	…	…	…	…
Futures	3		-39.3	…	…	…	…
Bivariate correlations	3			-3.1	…	…	…
BVAR	3				-48.7	…	…
OEM	3			*		-77.1	…
Machine learning	3		*	*			-115.2
Consensus	6	-74.6	…	…	…	…	…
Futures	6		-36.3	…	…	…	…
Bivariate correlations	6			138	…	…	…
BVAR	6				-68.1	…	…
OEM	6					-84.9	…
Machine learning	6		*				-124.3
Consensus	9	-76.3	…	…	…	…	…
Futures	9		-34	…	…	…	…
Bivariate correlations	9	*	*	211.2	…	…	…
BVAR	9				-81.6	…	…
OEM	9					-89.1	…
Machine learning	9		*				-125.1
Consensus	12	-76.1	…	…	…	…	…
Futures	12		-32.3	…	…	…	…
Bivariate correlations	12	*	*	221.9	…	…	…
BVAR	12			*	-49.9	…	…
OEM	12			*		-81.3	…
Machine learning	12		*				-119.4
Consensus	15	-77.7	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15		…	…	-67.8		…
OEM	15		…	…		-87.7	…
Machine learning	15		…	…			-100.2
Consensus	18	-81.0	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18		…	…	-35	…	…
OEM	18		…	…		-75.5	…
Machine learning	18		…	…			-78
Consensus	21	-79.2	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21		…	…	68.6	…	…
OEM	21		…	…		-67.8	…
Machine learning	21		…	…			-87.5
Consensus	24	-78.4	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24		…	…	103	…	…
OEM	24		…	…		-21.4	…
Machine learning	24		…	…			-87.2

Source: World Bank.

Note: Bias is defined as the difference between the actual price and the predicted price. The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate estimates of average bias for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to a t-test (bias).

Table 7.B. Model comparison: Root mean squared error of lead price forecasts

(U.S. dollars per metric tonne)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	153.8	…	…	…	…	…
Futures	3	*	103.7	…	…	…	…
Bivariate correlations	3			105.7	…	…	…
BVAR	3	*	*	*	252.9	…	…
OEM	3				*	94.5	…
Machine learning	3				*		133.8
Consensus	6	156.9	…	…	…	…	…
Futures	6	*	99.8	…	…	…	…
Bivariate correlations	6	*	*	303.5	…	…	…
BVAR	6	*	*		266.5	…	…
OEM	6			*	*	104.7	…
Machine learning	6			*	*		125.8
Consensus	9	164.5	…	…	…	…	…
Futures	9	*	97.1	…	…	…	…
Bivariate correlations	9	*	*	300.2	…	…	…
BVAR	9	*	*		336.8	…	…
OEM	9			*	*	131.2	…
Machine learning	9		*	*	*		149.9
Consensus	12	176	…	…	…	…	…
Futures	12	*	94.7	…	…	…	…
Bivariate correlations	12		*	241	…	…	…
BVAR	12	*	*	*	358.1	…	…
OEM	12		*		*	166.5	…
Machine learning	12		*		*		208.1
Consensus	15	187.8	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15	*	…	…	360.7		…
OEM	15		…	…	*	204.4	…
Machine learning	15	*	…	…			278.3
Consensus	18	194	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18	*	…	…	338.8	…	…
OEM	18		…	…		224.4	…
Machine learning	18	*	…	…			331.3
Consensus	21	195.1	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21	*	…	…	352.7	…	…
OEM	21		…	…		236	…
Machine learning	21	*	…	…			364.3
Consensus	24	204	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24	*	…	…	370.8	…	…
OEM	24		…	…	*	200.2	…
Machine learning	24	*	…	…		*	379.2

Source: World Bank.
Note: The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate average RMSE estimates for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to an F-test (RMSE).

Table 8.A. Model comparison: Bias of nickel price forecasts

(U.S. dollars per metric tonne)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	-545.8	…	…	…	…	…
Futures	3		-221	…	…	…	…
Bivariate correlations	3			-125.2	…	…	…
BVAR	3				-309.1	…	…
OEM	3					-222	…
Machine learning	3		*	*		*	-1279.4
Consensus	6	-343.7	…	…	…	…	…
Futures	6		-193.4	…	…	…	…
Bivariate correlations	6			1122.9	…	…	…
BVAR	6				-266.7	…	…
OEM	6					-126.6	…
Machine learning	6						-1089.7
Consensus	9	-167.4	…	…	…	…	…
Futures	9		-161.6	…	…	…	…
Bivariate correlations	9		*	1056.8	…	…	…
BVAR	9				-333	…	…
OEM	9			*		-127.5	…
Machine learning	9						-1143.2
Consensus	12	21.6	…	…	…	…	…
Futures	12		-131.9	…	…	…	…
Bivariate correlations	12	*	*	1950	…	…	…
BVAR	12			*	-154.6	…	…
OEM	12			*		22.6	…
Machine learning	12		*				-1276.2
Consensus	15	180.7	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15		…	…	-393		…
OEM	15		…	…		252.5	…
Machine learning	15		…	…			-1371.1
Consensus	18	295.9	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18		…	…	-371	…	…
OEM	18		…	…		441.9	…
Machine learning	18		…	…			-1407.3
Consensus	21	425.4	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21		…	…	623.9	…	…
OEM	21		…	…		685.3	…
Machine learning	21		…	…			-1386.5
Consensus	24	597.6	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24		…	…	1136.4	…	…
OEM	24		…	…		477.3	…
Machine learning	24		…	…			-1314.8

Source: World Bank.
Note: Bias is defined as the difference between the actual price and the predicted price. The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate estimates of average bias for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to a t-test (bias).

Table 8.B. Model comparison: Root mean squared error of nickel price forecasts

(U.S. dollars per metric tonne)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	1442.5	…	…	…	…	…
Futures	3	*	877.7	…	…	…	…
Bivariate correlations	3			989.6	…	…	…
BVAR	3	*	*	*	2763	…	…
OEM	3				*	958.8	…
Machine learning	3	*	*	*		*	2327.3
Consensus	6	1699.8	…	…	…	…	…
Futures	6	*	828.2	…	…	…	…
Bivariate correlations	6	*	*	3134.4	…	…	…
BVAR	6	*	*		2933.8	…	…
OEM	6	*		*	*	741.8	…
Machine learning	6		*			*	2380.1
Consensus	9	2044.1	…	…	…	…	…
Futures	9	*	793.7	…	…	…	…
Bivariate correlations	9		*	2064	…	…	…
BVAR	9	*	*	*	3307.3	…	…
OEM	9	*		*	*	762.6	…
Machine learning	9		*			*	2634.6
Consensus	12	2336.7	…	…	…	…	…
Futures	12	*	769.3	…	…	…	…
Bivariate correlations	12		*	2519.9	…	…	…
BVAR	12	*	*		3662.1	…	…
OEM	12	*	*	*	*	1220.5	…
Machine learning	12		*			*	2883.9
Consensus	15	2528.9	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15	*	…	…	4032.9		…
OEM	15		…	…	*	1600.7	…
Machine learning	15		…	…		*	3052.6
Consensus	18	2755.7	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18		…	…	4015.5	…	…
OEM	18		…	…	*	1783.7	…
Machine learning	18		…	…		*	3161.7
Consensus	21	2886.1	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21		…	…	3196.3	…	…
OEM	21		…	…		1999.8	…
Machine learning	21		…	…			3208.3
Consensus	24	3011	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24		…	…	3376.5	…	…
OEM	24		…	…		2233	…
Machine learning	24		…	…			3173.2

Source: World Bank.

