computer sciences and mathematics forum

MDPI

Proceeding Paper

The Forecasting of Aluminum Prices: A True Challenge for Econometric Models †

Krzysztof Drachal * and Joanna Jędrzejewska

Faculty of Economic Sciences, University of Warsaw, 00-241 Warszawa, Poland; j.jedrzejews@student.uw.edu.pl
* Correspondence: kdrachal@wne.uw.edu.pl
† Presented at the 11th International Conference on Time Series and Forecasting, Canaria, Spain, 16–18 July 2025.

Abstract

This paper explores the forecasting of aluminum prices using various predictive models dealing with variable uncertainty. A diverse set of economic and market indicators is considered as potential price predictors. The performance of models including LASSO, RIDGE regression, time-varying parameter regressions, LARS, ARIMA, Dynamic Model Averaging, Bayesian Model Averaging, etc., is compared according to forecast accuracy. Despite the initial expectations that Bayesian dynamic mixture models would provide the best forecast accuracy, the results indicate that forecasting by futures prices and with Dynamic Model Averaging outperformed all other methods when monthly average prices are considered. Contrary, when monthly closing spot prices are considered, Bayesian dynamic mixture models happen to be very accurate compared to other methods, although beating the no-change method is still a hard challenge. Additionally, both revised and originally published macroeconomic time-series data are analyzed, ensuring consistency with the information available during real-time forecasting by mimicking the perspective of market players in the past.

Keywords: commodity forecast accuracy; price forecasting; model uncertainty; real-time forecasting; time-varying parameters

check for updates

Academic Editors: Olga Valenzuela, Fernando Rojas, Luis Javier Herrera, Hector Pomares and Ignacio Rojas

Published: 31 July 2025

Citation: Drachal, K.; Jędrzejewska, J. The Forecasting of Aluminum Prices: A True Challenge for Econometric Models. *Comput. Sci. Math. Forum* **2025**, *11*, 13. https://doi.org/10.3390/cmsf2025011013

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Forecasting commodity prices, particularly aluminum, is essential for producers, consumers, investors, and policymakers. Aluminum is a key material for industries such as construction, aerospace, electronics, and automotive manufacturing. Accurate price forecasting is crucial for making informed decisions regarding investments, production, and trade. However, this is a complex task due to the volatility and unpredictability inherent in commodity markets in general, especially because prices are influenced by various economic, geopolitical, and financial factors, which makes forecasting a very challenging task [1–3].

Commodity markets, including aluminum, are affected by supply disruptions, demand fluctuations, geopolitical tensions, and changes in macroeconomic conditions. For example, variations in global industrial production, particularly in major aluminum-consuming countries like China, can significantly affect aluminum demand and prices. Additionally, the claimed financialization of commodities has introduced further complexity, where futures markets and speculative pressures can influence price movements alongside fundamental economic factors [4–6]. These dynamics make forecasting commodity prices even more complex, as it requires consideration of both market fundamentals

and numerous other factors, leading to model uncertainty or feature selection problems. Furthermore, the relationship between a commodity price and its predictors can vary over time. On the other hand, in many cases, it is difficult to beat the simple no-change (NAIVE) or ARIMA method by generating more accurate forecasts [7–12]. Moreover, it is real-time data, which is quickly available to decisionmakers. In contrast, revised data provides a clearer picture of the past, but it is available with a time delay. Constructed models are usually trained using revised data, rather than the one that was truly available to market players in the past. As a result, this can generate another issue in econometric forecasting [13].

A promising approach to the problems above can be obtained through the application of Bayesian dynamic mixture models (BDMMs). These models incorporate model and parameter uncertainty, allowing forecasting adaptation as new information becomes available. However, these methods have not been studied much in applications to commodity prices. Therefore, this study contributes to the literature by comparing the accuracy of forecasts derived from various methods dealing with time-varying parameters and feature selection, as well as originally employing BDMMs in various specifications [14,15]. Furthermore, both real-time and revised datasets are employed.

2. Materials and Methods

2.1. Data

Monthly data from 04/2008 and 10/2024 were analyzed. LME (London Metal Exchange) spot prices (in USD per metric ton; monthly averages) for aluminum (unalloyed primary ingots, high grade, minimum 99.7% purity) were taken from World Bank [16]. This variable was denoted by p_aluminum.

As explanatory variables, excess demand (d_aluminum) was adopted following, for instance, Buncic and Moretto [2]. It was constructed as the difference between global consumption and the sum of global production and inventories. Consumption was proxied from the 2006 global estimates and assumed to mimic the time path of the OECD index of industrial production for China (2015 = 100, calendar and seasonally adjusted), due to a lack of exact data. Indeed, industrial production can be an effective proxy for this variable, and the role of China is huge on the analyzed market [17,18]. Inventories were taken as monthly averages of stocks on LME from Westmetall [19] in thousand metric tonnes. They were also taken as another explanatory variable (stocks_aluminum), as they can proxy speculative pressures. Production was taken in thousand metric tonnes from the International Aluminium Institute [20]. Following Buncic and Moretto [2], a convenience yield (cy_aluminum) was also taken (in percentages). It is constructed as

$$\text{r_short} - (\text{f_aluminum} - \text{p_aluminum})/\text{p_aluminum},$$

where r_short is the 3-month risk-free rate (3-month treasury bill: secondary market rate, TB3MS, monthly averages) taken from ALFRED and FRED [21,22], and f_aluminum denotes the official 3-month aluminum prices from LME in USD (^AH_3.F) taken from Stooq [23].

