



Article

# Forecasting Selected Commodities' Prices with the Bayesian Symbolic Regression

Krzysztof Drachal \* and Michał Pawłowski

Faculty of Economic Sciences, University of Warsaw, 00-241 Warszawa, Poland

\* Correspondence: kdrachal@wne.uw.edu.pl

**Abstract:** This study firstly applied a Bayesian symbolic regression (BSR) to the forecasting of numerous commodities' prices (spot-based ones). Moreover, some features and an initial specification of the parameters of the BSR were analysed. The conventional approach to symbolic regression, based on genetic programming, was also used as a benchmark tool. Secondly, various other econometric methods dealing with variable uncertainty were estimated including Bayesian Model Averaging, Dynamic Model Averaging, LASSO, ridge, elastic net, and least-angle regressions, etc. Therefore, this study reports a concise and uniform comparison of an application of several popular econometric models to forecasting the prices of numerous commodities. Robustness checks and statistical tests were performed to strengthen the obtained conclusions. Monthly data beginning from January 1988 and ending in August 2021 were analysed.

**Keywords:** Bayesian econometrics; commodities markets; financial time-series forecasting; price forecasting; spot prices; symbolic regression; variable uncertainty



**Citation:** Drachal, Krzysztof, and Michał Pawłowski. 2024. Forecasting Selected Commodities' Prices with the Bayesian Symbolic Regression.

*International Journal of Financial Studies* 12: 34. <https://doi.org/10.3390/ijfs12020034>

Academic Editor: Konstantinos Baltas

Received: 18 December 2023

Revised: 13 March 2024

Accepted: 25 March 2024

Published: 29 March 2024



**Copyright:** © 2024 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

The objective of this paper is to apply a novel econometric tool in forecasting various commodities' prices. The novelty of the applied method lies in its fusion of two distinct approaches: symbolic regression (Koza 1998) and Bayesian econometrics (Koop 2017). The rationale behind such an idea stems from the challenges faced by researchers when constructing forecasting models, particularly when selecting explanatory variables for methods such as multilinear regressions. This is a non-trivial task (Tapia Cortez et al. 2018), especially when adopting the (standard) frequentist approach and employing the commonly used ordinary least squares method to estimate regression coefficients.

Of course, there are some classical techniques available for selecting explanatory variables, but the issue of model uncertainty still remains a serious challenge. Researchers sometimes address this problem by using conventional tools such as model averaging techniques (Steel 2020; Burnham and Anderson 2002). However, applying the frequentist approach still requires careful analysis and is susceptible to human bias. Moreover, when dealing with data sets consisting of more variables than the number of observations, the Bayesian approach often becomes a useful alternative (Koop 2017).

Indeed, a viable solution to this issue is transitioning from the frequentist approach to the Bayesian one, which generally is an efficient tool in forecasting commodity prices (Nurmakhanova 2020). In the Bayesian framework, a researcher starts from some prior knowledge or assumptions on estimated coefficients, which are then updated with the new information forming the posterior knowledge. Such an approach offers several advantages. Mathematically, it enables dealing with a higher number of explanatory variables than observations for each time series. Moreover, it closely resembles the real-life market scenarios, in which investors can continuously update their econometric models with the new information in each session, re-estimating coefficients accordingly (Koop 2017).

In other words, the Bayesian approach starts with an initial (prior) belief about the parameter of interest. As the new information becomes available, this belief is updated,

leading to the posterior belief. Consequently, a value of the parameter of interest evolves over time as this new information emerges. It is worth noting that these considerations not only apply to the technical challenge of selecting explanatory variables for regression models but are also valid for more complex models (Koop 2017).

Notably, the combination of symbolic regression with Bayesian principles has not been extensively explored yet, despite both approaches being philosophically well-suited to address real market problems, as already described (Jin et al. 2019). Thus, the proposed combination of these two formalisms represents an intriguing endeavour. This can introduce a new econometric tool and contribute to the ongoing scientific discourse surrounding model uncertainty and variable selection. Surprisingly, limited attempts to integrate these formalisms have been made yet (Vazquez et al. 2022; Guimera et al. 2020; Jin et al. 2019; Regolin and Pozo 2005; Zhang 2000).

Nevertheless, it should be noted that Bayesian methods are not always computationally “cheap” in this context (Lan et al. 2022). Moreover, in economics or finance, symbolic regression has not yet been extensively applied. There are some studies on forecasting production or emission quotas, but no application to price forecasting (Yang et al. 2015a, 2015b). Symbolic regression itself, which is a well-known tool (Koza 1998; Sinha et al. 2015), relies heavily on evolutionary algorithms, especially genetic ones (Koza 1998; Bhattacharya et al. 2016; Sinha et al. 2015; Eiben and Smith 2015).

Interestingly, genetic algorithms have been used in various econometric models, and were claimed to be very useful (Claveria et al. 2022; Garcia and Kristjanpoller 2019; Claveria et al. 2016, 2017; Mostafa and El-Masry 2016; Aguilar-Rivera et al. 2015; Sermpinis et al. 2015; Sheta et al. 2013; Hasheminia and Niaki 2006). While widely popular in technical fields, such as engineering, and nature-oriented sciences, such as ecology and medicine (Dimoulkas et al. 2018; Klotz et al. 2017; Golafshani and Ashour 2016; Ceperic et al. 2014; Narotam et al. 2014; Sarradj and Geyer 2014), they have not yet been extensively applied in economics or finance, especially in the context of variable uncertainty, and particularly for forecasting commodity prices.

Indeed, if applied to economics and finance, it was rather as modelling, optimisation, or discovering the “true” equation over some “static” data set (Brabazon et al. 2020). The true forecasting aim, moreover, with (dynamic, temporal changing) time-series, was rarely tackled. As a result, these issues create an important research gap, which this paper aims to fill.

Additionally, in this paper, BSR is compared with other popular forecasting methods dealing with variable uncertainty. These include, for example, Bayesian-based model averaging methods, such as Bayesian Model Averaging and Dynamic Model Averaging (Belmonte and Koop 2014; Koop and Korobilis 2011; Onorante and Raftery 2016; Cross and Nguyen 2017), and shrinkage methods, such as LASSO regression, which resembles the ordinary least squares method, but with the optimisation performed over a penalized function (Steel 2020; Burnham and Anderson 2002). In particular, these methods are compared in the context of their ability to forecast commodity prices, as each of them has their own strengths and limitations in this context.

When undertaking this research, some hypotheses were also claimed. First, that combining symbolic regression with the Bayesian approach can lead to some improvement in forecasting performance compared to using each approach separately. Secondly, that the proposed novel method can outperform other currently developed methods, such as LASSO and RIDGE regressions, Dynamic Model Averaging, Bayesian Model Averaging, and common conventional approaches such as ARIMA models, etc., in terms of forecast accuracy. Thirdly, that the recursive implementation of the proposed algorithm, which aligns with the flow of information from real markets, can further enhance forecasting performance, as model coefficients are re-estimated and updated in each subsequent period (Zhao et al. 2021; Tashman 2000).

This paper is organised in the following way. The next section contains a short review on commodity price determinants and modelling. It also provides references for important

and interesting reviews about the topics motivating the study and closely linked with the research described herein. The next section is devoted to a description of the applied data set. The next section briefly describes the BSR, benchmark models, forecast quality measures, and other methods applied in this research. The last sections present the obtained results and conclusions. The additional main advantage of this research is that a large number of commodities, and a large number of forecasting models suitable to deal with variable uncertainty issue, are estimated in a unified and consistent way and compared with each other, contrary to some more tailored research focusing on relatively few methods each and just some narrowed time-series usually covering different time spans. As a result, quite a wide insight over many methods is provided in this paper, and they are tested using many different commodities (time-series) at once in a unified way.

## 2. Literature Review

Forecasting commodity prices is a challenging but essential task (Drachal 2018b; Harvey et al. 2018; Arango et al. 2012; Byrne et al. 2013; Gargano and Timmermann 2014). There is an obvious practical need and a pure scientific interest in exploring novel econometric methods for this purpose (Herrera et al. 2019). However, despite numerous attempts, it remains difficult to propose an econometric method that consistently outperforms even simple methods, such as ARIMA models, or the naive method (i.e., the method in which the last observed value is taken as a one-step ahead forecast). Nevertheless, the ability to predict commodity prices is highly desirable for investors, policymakers, and governments. Obtaining a good forecast accuracy is indeed a common challenge in finance, and it raises further questions about the extent to which commodities prices are predictable at all (Wang et al. 2015a).

### 2.1. Forecasting Methods Challenges

In particular, the Bayesian symbolic regression (BSR) described by Jin et al. (2019) seems to be quite a promising forecasting tool. It starts from the assumption that an output function can be expressed as a linear combination of quite simple component functions. These functions are encoded (Weiss 2014) using symbolic trees (i.e., binary expression trees). Bayesian inference with the Markov chain Monte Carlo (MCMC) method is applied to describe the evolution of these trees structures. Jin et al. (2019) claimed that such a method leads to an improvement in the forecast accuracy and reduces complexity and computational issues.

Indeed, the conventional frequentist approach requires certain conditions to be met by the data set. For instance, the ordinary least squares method assumes the availability of enough observations to yield “reasonable” estimates of regression parameters. The solution is obtained by minimizing a specific objective function, requiring the existence and uniqueness of a solution. This process involves computing the Moore–Penrose pseudoinverse matrix, which requires certain matrices to be invertible (subject to specific restrictions). However, these methods become simply infeasible when the number of explanatory variables exceeds the number of observations in each of the considered time-series (Burnham and Anderson 2002).

Thus, conventional methods begin with a rigorous pre-selection process of explanatory variables, heavily reliant on a researcher’s subjective approach. Of course, a researcher is required to conduct a thorough literature review and carefully choose explanatory variables for an econometric model based on previous findings and conclusions. Then, the number of variables is constrained (Burnham and Anderson 2002), but this selection process is a crucial and challenging task.

Indeed, two significant challenges are common. First, the pre-selection of explanatory variables is influenced by the researcher’s subjective approach. Although researchers use their intuition and prior experience to strike a balance between preferences and objectivity, this remains a highly subjective aspect of any research and is naturally biased by human preferences. Therefore, seeking a more “automated” tool for this process is a very desirable

aim. Secondly, in modern research, dealing with “big data” has become quite a norm, particularly in forecasting commodity prices (Zhang et al. 2023; Liu and Lv 2020). This means that even after conducting a rigorous literature review, a considerable number of potentially relevant explanatory variables still remain (Koop 2017). As a result, the conventional frequentist approach becomes inadequate. Indeed, in the context of forecasting commodity prices, if a researcher would like to incorporate all the explanatory variables identified through a thorough literature review (Kaya 2016), then the frequentist approach would still face some significant challenges.

However, another important challenge faced with real markets is determining the appropriate model structure even after the potentially important explanatory variables have been pre-selected. For example, researchers must decide whether to construct a multiple linear regression model or apply certain transformations to the variables. Another example is making a decision involving whether to consider linear models versus including powers or logarithms of variables to capture potential non-linear relationships. Unlike the common econometrics practice that seeks data stationarity, the purpose of these transformations is to account purely for possible non-linearities (Caginalp and DeSantis 2011).

A common approach to this issue is considering various models and averaging their results. However, a less explored method is symbolic regression (Koza 1998). This regression analysis automatically explores numerous mathematical expressions to find the “best fitting” model while balancing accuracy and simplicity to address overfitting and over-parametrisation issues. The key advantage of symbolic regression, in the context of the described problems, is its immunity to human bias. It provides an automatic algorithm that simultaneously handles variable selection and model specification, including variable transformations. To be more precise, symbolic regression starts with an initial set of model specifications and later expands or explores more complex models derived from this initial set. Evolutionary algorithms are employed to construct new models (Eiben and Smith 2015).

These algorithms begin with a “population” of models, which, similar to the Darwinian theory of evolution, “crossover” with each other. This process requires specific probabilities to be set up. For example, let  $f_1$  and  $f_2$  be functions representing two model specifications (i.e., two multiple regression functions). First, both  $f_1$  and  $f_2$  can slightly change their functional forms through “mutation” with a certain probability. Secondly, a new function can be created with a “crossover” probability by combining a part (e.g., the first few symbols) from function  $f_1$  with another part (e.g., the last few symbols) from function  $f_2$ . Subsequently, the “population” of regression models is expanded with the modified  $f_1$  and the modified  $f_2$ , and their “crossover child”. The most simple approach is to keep these two probabilities fixed throughout the process (Koza 1998).

Concerning symbolic regression, there is quite a limited number of studies in which this method would be applied to economic or financial forecasting, especially in a direct way to forecast commodity prices. Symbolic regression—in this field of science—has been used mostly in forecasting production (Yang et al. 2015b) or emission quotas (Yang et al. 2015a), and modelling energy consumption (Rueda et al. 2019a, 2019b), etc. Senkerik et al. (2017a, 2017b) applied a certain hybridisation of symbolic regression based on analytical programming and differential evolution to forecast exchange rates. Furthermore, some applications in macroeconomic modelling (i.e., unemployment, inflation, etc.) were made (Kronberger et al. 2011). Still, computational issues are an important obstacle when applying symbolic regression and genetic algorithms to certain data sets (Huang et al. 2020).

Orzechowski et al. (2018) provided an extensive benchmarking review of several genetic programming approaches to symbolic regression based on up-to-date machine learning methods (e.g., least-angle regression). They explored over 100 regression problems. In particular, they concluded that symbolic regression outperforms gradient boosting algorithms, despite its relatively high computational price. Their research has been continuously updated to include the newest improvements of algorithms (La Cava et al. 2021). Still, the evolution of processes in symbolic regression optimisation (i.e., the desirable genetic



algorithm employed in symbolic regression) has been improved by several researchers (Landajuela et al. 2022; Zegklitz and Posik 2021; Kubalik et al. 2020; Hara et al. 2019). For instance, Haeri et al. (2017) proposed certain modifications to the mutation and crossover procedures (based on coefficients computed over the data set) in order to reduce bloat (i.e., uncontrollable growth of the average tree size in symbolic regression). These issues, as well as a concise review of the advantages and challenges of symbolic regression, are presented by Smits and Kotanchek (2005). Recently, Haider et al. (2023) continued improving symbolic regression algorithms, in particular by focusing on issues with the shape the regression functions and including prior knowledge about it.

In case of variable selection and model uncertainty problems, currently, the most extensively developed approaches are those involving the use of Bayesian methods. These methods are based on the assumption that among numerous potentially interesting models, there exists a “true” model that can be identified. Examples of such approaches are Bayesian Model Averaging, which has been found useful in macroeconomic modelling, and its extension, i.e., Dynamic Model Averaging (Steel 2020; Wang et al. 2017; Raftery et al. 2010). These Bayesian methods rely on model averaging techniques. Multiple regression models are constructed from a set of interesting, potentially important, variables, and each model is assigned a certain (posterior) probability. These probabilities are then used as weights in the model averaging procedure. The final forecast is a combination of forecasts obtained from individual models, resulting in an averaged forecast. There are arguments in favour of model averaging over simply selecting the model with the highest posterior probability (Baumeister and Kilian 2015; Wang et al. 2017). Moreover, some researchers argue that selecting a model with the highest posterior probability is not always the optimal solution, but, in certain conditions, the “median probability model” may perform better (Barbieri and Berger 2004).