Note: The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate average RMSE estimates for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to an F-test (RMSE).

Table 9.A. Model comparison: Bias of tin price forecasts

(U.S. dollars per metric tonne)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	-910.2	…	…	…	…	…
Futures	3		-478.9	…	…	…	…
Bivariate correlations	3			197.4	…	…	…
BVAR	3				-1091.5	…	…
OEM	3					21.4	…
Machine learning	3						-37.7
Consensus	6	-1168.2	…	…	…	…	…
Futures	6		-675.4	…	…	…	…
Bivariate correlations	6			691	…	…	…
BVAR	6				-1636.4	…	…
OEM	6				*	-244.5	…
Machine learning	6				*		150.8
Consensus	9	-1357.1	…	…	…	…	…
Futures	9		-814.1	…	…	…	…
Bivariate correlations	9			1541.5	…	…	…
BVAR	9				-2613.5	…	…
OEM	9				*	-497.3	…
Machine learning	9				*		290
Consensus	12	-1402.1	…	…	…	…	…
Futures	12		-939.4	…	…	…	…
Bivariate correlations	12		*	2621.6	…	…	…
BVAR	12		*		-3148.8	…	…
OEM	12			*	*	-661.4	…
Machine learning	12			*	*		435.5
Consensus	15	-1422.2	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15		…	…	-4091.6		…
OEM	15		…	…	*	-831.2	…
Machine learning	15		…	…	*		682.9
Consensus	18	-1430.2	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18		…	…	-4286.7	…	…
OEM	18		…	…	*	-838.7	…
Machine learning	18		…	…			890.3
Consensus	21	-1380.7	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21		…	…	-4027.8	…	…
OEM	21	*	…	…	*	-17103	…
Machine learning	21		…	…		*	1102.4
Consensus	24	-1379.6	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24		…	…	-4202	…	…
OEM	24		…	…		-901.2	…
Machine learning	24		…	…			1346.6

Source: World Bank.
Note: Bias is defined as the difference between the actual price and the predicted price. The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate estimates of average bias for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to a t-test (bias) or F-test (RMSE).

Table 9.B. Model comparison: Root mean squared error of tin price forecasts

(U.S. dollars per metric tonne)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	2171	…	…	…	…	…
Futures	3	*	1425.1	…	…	…	…
Bivariate correlations	3			1575.7	…	…	…
BVAR	3		*	*	2409.7	…	…
OEM	3					1959.1	…
Machine learning	3						1805.3
Consensus	6	2537.4	…	…	…	…	…
Futures	6	*	1280.3	…	…	…	…
Bivariate correlations	6	*		1722.6	…	…	…
BVAR	6		*	*	3241.5	…	…
OEM	6	*			*	1166.9	…
Machine learning	6		*		*		1969.2
Consensus	9	3021.2	…	…	…	…	…
Futures	9	*	1217.4	…	…	…	…
Bivariate correlations	9		*	2376.8	…	…	…
BVAR	9	*	*	*	4907.3	…	…
OEM	9	*		*	*	1395.1	…
Machine learning	9		*		*	*	2887
Consensus	12	3456.8	…	…	…	…	…
Futures	12	*	1180.5	…	…	…	…
Bivariate correlations	12		*	3080.5	…	…	…
BVAR	12	*	*	*	5481.7	…	…
OEM	12	*	*		*	1989.4	…
Machine learning	12		*			*	4114.3
Consensus	15	3742.7	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15	*	…	…	6794.4		…
OEM	15		…	…	*	2672.3	…
Machine learning	15		…	…		*	5243.3
Consensus	18	3955.5	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18	*	…	…	7065.7	…	…
OEM	18		…	…	*	3277	…
Machine learning	18	*	…	…		*	6103.1
Consensus	21	4077.3	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21	*	…	…	6695.9	…	…
OEM	21	*	…	…	*	2368.6	…
Machine learning	21	*	…	…		*	6639.7
Consensus	24	4188.2	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24	*	…	…	6632.2	…	…
OEM	24		…	…		4022.5	…
Machine learning	24	*	…	…		*	6879.6

Source: World Bank.

Note: The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate average RMSE estimates for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to an F-test (RMSE).

Table 10.A. Model comparison: Bias of zinc price forecasts

(U.S. dollars per metric tonne)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	-98.1	…	…	…	…	…
Futures	3		-44.7	…	…	…	…
Bivariate correlations	3			-19.7	…	…	…
BVAR	3				-102.3	…	…
OEM	3					-54.6	…
Machine learning	3						-55.9
Consensus	6	-103.2	…	…	…	…	…
Futures	6		-52	…	…	…	…
Bivariate correlations	6			129.5	…	…	…
BVAR	6				-130.1	…	…
OEM	6					-64.1	…
Machine learning	6						-34.2
Consensus	9	-124.6	…	…	…	…	…
Futures	9		-61.7	…	…	…	…
Bivariate correlations	9		*	227.3	…	…	…
BVAR	9				-126.8	…	…
OEM	9					-73.2	…
Machine learning	9			*			-16.3
Consensus	12	-130.6	…	…	…	…	…
Futures	12		-75.4	…	…	…	…
Bivariate correlations	12	*	*	331.2	…	…	…
BVAR	12			*	-105.6	…	…
OEM	12			*		-83.8	…
Machine learning	12			*			1.6
Consensus	15	-134.5	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15		…	…	-176.8		…
OEM	15		…	…		-86.2	…
Machine learning	15		…	…			27.2
Consensus	18	-140.6	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18		…	…	-161.3	…	…
OEM	18		…	…		-75.2	…
Machine learning	18		…	…			57
Consensus	21	-146.9	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21		…	…	-105.1	…	…
OEM	21		…	…		-68.9	…
Machine learning	21		…	…			68.3
Consensus	24	-157.2	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24		…	…	-109.5	…	…
OEM	24		…	…		12.1	…
Machine learning	24		…	…			70.2

Source: World Bank.

Note: Bias is defined as the difference between the actual price and the predicted price. The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate estimates of average bias for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to a t-test (bias) or F-test (RMSE).