The US Industrial Production Index (INDPRO) was also taken into account [21,22]. It was denoted by ind_prod. The Index of Global Real Economic Activity (IGREA) proposed by Kilian [24] was obtained from ALFRED and FRED [21,22]. This was denoted by ec_act. Both ind_prod and ec_act were taken as explanatory variables resembling economic conditions.

Market stress was measured by two variables. First, the term spread, denoted by term_spread (i.e., U.S. 10-year treasury constant maturity minus U.S. 2-year treasury

constant maturity, T10Y2YM, monthly averages) was used [21,22]. The second, monthly averages of the VIX index of implied volatility [25] were denoted by VIX.

Stock market behavior was measured by the S&P 500 index (^SPX) and the Shanghai Composite Index (^SHC). They were denoted by SP500 and SSE, respectively. This allows for including both the U.S. and Chinese economy—to capture global changes in economies and commodity markets [23]. Additionally, two leading aluminum sector companies' stock prices were included, i.e., Rio Tinto Plc (RIO.UK, in GBX) and Alcoa Corp (AA.US, in USD), denoted by Rio_Tinto and Alcoa [26,27].

Moreover, two exchange rate currencies to USD of the largest primary aluminum producers were included, i.e., China (USDCNY) and Russia (USDRUB). They were denoted by fx_CNY and fx_RUB, respectively [23,28,29].

Finally, to capture the particular market past or autoregressive behaviors, the World Bank Base Metals Price Index (2010 = 100, USD, including aluminum, copper, lead, nickel, tin, and zinc), denoted by p_metals, was used [16].

Unless otherwise specified, for the above variables, the last observations were always taken into account (and closing and adjusted, if suitable).

2.2. Data Transformations

Similar to, for example, Buncic and Moretto [2], ordinary returns were assumed for most variables. In other words, time-series y_t was transformed into $(y_t - y_{t-1})/y_t$ where t denotes time. However, cy_aluminum, ec_act, term_spread, and VIX were not differentiated.

Next, following, for example, Koop and Korobilis [30], all variables were standardized into approximately stationary time-series. In other words, y_t is a time-series, μ is its mean of the first T observations, and σ is its standard deviation of the first T observations. Then standardization is understood as the transformation of $y_t = (y_t - \mu)/\sigma$. In particular, T = 100 was taken (and it was kept throughout the entire study as the in-sample period, for computing forecast accuracy, etc.). A whole sample was not used in order to omit the forward-looking bias. Furthermore, such a transformation allows all the time-series to have similar magnitudes, which improves computational issues in the applied models [31].

However, all forecast evaluation was carried out for raw (levels) aluminum prices; therefore, the predictions of returns obtained directly from the models (except ARIMA, NAÏVE, futures-based, and historical averages, which were computed directly for levels, i.e., raw data) were transformed back into levels. In other words, the estimations of models and their predictions were made over transformed variables, but for evaluation, the predictions of aluminum price were taken.

Explanatory variables were delayed one period back (i.e., first lags), except ec_act, where two periods back (i.e., second lags) were taken. This procedure resembles real-time information availability on the market [21].

Table 1 reports descriptive statistics, and Table 2 shows stationarity tests [32–35]. When assuming a 5% significance level for all variables, except for cy_aluminum, it can be assumed that they are stationary. Usual tests were applied: augmented Dickey–Fuller (ADF), Phillips–Perron (PP), and (a stationary-level version of) Kwiatkowski–Phillips–Schmidt–Shin (KPSS).

Table 1. Descriptive statistics.

	Min	Max	Mean	Median	Standard Deviation	Coefficient of Variation	Skewness
p_aluminum	−3.3049	2.7834	0.0661	0.0778	0.8876	13.4292	−0.3039
d_aluminum	−2.8146	3.3251	−0.0370	0.0080	0.9809	−26.4973	0.0032
stocks_aluminum	−2.0897	3.6537	−0.1230	−0.2754	0.9149	−7.4362	1.2928

Table 1. *Cont.*

	Min	Max	Mean	Median	Standard Deviation	Coefficient of Variation	Skewness
cy_aluminum	−0.5798	11.9777	2.0681	−0.0919	3.7896	1.8324	1.5975
ind_prod	−7.5606	3.4581	0.0337	0.1668	1.0862	32.2483	−3.8626
ec_act	−2.0417	2.6619	−0.0249	−0.0866	0.8256	−33.1113	0.6761
term_spread	−5.6916	1.7419	−1.5855	−1.4015	1.9364	−1.2213	−0.1627
VIX	−1.1556	4.1777	−0.1539	−0.3918	0.8850	−5.7491	2.2702
SP500	−3.7403	2.6016	0.0629	0.2084	0.9855	15.6561	−0.6114
SSE	−2.8843	2.4156	0.0153	0.0240	0.7945	51.8841	−0.3745
p_metals	−3.8620	2.8732	0.0765	0.0927	0.8618	11.2699	−0.6471
fx_CNY	−2.8497	4.0257	0.0650	−0.0421	1.1871	18.2563	0.5568
fx_RUB	−2.6425	3.8429	−0.0584	−0.1498	0.9966	−17.0747	0.9778
Rio_Tinto	−3.9567	2.7423	0.0754	0.0461	0.8487	11.2503	−0.4827
Alcoa	−4.5478	4.4837	0.1053	0.0472	1.2297	11.6781	−0.0903

Table 2. Stationarity tests.