While model averaging can be performed in the conventional frequentist approach (for example, using the Akaike Information Criterion as weights), it is still constrained by the limitations on the ratio between the number of variables and the number of observations, as already mentioned in the previous part of this paper. Other methods to deal with model uncertainty include stepwise regression, shrinkage methods (such as LASSO regression), extreme bounds analysis, *s*-values, general-to-specific modelling, the model confidence set approach, and the best subset regression (Steel 2020). However, the motivation behind the research described in this paper is rooted in the recognition that existing econometric methods used for forecasting commodities prices have both advantages and drawbacks, with no single method clearly superior to others. Moreover, the modern landscape for researchers and practitioners includes quite large and complex data sets. For instance, the FRED Monthly Database for Macroeconomic Research comprises over 150 time-series, necessitating specialised econometric methods beyond the conventional ones (FRED 2015).

Another desirable feature of a commodity price model is to be “flexible”. In particular, in the case of symbolic regression, such “dynamic” approaches are useful and can improve the forecasting performance. For instance, Wagner et al. (2007) already pointed out that dynamic (i.e., with a rolling window) estimations for symbolic regression can outperform static estimations because otherwise, in the case of a time-series, the changing environment issue is ignored. Secondly, they also considered forecast combination advantages over the simple selection of the best fitting solution. An improvement in forecasting with rolling windows was also noticed by Winkler et al. (2015a, 2015b). Furthermore, Lee (1999) advocated recursive computations for time-series modelling with symbolic regression. Besides that, in a more general context, Clark and McCracken (2009) advocated combining recursive and rolling forecasts when dealing with linear models. According to them, such a procedure might improve the forecast accuracy.

To sum up, the desired properties of an econometric model for commodity prices include the following features: the ability to handle a large number of variables on a theoretical basis; “adaptability”, meaning that model coefficients are continuously re-estimated (updated) as new market information becomes available; the ability to minimise

the bias towards human decisions, as human subjectivity can influence model outcomes; and the ability to capture the time-varying importance of different explanatory variables (Huang et al. 2021). Indeed, addressing these objectives is crucial in developing a robust econometric approach that can effectively forecast commodities prices in the presence of complex and dynamic data environments (Chai et al. 2018; Yin et al. 2018; Zhao et al. 2017).

The last property can be briefly justified by a quick look at crude oil prices as an example. Prior to the 1980s, most models relied mostly on supply and demand factors to explain oil price movements. On the other hand, a well-documented empirical and theoretical factor that influences oil prices is the interest rate (Baumeister and Kilian 2015; Arango et al. 2012). However, since the 1990s, more attention has been given to the impact of exchange rates (Basher et al. 2012; Reboredo 2012; Chen and Chen 2007). This is because exchange rates can indirectly influence supply and demand forces. The appreciation or depreciation of an exchange rate can impact the real price paid by importers (and thus their demand) and affect the real profit of commodity exporters (and thus their production motivation).

Further, recognizing that the global economic landscape is continuously evolving and influenced by the process of globalisation, researchers began to focus on the importance of global economic activity (Wang et al. 2015b). Indeed, commodity markets are usually global markets with a limited number of significant producing countries and numerous consuming (importing) countries. As a result, changes in the global economic activity can significantly impact demand forces, and should be taken into account in a modelling framework.

Furthermore, since the 2000s, researchers have focused on understanding the intricate links between futures and spot prices. Some have found that factors such as fluctuations in open interest (i.e., the total number of outstanding options or futures contracts that have not yet been settled) can serve as better proxies for the futures market than simply using futures prices (Hong and Yogo 2012). This has led to a growing interest in exploring the relationship between a given commodity price and various stock market indices. Stress market indices (for example, the VIX index) have also been found to be useful in this context. Additionally, policy uncertainty also began to be considered as an important commodity prices driver (Byun 2017; Chen et al. 2014; Gargano and Timmermann 2014; Arslan-Ayaydin and Khagaleeva 2013).

Consequently, in the example of the oil market, modelling data before the 1980s require more emphasis on demand and supply variables. In the 1990s, more variables representing fluctuations in exchange rates should be included, and after the 2000s, additional variables from the stock markets should be considered. In other words, the state-space of the model should be allowed to vary in time, in order to account for the changing dynamics of the commodity market (Nonejad 2019; Wang et al. 2017; Arouri et al. 2010; Cross and Nguyen 2017; Drachal 2016).

## 2.2. Crude Oil

The first group of oil price determinants is strictly connected to the stock markets. Several papers pointed out a link between exchange rates and the demand for oil: a change in a currency's value in comparison with a currency in which oil is denominated affects the demand positively (in the same direction). Currency appreciation increases the purchasing power, and as a consequence, decreases the relative price of oil (Bal and Rath 2015; Reboredo et al. 2014; Hartley and Medlock 2014; Aloui et al. 2013a, 2013b; Uddin et al. 2013; Tiwari et al. 2013; Akram 2009). This link has been confirmed for emerging markets as well as for developed markets (Chen and Chen 2007; Wang et al. 2004). Xu et al. (2019) found that this impact began especially in the early 2000s, and, moreover, the relationship is non-linear.

Byrne et al. (2013) analysed the role of demand and supply in determining oil prices, while other studies examined the connection between export in developed countries (Riggi and Venditti 2015) and policy uncertainty (Andreasson et al. 2016; Bekiros et al. 2015). Uncertainty shocks hit the real side of an economy, decreasing production and investment (Bloom 2009), which in turn negatively affects commodity prices. Haigh (2018) analysed

fundamentals and concluded that in the oil market, supply forces drive prices during geopolitical tensions, but during economic crises demand forces dominate. [Ghalayini \(2017\)](#) concluded that oil inventories impact the short-term oil prices. However, in the long-term the impact of demand, supply, the exchange rate, and speculation in the futures market is also important.

The effect of gross domestic product, stock market activity, and the volatility of the stock market on oil prices was also analysed in the literature ([Basher et al. 2012](#); [Aroui et al. 2011](#); [Bernabe et al. 2004](#); [Yousefi and Wirjanto 2004](#)). There is a clear correlation between economic growth and the growth in demand for oil. In recent years, the emerging markets were those where the demand for crude oil grew most rapidly ([Basher et al. 2012](#)). Additionally, [Mensi et al. \(2013\)](#) described the process of shock and volatility transmission between various markets in the globalised economy, finding the S&P's index significance in oil price prediction. A shock (e.g., panic, optimism) in a given market, in this case a stock market, can be easily transferred to a different market, say the commodities market.

[Du and He \(2015\)](#) also found a relationship between the volatility of stock markets and the oil market, similarly to [Smiech and Papiez \(2013\)](#), who noted such a link between exchange rates, stock market volatility, and fossil fuels. Recently, the role of oil futures markets' financialisation ([Duc Huynh et al. 2020](#)) and of speculative pressures has been discussed ([Diaz-Rainey et al. 2017](#); [Liu et al. 2016](#); [Yin and Yang 2016](#); [Kilian and Murphy 2014](#); [Fattouh and Scaramozzino 2011](#); [Hamilton 2009](#)). Financialisation implies a growing role for treating commodities and their derivatives as investment assets ([UNCTAD 2012](#)). The investigated mechanism would be the following: an increase (or decrease) in the speculative demand would affect commodity prices similarly to changes in the "regular" demand. Sometimes, the level of inventories was taken as a proxy for the pressure on the market ([Hamilton 2009](#)), yet the results were rather ambiguous ([Kim et al. 2017](#)). The very role of inventories might be twofold: they might serve as hedging instruments, used in order to decrease the risk in an investment portfolio through its diversification, or as a speculative tool to achieve gains ([Irwin et al. 2009](#)). It has been debated how strongly the financialisation of the oil futures market influenced the commodity's spot price, albeit without yet reaching clear results ([Carmona 2015](#); [Fattouh et al. 2013](#)).

It has also been pointed out that oil prices have been strongly affected by the demand growth in rapidly developing Asian countries, most notably by the change in China's character from a net petroleum exporter until 1992 to the world's top importer in 2019, surpassing the U.S. in 2017 ([EIA 2020](#); [Killian and Hicks 2013](#); [Kaufmann 2011](#); [Li and Leung 2011](#)). For example, [Wang and Sun \(2017\)](#) concluded that economic activity is the most significant factor driver of oil prices compared with other factors. New emerging markets boost the demand for oil further, and, consequently, contribute to enhancing the price growth pressure. [Abd Elaziz et al. \(2020\)](#) obtained promising results in their oil price forecasting model, employing 10 determinants of the commodity's prices: exchange rates (denominated to the U.S. dollar) of the Canadian dollar, Euro, and Chinese yuan, as well as coal, natural gas, copper, gold, silver, iron, and lagged oil prices. There was a clear positive correlation between oil prices and all other commodities (the strongest being for copper and iron: 0.92), while for exchange rates the correlation was negative. In this case, the most important correlation was found between the oil price and the Canadian dollar exchange rate (−0.75).

### 2.3. Natural Gas

The natural gas market is much more segmented than the markets for oil or coal ([Mohammadi 2011](#)) and is not as globalised. While comparing different countries, one must remember that geographical location, deregulation policies, technological advances, and trade agreements all play a crucial role in determining gas prices ([Olsen et al. 2015](#)). These factors, increasing transportation and transaction costs, can be impediments to arbitrage, thus restricting providing homogenous goods, in this case natural oil, at a particular price, according to the law of one price.

Due to this commodity's characteristics, for example, requirements for transport infrastructure (e.g., pipelines), access to it, and its prices strongly depend on geographical and infrastructural factors. Having said that, it is not at all surprising that differences between states can be large. [Ji et al. \(2014\)](#) observed that while North American gas prices depend on the condition of the global economy, gas prices in Europe and Asia are affected more by crude oil prices. According to [Ji et al. \(2014\)](#), Europe and Asia, as major gas consumers, are characterised by a rigid demand for the commodity, so global economic activity has a relatively small impact on their gas prices. At the same time, these gas markets are relatively immature compared to America. Namely, their pricing mechanisms are based to a much greater extent on oil prices, and are less determined by the internal demand/supply balance.

Some general determinants in the short term include temperature and supply shocks, while in the long term these include coal and long-term oil prices ([Nick and Thoenes 2014](#); [Mu 2007](#)), climate uncertainty ([Bistline 2014](#)), and even the release of gas storage reports ([Rubaszek and Uddin 2020](#); [Linn and Zhu 2004](#)). While temperature affects the amount of gas consumption and, thus, the demand for it, coal (discussed further herein) can be described as the substitute for this commodity ([Obadi and Korcek 2020](#)); the gas storage reports can be a determinant due to the previously discussed financialisation of commodity markets. Investors adjust their decisions to the market situation, actualised with the accessible information. Indeed, [Chiou-Wei et al. \(2020\)](#) found that underground storage is an important driver of natural gas prices in the U.S. (as well as the weather, oil price, and macroeconomic news).

Despite the fact that in certain conditions natural gas may be treated as a substitute to crude oil, their linkage is still sometimes found to be unclear ([Atil et al. 2014](#); [Ji et al. 2014](#); [Brown and Yucel 2008](#)). Here again, as in the case of oil price determinants, the changing position of the Asian economies is an important factor, affecting gas prices through their influence on the global supply/demand balance ([LaRose 2014](#); [Cornelius and Story 2007](#)). According to a general relationship, economic growth in Asian countries was followed by increased energy consumption. As [LaRose \(2014\)](#) pointed out, estimated forecasts give reason to believe that energy demand in Asia will double in the next 50 years.

#### 2.4. Coal

The basic characteristics of the coal market are the price of the commodity and its quality ([Li et al. 2014](#)). Furthermore, an important role is played by demand for energy and prices of other fuels ([LaRose 2014](#); [Dong et al. 2010](#)). Of course, when the price of substitute goods (other energy sources) falls, the demand for a given commodity based on the past price will also drop, and the price will have to adjust. [Regnier \(2007\)](#) found that coal prices are much less volatile than the price of the other energy commodities, and [Yang et al. \(2012\)](#) pointed out that much of the coal price volatility and its surges have been caused by China and its decision to decontrol the commodity's prices; uncontrolled prices then had to adjust to the value reflecting production costs and market conditions.

In contrast to past market liberalisations in many countries, the process was not sudden, but more gradual, starting in the early 2000s. As [Yang et al. \(2012\)](#) noted, one of the first steps towards coal market liberalisation in China was allowing producers to sell the excess output on a free market, the initial result of which was coal price increases and greater volatility; from 2003 to 2010, the price of coal more than tripled. [Arora and Tanner \(2013\)](#) found some evidence confirming Hotelling's hypothesis ([Hotelling 1931](#)) that non-renewable commodities are influenced by interest rates. According to [Hotelling \(1931\)](#), in an efficient market, owners of non-renewable resources will provide a supply only if it will be profitable in comparison with financial instruments yielding their interest rate.

Similar general observations were made in the case of energy commodities' dependency on economic growth and size of labour force ([Dogan 2016](#); [Wang et al. 2015b](#); [Apergis and Payne 2010](#)), and on policy uncertainty ([Wang et al. 2015a](#)). In the work of [Alameer et al. \(2020\)](#), the following factors proved to be important determinants of coal prices: copper,



natural gas, iron, silver, crude oil, and gold prices, as well as the Australian dollar, Indian rupee, and Chinese yuan exchange rates (denominated to the U.S. dollar). The Pearson correlation between the other commodities' prices and coal prices was positive and rather high (more than 0.80 on average), while for the U.S. dollar to Indian rupee exchange rates it was weaker (0.470), and for the U.S. dollar to the Australian dollar and the U.S. dollar to Chinese yuan it was negative ( $-0.666$  and  $-0.283$ , respectively).

## 2.5. Metals

Of course, fundamentals are agreed to play the major role in driving prices of metal commodities. However, non-fundamental factors can be important, especially in the short-term horizon (Guzman and Silva 2018). Similar to previously discussed commodities, metals prices are also significantly affected by the exchange rates (Chen et al. 2010). For example, Ciner (2017) confirmed that South African rand exchange rates can be used as a predictor of palladium, platinum, and silver prices. It is of note that South Africa is one of the leading producers of these resources. According to Prates (2007), especially in the early 2000s, there was a strong correlation between commodity prices and the global macroeconomic conditions. Chen (2016) found that the stock prices of resource-based companies improve metal price forecasting. On the other hand, Kaur and Dhiman (2017) stated that metal stock returns have a limited impact on metal commodities' prices. Brown and Hardy (2019) focused on the Chilean exchange rate and its ability to predict the prices of non-ferrous metals. Pincheira and Hardy (2021) concluded that the exchange rates of some commodity exporting countries can be successfully used in forecasting the aluminium price.

Furthermore, in the case of metal commodities, some attempts have been made to predict the commodity prices based on inventories quotas (Geman and Smith 2013). On the other hand, financialisation issues are not commonly agreed for metal markets (Mayer et al. 2019). Weng et al. (2018) employed several financial factors to build iron ore forecasts, while in a more recent study, Ewees et al. (2020) confirmed the following factors to be good predictors of iron ore prices: crude oil, gold, scrap, silver, copper, and lagged iron ore prices; Australian dollar to U.S. dollar and Chinese yuan to U.S. dollar exchange rates; and inflation rates in the U.S. and China. The correlation with other commodity prices was strong and positive (with the highest value for crude oil of 0.919), while for exchange and inflation rates, correlations were rather weaker and negative. For example, the correlation between the U.S. inflation rate and the iron ore price was  $-0.297$ . The authors justified the choice of variables corresponding to Australian and Chinese economic conditions with the fact that Australia is one of the biggest producers and exporters of iron ore, and China is the biggest importer.