Table 10.B. Model comparison: Root mean squared error of zinc price forecasts

(U.S. dollars per metric tonne)

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	213.6	…	…	…	…	…
Futures	3		150.5	…	…	…	…
Bivariate correlations	3	*		146.1	…	…	…
BVAR	3	*	*	*	401.9	…	…
OEM	3				*	205.9	…
Machine learning	3				*		202.1
Consensus	6	212.7	…	…	…	…	…
Futures	6		146	…	…	…	…
Bivariate correlations	6	*	*	312.8	…	…	…
BVAR	6	*	*		389.4	…	…
OEM	6				*	190.3	…
Machine learning	6			*	*		199.6
Consensus	9	225.9	…	…	…	…	…
Futures	9	*	142	…	…	…	…
Bivariate correlations	9		*	327	…	…	…
BVAR	9	*	*		394.8	…	…
OEM	9				*	210.4	…
Machine learning	9		*		*		243.1
Consensus	12	254.7	…	…	…	…	…
Futures	12	*	141.9	…	…	…	…
Bivariate correlations	12		*	342.7	…	…	…
BVAR	12	*	*		418.5	…	…
OEM	12		*			259.7	…
Machine learning	12		*				341.3
Consensus	15	281.7	…	…	…	…	…
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15	*	…	…	537.9		…
OEM	15		…	…	*	281.6	…
Machine learning	15	*	…	…			455
Consensus	18	308.9	…	…	…	…	…
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18	*	…	…	553.4	…	…
OEM	18		…	…	*	297.7	…
Machine learning	18	*	…	…		*	558.6
Consensus	21	330.1	…	…	…	…	…
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21	*	…	…	492.3	…	…
OEM	21		…	…		303.2	…
Machine learning	21	*	…	…		*	633.9
Consensus	24	356.8	…	…	…	…	…
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24		…	…	425.9	…	…
OEM	24		…	…		290	…
Machine learning	24	*	…	…	*	*	686

Source: World Bank.
Note: The forecast period is 2015Q1-2022Q1. Forecasts for OEM are only available at the semi-annual frequency; forecasts for all other approaches are available at the quarterly frequency. Diagonal entries indicate average RMSE estimates for the respective approach. * indicates significant difference between the forecasts of any pair of approaches, at the 5 percent significance level according to an F-test (RMSE).

Table 11.A. Model comparison: Diebold and Mariano (1995) test for oil prices

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	…			*		
Futures	3	*	…		*	*	
Bivariate correlations	3	†		…	*		
BVAR	3				…		
OEM	3					…	
Machine learning	3	†			*	†	…
Consensus	6	…			*		
Futures	6	†	…		*	†	
Bivariate correlations	6			…	**		
BVAR	6				…		
OEM	6				*	…	
Machine learning	6				*		…
Consensus	9	…			†		†
Futures	9	*	…		*		*
Bivariate correlations	9			…	*		
BVAR	9				…		
OEM	9				*	…	
Machine learning	9				†		…
Consensus	12	…			†		**
Futures	12	*	…		†		**
Bivariate correlations	12			…	*		*
BVAR	12				…		
OEM	12			†	*	…	*
Machine learning	12				†		…
Consensus	15	…	…	…	†		***
Futures	15	…	…	…	…	…	…
Bivariate correlations	15		…	…			
BVAR	15		…	…	…		
OEM	15	†	…	…	**	…	**
Machine learning	15		…	…			…
Consensus	18	…	…	…	*		***
Futures	18	…	…	…	…	…	…
Bivariate correlations	18		…	…			
BVAR	18		…	…	…		
OEM	18	*	…	…	***	…	***
Machine learning	18		…	…	†		…
Consensus	21	…	…	…	*		***
Futures	21	…	…	…	…	…	…
Bivariate correlations	21		…	…			
BVAR	21		…	…	…		
OEM	21	*	…	…		…	***
Machine learning	21		…	…	†		…
Consensus	24	…	…	…	**		***
Futures	24	…	…	…	…	…	…
Bivariate correlations	24		…	…			
BVAR	24		…	…	…		
OEM	24	*	…	…		…	***
Machine learning	24		…	…	*		…

Source: World Bank.
Note: Diebold Mariano statistic tests whether the forecasts in the model indicated in the row is more accurate than the forecast of the model indicated in the column. *** indicates statistically significantly more accurate model in the row at the 0.1 percent significance level, ** at the 1 percent level, * at the 5 percent level, and † at the 10 percent level.

Table 11.B. Model comparison: Diebold and Mariano (1995) test for aluminum prices

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	…	***		**		
Futures	3		…				
Bivariate correlations	3	**	***	…	**	†	†
BVAR	3		***		…		
OEM	3		***		†	…	
Machine learning	3	*	***		**	†	…
Consensus	6	…	***				
Futures	6		…				
Bivariate correlations	6	**	***	…	†	*	*
BVAR	6		†		…		
OEM	6	†	***			…	
Machine learning	6	*	***		†	*	…
Consensus	9	…	***		†		
Futures	9		…				
Bivariate correlations	9			…			
BVAR	9				…		
OEM	9	†	***		†	…	
Machine learning	9	*	***		†		…
Consensus	12	…	***		†		
Futures	12		…				
Bivariate correlations	12			…			
BVAR	12		†		…		
OEM	12	*		*	†	…	
Machine learning	12	*	***		*	***	…
Consensus	15	…	…	…	†		
Futures	15	…	…	…	…	…	…
Bivariate correlations	15	…	…	…	…	…	…
BVAR	15		…	…	…		
OEM	15	**	…	…	†	…	
Machine learning	15	**	…	…	†	†	…
Consensus	18	…	…	…	†		
Futures	18	…	…	…	…	…	…
Bivariate correlations	18	…	…	…	…	…	…
BVAR	18		…	…	…		
OEM	18	**	…	…	*	…	
Machine learning	18		…	…	*		…
Consensus	21	…	…	…	*		
Futures	21	…	…	…	…	…	…
Bivariate correlations	21	…	…	…	…	…	…
BVAR	21		…	…	…		
OEM	21	**	…	…	**	…	
Machine learning	21		…	…	*		…
Consensus	24	…	…	…	*		
Futures	24	…	…	…	…	…	…
Bivariate correlations	24	…	…	…	…	…	…
BVAR	24		…	…	…		
OEM	24		…	…		…	*
Machine learning	24		…	…	**		…

Source: World Bank.

Note: Diebold Mariano statistic tests whether the forecasts in the model indicated in the row is more accurate than the forecast of the model indicated in the column. *** indicates statistically significantly more accurate model in the row at the 0.1 percent significance level, ** at the 1 percent level, * at the 5 percent level, and † at the 10 percent level.

Table 11.C. Model comparison: Diebold and Mariano (1995) test for copper prices

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	…			***		
Futures	3	***	…		***	*	*
Bivariate correlations	3	***		…	***	*	†
BVAR	3				…		
OEM	3	†			**	…	
Machine learning	3	**			***	*	…
Consensus	6	…			*		
Futures	6	***	…		**	†	*
Bivariate correlations	6			…			
BVAR	6				…		
OEM	6	*			*	…	
Machine learning	6				*		…
Consensus	9	…			†		
Futures	9	**	…		*	*	**
Bivariate correlations	9			…			
BVAR	9				…		
OEM	9	***		†	†	…	**
Machine learning	9				†		…
Consensus	12	…		†	*		*
Futures	12	**	…	†	*	**	**
Bivariate correlations	12			…			
BVAR	12				…		
OEM	12	**		*	*	…	**
Machine learning	12				†		…
Consensus	15	…	…	…	†		**
Futures	15	…	…	…	…	…	…
Bivariate correlations	15		…	…			
BVAR	15		…	…	…		
OEM	15	***	…	…	*	…	**
Machine learning	15		…	…			…
Consensus	18	…	…	…	*		***
Futures	18	…	…	…	…	…	…
Bivariate correlations	18		…	…			
BVAR	18		…	…	…		
OEM	18	**	…	…	*	…	***
Machine learning	18		…	…			…
Consensus	21	…	…	…	*		***
Futures	21	…	…	…	…	…	…
Bivariate correlations	21		…	…			
BVAR	21		…	…	…		
OEM	21	***	…	…	***	…	***
Machine learning	21		…	…			…
Consensus	24	…	…	…	*		***
Futures	24	…	…	…	…	…	…
Bivariate correlations	24		…	…			
BVAR	24		…	…	…		
OEM	24		…	…		…	
Machine learning	24		…	…			…

Source: World Bank.