	ADF Statistic	ADF *p*-Value	PP Statistic	PP *p*-Value	KPSS Statistic	KPSS *p*-Value
p_aluminum	−5.5373	<0.01	−137.4591	<0.01	0.1050	>0.1
d_aluminum	−6.5077	<0.01	−238.4936	<0.01	0.1053	>0.1
stocks_aluminum	−3.8686	0.0168	−66.5018	<0.01	0.6401	0.0190
cy_aluminum	−2.9544	0.1764	−5.4427	0.8037	1.7829	<0.01
ind_prod	−5.9235	<0.01	−158.5210	<0.01	0.1406	>0.1
ec_act	−4.0650	<0.01	−23.2835	0.0320	0.6412	0.0189
term_spread	−3.4339	0.0505	−14.1665	0.3081	3.2267	<0.01
VIX	−3.4498	0.0488	−36.3286	<0.01	0.6446	0.0186
SP500	−6.1128	<0.01	−201.2408	<0.01	0.1750	>0.1
SSE	−5.4294	<0.01	−191.9523	<0.01	0.0866	>0.1
p_metals	−6.1435	<0.01	−118.8156	<0.01	0.0518	>0.1
fx_CNY	−6.1865	<0.01	−182.8249	<0.01	0.1750	>0.1
fx_RUB	−5.9000	<0.01	−151.7995	<0.01	0.0399	>0.1
Rio_Tinto	−6.5383	<0.01	−185.9790	<0.01	0.0488	>0.1
Alcoa	−5.1368	<0.01	−206.0026	<0.01	0.1573	>0.1

2.3. Models

As mentioned before, the first 100 observations were taken as in-samples. However, all models were estimated in a recursive manner (i.e., expanding window in-samples), meaning that a forecast for time t was generated from the model trained over all (monthly) data available up to time $t - 1$.

Firstly, various types of Bayesian dynamic mixture models (BDMMs) were estimated [14,15], both with state-space (SS) and normal regression (NR) components. Besides the original versions, weighted averaging schemes were also modified to include weighted averaged forecasting (A), forecasting from the component model with the highest posterior probability (H), and a median probability model (M) from Barbieri and Berger [36]. These were denoted as BDMM-SS, BDMM-SS-A, BDMM-SS-H, BDMM-SS-M, BDMM-NR, BDMM-NR-H, and BDMM-NR-M [37]. The details can be found in [14,15,37].

Next, Dynamic Model Averaging (DMA), Dynamic Model Selection (DMS), and DMS based on a median probability model (DMP) were estimated. For these models, the standard values of forgetting factors, i.e., 0.99, were taken [38]. Also, Bayesian Model Averaging (BMA), Bayesian Model Selection (BMS), and BMS based on the median probability model (BMP) were estimated. These are versions of the previous models, but with forgetting

factors equal to 1. Besides these models, which consider averaging over all possible multilinear regression models (i.e., 2^K, where K is the number of explanatory variables, herein K = 14), modified models—averaging over only one-variable simple linear regression models—were also estimated. This is because such models highly reduce computation time, as they consider K + 1 models only (due to the inclusion of a model with constant variable only as well). These were denoted as DMA-1VAR, DMS-1VAR, BMA-1VAR, and BMS-1VAR [39].

Also, LASSO, RIDGE, and Elastic Net (EL-NET) regressions were estimated [40]. Bayesian versions [41] were also estimated (B-LASSO and B-RIDGE). The penalty parameter was selected in a recursive manner (consistently with the spirit of this study) based on t-fold cross-validation (t being the time period) with respect to Mean Square Error (MSE). Mixing parameters {0.1, 0.2, . . ., 0.9} were applied.

Moreover, the least-angle regression (LARS) was estimated [42] with similar t-fold cross-validation based on MSE.

Additionally, time-varying regressions were estimated (which are equivalent to DMA with only one component model, i.e., the one with constant and all explanatory variables). This was denoted as TVP-FOR. Similarly, time-varying regression with no forgetting was estimated, i.e., the forgetting factor was set equal to 1 (TVP).

The recursively estimated ARIMA model [43] was also used in an automated way, as well as the no-change (NAÏVE) method. Historical-average forecasts were computed both as an expanding window, i.e., including all past historical observations (HA), and as a rolling window of the past 100 periods (HA-ROLL).

Finally, futures-based forecasts were computed [7,44], including the basic ones (FUT-SIMPLE), i.e., futures prices, which are taken directly as the subsequent forecast period, and those based on the spread (FUT-SPREAD).