Similarly, Alameer et al. (2019a) used the Chilean peso, Peruvian sol, and Chinese yuan exchange rates; the inflation rates of the U.S. and China; and the prices of gold, silver, iron, oil, and lagged copper prices in copper price forecasting. Alameer et al. (2019b) provided a similar list of gold price determinants; instead of Peruvian sol and Chilean peso exchange rates, they took South African rand and Indian rupee exchange rates (denominated to the U.S. dollar), and, of course, removed gold as a predictor (since it was a dependent variable). At the same time, Yuan et al. (2020) successfully applied the opinion score based on web scraping technology as one of the predictors of gold prices. An algorithm using text mining methods was employed to retrieve words and phrases describing market sentiments and an opinion score variable was created. Its application enabled reducing the mean absolute percentage error for predictions significantly, and thus improved gold price forecasts.

Gangopadhyay et al. (2016) used the stock market index, exchange rate, U.S. bond rates, oil prices, and the consumer price index to model gold prices in India. Buncic and Moretto (2015) analysed copper prices based on numerous potential determinants with Dynamic Model Averaging. Geopolitical risk was also found as an important factor in predicting, for example, gold prices. Banerjee et al. (2019) utilised the geopolitical risk index based on political events highlighted in the major newspapers around the world.

## 2.6. Agricultural Commodities

Agricultural products' prices are, of course, most of all determined by supply and demand. The latter, as [Rezitis and Sassi \(2013\)](#) noted, is strongly dependent on factors such as population growth, GDP growth, and changes in consumption. The former, on the other hand, is determined by weather and climate conditions, harvest areas, and the cost of fertilizers. Apart from these two groups of variables, crude oil prices and exchange rates may serve as predictors of agricultural goods prices ([Fernandez-Diaz and Morley 2019](#); [Osathanunkul et al. 2018](#); [Nazlioglu and Soytas 2012](#)).

The influence of oil prices on agricultural product prices might be twofold; first, higher oil prices result in production cost surges. Second, more expensive crude oil means a higher demand for biofuels. As [Sukcharoen and Leatham \(2018\)](#) have suggested, these correlations are particularly important during economic downturns, yet not in times of prosperity. [Chen et al. \(2012\)](#), as well as [Thiyagarajan et al. \(2015\)](#), added stock market indices to this list, and [Ribeiro and Oliveira \(2011\)](#) noted the significance of stock quotas and convenience yields in forecasting commodity price variations. Similarly, [Schewe et al. \(2017\)](#) identified storage dynamics as crucial in modelling global wheat prices. This factor was also identified by [Hamid and Shabri \(2017\)](#) for palm oil (amongst production quotas and substitute commodity prices). [Alam and Gilbert \(2017\)](#) emphasised the role of global economic conditions, monetary policy, and the U.S. dollar exchange rates, while [Hatzenbuehler et al. \(2016\)](#) provided an analysis supporting a claim regarding the importance of policy shifts.

[Frankel and Rose \(2010\)](#) proposed a model of agricultural goods prices combining three important groups of their determinants, namely, global economic activity, speculation, and monetary policy conditions. The role of economic growth and its influence on demand increase has been discussed previously, as well as the impact of speculation on commodities prices. With respect to monetary policy conditions, they noted that low interest rates, apart from having an effect (also previously discussed) on providing supply (high interest rates diminish supply, and, consequently, increased prices), have a converse effect of high interest rates on the cost of keeping inventories. This contributes to decreasing demand and prices. Such effects were also discussed by [Ouyang and Zhang \(2020\)](#), [Etienne et al. \(2018\)](#), and [Algieri et al. \(2017\)](#).

[Borychowski and Czyzewski \(2015\)](#) added one important factor, namely, conditions of trade policy, to some of the aforementioned determinants of food supply. They noted that export bans, tariffs, and export taxes decrease the supply and restrict providing an optimal amount of agricultural commodities beyond national borders. This, consequently, results in higher prices. Based on Finland's case, [Irz et al. \(2013\)](#) suggested that the most important determinant of food prices are farm prices, followed by wages in food retail, and then energy prices, the latter playing only a limited role.

## 2.7. General Remarks on Commodity Price Predictors

In general, certain common predictors for various commodities can be identified. Apart from the above-mentioned exchange rates, which were also analysed in the context of the overall commodity market ([Pincheira-Brown et al. 2022](#); [Souza et al. 2021](#); [Ayres et al. 2020](#); [Zhang et al. 2016](#)), and financial factors, it was found that crude oil and gold prices might enable good forecasts of other commodities' prices ([Lubbers and Posch 2016](#); [Chen 2015](#); [Al-Qudsi 2010](#)). As previously discussed, it seems that interest rates can also serve as commodity price predictors ([Byrne et al. 2013](#); [Arango et al. 2012](#)). [Gargano and Timmermann \(2014\)](#) provided a broad study, according to which the most important predictors were the investment-to-capital ratio and the growth of industrial production. [Ahumada and Cornejo \(2015\)](#) added supply and demand factors (in particular, with reference to emerging markets such as China), economic growth, and monetary policies. [Steermer \(2018\)](#) argued that in the long term, demand forces play a more important role than supply shocks. This was also confirmed by [Jacks and Stuermer \(2020\)](#). Other predictors common for various commodities are the U.S. inflation rate, world industrial production, and the world stock index ([Kagraoka 2016](#)). Furthermore, financialisation

issues were discussed in the context of various commodities (Fishe and Smith 2019; Yan and Yuan 2019). Moreover, the recent economic growth of several emerging economies resulted in increasing the demand for commodities (Labys 2006). Tan and Ma (2017) analysed various commodities and confirmed that macroeconomic uncertainty significantly impacts their prices.

The above considerations are summarised in Table 1.

**Table 1.** Most important determinants of commodity prices (except supply, demand, and stocks).

Crude Oil	Natural Gas	Coal	Metals	Agricultural Commodities
<ul style="list-style-type: none"> <li>• economic boost in Asia</li> <li>• economic growth</li> <li>• exchange rates</li> <li>• financialisation of the market</li> <li>• interest rates</li> </ul>	<ul style="list-style-type: none"> <li>• crude oil prices</li> <li>• economic growth in Asian countries</li> <li>• geographical conditions</li> <li>• nature of gas price mechanisms</li> <li>• trade agreements and infrastructure</li> <li>• weather and climate conditions</li> </ul>	<ul style="list-style-type: none"> <li>• crude oil and gas prices</li> <li>• economic growth</li> <li>• interest rates</li> <li>• past market liberalisation in China</li> </ul>	<ul style="list-style-type: none"> <li>• economic conditions (inflation, exchange rates) of main importers and producers</li> <li>• global macroeconomic conditions</li> <li>• market sentiments</li> </ul>	<ul style="list-style-type: none"> <li>• cost of fertilizers</li> <li>• crude oil prices</li> <li>• demand for biofuels</li> <li>• energy prices</li> <li>• farm prices</li> <li>• harvest area</li> <li>• interest rates</li> <li>• international trade agreements and policies</li> <li>• population growth</li> <li>• weather and climate conditions</li> </ul>

### 3. Data

Monthly data starting in January 1988 and ending in August 2021 were used. Such a time span was chosen due to data availability. The data frequency was selected in a way to satisfactorily capture changes on the market but to exclude some short-term fluctuations (e.g., due to speculative activities). On the other hand, a monthly frequency should allow us to include as explanatory variables some macro data, which are published in quite low frequencies (Alquist et al. 2013).

Commodity prices were taken from The World Bank (2022) and transformed to logarithmic differences. In particular, the logarithmic difference of the variable  $Y_t$  was defined as  $\log(Y_t) - \log(Y_{t-1})$ . Such a transformation is quite standard and common in econometric analysis. In the case of symbolic regression, data transformation is not obligatory. However, some benchmark models require, for example, stationary time-series. Secondly, even if not necessary, transformed data can often result in a better forecast accuracy of the final models (Coulombe et al. 2021; Medeiros et al. 2019; Drachal 2018a). The widest basket of commodities was attempted to be collected. However, due to missing observations, etc., 56 time-series of commodities prices were finally considered (monthly averages of spot prices). They are listed in Table A1 in Appendix A. A detailed description of the time-series can be found in the original source (The World Bank 2022). The set of explanatory variables was constructed in line with the already presented literature review. Furthermore, similar variables were used by Guidolin and Pedio (2021), Salisu et al. (2019), Gargano and Timmermann (2014), and Juvenal and Petrella (2014). The full list is presented in Table 2. In total, 39 explanatory variables were considered.

In particular, the dividend-to-price ratio was taken as the difference between the logarithm of U.S. stock dividends (Schiller 2000, 2022) and the logarithm of U.S. stock prices, i.e., the S&P 500 Index (Stooq 2022). Before taking logarithms, dividends were aggregated to 12-month moving sums. The price-to-earnings ratio was taken from Schiller (2000, 2022). It was taken as Cyclically Adjusted Price Earnings Ratio P/E10 (or CAPE). It is based on the average inflation-adjusted earnings from the previous 10 years.

**Table 2.** The list of explanatory variables.

Abbreviation	Description
dpr	Dividend-to-price ratio
pe	Price earnings ratio
str	Short-term interest rate
ltr_US	Long-term interest rate for the U.S.
ltr_EU	Long-term interest rate for the Euro area
ts	Term spread
drs	Default return spread
cpi	U.S. Consumer Price Index for all urban consumers
ppi	U.S. Producer Price Index
ip	U.S. industrial production
ee	U.S. average hourly earnings of production and nonsupervisory employees
M1	M1 money stock for U.S.
M2	M2 money stock for U.S.
gea	Killian's Index of Global Real Economic Activity
une	U.S. unemployment rate
AUD	Australian dollar to U.S. dollar exchange rate
CAD	Canadian dollar to U.S. dollar exchange rate
INR	Indian rupee to U.S. dollar exchange rate
reer_AUD	Real effective exchange rate based on manufacturing Consumer Price Index for Australia
reer_CAD	Real effective exchange rates based on manufacturing Consumer Price Index for Canada
reer_INR	Real effective exchange rates based on manufacturing Consumer Price Index for India
reer_US	Real effective exchange rates based on manufacturing Consumer Price Index for U.S.
tb_US	U.S. trade balance
GSCI	S&P GSCI Commodity Total Return Index
oi_USD	Dollar open interest
t_ind	Working's dollar T-index
VXO	VXO index (implied volatility based on 30-day S&P 100 index at-the-money options)
GPR	Global Geopolitical Risk Index (The Benchmark GPR Index)
stocks_US	S&P 500 Index
stocks_World	MSCI WORLD for developed markets index
stocks_G7	MSCI G7 index
stocks_EU	MSCI EU index
stocks_EM	MSCI EM for emerging markets index
stocks_CN	Hang Seng Index and Shanghai Composite Index glued and rescaled (in December 1990)
ts_BRICS	The share of BRIC countries trade in the total global trade
li_US	Leading indicator for U.S.
li_G7	Leading indicator for G7 countries
li_EU	Leading indicator for the euro area
li_CN	Leading indicator for China

The short-term interest rate was taken as the U.S. 3-month treasury bill rate on the secondary market (FRED 2022). The long-term interest rate was measured using the 10-year government bond yields for the U.S. and the Euro area (FRED 2022; OECD 2022). The term spread was measured as the difference between the U.S. long-term and U.S. short-term interest rates. The default return spread was computed as the difference between the Moody's seasoned Aaa corporate bond yield, based on bonds with maturities of 20 years and above (FRED 2022; Moody's 2022), and the short-term interest rate (understood as above).

Inflation was measured using the U.S. Consumer Price Index for all urban consumers and the U.S. Producer Price Index (FRED 2022). Following, for instance, Nonejad (2020), both of these indices can be important explanatory variables. They were transformed into logarithmic differences. Additionally, the U.S. average hourly earnings of production and nonsupervisory employees were taken (FRED 2022). They were also transformed into logarithmic differences.



U.S. money stocks (both real M1 and real M2, deflated by the U.S. Consumer Price Index) were also taken (FRED 2022). These time-series were transformed into logarithmic differences.

Industrial production growth was taken and measured using logarithmic differences of the U.S. industrial production (FRED 2022). Furthermore, the U.S. unemployment rate was taken (FRED 2022). The economic growth was measured using the commonly used (if monthly frequency is desired) Kilian's Index of Global Real Economic Activity (FRED 2022; Kilian 2009, 2019; Kilian and Zhou 2018). Despite some recent concerns, this index is, indeed, a valid and useful measure (Funashima 2020).

Following, for instance, Cuaresma et al. (2018, 2021), leading indicators were used (CLI, amplitude adjusted, except for China, for which the normalised index was taken due to data availability). The U.S., G7, Euro area, and China were considered (OECD 2022).

The trade balance (i.e., exports less imports) for the U.S. was computed in U.S. dollars (United Nations Statistics Division 2022). Due to the existence of seasonal patterns, 12-month differences were taken. Furthermore, the share of BRIC (Brazil, Russia, India, and China) countries' trade in the total global trade was computed (United Nations Statistics Division 2022), as these countries are important players on commodities markets (Ghoshray and Pundit 2021). Similarly, 12-month differences were taken.

Exchange rates impacts were measured using real effective exchange rates based on the manufacturing Consumer Price Index for Australia, Canada, India, and the U.S. Furthermore, the Australian dollar to U.S. dollar exchange rate was taken, as well as the Canadian dollar to U.S. dollar and Indian rupee to U.S. dollar exchange rates (Stooq 2022; OECD 2022; FRED 2022). The selection of countries was made with a focus on the largest exporters and importers of commodities and to include so-called "commodities currencies". Indeed, according to the WTO (2022), amongst the largest commodities exporters and importers in 2020 and 2019 were Australia, Brazil, Canada, China, Germany, India, Japan, Russia, the United Arab Emirates, and the U.S. A similar set of variables was used by Cuaresma et al. (2018, 2021), Gargano and Timmermann (2014), Chen et al. (2010), Clements and Fry (2008), and Cashin et al. (2004). These variables were transformed into logarithmic differences.

Open interest data were taken from the Commodity Futures Trading Commission (2022). Futures-only based data were used. This data set required some cleaning, for instance, due to overlapping commodities codes in some cases. Secondly, contracts are listed in various quantities, which needs to be considered in the computation of the dollar open interest representing the capital engaged. Finally, open interest data must be consistent with price time-series data (The World Bank 2022; Hong and Yogo 2012; Shilling 1996). In particular, each contract was aggregated to its monthly average, and then the sum of contracts of all types in a month was taken. The obtained time-series were transformed into logarithmic differences. Furthermore, Working's dollar T-index was also computed (Working 1960). This index measures the excess of speculative and hedging positions (Buyuksahin and Robe 2014). In particular, if  $CL > CS$ , where CL denotes long positions of commercial traders and CS denotes short positions of commercial traders, then  $T = 1 + NCS / (CL + CS)$ , where NCS denotes short positions of non-commercial traders. In the opposite case,  $T = 1 + NCL / (CL + CS)$ , where NCL denotes long positions of non-commercial traders. Non-commercial traders are perceived as a source of speculation, whereas commercial ones are perceived as a source of hedging activities.