Note: Diebold Mariano statistic tests whether the forecasts in the model indicated in the row is more accurate than the forecast of the model indicated in the column. *** indicates statistically significantly more accurate model in the row at the 0.1 percent significance level, ** at the 1 percent level, * at the 5 percent level, and † at the 10 percent level.

Table 11.D. Model comparison: Diebold and Mariano (1995) test for lead prices

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	…			**		
Futures	3	*	…		**		***
Bivariate correlations	3	*		…	**		***
BVAR	3				…		
OEM	3	†			*	…	**
Machine learning	3				*		…
Consensus	6	…			**		
Futures	6	†	…		**		**
Bivariate correlations	6			…			
BVAR	6				…		
OEM	6				**	…	†
Machine learning	6				*		…
Consensus	9	…			**		
Futures	9	†	…		***		†
Bivariate correlations	9			…			
BVAR	9				…		
OEM	9	*			**	…	*
Machine learning	9				**		…
Consensus	12	…			**		
Futures	12	*	…	†	***	†	†
Bivariate correlations	12			…			
BVAR	12				…		
OEM	12				*	…	*
Machine learning	12				†		…
Consensus	15	…	…	…	**		*
Futures	15	…	…	…	…	…	…
Bivariate correlations	15		…	…			
BVAR	15		…	…	…		
OEM	15		…	…	*	…	
Machine learning	15		…	…			…
Consensus	18	…	…	…	***		***
Futures	18	…	…	…	…	…	…
Bivariate correlations	18		…	…			
BVAR	18		…	…	…		
OEM	18		…	…	***	…	
Machine learning	18		…	…			…
Consensus	21	…	…	…	***		***
Futures	21	…	…	…	…	…	…
Bivariate correlations	21		…	…			
BVAR	21		…	…	…		
OEM	21		…	…	***	…	
Machine learning	21		…	…			…
Consensus	24	…	…	…	***		***
Futures	24	…	…	…	…	…	…
Bivariate correlations	24		…	…			
BVAR	24		…	…	…		
OEM	24		…	…		…	
Machine learning	24		…	…			…

Source: World Bank.

Note: Diebold Mariano statistic tests whether the forecasts in the model indicated in the row is more accurate than the forecast of the model indicated in the column. *** indicates statistically significantly more accurate model in the row at the 0.1 percent significance level, ** at the 1 percent level, * at the 5 percent level, and † at the 10 percent level.

Table 11.E. Model comparison: Diebold and Mariano (1995) test for nickel prices

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	…			**		**
Futures	3	*	…		***		***
Bivariate correlations	3	*		…	***		***
BVAR	3				…		
OEM	3	*		†	*	…	**
Machine learning	3						…
Consensus	6	…			**		*
Futures	6	**	…		**		***
Bivariate correlations	6			…			
BVAR	6				…		
OEM	6	*			*	…	**
Machine learning	6						…
Consensus	9	…			**		*
Futures	9	**	…	†	**		***
Bivariate correlations	9			…	†		
BVAR	9				…		
OEM	9	**		†	*	…	**
Machine learning	9						…
Consensus	12	…			**		**
Futures	12	**	…	†	**	*	***
Bivariate correlations	12			…			
BVAR	12				…		
OEM	12	***		†	***	…	**
Machine learning	12						…
Consensus	15	…	…	…	**		**
Futures	15	…	…	…	…	…	…
Bivariate correlations	15		…	…			
BVAR	15		…	…	…		
OEM	15	**	…	…	**	…	***
Machine learning	15		…	…	†		…
Consensus	18	…	…	…	***		**
Futures	18	…	…	…	…	…	…
Bivariate correlations	18		…	…			
BVAR	18		…	…	…		
OEM	18	**	…	…	**	…	
Machine learning	18		…	…	†		…
Consensus	21	…	…	…	***		*
Futures	21	…	…	…	…	…	…
Bivariate correlations	21		…	…			
BVAR	21		…	…	…		
OEM	21	**	…	…	***	…	
Machine learning	21		…	…			…
Consensus	24	…	…	…	***		*
Futures	24	…	…	…	…	…	…
Bivariate correlations	24		…	…			
BVAR	24		…	…	…		
OEM	24		…	…		…	
Machine learning	24		…	…	**		…

Source: World Bank.
Note: Diebold Mariano statistic tests whether the forecasts in the model indicated in the row is more accurate than the forecast of the model indicated in the column. *** indicates statistically significantly more accurate model in the row at the 0.1 percent significance level, ** at the 1 percent level, * at the 5 percent level, and † at the 10 percent level.

Table 11.F. Model comparison: Diebold and Mariano (1995) test for tin prices

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	…			†		
Futures	3	*	…		***		
Bivariate correlations	3	*		…	***		
BVAR	3				…		
OEM	3				†	…	
Machine learning	3	*			*		…
Consensus	6	…					
Futures	6	*	…		†		
Bivariate correlations	6	*		…	*		
BVAR	6				…		
OEM	6	†				…	
Machine learning	6				†		…
Consensus	9	…					
Futures	9	†	…	†	†		†
Bivariate correlations	9			…			
BVAR	9				…		
OEM	9	†				…	†
Machine learning	9						…
Consensus	12	…					
Futures	12	†	…	†		†	*
Bivariate correlations	12			…			
BVAR	12				…		
OEM	12	†		†	†	…	*
Machine learning	12						…
Consensus	15	…	…	…	†		
Futures	15	…	…	…	…	…	…
Bivariate correlations	15		…	…			
BVAR	15		…	…	…		
OEM	15	*	…	…	†	…	**
Machine learning	15		…	…			…
Consensus	18	…	…	…	†		*
Futures	18	…	…	…	…	…	…
Bivariate correlations	18		…	…			
BVAR	18		…	…	…		
OEM	18	*	…	…	*	…	***
Machine learning	18		…	…			…
Consensus	21	…	…	…	*		***
Futures	21	…	…	…	…	…	…
Bivariate correlations	21		…	…			
BVAR	21		…	…	…		
OEM	21		…	…		…	
Machine learning	21		…	…		***	…
Consensus	24	…	…	…	*		***
Futures	24	…	…	…	…	…	…
Bivariate correlations	24		…	…			
BVAR	24		…	…	…		
OEM	24	***	…	…		…	
Machine learning	24		…	…			…

Source: World Bank.

Note: Diebold Mariano statistic tests whether the forecasts in the model indicated in the row is more accurate than the forecast of the model indicated in the column. *** indicates statistically significantly more accurate model in the row at the 0.1 percent significance level, ** at the 1 percent level, * at the 5 percent level, and † at the 10 percent level.