2.4. Methods

Estimations were carried out in R [45]. The forecast accuracy was measured by Root Mean Square Error (RMSE) and its normalized version (N-RMSE). For N-RMSE, the value of RMSE is divided by the mean of the forecasted time-series. Also, Mean Absolute Error (MAE) and Mean Absolute Scaled Error (MASE) were computed [46]. Competing forecasts were compared pairwise with the Diebold–Mariano test [43,47] with Harvey et al.'s [48] correction method. The full set of forecasts was evaluated with the Model Confidence Set (MCS) from Hansen et al. [49]. Squared error loss functions were applied [50]. Additionally, the Giacomini and Rossi fluctuation test [51] was employed. The rolling parameter $\mu = 0.6$ was used, which corresponds to approximately 5-year periods for the analyzed data [52].

3. Results

3.1. Main Results

Table 3 reports values of forecast accuracy measures. It can be seen that FUT-SPREAD generated the most accurate forecast. Also, FUT-SIMPLE is the second-most accurate forecast. The third best is DMA. This is according to each of the considered metrics. In the case of RMSE or N-RMSE, BDMMs do not perform very well, but if MAE or MASE is applied, BDMM-SS-A seems to be in top positions. Still, futures-based forecasts are the most accurate, DMA places next, and Bayesian versions of penalty regression models follow afterwards.

Table 4 reports outcomes from the Diebold–Mariano test. The forecast from the most accurate model, i.e., FUT-SPREAD, is compared with those from other methods. The same procedure is performed for the most accurate BDMM model, i.e., BDMM-SS-A, and for DMA. Assuming a 5% significance level, it can be seen that in the majority of cases, the

competing method can be assumed to be significantly less accurate than FUT-SPREAD, except for FUT-SIMPLE, BDMM-SS-A, and DMA. In the case of the BDMM-SS-A method, only BDMM-SS and historical-average methods can be assumed to be less accurate. In the case of DMA, models such as BDMM-SS-M, BMS-1VAR, and TVP can additionally be assumed to be less accurate. If the absolute scaled error loss function, i.e., the one corresponding to the MASE metric, is considered, then the outcomes are even stronger. All models are less accurate than FUT-SPREAD. In the case of BDMM-SS-A, both time-varying parameter regressions (TVP and TVP-FOR) can additionally be assumed to be less accurate. But in the case of DMA, BDMM-SS-M cannot be assumed to be less accurate. Instead, BDMM-NR-H and BDMM-NR-M can be assumed as such, as well as both time-varying parameter regressions. (Due to the limited space herein, detailed outcomes are not included.)

Table 3. Forecast accuracy measures.

	RMSE	**MAE**	**N-RMSE**	**MASE**
BDMM-SS	146.0506	98.3988	0.0681	1.2168
BDMM-SS-A	110.6545	74.8306	0.0516	0.9254
BDMM-SS-H	120.5796	77.9083	0.0562	0.9634
BDMM-SS-M	122.4366	79.9203	0.0571	0.9883
BDMM-NR	113.8516	81.6585	0.0531	1.0098
BDMM-NR-H	115.3268	83.6339	0.0538	1.0343
BDMM-NR-M	115.5588	83.0397	0.0539	1.0269
DMA	103.0514	73.9686	0.0480	0.9147
DMS	105.1728	75.8029	0.0490	0.9374
DMP	103.6592	76.0264	0.0483	0.9402
DMA-1VAR	104.5979	76.1900	0.0488	0.9422
DMS-1VAR	105.6150	77.7138	0.0492	0.9610
BMA	103.3669	75.3922	0.0482	0.9323
BMS	103.1626	75.3931	0.0481	0.9323
BMP	104.6967	77.2049	0.0488	0.9548
BMA-1VAR	110.1101	78.8818	0.0513	0.9755
BMS-1VAR	112.8305	80.7467	0.0526	0.9986
LASSO	106.5336	75.3249	0.0497	0.9315
RIDGE	105.1423	75.1496	0.0490	0.9293
EL-NET	106.6874	75.8148	0.0497	0.9376
B-LASSO	103.2737	74.5742	0.0481	0.9222
B-RIDGE	103.2769	74.7000	0.0481	0.9238
LARS	105.9409	75.1329	0.0494	0.9291
TVP	116.5330	89.4348	0.0543	1.1060
TVP-FOR	112.9689	84.2938	0.0527	1.0424
ARIMA	109.1411	76.9269	0.0509	0.9513
NAIVE	113.4765	80.8639	0.0529	
FUT-SIMPLE	89.3698	62.7108	0.0417	0.7755
FUT-SPREAD	**88.9340**	**62.1025**	**0.0415**	**0.7680**
HA	412.0597	306.5365	0.1921	3.7908
HA-ROLL	431.3150	306.8770	0.2011	3.7950

Surprisingly, the MCS test did not reject any of the models (p-value 0.0836), meaning that no statistically significant difference in forecast accuracies can be detected by this test. This was achieved with a 5% significance level, a "Tmax" version of the test statistic, and 5000 bootstrapped samples, which was used to construct the test statistic. However, this test with respect to absolute scaled error loss function, i.e., the one corresponding to the MASE metric, rejected (p-value 0.1892) both models based on historical averages (HA and HA-ROLL).