Market stress was measured using the VXO index (CBOE 2022; FRED 2022). This index is a measure of implied volatility computed with the 30-day S&P 100 index at-the-money options. The currently more popular VIX index was not chosen because this new volatility index has only been reported and computed since 1990. Additionally, the global Geopolitical Risk Index (The Benchmark GPR Index) was taken. It is based on counting the occurrence of words related to geopolitical tensions in 11 leading newspapers (Caldara and Iacoviello 2022a, 2022b).

The overall behaviour of prices of a wide basket of various commodities was proxied using the S&P GSCI Commodity Total Return Index (Bloomberg 2022). It was transformed into logarithmic differences. Indeed, this index is a common measure of general commodity price movements in the world economy. It is based on the principal physical commodity futures contracts. In other words, the returns are calculated on fully collateralised contracts with full reinvestment (i.e., ones in which the buyers and sellers of a contract make an additional investment in the underlying asset with a value equal to the futures price). It is also a broadly diversified (across the spectrum of commodities) composite index of commodity sector returns. As a result, it aims to represent realizable returns attainable in the commodities markets. In particular, it consists of 24 commodities: energy products, industrial metals, agricultural products, livestock products, and precious metals (Downes and Goodman 2018).

Stock price movements were measured using the S&P 500 Index (Stooq 2022). Additionally, in order to capture the developing economies' stock markets, the Hang Seng Index was taken before December 1990, and the Shanghai Composite Index afterwards (Stooq 2022). Indeed, for instance, China became the biggest oil importer in 2017, overtaking the position of the U.S., and the trend is going to continue (EIA 2022; Wang et al. 2018). Furthermore, the MSCI stock market indices were taken (MSCI 2022). In particular, the MSCI WORLD for developed markets, the MSCI G7 INDEX, and the MSCI EU were taken. Furthermore, the MSCI EM for emerging markets was taken. All stock market indices were transformed into logarithmic differences.

If not stated otherwise already, time-series were collected to represent the last observed monthly value, as this can lead to a better forecast accuracy than, for example, the use of the mean values from a given month. Furthermore, if time-series follow a random walk, then, by construction, the aggregated time-series derived from the original one (e.g., averages or sums) may not follow a random walk. These features were studied in detail in terms of oil prices by Benmoussa et al. (2020).

Finally, following Koop and Korobilis (2013), the variables were standardised. In other words, before inserting them into the modelling scheme, the mean was subtracted and the outcome was divided by the standard deviation. These statistics were estimated on the basis of the first 100 observations. As a result, the transformed time-series were approximately stationary, but forward-looking bias was omitted. Moreover, the obtained time-series had similar magnitudes, which is an important and helpful feature improving numerical estimations.

The descriptive statistics are presented in Table A2 in Appendix A. Augmented Dickey–Fuller (ADF), Phillips–Perron (PP), and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) stationarity tests outcomes are presented in Table A3 in Appendix A. Assuming a 10% significance level (but for the majority of variables even 5% would be enough), most variables, with few exceptions, can be assumed to be stationary. Nevertheless, for example, *gea* is stationary by construction. This observed discrepancy is because, herein, the time-series covering a long-term period was trimmed to the shortened period of the analysis.

#### 4. Methodology

Numerical computations were performed in R (R Core Team 2018) and Python (Van Rossum and Drake 1995). Furthermore, a few packages and libraries were very useful in this regard, i.e., “NumPy”, “pandas”, and “SciPy” (Harris et al. 2020; The Pandas Development Team 2020; McKinney 2010).

##### 4.1. Bayesian Symbolic Regression

Bayesian symbolic regression (BSR) was introduced by Jin et al. (2019) and implemented by Jin (2021). This novel approach to symbolic regression aims to overcome certain difficulties (Korns 2011) with incorporating prior knowledge to genetic programming, deals with complexity issues in outcomes expressions, and improves interpretability of the outcomes

Herein, two approaches were applied. First, the estimation of parameters (for example, regression coefficients) was performed over some in-sample period (i.e., first 100 observations), kept fixed, and then applied to out-of-sample forecasting. Second, the in-sample period was recursively expanded, and in each step a forecast for just one period ahead was made. In other words, BSR forecast for the period  $t + 1$  was estimated based on the explanatory variables data set available up to period  $t$ . Next, BSR forecast for the period  $t + 2$  was estimated based on the data set expanded with the new data, i.e., the one available up to the period  $t + 1$ , etc. Such a recursive implementation resembles the real-life market situation and perspective.

The second crucial aspect of BSR involves enhancing the understandability of the derived expressions. To achieve this goal, BSR strives to capture succinct yet informative signals, assuming their structure to be both linear and additive. The prior distributions describing these components are designed to control the complexity of the obtained expressions, which are represented using symbolic trees (Weiss 2014).

At the heart of BSR lies the utilisation of Markov chain Monte Carlo (MCMC) sampling. This technique is employed to draw samples of these symbolic trees from the posterior distribution. Despite its computational intensity, Jin et al. (2019) demonstrated that this approach can even enhance memory utilisation in comparison with the standard genetic programming methods for symbolic regression. Furthermore, simulations conducted by Jin et al. (2019) showed robustness of BSR across various parameter settings. Notably, the method exhibited an ability to enhance predictive accuracy when contrasted with conventional symbolic regression algorithms, specifically those founded on genetic programming principles.

Herein, only a short outlook on BSR is provided. The full description can be found in the original paper (Jin et al. 2019). Let  $y_t$  be the forecasted time series, i.e., the given commodity price (possibly transformed as described in the previous section). Let  $x_{1,t}, \dots, x_{n,t}$  be the explanatory time series (also possibly transformed). Then, it is assumed that  $y_t = \beta_0 + \beta_1 * f_1(x_{1,1,t-1}, \dots, x_{1,i,t-1}) + \dots + \beta_k * f_k(x_{k,1,t-1}, \dots, x_{k,i,t-1})$ , with  $x_{i,j,t}$  standing for those of explanatory variables (out of  $n$  possible ones) which are present in the  $i$ -th component expression, i.e.,  $f_i$ , with  $j = \{1, \dots, n\}$  and  $i = \{1, \dots, k\}$ . The number of components,  $k$ , is fixed and must be set up during the initial stage. Coefficients  $\beta_i$  are estimated with the ordinary least squares method. Jin et al. (2019) claimed that higher values of  $k$  lead to better forecast accuracy, but that this gain diminishes when  $k$  becomes large enough.

Each component expression  $f_i$  is represented by the symbolic tree constructed from operators (such as  $+$ ,  $*$ , and  $1/x$ , etc.). Nicolau and Agapitos (2021) and Keijzer (2004) claimed that the operator  $lt(x) = a * x + b$ , with  $a$  and  $b$  being some real numbers, can improve the set of construable expressions in a noticeable manner. Indeed, the set of operators must be specified during the initial stage of BSR. For this purpose, 6 sets of operators were considered, denoted by  $F = \{1, \dots, 6\}$ , and  $k = 10$  (i.e.,  $k = 10$  linear components, denoted by  $K = \{1, \dots, 10\}$ ) were considered. For each commodity, models for all combinations of  $F$  and  $K$  were estimated over the data consisting of first 100 observations (i.e., the in-sample period). Next, the combination minimising root-mean-square error (RMSE) was selected for further estimations. For robustness, the mean absolute error (MAE) and mean absolute scaled error (MASE) were analysed (Hyndman and Koehler 2006), but usually the conclusions were the same as those based on RMSE.

In particular,  $F = 1$  represents the set consisting of unary  $neg(x_{i,t}) = -x_{i,t}$  and binary  $add(x_{i,t}, x_{j,t}) = x_{i,t} + x_{j,t}$  operators.  $F = 2$  expands  $F = 1$  with unary  $square(x_{i,t}) = (x_{i,t})^2$ .  $F = 3$  expands  $F = 1$  with unary 12 periods back moving average, i.e.,  $ma12(x_{i,t}) = (x_{i,t} + \dots + x_{i,t-11})/12$ , and unary  $lag(x_{i,t}) = x_{i,t-1}$ .  $F = 4$  expands  $F = 2$  with binary  $mul(x_{i,t}, x_{j,t}) = x_{i,t} * x_{j,t}$ .  $F = 5$  expands  $F = 4$  with unary  $inv(x_{i,t}) = 1/x_{i,t}$ , unary  $cubic(x_{i,t}) = (x_{i,t})^3$ , unary  $sqrt(x_{i,t}) = \sqrt{x_{i,t}}$ , unary  $log(x_{i,t}) = \ln(|x_{i,t}|)$ , unary  $ma12$ , and unary  $lag$ .  $F = 6$  expands  $F = 1$  with the unary operator  $lt(x_{i,t}) = a * x_{i,t} + b$ , with  $a$  and  $b$  being some real numbers. For example, Yang et al. (2015b) concluded that narrowing to just simple operators can

save computational time, but does not weaken the power of symbolic regression, keeping it effective enough in discovering useful model structures. On the other hand, it can be interesting to consider operators representing some non-linear structures specific to financial and economic time-series.

The Bayesian approach is employed by considering the Bayesian inference over the symbolic trees. In particular, the Bayesian regression tree models of [Chipman et al. \(1998a, 1998b\)](#) were implemented, and the methods of [Hastie and Tibshirani \(2000\)](#). A symbolic tree is represented by  $g(\cdot; T, M, \theta)$ , with  $g$  being some function as above, i.e.,  $g = f_1 + \dots + f_k$ .  $T$  denotes the set of nodes,  $M$  denotes their features, and  $\theta$  denotes their parameters. Initially, uniform priors are taken as they correspond to equal probabilities of selecting possible operators and node features. A node feature determines whether the given node is a terminal one, extends to a one, single, or child node, or splits into some two child nodes. The probability that a given node is terminal is  $1 - \alpha(1 + d)^{-\beta}$ , with  $\alpha$  and  $\beta$  being some parameters and  $d$  being the depth of the node ([Jin et al. 2019](#)). Following [Jin et al. \(2019\)](#),  $\alpha = 0.4$  and  $\beta = -1$  were used. High values of  $\beta$  control depth of trees and  $\alpha$  controls the symmetric shape of the distribution. The priors for  $a$  and  $b$  of operators  $lt$  were Gaussian and centred around the identity function ([Jin et al. 2019](#)).

The prior-posterior inferences in BSR model were performed with the Metropolis-Hastings algorithm ([Green 1995](#); [Hastings 1970](#); [Metropolis et al. 1953](#)). It was implemented in such a way that the transition structure penalised high complexity of the outcomes. Following [Jin et al. \(2019\)](#),  $M = 50$  iterations were performed, as the simulations based on various data sets suggest that this is large enough to stabilize the structure of the sought expression ([Chen et al. 2016](#)).

Additionally, model averaging schemes were employed. In the basic BSR version, the outcome is taken from the last iteration. However, let  $y_1, \dots, y_{50}$  be the forecasts obtained from  $M = 50$  iterations. Let  $w_1, \dots, w_{50}$  be some weights (such that  $w_1 + \dots + w_{50} = 1$ ) ascribed to each of these forecasts. The weighted average forecast is defined as  $w_1 * y_1 + \dots + w_{50} * y_{50}$ . Following [Steel \(2020\)](#) and [Stock and Watson \(2004\)](#), two schemes were considered. The first one considers weights inversely proportional to the mean-squared errors (MSEs) of the component models. The second considers equal weights for the component models. In order to sum up to 1, the initial weights were normalised (i.e., divided by the sum of all the individual weights).

The weights constructed in the above way, except forecasting, can be used to construct relative variable importance (RVI). In particular, after the mentioned rescaling, they sum up to 1. One can sum up the weights of exactly those models which contain a given explanatory variable. Such a sum defines the RVI of this variable. It can be used as some rough measure of the importance of a variable as the commodity price predictor ([Burnham and Anderson 2002](#)). Of course, RVI is a number between 0 and 1 by construction. In case of model selection schemes, one can simply indicate just whether a given explanatory variable is present or is not present in the selected individual model. Moreover, weighted average coefficients can be constructed. In particular, if a model averaging scheme would be narrowed only to linear component models (this is for a given explanatory variable to be used in all component models in exactly the same functional form), then  $w_1 * \theta_1 + \dots + w_{50} * \theta_{50}$  can be considered, with  $\theta_i$  being the regression coefficient corresponding to the given explanatory variable in the  $i$ -th component model ([Drachal 2020](#); [Banner and Higgs 2016](#); [Cade 2015](#); [Burnham and Anderson 2002](#)).

#### 4.2. Benchmark Models

BSR forecasts were compared with some alternative models. Of course, the standard symbolic regression with genetic programming ([Stephens 2021](#); [Koza 1998](#)) was employed. Due to computational issues, the population size was taken as 50 and generations were reduced to 10. Earlier pre-simulations with some selected commodities' time-series indicated that these numbers were high enough and there was no significant gain in forecast accuracy from taking higher values. On the other hand, lower values reduce the computational time.



The cross-over probability 0.95 was taken; subtree, hoist, and point mutations probabilities were set up at 0.01. RMSE was applied as a metric. All in all, these are quite standard and common specifications (Stephens 2021; Hassanat et al. 2019; Fuad and Hussain 2015). The set of operators indicated by BSR model was considered for each of the commodity. (Indeed, an estimation of the recursive BSR model for the applied data took approximately 1.6 h on average; whereas, for example, Dynamic Model Averaging with Occam window for the same data took approximately 20 min only.)

Bayesian model combination schemes were also considered (Raftery et al. 2010). In particular, Dynamic Model Averaging (DMA) and Bayesian Model Averaging (BMA). Following Onorante and Raftery (2016), a dynamic Occam window was applied because of the large number of explanatory variables. In particular, the cut-off limit was set at 0.25 and the number of models in the combination scheme was reduced to 100 (Drachal 2020). Dynamic Model Selection (DMS) and Bayesian Model Selection (BMS) schemes were also estimated. Following Koop and Korobilis (2012), an exponentially weighted moving average method with the parameter  $\kappa = 0.97$  was used to update the state-space equation variance. Furthermore, the mentioned dynamic Occam window, also averaging over models with exactly one explanatory variable, was performed (Drachal 2020). DMA, as described by Raftery et al. (2010), involves averaging over some time-varying parameters regressions. In particular, the time-varying parameters regression (with all 39 explanatory variables) is a special case of DMA, so it is reasonable to consider it as an additional benchmark model. Two versions were considered: one with a forgetting factor equal to 1 (i.e., no forgetting), and one with the (standard recommendation) forgetting factor equal to 0.99 (Raftery et al. 2010).

Additionally, LASSO and RIDGE regressions were estimated in a recursive way (Friedman et al. 2010). The  $\lambda$  parameter was separately selected in each recursive step, with t-fold cross-validation using MSE measure, where t is the time period. Elastic net regression was also employed. The following mixing parameters {0.1, 0.2, ..., 0.9} were used. Moreover, Bayesian versions of LASSO and RIDGE regressions were estimated (Gramacy 2019).

Finally, the least-angle regression (LARS) was estimated (Hastie and Efron 2013). Similarly, as before, t-fold cross-validation with MSE was used.

Furthermore, some commonly used models were also employed. In particular, these were the ARIMA model (in a recursive way), the no-change (NAÏVE) method, and the historical average. The number of lags for the ARIMA models was specified with the automatic procedure described by Hyndman and Khandakar (2008).

The list of all estimated models is reported in Table 3.