Table 11.G. Model comparison: Diebold and Mariano (1995) test for zinc prices

Approach	Horizon	Consensus	Futures	Bivariate correlations	BVAR	OEM	Machine learning
Consensus	3	…			***		
Futures	3	*	…		***	†	*
Bivariate correlations	3	*		…	***	†	*
BVAR	3				…		
OEM	3				**	…	
Machine learning	3				***		…
Consensus	6	…			**		
Futures	6	*	…		***		*
Bivariate correlations	6			…			
BVAR	6				…		
OEM	6				**	…	
Machine learning	6				***		…
Consensus	9	…			***		
Futures	9	**	…		***		**
Bivariate correlations	9			…			
BVAR	9				…		
OEM	9	**			**	…	
Machine learning	9				***		…
Consensus	12	…			***		†
Futures	12	***	…	†	***	*	***
Bivariate correlations	12			…			
BVAR	12				…		
OEM	12					…	
Machine learning	12				**		…
Consensus	15	…	…	…	**		**
Futures	15	…	…	…	…	…	…
Bivariate correlations	15		…	…			
BVAR	15		…	…	…		
OEM	15		…	…	*	…	**
Machine learning	15		…	…			…
Consensus	18	…	…	…	**		***
Futures	18	…	…	…	…	…	…
Bivariate correlations	18		…	…			
BVAR	18		…	…	…		
OEM	18		…	…	**	…	***
Machine learning	18		…	…			…
Consensus	21	…	…	…	***		***
Futures	21	…	…	…	…	…	…
Bivariate correlations	21		…	…			
BVAR	21		…	…	…		
OEM	21	***	…	…	***	…	
Machine learning	21		…	…			…
Consensus	24	…	…	…	*		***
Futures	24	…	…	…	…	…	…
Bivariate correlations	24		…	…			
BVAR	24		…	…	…		
OEM	24		…	…		…	
Machine learning	24		…	…			…

Source: World Bank.

Note: Diebold Mariano statistic tests whether the forecasts in the model indicated in the row is more accurate than the forecast of the model indicated in the column. *** indicates statistically significantly more accurate model in the row at the 0.1 percent significance level, ** at the 1 percent level, * at the 5 percent level, and † at the 10 percent level.

Table 12. Approaches with lowest bias and RMSEs

	Quarters				
Commodity	1	2	3	4	5+
Aluminum	Bivariate correlations	Bivariate correlation s	OEM and machine learning techniques	OEM and machine learning techniques	OEM, consensus, machine learning techniques
Copper	Bivariate correlations	Futures	Futures	Futures	OEM and consensus
Lead	Bivariate correlations	OEM and futures	OEM and futures	Futures	OEM and consensus
Nickel	Futures and bivariate correlations	OEM and futures	OEM and futures	Futures	OEM and consensus
Oil	Futures	Futures	Futures	Futures	OEM and consensus
Tin	Bivariate correlations or futures	Any except BVAR	Any except BVAR	Any except BVAR	OEM
Zinc	Any except BVAR	Futures	Futures	Futures	OEM and consensus

Note: BVAR (Bayesian Vector Autoregression); OEM (Oxford Economic Model).

Table 13. Features of approaches

Forecasting approach	RMSE and bias	Appropriate for scenario analysis	Data requirements
Futures	For most commodities (except aluminum), lowest RMSEs or bias for forecasts up to four quarters but reliable data unavailable for longer horizons	No	Low
Consensus forecasts	For most commodities (except tin and zinc), lowest RMSEs or bias for forecasts of more than a year but poorer short-term performance	No	Low
Bivariate correlations	For metal commodities, lowest RMSE or bias but only at the very shortest horizon	No	Medium
BVAR	For all commodities at all horizons, higher bias and RMSE than other approaches	Yes	Medium
OEM	For all commodities, lowest RMSEs or bias for forecast horizons above one year	Yes	High
Machine learning techniques	For all commodities (except for nickel), intermediate bias and RMSEs.	No	Medium

Note: BVAR (Bayesian Vector Autoregression); OEM (Oxford Economic Model); RMSE (Root Mean Squared Error).

References

Abosedra, S. 2006. “Futures versus Univariate Forecast of Crude Oil Prices.” *OPEC Review* 29 (4): 231–241.

Abosedra, S., and H. Baghestani. 2004. “On the Predictive Accuracy of Crude Oil Futures Prices.” *Energy Policy* 32 (12): 1389–1393.

Ahmad, W., M. Aamir, U. Khalil, M. Ishaq, N. Iqbal, and M. Khan. 2021. “A New Approach for Forecasting Crude Oil Prices Using Median Ensemble Empirical Mode Decomposition and Group Method of Data Handling.” *Mathematical Problems in Engineering* 2021: 1389–1393.

Alipour, A., A. Khodaiari, and A. Jafari. 2019. “Modeling and Prediction of Time-Series of Monthly Copper Prices.” *Int. J. Min. Geo-Eng* 53 (1): 91–97.

Alquist, R., and E. Arbatli. 2010. “Crude Oil Futures: A Crystal Ball?” *Bank of Canada Review* 2010 (Spring): 3-11.

Alquist, R., and L. Kilian. 2010. “What Do We Learn from the Price of Crude Oil Futures?” *Journal of Applied Econometrics* 25 (4): 539-573.

Alquist, R., L. Kilian, and R. J. Vigfusson. 2013. “Forecasting the Price of Oil.” In *Handbook of Economic Forecasting*, edited by G. Elliott, C. W. J. Granger, and A. Timmermann, 2 (A): 427-507. Amsterdam: Elsevier.

Astudillo, G., R. Carrasco, C. Fernández-Campusano, and M. Chacón. 2020. “Copper Price Prediction Using Support Vector Regression Technique.” *Applied Sciences* 10 (19): 1-12.

Baffes, J., and P. Nagle. 2022. *Commodity Markets: Evolution, Challenges and Policies*. Washington, DC: World Bank.

Baumeister, C. 2022. “Measuring Market Expectations.” In *Handbook of Economic Expectations*, edited by Bachmann, R., G. Topa, and W. van der Klaauw, November (2022): 413-442. Academic Press: Elsevier.

Baumeister, C., D. Korobilis and T. Lee. 2022. “Energy Markets and Global Economic Conditions.” *Review of Economics and Statistics* 104 (4): 828-844.

Baumeister, C., and L. Kilian. 2012. “Real-Time Forecasts of the Real Price of Oil.” *Journal of Business and Economic Statistics* 30 (2): 326–36.

Baumeister, C., and L. Kilian. 2014. “What Central Bankers Need to Know about Forecasting Oil Prices.” *International Economic Review* 55 (3): 869-889.

Baumeister, C., and L. Kilian. 2015. “Forecasting the Real Price of Oil in A Changing World: A Forecast Combination Approach.” *Journal of Business and Economic Statistics* 33 (3): 338-351.

Behmiri, N. B., and J. R. P. Manso. 2013. “Crude Oil Price Forecasting Techniques: A Comprehensive Review of Literature.” *Alternative Investment Analyst Review* 2013 (2): 29-48.

Berk, C. 2016. “Indexing Oil from a Financial Point of View: A Comparison between Brent and West Texas Intermediate.” *International Journal of Energy Economics and Policy* 6 (2): 152-158.

Bopp, A. E., and G. M. Lady. 1991. “A comparison of petroleum futures versus spot prices as predictors of prices in the future.” *Energy Economics* 13, 274–282.