Table 4. Diebold–Mariano tests.

	FUT-SPREAD Statistic	FUT-SPREAD *p*-Value	BDMM-SS-A Statistic	BDMM-SS-A *p*-Value	DMA Statistic	DMA *p*-Value
BDMM-SS	2.8352	**0.0028**	2.3765	**0.0097**	2.6824	**0.0043**
BDMM-SS-A	1.6532	0.0507			1.4548	0.0745
BDMM-SS-H	1.9960	**0.0244**	1.0990	0.1372	1.6107	0.0552
BDMM-SS-M	2.1689	**0.0163**	1.2686	0.1038	1.7890	**0.0384**
BDMM-NR	2.5172	**0.0067**	0.3245	0.3731	1.3740	0.0863
BDMM-NR-H	2.6589	**0.0046**	0.4823	0.3153	1.5915	0.0574
BDMM-NR-M	2.6154	**0.0052**	0.4820	0.3154	1.5204	0.0658
DMA	1.5352	0.0640	−1.4548	0.9255		
DMS	2.0453	**0.0218**	−0.7672	0.7776	0.7592	0.2248
DMP	1.7663	**0.0402**	−1.0054	0.8414	0.2222	0.4123
DMA-1VAR	2.0490	**0.0216**	−0.8266	0.7948	0.5079	0.3063
DMS-1VAR	2.1926	**0.0154**	−0.6778	0.7503	0.7893	0.2159
BMA	1.9797	**0.0253**	−0.9506	0.8279	0.1068	0.4576
BMS	2.0254	**0.0228**	−0.9209	0.8203	0.0306	0.4878
BMP	2.2205	**0.0143**	−0.7355	0.7681	0.4519	0.3262
BMA-1VAR	2.5783	**0.0057**	−0.0647	0.5257	1.3109	0.0965
BMS-1VAR	2.8562	**0.0026**	0.2591	0.3981	1.7349	**0.0430**
LASSO	1.9072	**0.0297**	−0.7262	0.7653	1.4656	0.0730
RIDGE	1.7379	**0.0427**	−0.9509	0.8280	0.8529	0.1979
EL-NET	1.9279	**0.0284**	−0.7005	0.7574	1.5423	0.0631
B-LASSO	1.9496	**0.0270**	−0.9554	0.8291	0.0706	0.4719
B-RIDGE	1.8973	**0.0304**	−0.9817	0.8357	0.0764	0.4696
LARS	1.7662	**0.0402**	−0.8308	0.7959	0.9469	0.1730
TVP	2.9989	**0.0017**	0.6364	0.2630	1.9820	**0.0251**
TVP-FOR	2.4348	**0.0084**	0.2701	0.3938	1.5100	0.0671
ARIMA	1.8677	**0.0324**	−0.1654	0.5655	0.8870	0.1886
NAIVE	2.4360	**0.0083**	0.2945	0.3845	1.3396	0.0917
FUT-SIMPLE	1.2551	0.1062	−1.6325	0.9471	−1.4954	0.9310
FUT-SPREAD			−1.6532	0.9493	−1.5352	0.9360
HA	4.6742	**0.0000**	4.6447	**0.0000**	4.6705	**0.0000**
HA-ROLL	4.5022	**0.0000**	4.4755	**0.0000**	4.4990	**0.0000**

3.2. Robustness Checks

The Diebold–Mariano test, when applied pairwise to all the considered models, but firstly estimated with data as they were released, and secondly with revised data (as in the last period of the time span of this research), does not detect any significant differences for any models.

In the case of the Giacomini and Rossi fluctuation test (BDMM-SS-A vs. FUT-SPREAD and DMA vs. FUT-SPREAD), assuming a 5% significance level, the null hypothesis that the models' forecasting performance (i.e., BDMM-SS-A or DMA) is the same as that of FUT-SPREAD cannot be rejected (in favor of the alternative that a model performs worse than FUT-SPREAD). Details are presented in Figure 1.

The estimated models were also constructed by replacing the dependent variable with the aluminum LME cash official closing prices (AH_C.F) values of the month (instead of monthly average) [23]. In such a case, the two top minimizing RMSE models were NAÏVE and ARIMA. Futures-based forecasts were third and fourth. Interestingly, the next most accurate models were BDMM-SS and BDMM-SS-A. In the case of MAE metric, NAÏVE and ARIMA were still the most accurate, but BDMM-SS was the third most accurate model. Similarly, considering MASE, ARIMA was most accurate, and BDMM-SS was second best. Details are presented in Table 5.

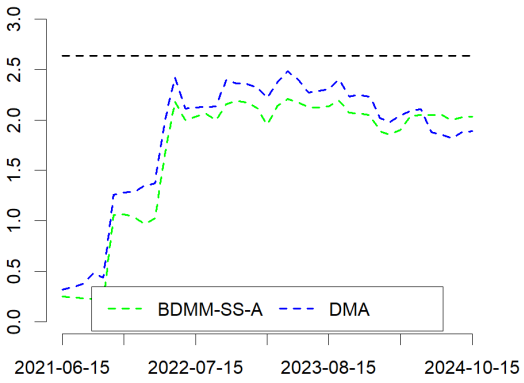


Figure 1. Giacomini and Rossi fluctuation test. Horizontal black line represents critical value.