**Table 3.** Estimated models.

Abbreviation	Description
BSR rec	Bayesian symbolic regression (recursive)
BSR av MSE rec	Bayesian symbolic regression (recursive) with averaging and weights inversely proportional to MSE
BSR av EW rec	Bayesian symbolic regression (recursive) with equal weights
GP rec	Symbolic regression with genetic programming (recursive)
BSR fix	Bayesian symbolic regression (fixed parameters)
BSR av MSE fix	Bayesian symbolic regression (fixed parameters) with averaging and weights inversely proportional to MSE
BSR av EW fix	Bayesian symbolic regression (fixed parameters) with equal weights
GP fix	Symbolic regression with genetic programming (fixed parameters)
DMA	Dynamic Model Averaging with Occam window
BMA	Bayesian Model Averaging with Occam window
DMA 1V	Dynamic Model Averaging over one-variable component models

Table 3. Cont.

Abbreviation	Description
DMS 1V	Dynamic Model Selection over one-variable component models
BMA 1V	Bayesian Model Averaging over one-variable component models
BMS 1V	Bayesian Model Selection over one-variable component models
LASSO	LASSO regression (recursive)
RIDGE	RIDGE regression (recursive)
EN	Elastic net regression (recursive)
B-LASSO	Bayesian LASSO regression (recursive)
B-RIDGRE	Bayesian RIDGE regression (recursive)
LARS	Least-angle regression
TVP	Time-Varying Parameters regression with forgetting factor equal to 1
TVP f	Time-Varying Parameters regression with forgetting factor equal to 0.99
ARIMA	Automatic ARIMA (recursive)
HA	Historical average
NAÏVE	No-change method

#### 4.3. Forecast Evaluation

The forecast accuracy was evaluated with nRMSE (normalised root-mean-square error) and RMSE. In case of comparison between various models for a given commodity, RMSE was used. However, for comparisons between various commodities, nRMSE was used. Additionally, mean absolute error (MAE) and mean absolute scaled error (MASE) were considered (Hyndman and Koehler 2006). nRMSE is understood as RMSE divided by the mean of the analysed time-series.

When two forecasts from two competing models were compared, it was performed with the Diebold–Mariano test (Diebold and Mariano 1995) with Harvey et al. (1997) modification. Multiple forecasts were evaluated with the Model Confidence Set (MCS) of Hansen et al. (2011). In order to be consistent with RMSE measure, the squared errors loss functions were used in these tests (Bernardi and Catania 2018).

However, these tests evaluate a forecast's behaviour over the whole analysed period. On the other hand, the relative forecast's accuracies may vary over time. The Giacomini and Rossi (2010) fluctuation test deals with this issue. As before, squared errors loss function was applied. For the rolling procedure, the parameter  $\mu = 0.3$  was used, which corresponds to approximately 7.5-year periods.

## 5. Results

As 39 explanatory variables and 56 commodities were analysed, this section is divided into sub-sections for reasons of clarity.

### 5.1. Forecast Accuracy—Measures

Table 4 reports the nRMSE of various estimated models. It can be seen that in few cases the methods based on symbolic regression resulted in very high errors. GP fix, GP rec, and BSR rec were most robust against generating such outlier results. ARIMA and DMA were the methods that most often minimised the nRMSE. In this regard, it should be noticed that BMA is a special case of DMA (Raftery et al. 2010). The methods based on symbolic regression rarely minimised the nRMSE. These conclusions are more or less the same when RMSE, MAE, or MASE are considered (not reported herein).

Table 4. nRMSE of the estimated models.

	BSR rec	BSR av MSE rec	BSR av EW rec	GP rec	BSR fix	BSR av MSE fix	BSR av EW fix	GP fix	DMA	BMA	DMA 1V	DMS 1V	BMA 1V	BMS 1V	LASSO	RIDGE	EN	B-LASSO	B-RIDGRE	LARS	TVP	TVP f	ARIMA	HA	NAIVE
Brent	0.0992	0.0848	0.0840	0.1099	0.0969	0.0824	0.0882	0.0924	0.0721	0.0738	0.0751	0.0749	0.0742	0.0742	0.0778	0.0815	0.0774	0.0761	0.0781	0.0774	0.0870	0.0875	0.0877	0.6498	0.0925
Dubai	0.0842	0.0761	0.0766	0.1097	0.1576	0.0980	0.0909	0.0908	0.0678	0.0686	0.0707	0.0706	0.0700	0.0700	0.0733	0.0781	0.0743	0.0725	0.0745	0.0815	0.0833	0.0834	0.0843	0.6647	0.0908
WTI	0.1113	1047.6466	0.0937	0.1041	0.1175	3264.9070	2771.7826	0.1352	0.0776	0.0755	0.0768	0.0766	0.0760	0.0760	0.0790	0.0837	0.0792	0.0794	0.0816	0.0798	0.0898	0.0928	0.0905	0.6029	0.0947
Coal_AU	0.1036	0.1102	0.1087	0.0955	0.1063	4014.9355	0.6056	0.1565	0.0921	0.0946	0.0994	0.1008	0.0998	0.0998	0.0926	0.0936	0.0925	0.0927	0.0919	0.0983	0.1038	0.1092	0.1027	0.5490	0.0994
Coal_ZA	0.0925	0.0808	0.0786	0.0851	0.0955	0.0893	0.0878	0.0839	0.0855	0.0871	0.0855	0.0851	0.0846	0.0838	0.0810	0.0822	0.0813	0.0815	0.0808	0.0826	0.0885	0.0869	0.0821	0.5423	0.0872
Gas_US	0.1841	0.1872	1.6213	0.4869	0.1821	41,824.3614	71,166.5600	1.1844	0.1742	0.1743	0.1790	0.1785	0.1754	0.1752	0.1780	0.1838	0.1786	0.1798	0.1816	0.1785	0.2033	0.2131	0.1854	0.5950	0.1812
Gas_EU	0.0874	0.1635	0.0862	0.0894	0.5944	17.9480	36,101.1264	0.2850	0.0751	0.0771	0.0806	0.0799	0.0804	0.0802	0.0755	0.0760	0.0755	0.0765	0.0767	0.0788	0.0825	0.0963	0.0756	0.5984	0.0839
Gas_JP	0.0607	6756.0010	6756.0010	0.0595	0.1355	13,511.9823	28,114.6024	0.1018	0.0546	0.0552	0.0567	0.0562	0.0585	0.0591	0.0875	0.0590	0.0589	0.0569	0.0583	0.0822	0.1010	0.1058	0.0552	0.5776	0.0586
Cocoa	0.0617	0.0587	0.0586	0.0704	0.1078	0.0612	0.0611	0.3008	0.0585	0.0574	0.0584	0.0613	0.0581	0.0582	0.0585	0.0594	0.0591	0.0580	0.0579	0.0585	0.0772	0.0849	0.0572	0.3769	0.0578
Coffee_Arabica	0.0886	0.0750	0.0740	0.0925	0.3330	0.1368	0.0809	0.0768	0.0738	0.0754	0.0755	0.0808	0.0751	0.0778	0.0758	0.0759	0.0757	0.0742	0.0741	0.0799	0.0858	0.0992	0.0740	0.3818	0.0741
Coffee_Robusta	0.7544	35,128.2980	60,851.9278	0.4188	49,684.2213	167,366.9005	182,580.9262	0.0692	0.0556	0.0597	0.0562	0.0583	0.0566	0.0572	0.0599	0.0576	0.0590	0.0560	0.0568	0.0629	0.0663	0.0656	0.0550	0.3341	0.0559
Tea_Colombo	0.0559	0.0530	0.0526	0.0906	0.0666	0.0642	0.0666	0.0602	0.0588	0.0502	0.0524	0.0535	0.0511	0.0515	0.0507	0.0518	0.0507	0.0512	0.0514	0.0509	0.1027	0.1349	0.0508	0.3743	0.0508
Tea_Kolkata	0.1283	0.1271	0.1260	0.1461	0.1384	0.1207	0.1206	0.1270	0.1510	0.1201	0.1270	0.1288	0.1269	0.1280	0.1266	0.1273	0.1269	0.1264	0.1271	0.1287	0.1518	0.6858	0.1215	0.2654	0.1270
Tea_Mombasa	0.0869	0.0676	0.0999	0.0772	0.0763	81,392.4276	130.0489	0.2088	0.0687	0.0675	0.0688	0.0695	0.0679	0.0686	0.0681	0.0685	0.0684	0.0681	0.0681	0.0729	0.2347	0.2229	0.0685	0.2842	0.0680
Coconut_oil	0.1117	0.0902	0.0899	0.1168	0.1088	0.0947	0.0954	0.0975	0.0879	0.0870	0.0890	0.0900	0.0890	0.0888	0.0897	0.0901	0.0901	0.0901	0.0906	0.0903	0.1012	0.1043	0.0895	0.4950	0.0910
Groundnuts	0.0665	0.0637	0.0643	0.0987	0.0775	0.0821	0.0694	0.1828	0.0582	0.0612	0.0624	0.0628	0.0633	0.0656	0.0631	0.0632	0.0632	0.0630	0.0628	0.0638	0.0790	0.0752	0.0596	0.3458	0.0629
Fish_meal	53.4523	0.1942	25.4774	53.3846	0.1465	106.7756	141.3178	0.0563	0.0563	0.0557	0.0560	0.0583	0.0555	0.0563	0.0556	0.0553	0.0555	0.0550	0.0550	0.0593	0.0627	0.0705	0.0570	0.5068	0.0550
Palm_oil	0.1380	0.1923	84.1373	0.0753	382.0911	188.1413	0.2427	0.1500	0.0636	0.0660	0.0663	0.0660	0.0668	0.0674	0.0679	0.0686	0.0693	0.0675	0.0682	0.0728	0.0780	0.0806	0.0634	0.4160	0.0685
Soybeans	0.0894	1.5004	12.5538	346.5396	0.0902	1.7229	219.1701	0.0620	0.0584	0.0577	0.0595	0.0607	0.0588	0.0597	0.0586	0.0588	0.0585	0.0589	0.0597	0.0590	0.0669	0.0687	0.0593	0.3699	0.0598
Soybean_oil	0.0558	0.0561	0.0560	0.0622	0.0590	0.0566	0.0563	0.2619	0.0522	0.0541	0.0543	0.0553	0.0554	0.0558	0.0556	0.0550	0.0550	0.0550	0.0554	0.0553	0.0622	0.0718	0.0526	0.4172	0.0568
Soybean_meal	0.0695	0.0615	0.0613	0.0773	0.0965	0.0676	0.0688	0.0664	0.0622	0.0622	0.0624	0.0634	0.0623	0.0634	0.0624	0.0622	0.0622	0.0620	0.0622	0.0634	0.0738	0.0905	0.0583	0.3990	0.0622
Maize	0.0752	0.0713	0.0710	0.1380	0.1018	0.0796	0.0822	0.2767	0.0711	0.0714	0.0720	0.0729	0.0722	0.0726	0.0713	0.0712	0.0714	0.0718	0.0718	0.0722	0.0784	0.0860	0.0718	0.4252	0.0725
Rice_5	0.0866	0.0786	0.0782	0.1094	0.1064	0.0882	0.0871	10.5696	0.0844	0.0774	0.0776	0.0782	0.0768	0.0783	0.0780	0.0774	0.0777	0.0767	0.0757	0.0873	0.0870	0.0914	0.0749	0.3829	0.0755
Rice_100	0.0878	0.0793	0.0780	0.0882	0.1555	0.0844	0.0864	0.0784	0.0827	0.0780	0.0788	0.0797	0.0781	0.0798	0.0819	0.0794	0.0800	0.0792	0.0800	0.0866	0.0889	0.1108	0.0793	0.4523	0.0768
Wheat_SRW	0.0823	0.0805	0.0806	0.1056	0.1118	0.0832	0.0828	0.0830	0.0809	0.0797	0.0812	0.0829	0.0810	0.0814	0.0809	0.0811	0.0808	0.0809	0.0809	0.0819	0.0885	0.0925	0.0795	0.3694	0.0804
Wheat_HRW	0.0781	0.0751	0.0750	0.0784	0.0895	0.0844	0.0926	0.0863	0.0775	0.0742	0.0753	0.0757	0.0751	0.0756	0.0754	0.0752	0.0752	0.0749	0.0752	0.0782	0.0824	0.0867	0.0744	0.3714	0.0746
Banana	0.0961	0.0947	0.0944	0.1677	0.1030	0.0976	0.1085	0.4457	0.0912	0.0917	0.0930	0.0972	0.0923	0.0938	0.0924	0.0918	0.0922	0.0922	0.0917	0.0922	0.1210	0.1249	0.0919	0.3993	0.0915
Orange	0.5287	77,708.5698	109,896.5186	0.2958	0.1121	495,972.8294	544,551.3422	747.4920	0.1114	0.1119	0.1133	0.1155	0.1127	0.1127	0.1136	0.1124	0.1138	0.1124	0.1121	0.1144	0.1358	0.1511	0.1106	0.3571	0.1120
Beef	0.0497	0.0485	0.0488	0.0507	0.1451	0.0506	0.0496	0.0859	0.0476	0.0478	0.0484	0.0486	0.0483	0.0485	0.0483	0.0487	0.0487	0.0487	0.0489	0.0494	0.0526	0.0550	0.0456	0.3768	0.0487
Chicken	0.0484	0.0465	0.0460	0.0483	0.0474	0.0455	0.0456	0.0465	0.0457	0.0454	0.0461	0.0463	0.0458	0.0457	0.0460	0.0462	0.0461	0.0465	0.0469	0.0469	0.0482	0.0521	0.0435	0.2619	0.0460
Shrimps	0.0445	0.0434	0.0432	0.0522	0.0476	0.0467	0.0478	0.0719	0.0421	0.0426	0.0429	0.0439	0.0431	0.0434	0.0429	0.0430	0.0430	0.0428	0.0429	0.0469	0.0535	0.0531	0.0392	0.1930	0.0428
Sugar_EU	0.0346	0.0323	0.0323	0.0401	0.0385	0.0345	0.0333	0.0784	0.0326	0.0322	0.0342	0.0339	0.0339	0.0339	0.0322	0.0321	0.0322	0.0328	0.0328	0.0331	0.0392	0.0387	0.0340	0.2482	0.0325
Sugar_US	0.0417	0.0413	0.0416	0.0422	0.0417	0.0415	0.0413	0.0419	0.0430	0.0406	0.0417	0.0422	0.0410	0.0408	0.0421	0.0416	0.0421	0.0417	0.0415	0.0435	0.0440	0.0444	0.0400	0.2106	0.0413
Sugar_World	162.3136	0.1653	0.2684	0.1072	195,035.8825	283,162.9708	516,035.5325	0.6007	0.0825	0.0818	0.0828	0.0866	0.0824	0.0828	0.0821	0.0826	0.0824	0.0822	0.0822	0.0819	0.0933	0.0981	0.0780	0.4136	0.0821
Tobacco	0.0242	0.0171	0.0171	0.0202	0.0458	0.0296	0.0255	0.0201	0.0167	0.0178	0.0166	0.0171	0.0171	0.0174	0.0186	0.0178	0.0184	0.0168	0.0171	0.0182	0.0216	0.0226	0.0162	0.2379	0.0165
Logs_CM	0.0342	0.0316	0.0318	0.0361	0.0384	0.0371	0.0370	0.0594	0.0320	0.0304	0.0320	0.0324	0.0318	0.0322	0.0320	0.0326	0.0322	0.0317	0.0321	0.0344	0.0368	0.0434	0.0317	0.2294	0.0328
Logs_MY	0.0661	458.1314	323.9475	0.1741	0.0370	0.1722	0.0354	0.0461	0.0319	0.0337	0.0332	0.0338	0.0333	0.0336	0.0334	0.0332	0.0334	0.0331	0.0330	0.0444	0.0428	0.0461	0.0308	0.2390	0.0330
Sawnwood	0.0271	0.0269	0.0269	0.1205	0.0427	0.0348	0.0306	0.1523	0.0244	0.0248	0.0260	0.0277	0.0260	0.0264	0.0263	0.0260	0.0261	0.0262	0.0266	0.0268	0.0352	0.0352	0.0259	0.2006	0.0258
Plywood	0.0284	0.0759	0.1058	0.0384	0.0389	112.6283	424.0129	0.0379	0.0219	0.0224	0.0227	0.0239	0.0229	0.0236	0.0233	0.0224	0.0230	0.0223	0.0223	0.0280	0.0310	0.0348	0.0222	0.1541	0.0222
Cotton	0.0844	0.0934	5.7991	0.0806	0.0889	116,060.2346	267.7386	0.1411	0.0778	0.0790	0.0782	0.0787	0.0778	0.0779	0.0808	0.0791	0.0807	0.0784	0.0789	0.0793	0.0816	0.0845	0.0638	0.3293	0.0769
Rubber	0.1118	0.0995	0.1003	0.1199	0.2225	0.1372	0.1098	0.1014	0.1043	0.0983	0.1022	0.1055	0.1013	0.1011	0.1022	0.1007	0.1014	0.1016	0.1027	0.1027	0.1103	0.1102	0.0995	0.6144	0.1016

Table 4. Cont.