Bowman, C., and A. Husain. 2004. “Forecasting Commodity Prices: Futures Versus Judgment.” Working Paper No. 04/41. International Monetary Fund, Washington, DC.

Buncic, D., and C. Moretto. 2015. “Forecasting Copper Prices with Dynamic Averaging and Selection Models.” *The North American Journal of Economics and Finance* 33 (2015): 1-38.

Castro, J., A. Araujo, and A. A. Montini. 2013. “Comparative Modeling Assessment of Aluminum Price Forecast.” Conference proceedings, XVI Seminários em Administração, São Paulo, October 2013.

Chen, S-S. 2014. “Forecasting Crude Oil Price Movements with Oil-Sensitive Stocks.” *Economic Inquiry* 52 (2): 830-844.

Chen, Y., K. He, and C. Zhang. 2016. “A Novel Grey Wave Forecasting Method for Predicting Metal Prices.” *Resources Policy* 49 (C): 323-331.

Cheng, F., T. Li, Y-M. Wei, and T. Fan. 2019. “The VEC-NAR Model for Short-Term Forecasting of Oil Prices.” *Energy Economics* 78 (C): 656-667.

Cheong, C. W. 2009. “Modeling and Forecasting Crude Oil Markets using ARCH-Type Models.” *Energy Policy* 37 (6): 2346–2355.

Chernenko, S. V., K. B. Schwarz, and J. H. Wright. 2004. “The Information Content of Forward and Futures Prices: Market Expectations and the Price of Risk.” International Finance Discussion Papers 808, Board of Governors of the Federal Reserve System (U.S.), Washington, DC.

Chinn, M. D., and O. Coibion. 2014. “The Predictive Content of Commodity Futures.” *Journal of Futures Markets* 34 (7): 607-636.

Chinn, M., M. LeBlanc, and O. Coibion. 2005. “The Predictive Content of Energy Futures: An Update on Petroleum, Natural Gas, Heating Oil and Gasoline.” National Bureau of Economic Research Working Paper 11033, Cambridge, M. A.

Chiroma, H., S. Abdulkareem, and T. Herawan. 2015. “Evolutionary Neural Network Model for West Texas Intermediate Crude Oil Price Prediction.” *Applied Energy* 142 (2015): 266–273.

Chu, P. K., K. Hoff, P. Molnár, and M. Olsvik. 2022. “Crude Oil: Does the Futures Price Predict the Spot Price?” *Research in International Business and Finance* 60 (2022): 1-7.

Coimbra C., and P. Esteves. 2004. “Oil Price Assumptions in Macroeconomic Forecasts: Should we follow Futures Market Expectations?” *OPEC Energy Review* 28 (2): 87-106.

Coppola, A. 2008. “Forecasting Oil Price Movements: Exploiting the Information in the Futures Market.” *Journal of Futures Markets* 28 (1): 34–56.

Cortazar, G., H. Ortega, and C. Valencia. 2021. “How Good are Analyst Forecasts of Oil Prices?” *Energy Economics* 102 (C): 1-11.

Dées, S., P. V. Karadeloglou, R. K. Kaufmann, and M. Sánchez. 2007. “Modelling the World Oil Market: Assessment of a Quarterly Econometric Model.” *Energy Policy* 35 (2007): 178-191.

Degiannakis, S., and G. Filis. 2017. “Forecasting Oil Price Realized Volatility Using Information Channels from Other Asset Classes.” *Journal of International Money and Finance*, 76 (2017): 28-49.

Dehghani, H. 2018. “Forecasting Copper Price Using Gene Expression Programming.” *Journal of Mining and Environment* 9 (2): 349–360.

Diebold, F.X. 2015. “Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests.” *Journal of Business and Economic Statistics*, 33(1): 1-1.

Diebold, F.X., and R.S. Mariano. 1995. “Comparing Predictive Accuracy.” *Journal of Business and Economic Statistics* 13: 253–263.

Dooley, G., and H. Lenihan. 2005. “An Assessment of Time Series Methods in Metal Price Forecasting.” *Resources Policy* 30 (2005): 208–217.

Drachal, K. 2016. “Forecasting Spot Oil Price in a Dynamic Model Averaging Framework — Have the Determinants Changed Over Time?” *Energy Economics* 60 (2016): 35–46.

Du, P., J. Guo, S. Sun, S. Wang, and J. Wu. 2021. “Multi-Step Metal Prices Forecasting Based on a Data Preprocessing Method and an Optimized Extreme Learning Machine by Marine Predators Algorithm.” *Resources Policy* 74 (2021): 1-10.

Edmundo, G., L. Luiz, and A. Edmundo. 2010. “Forecasting Oil Price Trends Using Wavelets and Hidden Markov Models.” *Energy Economics* 32 (6): 1507–1519.

Ellwanger, R., and S. Snudden. 2023. “Futures Prices are Useful Predictors of the Spot Price of Crude Oil.” The Energy Journal 44 (4): 45-62.

Emmons, W. R., and T. J. Yeager. 2002. “The Futures Market Is an Imperfect Forecasting Tool.” Federal Reserve Bank, St. Louis. Accessed September 29, 2022.

https://www.stlouisfed.org/publications/regional-economist/january-2002/the-futures-market-as-forecasting-tool-an-imperfect-crystal-ball

Fernandez, V. 2007. "Forecasting Commodity Prices by Classification Methods." Working Paper, Department of Industrial Engineering, University of Chile, Santiago.

Frey, G., M. Manera, A. Markanday, and A. Scarpa. 2009. "Econometrics models for oil price forecasting: A critical survey." *Center for Economic Studies Forum* 10 (1): 29-44.

Fritz, A., and C. Weber. 2012. "Informational Efficiency in Futures Markets for Crude Oil." Working Paper, No. 03/11, Chair for Management Science and Energy Economics, University of Duisburg, Essen.

Gabralla, L. A., R. Jammazi, and A. Abraham. 2013. "Oil Price Prediction Using Ensemble Machine Learning." Conference proceedings, International Conference on Computing, Electrical and Electronic Engineering, Khartoum, August 2013.

Godarzi, A., R. Amiri, A. Talaei, and T. Jamasb. 2014. "Predicting Oil Price Movements: A Dynamic Artificial Neural Network Approach." *Energy Policy* 68 (2014): 371–382.

Gong, X., and B. Lin. 2018. "Structural Breaks and Volatility Forecasting in the Copper Futures Market." *Journal of Futures Markets* 38 (3): 290–339.

Guenette, J. D., and T. Yamazaki. 2021. "Projecting the Economic Consequences of the COVID-19 Pandemic." Policy Research Working Paper 9589. Washington, DC: World Bank.

Hamilton, J. D., and J. C. Wu. 2014. "Risk Premia in Crude Oil Futures Prices." *Journal of International Money and Finance* 42 (C): 9-37.

Haugom, E., H. Langeland, P. Molnár, and S. Westgaard. 2014. "Forecasting Volatility of The US Oil Market." *Journal of Banking and Finance* 47 (C): 1–14.

He, K., X. Lu, Y. Zou, and K. Lai. 2015. "Forecasting Metal Prices with a Curvelet Based Multiscale Methodology." *Resources Policy* 45 (2015): 144–150.

He, X. 2018. "Crude Oil Prices Forecasting: Time Series vs. SVR Models." *Journal of International Technology and Information Management* 27 (2): 25-42.