Table 5. Forecast accuracy measures (forecasted monthly closing prices).

	RMSE	MAE	N-RMSE	MASE
BDMM-SS	206.8595	142.5773	0.0957	1.4725
BDMM-SS-A	209.1845	149.7094	0.0968	1.5461
BDMM-SS-H	219.1199	158.5795	0.1014	1.6377
BDMM-SS-M	217.7943	157.7649	0.1008	1.6293
BDMM-NR	220.4677	161.5014	0.1020	1.6679
BDMM-NR-H	219.0531	160.6152	0.1014	1.6588
BDMM-NR-M	220.8676	161.8159	0.1022	1.6712
DMA	218.9025	157.9856	0.1013	1.6316
DMS	218.0156	159.3862	0.1009	1.6461
DMP	218.8284	160.4063	0.1013	1.6566
DMA-1VAR	219.4383	160.4648	0.1016	1.6572
DMS-1VAR	218.5520	159.8007	0.1012	1.6504
BMA	218.6148	159.3702	0.1012	1.6459
BMS	219.7133	161.2518	0.1017	1.6653
BMP	219.6996	161.3329	0.1017	1.6662
BMA-1VAR	219.7482	161.0409	0.1017	1.6632
BMS-1VAR	219.3107	160.6929	0.1015	1.6596
LASSO	217.7136	158.5958	0.1008	1.6379
RIDGE	217.8733	157.8599	0.1008	1.6303
EL-NET	218.0318	158.8263	0.1009	1.6403
B-LASSO	219.2751	160.0091	0.1015	1.6525
B-RIDGE	218.5653	159.3973	0.1012	1.6462
LARS	217.9196	158.2746	0.1009	1.6346
TVP	226.3519	159.9796	0.1048	1.6522
TVP-FOR	223.6009	158.2434	0.1035	1.6343
ARIMA	134.7167	99.3532	0.0624	**1.0261**
NAIVE	**132.1674**	**96.8283**	**0.0612**	
FUT-SIMPLE	205.0920	148.4040	0.0949	1.5327
FUT-SPREAD	204.3735	148.3078	0.0946	1.5317
HA	420.2411	311.4823	0.1945	3.2169
HA-ROLL	436.6985	311.4661	0.2021	3.2167

However, in the case of MASE, none of the models obtained an MASE less than 1. Unfortunately, Diebold–Mariano tests (not reported herein, but available upon request) suggest that competing methods are less accurate than NAÏVE, and BDMM-SS and DMA, at least in most cases, are not significantly more accurate than other non-trivial models. However, the MCS test did not eliminate any model (p-value 0.0688).

Another modification was to replace the exchange rates by the exchange rates of countries that are biggest producers of ore (bauxite), i.e., Australia (AUDUSD) and the

Republic of Guinea (GNF) [23]. For GNF, monthly averages were taken from FXTOP [53]. In such a case, futures-based forecasts were still the most accurate according to all metrics. However, for the next most accurate models, these included BMS, BMA, DMS, and BMP, i.e., versions of the general DMA scheme [38,39]. Interestingly, in the case of MAE and MASE, it was LARS, EL-NET, LASSO, and RIDGE that were the most accurate. Still, the MCS test did not reject any model (p-value 0.0996). Detailed outcomes are not reported herein, due to the limited space.

4. Discussion

This study compared various aluminum price forecasting models, which are able to deal with variable selection problems and time-varying regression coefficients. In other words, they can deal with problems where the importance of price predictor changes over time, as well as its relationship strength with the predicted price changes. It was found that futures-based forecasts (in particular, end-of-month closing prices) can be a very accurate method when the monthly average aluminum price is predicted. Also, Dynamic Model Averaging is a promising forecasting scheme, whereas Bayesian dynamic mixture models performed less accurately than they were expected to. This adds some new value towards research into the usability of futures-based forecasts on commodity markets. However, in the case of predicting closing monthly prices, futures-based forecasts (also end-of-month closing prices) are still performing well, but Bayesian mixture models also become an interesting method, especially the state-space component version, as well as the one with an averaging scheme incorporated, outperforming Dynamic Model Averaging. On the other hand, for closing price forecasting, beating the no-change or ARIMA method is still a hard challenge. All of these results are in line with previous studies [2,54–57].

Author Contributions: Conceptualization, K.D.; methodology, K.D.; software, K.D.; validation, K.D. and J.J.; formal analysis, K.D.; investigation, K.D. and J.J.; resources, K.D.; data curation, K.D. and J.J.; writing—original draft preparation, K.D. and J.J.; writing—review and editing, K.D. and J.J.; visualization, K.D.; supervision, K.D.; project administration, K.D.; funding acquisition, K.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in whole by National Science Centre, Poland, grant number 2022/45/B/HS4/00510.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Republishing the full data is unavailable due to copyright restrictions. The majority of the data applied in the study are openly available as cited in the text (suitable tickers provided) and in the References. The forecast data generated in the study are openly available in RepOD at https://doi.org/10.18150/ERBJKK.