	BSR rec	BSR av MSE rec	BSR av EW rec	GP rec	BSR fix	BSR av MSE fix	BSR av EW fix	GP fix	DMA	BMA	DMA 1V	DMS 1V	BMA 1V	BMS 1V	LASSO	RIDGE	EN	B- LASSO	B- RIDGRE	LARS	TVP	TVP f	ARIMA	HA	NAIVE
Phosphate_rock	0.2018	0.1923	0.1918	0.1943	0.1935	0.1930	0.1924	0.1915	0.2253	0.1951	0.1925	0.1925	0.1925	0.1925	0.1989	0.1954	0.1954	0.1989	0.2047	0.2331	0.2076	0.2502	0.2393	0.8601	0.1934
Dap	0.1093	250.0493	0.5955	0.1107	1.8048	532.8508	569.0750	0.2111	0.0998	0.1010	0.1011	0.1020	0.1019	0.1014	0.1002	0.0977	0.0972	0.0999	0.1010	0.1027	0.1024	0.1020	0.0942	0.5790	0.1042
Tsp	0.1079	0.1075	0.1059	0.1157	0.1399	0.1186	0.1192	0.1140	0.1024	0.1137	0.1102	0.1099	0.1139	0.1127	0.1067	0.1044	0.1053	0.1062	0.1058	0.1100	0.1110	0.1109	0.0947	0.6395	0.1147
Urea	0.1559	0.1424	0.1469	0.2334	0.2902	0.1812	0.2412	0.1426	0.1487	0.1445	0.1529	0.1504	0.1520	0.1537	0.1522	0.1523	0.1520	0.1510	0.1526	0.1533	0.1722	0.1762	0.1483	0.5829	0.1524
Potash	0.1239	0.1151	0.1149	0.1277	0.1266	0.1166	0.1153	0.1148	0.1170	0.1142	0.1151	0.1151	0.1145	0.1145	0.1153	0.1157	0.1161	0.1151	0.1147	0.1218	0.1165	0.1196	0.1271	0.6299	0.1142
Aluminium	0.0570	0.0471	0.0472	0.0631	0.0663	0.0646	0.0540	0.0515	0.0439	0.0456	0.0466	0.0463	0.0469	0.0480	0.0459	0.0468	0.0460	0.0463	0.0468	0.0461	0.0557	0.0568	0.0499	0.2351	0.0508
Iron	0.1112	0.1111	0.1106	0.1127	0.1147	0.1140	0.1139	0.1206	0.1186	0.1218	0.1190	0.1189	0.1201	0.1201	0.1149	0.1097	0.1115	0.1113	0.1108	0.1144	0.1621	0.1215	0.1126	0.6999	0.1148
Copper	0.0709	0.0692	0.0696	0.0765	0.1273	0.0852	0.0820	0.2231	0.0666	0.0675	0.0687	0.0707	0.0683	0.0679	0.0686	0.0695	0.0686	0.0686	0.0695	0.0722	0.0768	0.0797	0.0702	0.5706	0.0718
Lead	0.0899	0.0831	0.0836	0.0978	0.1803	0.1050	0.1081	0.0861	0.0847	0.0821	0.0839	0.0852	0.0830	0.0825	0.0846	0.0849	0.0842	0.0839	0.0852	0.0835	0.0933	0.0989	0.0818	0.5983	0.0845
Tin	0.0711	0.0711	0.0709	0.0743	0.0762	0.0755	0.0763	0.1300	0.0683	0.0703	0.0710	0.0716	0.0709	0.0712	0.0720	0.0711	0.0710	0.0707	0.0714	0.0717	0.0742	0.0813	0.0717	0.6120	0.0735
Nickel	0.1114	0.1106	0.1106	0.1297	0.1128	0.1168	0.1172	0.2589	0.1048	0.1105	0.1108	0.1124	0.1107	0.1107	0.1100	0.1106	0.1101	0.1097	0.1098	0.1125	0.1187	0.1150	0.1040	0.5837	0.1110
Zinc	0.0737	0.0751	0.0754	0.0814	0.1548	0.0786	0.0792	0.2273	0.0743	0.0743	0.0742	0.0742	0.0738	0.0734	0.0768	0.0752	0.0769	0.0746	0.0750	0.0783	0.0877	0.0935	0.0738	0.4652	0.0761
Gold	0.0415	0.0418	0.0421	0.0422	0.0447	0.0443	0.0436	0.0584	0.0413	0.0406	0.0412	0.0427	0.0418	0.0424	0.0464	0.0467	0.0464	0.0413	0.0419	0.0448	0.0522	0.0490	0.0421	0.6476	0.0418
Platinum	0.0819	0.0633	0.0628	0.0644	0.0949	0.0676	0.0689	0.0664	0.0656	0.0620	0.0639	0.0636	0.0642	0.0643	0.0625	0.0626	0.0628	0.0628	0.0629	0.0655	0.0712	0.0714	0.0664	0.5191	0.0664
Silver	0.0926	1.0462	0.1473	0.0895	0.2114	21,752.5782	8168.0694	0.2819	0.0900	0.0890	0.0897	0.0890	0.0896	0.0895	0.0900	0.0901	0.0899	0.0897	0.0908	0.0907	0.1178	0.1043	0.0912	0.6978	0.0915



### 5.2. Forecast Accuracy—Testing

Table A4 in Appendix A reports the outcomes from the Diebold–Mariano test and which model, for each commodity, minimised the RMSE out of all considered models. As mentioned before, finding the model which would generate more accurate forecasts than the ARIMA or the no-change method is often a challenging task for commodity prices. The null hypothesis of the test is that the forecast accuracy of both methods is the same. The alternative is that the forecasts generated by the “best” model are more accurate than those from the competing model (ARIMA or NAÏVE). Assuming a 5% significance level, it can be concluded that only in 13% of cases some other method than the ARIMA method minimised the RMSE, and this difference was statistically significant. Changing the significance level to 10% increased this ratio to 18% of commodities. However, when the NAÏVE method was taken as the benchmark, this was 34% and 48% of commodities, respectively. In 64% of cases, the model minimising the RMSE was neither the ARIMA nor the NAÏVE method. Assuming a 10% significance level, it can be also concluded that for 11% of commodities the model minimising the RMSE was neither the ARIMA nor the NAÏVE method, and, moreover, the model generated statistically significantly more accurate forecasts than both the ARIMA and NAÏVE methods.

Table A5 in Appendix A reports the outcomes from the Diebold–Mariano test, in which forecasts generated using the BSR rec method are tested against those generated using the ARIMA and NAÏVE methods, and the method which minimised the RMSE (named, as previously, “best”). Here, the alternative hypothesis was that forecasts generated using the BSR rec model are less accurate than those of the competing model. The null hypothesis was that forecasts generated using both methods would have the same accuracy. Assuming a 5% significance level, in 46% cases it could not be concluded that the BSR rec model generated statistically significantly less accurate forecasts than the “best” method. In 64% of cases, the BSR rec forecasts could not be said to be significantly less accurate than those of the ARIMA method, and in 71% of cases than those of the NAÏVE method.

Table A6 in Appendix A presents outcomes from the similar test in Table A5, which was previously described, but the BSR rec is replaced by the GP rec. It can be seen that GP rec performed much worse than BSR rec, when considering the forecast accuracy. Only in 21% of cases can it not be said that GP rec generated statistically significantly less accurate forecasts than the “best” method, if a 5% significance level is assumed. If the competing model was ARIMA, then this was only in 32% of cases, and for NAÏVE this was in 36% of cases. In this regard, GP rec performed much worse than BSR rec, and the above conclusions can advocate the use of Bayesian methods in symbolic regression over genetic programming.

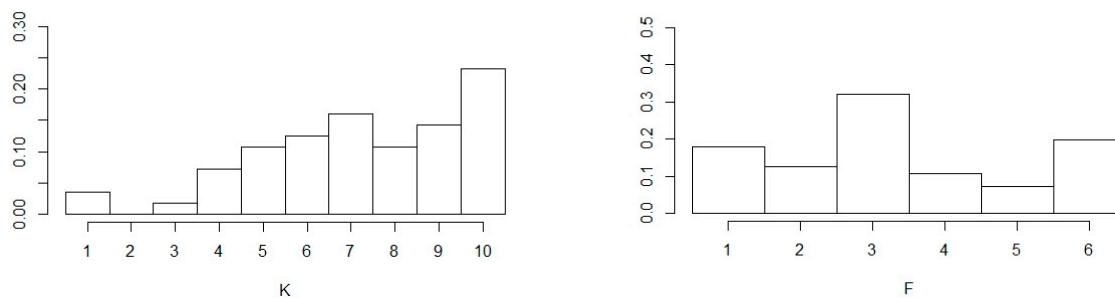
Table A7 in Appendix A reports the outcomes from the Diebold–Mariano test, in which forecasts generated using fixed versions of the selected models (“fix”) are compared with those generated using the recursive versions (“rec”). This was performed for BSR models (the original one and the two considered averaging schemes) and the GP model. The null hypothesis was that the “fix” and “rec” versions’ generated forecasts would have the same accuracy. The alternative hypothesis was that the “rec” version would generate a more accurate forecast than the “fix” version. Even assuming a 10% significance level, only in 1 case did BSR rec generate statistically significantly more accurate forecasts than BSR fix. However, if model averaging schemes are considered, then “rec” models seemed to improve the forecast accuracy more than the “fix” models. Assuming a 5% significance level, BSR av MSE rec generated statistically significantly more accurate forecasts than BSR av MSE fix for 57% of commodities. In the case of the BSR av EW scheme, this was for 68% of commodities. In the case of the GP method, this was for 45% of commodities.

Table A8 in Appendix A reports the outcomes from the Diebold–Mariano test, which compares forecasts generated using the BSR models with those generated using the GP models. The null hypothesis was that the BSR and GP models’ generated forecasts would have the same accuracy. The alternative hypothesis was that the BSR forecasts would be more accurate than the GP ones. Assuming a 5% significance level, BSR rec generated

statistically significantly more accurate forecasts than the GP rec method for 38% of commodities, and BSR fix generated statistically significantly more accurate forecasts than the GP fix method for 32% of commodities.

### 5.3. Selection of Parameters for BSR

Another considered problem was the selection of parameters for BSR. As mentioned before, it was performed on the basis of in-sample data. Figure 1 presents histograms of the selected K-s and F-s for BSR, if minimisation of RMSE is chosen as the criterion. For most commodities  $K = 10$ , i.e., the highest considered value, was selected. The second most often selected value was  $K = 7$ . Generally, higher values were preferred. Very small values, such as  $K = 1$  and  $K = 3$ , were rarely selected. For example,  $K = 2$  was never selected. The tendency to select high values of K-s may be due to an overfitting issue (which, by the way, was the background for developing LASSO and other model reduction methods). According to Jin et al. (2019), the improvement in forecast accuracy from an increasing K may not be significant if K is already high enough. Indeed, the MCS procedure reported further herein (and some pre-testing over the whole sample, not reported herein) confirmed this statement. Moreover, Jin et al. (2019) argued that if K is too large, then the regression coefficients in the linear combination are close to 0, making these extra components redundant.



**Figure 1.** The selected parameters for BSR.

In the case of the set of operators,  $F = 3$  was most often selected. This was the set consisting of simple operators (negation and addition) expanded with a 12-month moving average and 1st lag operator. In other words, these were the operators representing variable selection and transformations usually applied in economics and finance.

Nevertheless, just simple operators ( $F = 1$ ) or simple operators expanded with an “ln” operator ( $F = 6$ ) were also selected. However, sets consisting of operators representing non-linearities were not selected often. Similar conclusions were derived if MAE or MASE was considered (not reported herein). Nevertheless, this most often selected combination of parameters cannot be used as a general advice. The Diebold–Mariano test for forecasts generated (over the in-sample period) using the model with  $K = 10$  and  $F = 3$ , and the model indicated as that minimising RMSE, rejected the null hypothesis that both forecasts had the same accuracy, in favour of the alternative, that forecasts from the model minimising RMSE would be more accurate, for 32% of commodities (assuming a 5% significance level).

On the other hand, the selection of F seems to be less important. In particular, the Diebold–Mariano test was used to compare forecasts (over the in-sample period) obtained using the model with  $F = 1$  and the model with another F (with both models having the same K parameter). This test was performed for all commodities. As a result, there were  $5 \times 10 \times 56 = 2800$  pairwise comparisons. Assuming a 5% significance level, only in 6% of cases was the null hypothesis (that both forecasts have the same accuracy) rejected, and the alternative (that forecasts from the model with  $F = 1$  is less accurate) was assumed. This suggests that if the parameter K is properly chosen, then the set of operators is less important for the forecast accuracy, and the simple set of them can also lead to acceptable forecasts.

Additionally, for each commodity, forecasts generated using all BSR models with all K-s and F-s over the in-sample period were tested using the MCS procedure (with

1000 bootstrapped samples used to construct the statistic test, and with a “TR” statistic and quadratic loss function corresponding to RMSE, and with 90% confidence intervals). Indeed, there might be no statistically significant differences between the forecast accuracy from various BSR models with different K-s and F-s. The particular model, which was most often surviving the MCS procedure, was the one with K = 7 and F = 3, and the one with K = 9 and F = 3. In the case of the set of operators, this was consistent with previous outcomes, but for K a more moderate value is preferred.