Inter-American Development Bank. 2022. *From Recovery to Renaissance: Turning Crisis into Opportunity.* Washington, DC: Inter-American Development Bank.

International Monetary Fund. 2014. *Commodity Prices and Forecasts*. World Economic Outlook. Washington, DC: International Monetary Fund.

Issler, J. V., C. Rodrigues, and R. Burjack. 2014. “Using Common Features to Understand the Behavior of Metal-Commodity Prices and Forecast them at Different Horizons.” *Journal of International Money and Finance* 42 (C): 310-335.

Jammazi, R., and C. Aloui. 2012. “Crude Oil Price Forecasting: Experimental Evidence from Wavelet Decomposition and Neural Network Modeling.” *Energy Economics* 34 (3): 828–841.

Jiang, Z.-Q., W.-J. Xie, and W.-X. Zhou. 2014. “Testing the Weak-Form Efficiency of the WTI Crude Oil Futures Market.” Physica A: Statistical Mechanics and its Applications 405 (C): 235-244.

Jin, X. 2017. “Do futures prices help forecast the spot price?” *Journal of Futures Markets* 37 (12): 1205-1225.

Kaboudan, M. A. 2001. “Compumetric Forecasting of Crude Oil Prices.” Conference proceedings, 2001 Congress on Evolutionary Computation, Seoul, 2001.

Kahraman, E., and O. Akay. 2022. “Comparison of exponential smoothing methods in forecasting global prices of main metals.” *Miner Econ (2022). https://doi.org/10.1007/s13563-022-00354-y*

Karlsson, S. 2013. “Forecasting with Bayesian Vector Autoregression.” In *Handbook of Economic Forecasting Volume 2*, Part B, 791-897, edited by G. Elliott and A. Timmermann. Amsterdam: Elsevier.

Kaufmann, R. K., S. Dees, P. Karadeloglou, and M. Sanchez. 2004. “Does OPEC matter? An Econometric Analysis of Oil Prices.” *The Energy Journal* 25 (4): 67-90.

Khoshalan, H. A., J. Shakeri, I. Najmoddini, and M. Asadizadeh. 2021. “Forecasting Copper Price by Application of Robust Artificial Intelligence Techniques.” *Resources Policy* 73 (C): 1-11.

Kilian, L., and D. P. Murphy. 2014. “The Role of Inventories and Speculative Trading in the Global Market for Crude Oil.” *Journal of Applied Econometrics* 29 (3): 454-478.

Kriechbaumer, T., A. Angus, D. Parsons, and M. C. Rivas. 2014. “An Improved Wavelet–ARIMA Approach for Forecasting Metal Prices.” *Resources Policy* 39 (C): 32-41.

Kwas, M., and M. Rubascezk. 2021. “Forecasting Commodity Prices: Looking for a Benchmark.” *Forecasting* 3 (2): 447-459.

Lalonde, R., Z. Zhu, and F. Demers. 2003. “Forecasting and Analyzing World Commodity Prices.” Bank of Canada Working Paper 03–24, Bank of Canada: Ottawa.

Lasheras, S., F. Juez, A. S. Sánchez, and A. Krzemień. 2015. “Forecasting the COMEX Copper Spot Price by Means of Neural Networks and ARIMA Models” *Resources Policy* 45 (C): 37–43.

Lei, Y., J. Lin, Z. He, and M. J. Zuo. 2013. "A Review on Empirical Mode Decomposition in Fault Diagnosis of Rotating Machinery." *Mechanical systems and signal processing* 35 (1-2): 108-126.

Lenza, M. and G. E. Primiceri. 2022. "How to estimate a VAR after March 2020." *Journal of Applied Econometrics* 37 (4): 688–699.

Li, T., Z. Hu, Y. Jia, J. Wu, and Y. Zhou. 2018. "Forecasting Crude Oil Prices Using Ensemble Empirical Mode Decomposition and Sparse Bayesian Learning." *Energies* 11 (7): 1-23.

Lin, H. and Q. Sun. 2020. "Crude Oil Prices Forecasting: An Approach of Using CEEMDAN-Based Multi-Layer Gated Recurrent Unit Networks." *Energies* 13 (7): 1-21.

Lu, Q., S. Sun, H. Duan, and S. Wang. 2021. "Analysis and Forecasting of Crude Oil Price Based on the Variable Selection-LSTM Integrated Model." *Energy Informatics* 4 (2): 1-20.

Marioli, F.A. and Letelier, F., 2021. Commodities Fundamental Model. Banco Central

Manescu, C., and I. Van Robays. 2017. "Forecasting the Brent Oil Price: Addressing Time-Variation in Forecast Performance." ECB Working Paper No. 1735, European Central Bank, Frankfurt.

Miao, H., S. Ramchander, T. Wang, and D. Yang. 2017. "Influential Factors in Crude Oil Price Forecasting." *Energy Economics* 68 (C): 77-88.

Mirmirani, S., and H. C. Li. 2004. "A Comparison of VAR and Neural Networks with Genetic Algorithm in Forecasting Price of Oil." *Advances in Econometrics* 19 (2004): 203–223.

Mohammadi, H., and L. Su. 2010. "International Evidence on Crude Oil Price Dynamics: Applications of ARIMA–GARCH Models." *Energy Economics* 32 (5): 1001–1008.

Moosa, I. A., and N. E. Al-Loughani. 1994. "Unbiasedness and Time Varying Risk Premia in the Crude Oil Futures Market." *Energy Economics* 16 (2): 99–105.

Moshiri, S., and F. Foroutan. 2006. "Forecasting Nonlinear Crude Oil Futures Prices." *The Energy Journal* 27 (4): 81–95.

Mostafa, M. M., and A. A. El-Masry. 2016. "Oil Price Forecasting Using Gene Expression Programming and Artificial Neural Networks." *Economic Modelling* 54 (C): 40–53.

Naser, H. 2016. "Estimating and Forecasting the Real Prices of Crude Oil: A Data Rich Model Using a Dynamic Model Averaging (DMA) Approach." *Energy Economics* 56 (2016): 75–87.

Ng, S. 2021. "Modeling Macroeconomic Variations After COVID-19." NBER Working Paper 29060, National Bureau of Economic Research, Cambridge, MA.

Nixon, D., and T. Smith. 2012. “What Can the Oil Futures Curve Tell Us About the Outlook for Oil Prices?” *Bank of England Quarterly Bulletin* 52(1): 39-47.

Oxford Economics. 2019. “The Oxford Global Economic Model.” July. Oxford Economics, Oxford, UK.

Pakko, M. 2005. “Future Oil.” *Economic Synopses* No. 17, Federal Reserve Bank of St. Louis.

Peng, Y., C. Yuan, X. Qin, J. Huang, and Y. Shi. 2014. “An Improved Gene Expression Programming Approach for Symbolic Regression Problems.” *Neurocomputing* 137 (2014): 293-301.

Pesaran, M.H. and A. Timmermann. 1992. “A Simple Nonparametric Test of Predictive Performance.” *Journal of Business and Economic Statistics* 10(4): 461-465.

Pincheira-Brown, P., and N. Hardy. 2019. “Forecasting Base Metal Prices with the Chilean Exchange Rate.” *Resources Policy* 62 (C): 256–281.