Acknowledgments: This research was funded in whole by National Science Centre, Poland, grant number 2022/45/B/HS4/00510. For the purpose of Open Access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Alquist, R.; Kilian, L. What do we learn from the price of crude oil futures? *J. Appl. Econom.* **2010**, *25*, 539–573. [CrossRef]
2. Buncic, D.; Moretto, C. Forecasting copper prices with dynamic averaging and selection models. *N. Am. J. Econ. Financ.* **2015**, *33*, 1–38. [CrossRef]

3. Reynolds, D.B.; Umekwe, M.P. Shale-oil development prospects: The role of shale-gas in developing shale-oil. *Energies* **2019**, *12*, 3331. [CrossRef]
4. Kahraman, E.; Akay, O. Comparison of exponential smoothing methods in forecasting global prices of main metals. *Miner. Econ.* **2023**, *36*, 427–435. [CrossRef]
5. Kwas, M.; Paccagnini, A.; Rubaszek, M. Common factors and the dynamics of industrial metal prices: A forecasting perspective. *Resour. Policy* **2021**, *74*, 102319. [CrossRef]
6. Liu, Q.; Liu, M.; Zhou, H.; Yan, F. A multi-model fusion based non-ferrous metal price forecasting. *Resour. Policy* **2022**, *77*, 102714. [CrossRef]
7. Alquist, R.; Kilian, L.; Vigfusson, R.J. Forecasting the price of oil. In *Handbook of Economic Forecasting*; Elliott, G., Timmermann, A., Eds.; Elsevier: Amsterdam, Netherlands, 2013; Volume 2, Part A, pp. 427–507. [CrossRef]
8. Gargano, A.; Timmermann, A. Forecasting commodity price indexes using macroeconomic and financial predictors. *Int. J. Forecast.* **2014**, *30*, 825–843. [CrossRef]
9. He, Z.; Huang, J. A novel non-ferrous metal price hybrid forecasting model based on data preprocessing and error correction. *Resour. Policy* **2023**, *86*, 104189. [CrossRef]
10. Ma, F.; Wahab, M.I.; Huang, D.; Xu, W. Forecasting the realized volatility of the oil futures market: A regime switching approach. *Energy Econ.* **2017**, *67*, 136–145. [CrossRef]
11. Nguyen, D.K.; Walther, T. Modeling and forecasting commodity market volatility with long-term economic and financial variables. *J. Forecast.* **2020**, *39*, 126–142. [CrossRef]
12. Pincheira, P.; Hardy, N. Forecasting aluminum prices with commodity currencies. *Resour. Policy* **2021**, *73*, 102066. [CrossRef]
13. Croushore, D. Frontiers of real-time data analysis. *J. Econ. Lit.* **2011**, *49*, 72–100. [CrossRef]
14. Nagy, I.; Suzdaleva, E. Mixture estimation with state-space components and Markov model of switching. *Appl. Math. Model.* **2013**, *37*, 9970–9984. [CrossRef]
15. Nagy, I.; Suzdaleva, E.; Karny, M.; Mlynarova, T. Bayesian estimation of dynamic finite mixtures. *Int. J. Adapt. Control Signal Process.* **2011**, *25*, 765–787. [CrossRef]
16. World Bank. Commodity Markets. 2025. Available online: https://www.worldbank.org/en/research/commodity-markets (accessed on 3 February 2025).
17. Menzie, W.D.; Barry, J.J.; Bleiwas, D.I.; Bray, E.L.; Goonan, T.G.; Matos, G. *The Global Flow of Aluminum from 2006 Through 2025: Open-File Report 2010–1256*; U.S. Geological Survey: Reston, VA, USA, 2010. Available online: https://pubs.usgs.gov/of/2010/1256 (accessed on 3 February 2025).
18. OECD. OECD Data Explorer. 2025. Available online: https://data-explorer.oecd.org (accessed on 3 February 2025).
19. Westmetall. Market Data. 2025. Available online: https://www.westmetall.com/en/markdaten.php (accessed on 3 February 2025).
20. International Aluminium Institute. Statistics. 2025. Available online: https://international-aluminium.org (accessed on 3 February 2025).
21. ALFRED. Archival FRED. 2025. Available online: https://alfred.stlouisfed.org (accessed on 3 February 2025).
22. FRED. Economic Data. 2025. Available online: https://fred.stlouisfed.org (accessed on 3 February 2025).
23. Stooq. Stooq. 2025. Available online: https://stooq.pl/index.html (accessed on 3 February 2025).
24. Kilian, L. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *Am. Econ. Rev.* **2009**, *99*, 1053–1069. [CrossRef]
25. Chicago Board Options Exchange. Historical Data. 2025. Available online: https://www.cboe.com/tradable_products/vix/vix_historical_data (accessed on 3 February 2025).
26. Yahoo. Yahoo!finance. 2025. Available online: https://finance.yahoo.com (accessed on 3 February 2025).
27. Shrivastava, P.; Vidhi, R. Pathway to sustainability in the mining industry: A case study of Alcoa and Rio Tinto. *Resources* **2020**, *9*, 70. [CrossRef]
28. Barry, J.J.; Matos, G.R.; Menzie, W.D.U.S. *Mineral Dependence—Statistical Compilation of U.S. and World Mineral Production, Consumption, and Trade, 1990–2010, Open-File Report 2013-1184*; U.S. Geological Survey: Reston, VA, USA, 2013. Available online: https://pubs.usgs.gov/of/2013/1184 (accessed on 3 February 2025).
29. Idoine, N.E.; Raycraft, E.R.; Hobbs, S.F.; Everett, P.; Evans, E.J.; Mills, A.J.; Currie, D.; Horn, S.; Shaw, R.A. *World Mineral Production 2018–22*; British Geological Survey: Keyworth, Nottingham, UK, 2024. Available online: https://nora.nerc.ac.uk/id/eprint/537241/1/World%20Mineral%20Production%202018%20to%202022.pdf (accessed on 3 February 2025).
30. Koop, G.; Korobilis, D. Large time-varying parameter VARs. *J. Econom.* **2013**, *177*, 185–198. [CrossRef]
31. Drachal, K. Determining time-varying drivers of spot oil price in a Dynamic Model Averaging framework. *Energies* **2018**, *11*, 1207. [CrossRef]
32. Dickey, D.A.; Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **1979**, *74*, 427–431. [CrossRef]