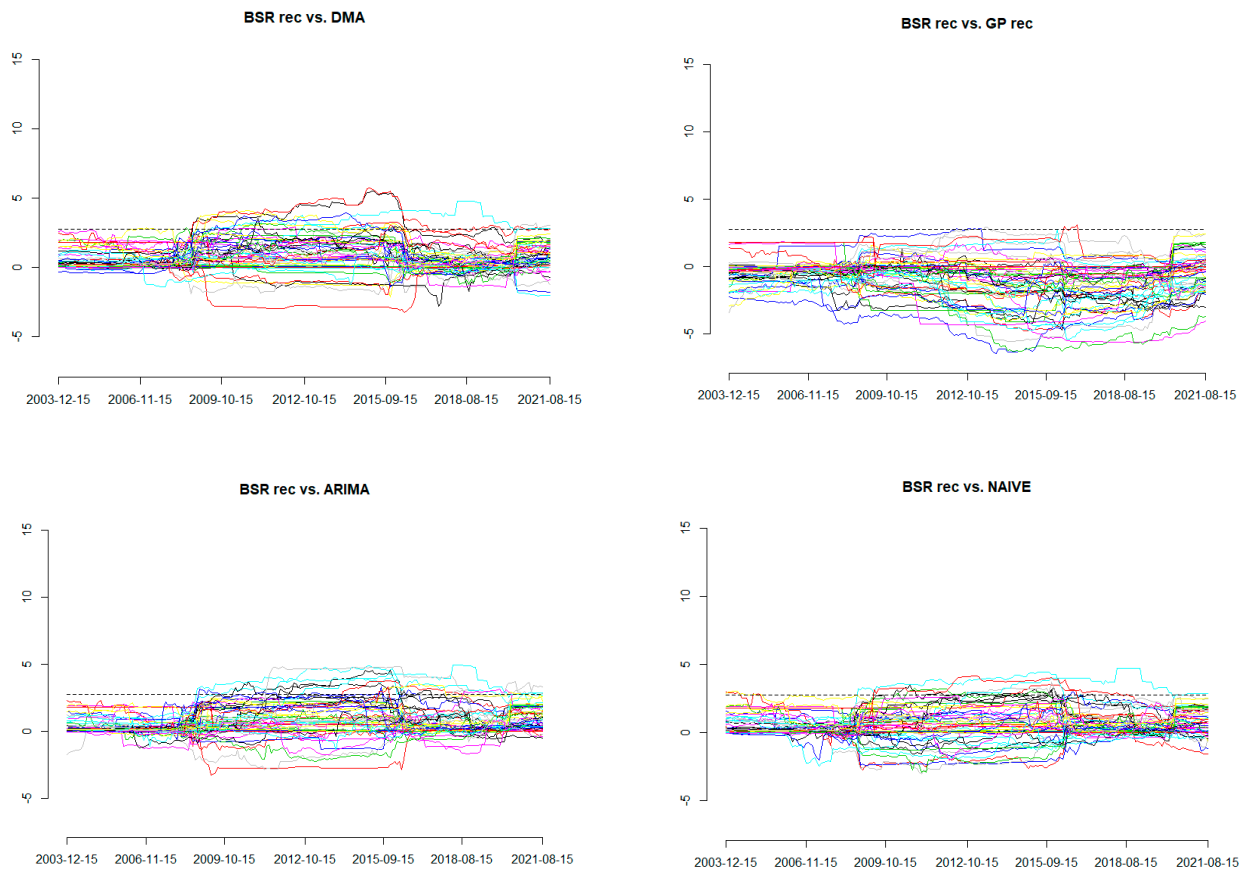
#### 5.4. Comparison of Models Performances

Another interesting piece of information derived from the conducted analysis was to see how, over the out-of-sample period, the MCS procedure selected models for each commodity. The same parameters for this procedure were set as previously. Table 5 presents how often (amongst all the analysed commodities) a given model survived the MCS procedure (with the parameters set as previously). It can be seen that ARIMA, DMA, and BMA were most often kept. These outcomes are consistent with those already reported herein. However, some types of BSR models were also kept for approximately 15% of commodities. Of course, the conventional benchmarks, such as ARIMA and NAÏVE, were found useful. DMA and its variations were also found useful. However, BSR-type models were found to be the next best ones, outperforming (in the sense of forecast accuracy), for example, LASSO and RIDGE regressions, and LARS and GP symbolic regression.

**Table 5.** Outcomes of the MCS procedure over out-of-sample period.

Model	Frequency
ARIMA	65.45%
DMA	60.00%
BMA	56.36%
NAÏVE	25.45%
BMS 1V	18.18%
BMA 1V	16.36%
BSR av MSE rec	14.55%
BSR av EW rec	14.55%
B-LASSO	14.55%
B-RIDGRE	14.55%
RIDGE	12.73%
EN	12.73%
BSR rec	10.91%
BSR av EW fix	10.91%
DMA 1V	10.91%
DMS 1V	10.91%
LASSO	9.09%
LARS	9.09%
GP fix	7.27%
GP rec	5.45%
BSR av MSE fix	5.45%
TVP	3.64%
TVP f	3.64%
BSR fix	0.00%
HA	0.00%

Furthermore, both the Diebold–Mariano test and the MCS procedure provide conclusions based on the whole analysed period (the whole out-of-sample period). The previously mentioned Giacomini–Rossi fluctuation test was performed over approximately 7.5-year periods ( $\mu = 0.3$ ). In particular, forecasts generated using the BSR rec model were tested against forecasts generated using the DMA, GP rec, ARIMA, and NAÏVE models. The null hypothesis was that the two models’ forecast performance would be the same, and the alternative was that the BSR rec model forecasts would be worse than the competing model. A 5% significance level was assumed. Figure 2 presents the outcomes for all analysed commodities.



**Figure 2.** The Giacomini–Rossi fluctuation test. (Different colors represent different commodities.).

In particular, test statistics for all analysed commodities are plotted, and the dotted line represents the critical value of the statistic. It can be seen that for the majority of commodities the null hypothesis could not be rejected for all competing models. However, in the cases of the DMA, ARIMA, and NAÏVE models, it could be rejected for some commodities, and the evidence was especially seen for the period between 2008 and 2016. In particular, when DMA is considered, the null hypothesis was rejected for Brent, Dubai, Gas\_EU, Gas\_JP, Coffee\_Arabica, Groundnuts, Soybean\_meal, Banana, Logs\_MY, Rubber, Potash, Aluminium, Copper, Lead, and Nickel. For GP rec, the null hypothesis was rejected for Coal\_AU, Coal\_ZA, Gas\_JP, Phosphate\_rock, and Silver. For ARIMA, the null hypothesis was rejected for Brent, Coal\_ZA, Gas\_JP, Cocoa, Coffee\_Arabica, Tea\_Colombo, Groundnuts, Soybean\_oil, Soybean\_meal, Chicken, Shrimps, Logs\_CM, Logs\_MY, Cotton, Rubber, and Lead. For NAÏVE, the null hypothesis was rejected for Brent, Gas\_JP, Cocoa, Coffee\_Arabica, Tea\_Colombo, Soybean\_meal, Rice\_100, Wheat\_SRW, Banana, Shrimps, Logs\_MY, Rubber, Phosphate\_rock, and Potash.

### 5.5. Time-Varying Importance of Price Predictors

The regression coefficients obtained from various estimated models (i.e., DMA, BMA, DMA 1V, DMS 1V, BMA 1V, BMS 1V, LASSO, RIDGE, EN, B-LASSO, B-RIDGE, LARS, TVP, and TVP f) seemed to follow similar time paths (except for some small discrepancies in relatively few cases). In other words, the exact numerical values were different, but they seemed to rise or decline over time consistently with each other, as well as have similar signs. (To prevent the paper being too long, they are not reported in detail herein.) However, the behaviour over time of these coefficients seemed to differ for different commodities.

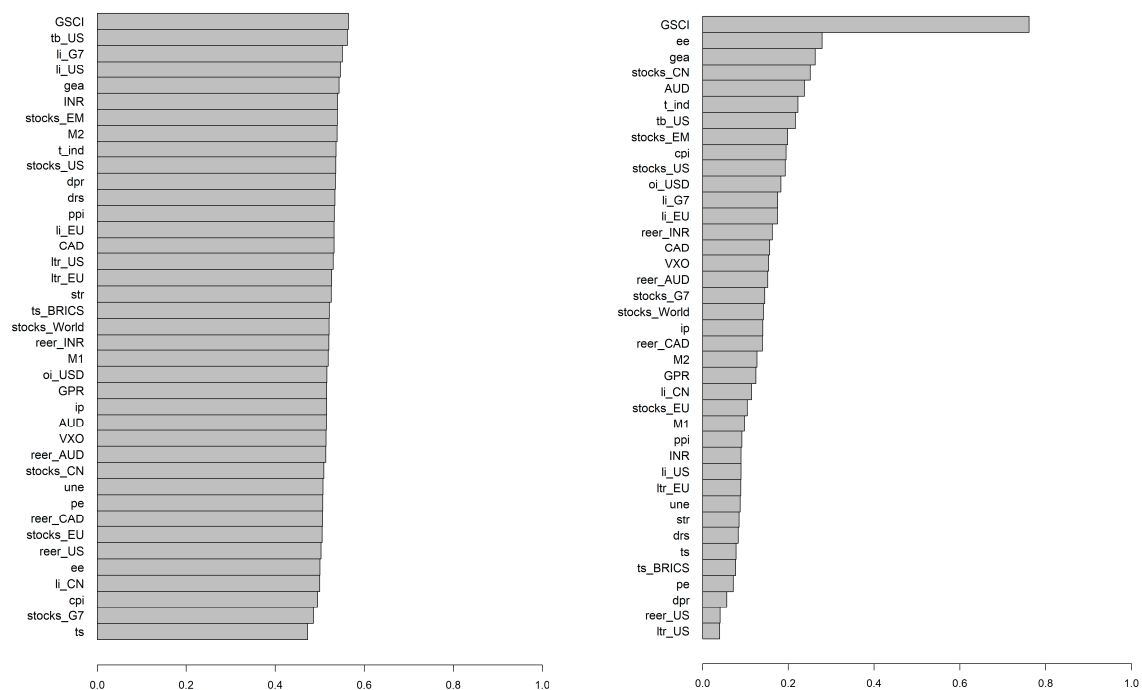
In order to keep the presentation clear and concise, only coefficients from the DMA model (i.e., the one that most often minimised the nRMSE and kept by the MCS procedure) are discussed in Appendix B.



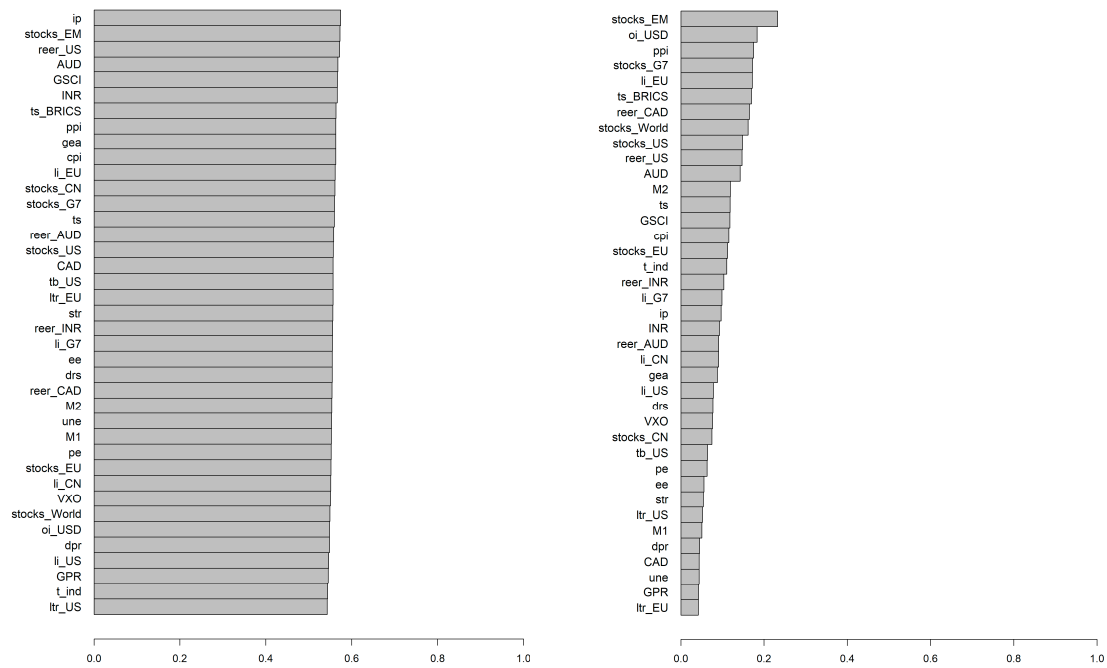
In order to detect the important impact of a given variable on a given commodity, the following procedure was performed (Burnham and Anderson 2002). First, it was assumed that the potential importance of a variable exists if the RVI exceeded 0.3 (Galipaud et al. 2014). Secondly, it was assumed that the absolute value of a regression coefficient (at a particular point of time) corresponding to this variable was greater than its standard deviation (computed over all values estimated for the out-of-sample period). If these two criteria were met in a given point of time, then it was assumed that the given variable has an important impact on a given commodity at that particular point of time. If the regression coefficient was positive in such a case, then it was assumed that this impact is positive; otherwise, it was considered negative.

### 5.6. Overall Importance of Price Predictors

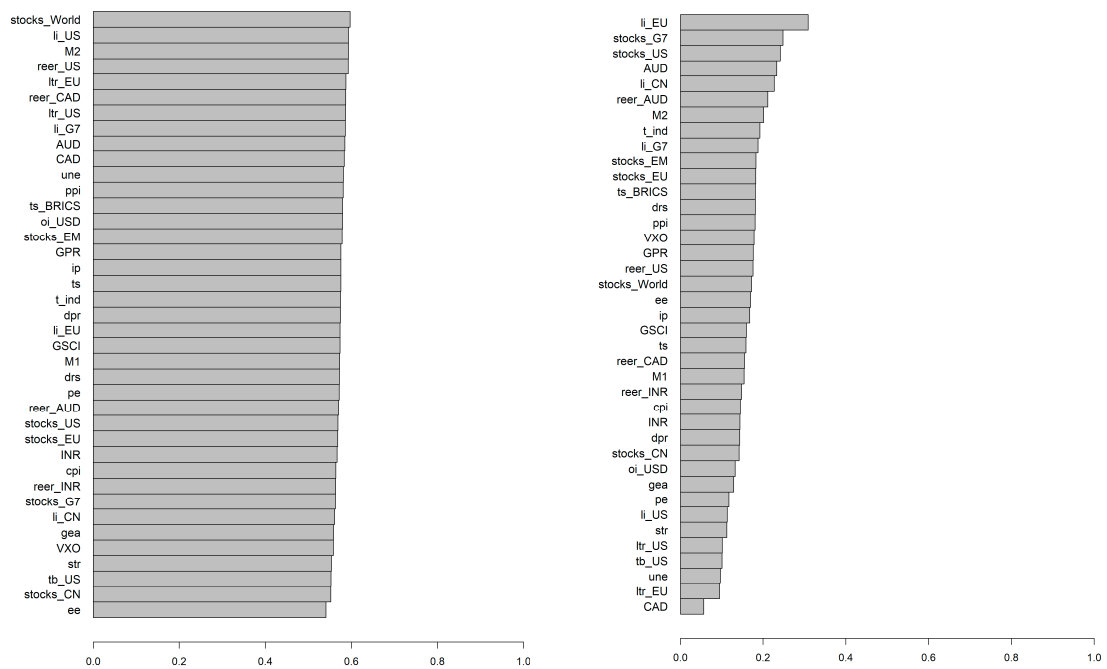
The analysed commodities were divided into four groups: energy, food, materials, and metals (i.e., Brent–Gas\_JP, Cocoa–Tobacco, Logs\_CM–Potash, and Aluminium–Silver, shown row-wise in Table A1 in Appendix A). Figures 3–6 show the frequency of how often a given explanatory variable was important in predicting a commodity's prices amongst these groups for the BSR av EW rec and DMA models. It can be seen that DMA was preferring more strongly certain variables, whereas BSR av EW rec was not strongly preferring any variable to such an extent. However, some conclusions can still be derived. In particular, for energy commodities, the GSCI seemed to be the most important price predictor. Interestingly, stocks\_CN was also quite often chosen, more often than other stock market indices and some financial indicators such as str, dpr, ts, and dpr. In the case of food commodities, those most often chosen were stocks\_EM and oi\_USD. Furthermore, ts\_BRICS was relatively more often chosen than, for example, some financial indicators and GPR. In the case of materials commodities, these were li\_EU, stocks\_G7, and stocks\_US; dpr, gea, and str were rarely chosen. In the case of metals commodities, these were stocks\_EM, stocks\_US, and stocks\_World; gea, t\_ind and dpr were rarely chosen.