Pincheira-Brown, P., and N. Hardy. 2021. “Forecasting Aluminum Prices with Commodity Currencies.” *Resources Policy* 73 (2021): 1-17.

Pincheira-Brown, P., N. Hardy, A. Bentancor, and N. Jarsún. 2022. “Forecasting Fuel Prices with the Chilean Exchange Rate: Going Beyond the Commodity Currency Hypothesis.” *Energy Economics* 106 (2022): 1-17.

Pincheira-Brown, P., N. Hardy, C. Antil, I. Pérez, and A. Bentancor. 2021. “Forecasting Base Metal Prices with an International Stock Index.” MPRA Paper 107828, University Library of Munich, Germany.

Ramyar, S., and F. Kianfar. 2017. “Forecasting Crude Oil Prices: A Comparison Between Artificial Neural Networks and Vector Autoregressive Models.” *Computational Economics* 53 (2019): 743–761.

Reeve, T. A., and R. J. Vigfusson. 2011. “Evaluating the Forecasting Performance of Commodity Futures Prices.” FRB International Finance Discussion Paper No. 1025, Board of Governors of the Federal Reserve System, United States.

Reichsfeld, D. A., and S. K. Roache. 2011. “Do Commodity Futures Help Forecast Spot Prices?” IMF Working Papers 2011/254, International Monetary Fund, Washington, DC.

Ru, Y., and H. J. Ren. 2012. “Application of ARMA Model in Forecasting Aluminum Price.” *Applied Mechanics and Materials* 155-156 (2012): 66-71.

Rubaszek, M., Z. Karolak, and M. Kwas. 2020. “Mean-Reversion, Non-Linearities and the Dynamics of Industrial Metal Prices: A Forecasting Perspective.” *Resources Policy* 65 (C): 1-8.

Rubaszek, M., Z. Karolak, M. Kwas, and G. Uddin. 2020. "The role of the threshold effect for the dynamics of futures and spot prices of energy commodities." *Studies in Nonlinear Dynamics and Econometrics* 24 (5): 1-20.

Schorfheide, F., and D. Song. 2021. "Real-Time Forecasting with a (Standard) Mixed-Frequency VAR During a Pandemic." CEPR Press Discussion Paper No. 16760.

Singleton, K. 2014. "Investor Flows and the 2008 Boom/Bust in Oil Prices." *Management Science* 60 (2): 300-318.

Shambora, W., and R. Rossiter. 2007. "Are there exploitable inefficiencies in the futures market for oil?" *Energy Economics* 29 (1): 18-27.

Svensson, L. 2006. "Oil Prices and ECB Monetary Policy." Briefing Paper, Committee on Economic and Monetary Affairs of the European Parliament.

Tang, S. 2010. *Forecasting Metals Prices with Regime Switching GARCH Models*. Ph.D. Dissertation. The Chinese University of Hong Kong, Ma Liu Shui, Hong Kong.

Tang, L., and S. Hammoudeh. 2002. "An Empirical Exploration of the World Oil Price Under the Target Zone Model." *Energy Economics* 24 (6): 577-596.

Villegas, B. 2021. *Mineral Commodity Price Forecasting Through Time-Series Modeling Techniques and Artificial Neural Networks: A Nickel Case Study*. Ph.D. Dissertation. McGill University, Montreal, Quebec.

Vo, M. T. 2009. "Regime-Switching Stochastic Volatility: Evidence from the Crude Oil Market." *Energy Economics* 31 (5): 779–788.

Wang, C., X. Zhang, M. Wang, M. K. Lim, and P. Ghadimi. 2019. "Predictive Analytics of the Copper Spot Price by Utilizing Complex Network and Artificial Neural Network Techniques." *Resources Policy* 63 (C): 1-17.

Wang, Y., C. Wu, and L. Yang. 2016. "Forecasting Crude Oil Market Volatility: A Markov Switching Multifractal Volatility Approach." *International Journal of Forecasting* 32 (1): 1–9.

Wen, F., X. Gong, and S. Cai. 2016. "Forecasting the Volatility of Crude Oil Futures Using HAR-Type Models with Structural Breaks." *Energy Economics* 59 (C): 400–413.

World Bank. 2019. "The Role of Substitution in Commodity Demand." In *Commodity Markets Outlook*, October. Washington, DC: World Bank.

Wu, T., and A. McCallum. 2005. "Do Oil Futures Prices Help Predict Future Oil Prices?" *Economic Letter* No. 2005-38, Federal Reserve Bank of San Francisco, San Francisco.

Xie, W., L. Yu, S. Xu, and S. Wang. 2006. "A New Method for Crude Oil Price Forecasting Based on Support Vector Machines." Conference proceedings, Computational Science–ICCS 2006: 6th International Conference, Reading, United Kingdom, May 28–31, 2006.

Xiong, T., Y. Bao, and Z. Hu. 2013. "Beyond One-Step-Ahead Forecasting: Evaluation of Alternative Multi-Step-Ahead Forecasting Models for Crude Oil Prices." *Energy Economics* 40 (C): 405–415.

Ye, M., J. Zyren, and J. Shore. 2002. "Forecasting Crude Oil Spot Price Using OECD Petroleum Inventory Levels." *International Advances in Economic Research* 8 (2002): 324–333.

Ye, M., J. Zyren, and J. Shore. 2005. "A Monthly Crude Oil Spot Price Forecasting Model Using Relative Inventories." *International Journal of Forecasting* 21 (3): 491–501.

Ye, M., J. Zyren, and J. Shore. 2006. "Forecasting Short-Run Crude Oil Price Using High-And Low Inventory Variables." *Energy Policy* 34 (17): 2736–2743.

Yousefi, S., I. Weinreich, and D. Reinarz. 2005. "Wavelet-Based Prediction of Oil Prices." *Chaos, Solitons and Fractals* 25 (2): 265-275.

Yu, L., S. Wang, and K. K. Lai. 2008. "Forecasting Crude Oil Price with an EMD-Based Neural Network Ensemble Learning Paradigm." *Energy Economics* 30 (5): 2623–2635.

Yu, L., Y. Zhao, and L. Tang. 2014. "A Compressed Sensing-Based AI Learning Paradigm for Crude Oil Price Forecasting." *Energy Economics* 46 (C): 236–245.

Zamani, M. 2004. "An Econometrics Forecasting Model of Short-Term Oil Spot Price." Conference proceedings, 6th International Association of Energy Economics European Conference, Tehran, 2004.

Zeng, T., and N. Swanson. 1998. "Predictive Evaluation of Econometric Forecasting Models in Commodity Future Markets." *Studies in Nonlinear Dynamics and Econometrics* 2 (4): 159–177.

Zhao, Y., J. Li, and L. Yu. 2017. "A Deep Learning Ensemble Approach for Crude Oil Price Forecasting." *Energy Economics* 66 (C): 9-16.

Zhang, J., Y-J. Zhang, and L. Zhang. 2015. "A Novel Hybrid Method for Crude Oil Price Forecasting." *Energy Economics* 49 (C): 649-659.

Zheng, J., J. Cheng, and Y. Yang. 2014. "Partly Ensemble Empirical Mode Decomposition: An Improved Noise-Assisted Method for Eliminating Mode Mixing." *Signal Processing* 96 (B): 362-374.