33. Phillips, P.C.B.; Perron, P. Testing for a unit root in time series regression. *Biometrika* **1988**, *75*, 335–346. [CrossRef]
34. Kwiatkowski, D.; Phillips, P.C.B.; Schmidt, P.; Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econom.* **1992**, *54*, 159–178. [CrossRef]
35. Trapletti, A.; Hornik, K. Tseries: Time Series Analysis and Computational Finance. 2024. Available online: https://CRAN.R-project.org/package=tseries (accessed on 3 February 2025).
36. Barbieri, M.M.; Berger, J.O. Optimal predictive model selection. *Ann. Stat.* **2004**, *32*, 870–897. [CrossRef]
37. Drachal, K. "Dynmix": An R package for the estimation of dynamic finite mixtures. *SoftwareX* **2023**, *22*, 101388. [CrossRef]
38. Raftery, A.E.; Karny, M.; Ettler, P. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics* **2010**, *52*, 52–66. [CrossRef] [PubMed]
39. Drachal, K. Dynamic Model Averaging in economics and finance with fDMA: A package for R. *Signals* **2020**, *1*, 47–99. [CrossRef]
40. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [CrossRef]
41. Gramacy, R.B. Monomvn: Estimation for MVN and Student-t Data with Monotone Missingness. 2019. Available online: https://CRAN.R-project.org/package=monomvn (accessed on 3 February 2025).
42. Hastie, T.; Efron, B. lars: Least Angle Regression, Lasso and Forward Stagewise. 2013. Available online: https://CRAN.R-project.org/package=lars (accessed on 3 February 2025).
43. Hyndman, R.J.; Khandakar, Y. Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* **2008**, *26*, 1–22. [CrossRef]
44. Ellwanger, R.; Snudden, S. Futures prices are useful predictors of the spot price of crude Oil. *Energy J.* **2023**, *44*, 65–82. [CrossRef]
45. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018; Available online: https://www.R-project.org (accessed on 3 February 2025).
46. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [CrossRef]
47. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. Available online: https://www.jstor.org/stable/1392155 (accessed on 3 February 2025). [CrossRef]
48. Harvey, D.; Leybourne, S.; Newbold, P. Testing the equality of prediction mean squared errors. *Int. J. Forecast.* **1997**, *13*, 281–291. [CrossRef]
49. Hansen, P.R.; Lunde, A.; Nason, J. The model confidence set. *Econometrica* **2011**, *79*, 453–497. [CrossRef]
50. Bernardi, M.; Catania, L. The model confidence set package for R. *Int. J. Comput. Econ. Econom.* **2018**, *8*, 144–158. [CrossRef]
51. Giacomini, R.; Rossi, B. Forecast comparisons in unstable environments. *J. Appl. Econom.* **2010**, *25*, 595–620. [CrossRef]
52. Jordan, A.; Krueger, F. Murphydiagram: Murphy Diagrams for Forecast Comparisons. 2019. Available online: https://CRAN.R-project.org/package=murphydiagram (accessed on 3 February 2025).
53. FXTOP. Historical Rates. 2025. Available online: https://fxtop.com/en/historical-exchange-rates.php (accessed on 3 February 2025).
54. Drachal, K. Forecasting selected energy commodities prices with Bayesian dynamic finite mixtures. *Energy Econ.* **2021**, *99*, 105283. [CrossRef]
55. Drachal, K. Forecasting prices of selected metals with Bayesian data-rich models. *Resour. Policy* **2019**, *64*, 101528. [CrossRef]
56. Farag, M.; Snudden, S.; Upton, G. Can futures prices predict the real price of primary commodities? *USAEE Work. Pap.* **2024**, *24*, 629. [CrossRef]
57. Kwas, M.; Rubaszek, M. Forecasting commodity prices: Looking for a benchmark. *Forecasting* **2021**, *3*, 447–459. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.