**Figure 3.** Frequency of being an important price predictor amongst energy commodities (BSR av EW rec on the left and DMA on the right).



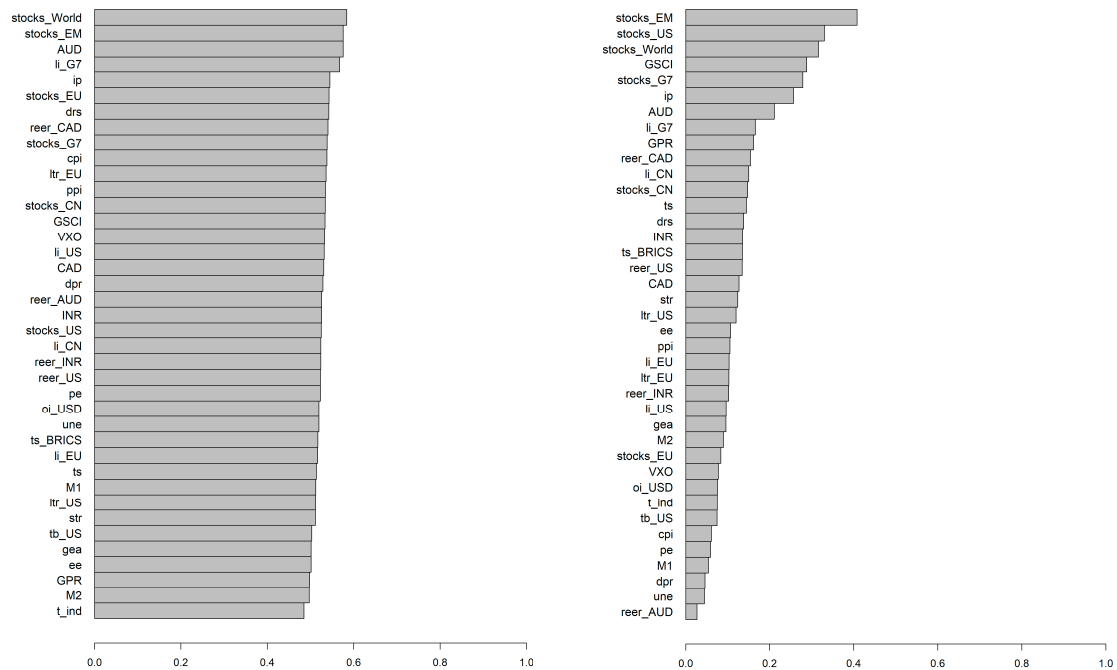
**Figure 4.** Frequency of being an important price predictor amongst food commodities (BSR av EW rec on the left and DMA on the right).



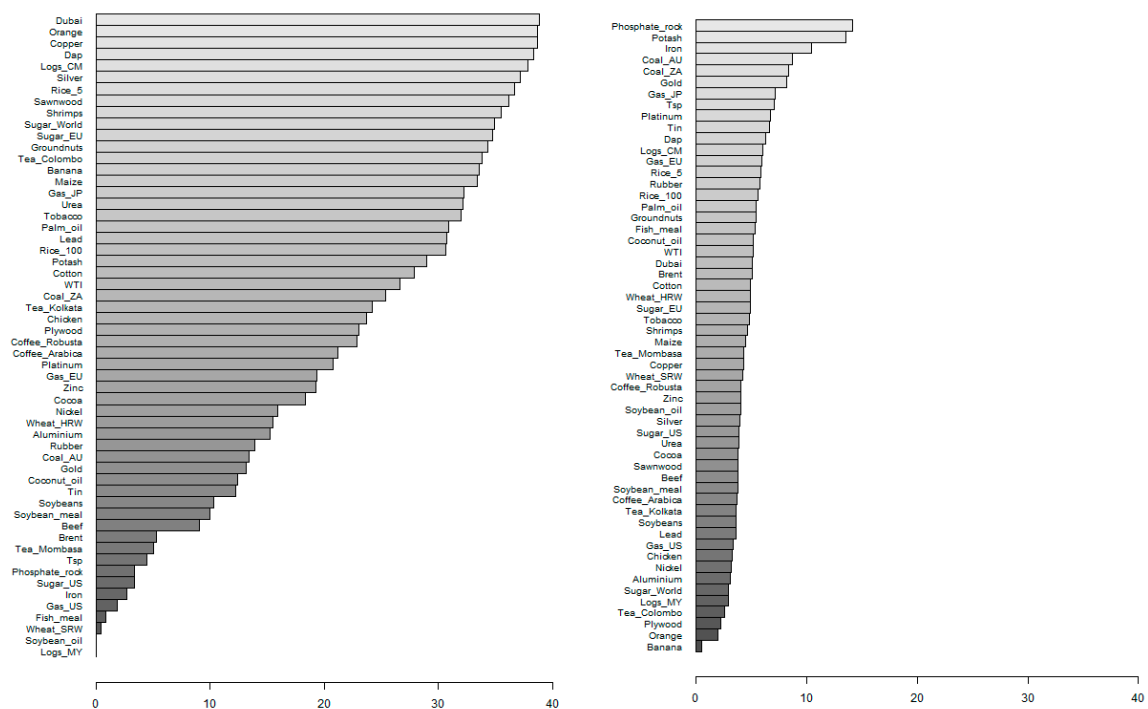
**Figure 5.** Frequency of being an important price predictor amongst materials commodities (BSR av EW rec on the left and DMA on the right).

Next, the already considered variable importances were further analysed. In particular, it was computed how often a given explanatory variable was important (according to the previously introduced two criteria in Section 5.5) for a given commodity over the out-of-sample period. Next, it was computed whether this frequency was over 50% (over the out-of-sample period). Figure 7 reports the number of explanatory variables for each commodity for which this condition was met. It can be seen that in the case of the BSR av EW rec method, Dubai, Orange, Copper, Dap, Logs\_CM, Silver, Rice\_5, Sawntwood, and Shrimps prices were impacted by a relatively high number of explanatory variables.

However, Brent, Tea\_Mombasa, Tsp, Phosphate\_rock, Sugar\_US, Iron, Gas\_US, Fish\_meal, Wheat\_SRW, Soybean\_oil, and Logs\_MY were impacted by relatively few explanatory variables. The BSR av EW rec method selected many more variables than the DMA method in general. In other words, the DMA method identified fewer variables as important price predictors. In particular, according to the DMA method, Phosphate\_rock, Potash, and Iron were affected by the highest number of predictors, whereas Tea\_Colombo, Plywood, Orange, and Banana were impacted by the smallest number of predictors.



**Figure 6.** Frequency of being an important price predictor amongst metals commodities (BSR av EW rec on the left and DMA on the right).



**Figure 7.** Number of important price predictors (BSR av EW rec on the left and DMA on the right).

## 6. Conclusions

In this study, a wide set of 56 commodities' prices were analysed and 39 potentially important explanatory variables were used to capture various market, financial, and economic influences. In addition to Bayesian symbolic regression, other methods dealing with variable uncertainty were considered, as well as some standard simple benchmark models. The obtained forecasts were evaluated with various measures and tested with various methods. In many cases, the developed econometric methods were able to generate more accurate forecasts than simple benchmark models.

Although no strong evidence was found that Bayesian symbolic regression itself outperforms many of the other methods, it was also not found to be significantly less accurate. However, the way that Bayesian symbolic regression selects explanatory variables was not able to clearly detect certain patterns in the time-varying importance of predictors, contrary to other developed econometric methods such as Dynamic Model Averaging.

In the case of comparing "fixed" vs. recursive computations with symbolic regression (Bayesian, the standard one based on genetic programming, and the ones including model averaging schemes), surprisingly recursive computations only in certain cases resulted in more accurate forecasts than "fixed" estimations. However, in a reasonable number of cases the recursive version of Bayesian symbolic regression generated significantly more accurate forecasts than the standard symbolic regression based on genetic programming. In the case of Bayesian symbolic regression, the previous hypothesis, that a higher number of linear components should be chosen if forecast accuracy is the aim, was confirmed. However, the selection of an initial set of operators (functions) was found to be a less important issue. For example, the set of operators capturing non-linear effects was not selected often. This might be due to the symbolic regression algorithm itself evolving functions into sufficiently complicated forms.

According to the model confidence set, Dynamic Model Averaging was the most often selected amongst all of the models dealing with the variable uncertainty problem; however, symbolic regression was still selected in a reasonable number of cases. In particular, the Bayesian version was chosen rather than the standard one based on genetic programming. The recursive version of Bayesian symbolic regression outperformed benchmark models for certain commodities, especially around the 2008–2016 period.

More detailed analysis showed that the set of important commodity price predictors differed amongst different groups of commodities (i.e., energy, food, materials, and metals). In particular, this evidence could be derived from the Dynamic Model Averaging analysis. Finally, there was a clear difference between the number of important explanatory variables indicated for each commodity using Bayesian symbolic regression and Dynamic Model Averaging. In particular, Dynamic Model Averaging was more restrictive in this manner.

This study confirmed that outperforming standard benchmark models, such as ARIMA, is still a challenging task when forecasting commodities' prices. It also confirmed the outcomes of some previous studies that Dynamic Model Averaging is a promising and high-performing method, outperforming, for example, LASSO, LARs, and RIDGE regressions. Nevertheless, Bayesian symbolic regression happened to also be an interesting and competing method with the standard symbolic regression based on genetic programming, and seems to be worth further studies and applications in economics and finance.

**Author Contributions:** Conceptualisation, K.D.; methodology, K.D.; software, K.D.; validation, K.D.; formal analysis, K.D.; investigation, K.D.; resources, K.D.; data curation, K.D.; writing—original draft preparation, K.D. and M.P.; writing—review and editing, K.D.; visualisation, 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:** Research funded by the grant of the National Science Centre, Poland, under the contract number DEC-2018/31/B/HS4/02021.

**Data Availability Statement:** The data presented in this study are available in repositories explicitly cited in the paper.

**Acknowledgments:** Research funded by the grant of the National Science Centre, Poland, under the contract number DEC-2018/31/B/HS4/02021.

**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 data; in the writing of the manuscript; or in the decision to publish the results.

## Appendix A

**Table A1.** The list of analysed commodities.

Abbreviation	Description
Brent	Brent oil
Dubai	Dubai oil
WTI	WTI oil
Coal_AU	Coal (Australia)
Coal_ZA	Coal (South Africa)
Gas_US	Gas (U.S.)
Gas_EU	Gas (Europe)
Gas_JP	Gas (Japan)
Cocoa	Cocoa
Coffee_Arabica	Coffee Arabic
Coffee_Robusta	Coffee Robusta
Tea_Colombo	Tea (Colombo)
Tea_Kolkata	Tea (Kolkata)
Tea_Mombasa	Tea (Mombasa)
Coconut_oil	Coconut oil
Groundnuts	Groundnuts
Fish_meal	Fish meal
Palm_oil	Palm oil
Soybeans	Soybeans
Soybean_oil	Soybean oil
Soybean_meal	Soybean meal
Maize	Maize
Rice_5	Rice 5% broken
Rice_100	Rice 100% broken
Wheat_SRW	U.S. soft red winter wheat
Wheat_HRW	U.S. hard red winter wheat
Banana	Banana
Orange	Orange
Beef	Beef
Chicken	Chicken
Shrimps	Shrimps
Sugar_EU	Sugar (Europe)
Sugar_US	Sugar (U.S.)
Sugar_World	Sugar (world)
Tobacco	Tobacco
Logs_CM	Logs (Cameroon)
Logs_MY	Logs (Malaysia)
Sawnwood	Sawnwood
Plywood	Plywood
Cotton	Cotton
Rubber	Rubber
Phosphate_rock	Phosphate rock
Dap	Diammonium phosphate
Tsp	Triple superphosphate
Urea	Urea
Potash	Potash
Aluminium	Aluminium
Iron	Iron ore
Copper	Copper



Table A1. Cont.

Abbreviation	Description
Lead	Lead
Tin	Tin
Nickel	Nickel
Zinc	Zinc
Gold	Gold
Platinum	Platinum
Silver	Silver

Table A2. Descriptive statistics.

Variable	Mean	Standard Deviation	Median	Min	Max	Skewness	Kurtosis
Brent	47.49	32.21	37.72	9.80	133.90	0.82	−0.48
Dubai	45.24	31.63	34.26	10.05	131.20	0.81	−0.52
WTI	46.09	28.70	37.77	11.31	133.90	0.77	−0.48
Coal_AU	58.10	31.02	47.70	22.25	180.00	1.10	0.68
Coal_ZA	54.67	29.46	46.62	21.25	167.80	0.95	0.20
Gas_US	3.55	2.14	2.84	1.19	13.52	1.75	3.68
Gas_EU	5.55	3.39	4.04	1.58	15.93	0.91	−0.23
Gas_JP	7.26	4.26	5.45	2.72	18.11	0.92	−0.32
Cocoa	1.92	0.70	1.69	0.86	3.53	0.48	−0.90
Coffee_Arabica	2.88	1.10	2.84	1.17	6.62	0.70	0.53
Coffee_Robusta	1.65	0.61	1.68	0.50	4.03	0.35	0.36
Tea_Colombo	2.35	0.86	1.96	1.18	4.27	0.40	−1.29
Tea_Kolkata	2.16	0.56	2.07	1.03	4.07	0.48	−0.29
Tea_Mombasa	1.96	0.53	1.83	1.12	3.39	0.67	−0.68
Coconut_oil	819.90	397.80	703.00	284.00	2256.00	1.05	0.46
Groundnuts	1192.00	405.40	1055.00	618.20	2528.00	1.37	1.65
Fish_meal	921.00	474.00	680.20	339.00	1926.00	0.37	−1.46
Palm_oil	618.10	252.60	576.60	234.00	1377.00	0.80	0.08
Soybeans	343.30	119.70	307.00	183.00	684.00	0.79	−0.36
Soybean_oil	707.50	283.80	626.00	286.90	1575.00	0.94	0.24
Soybean_meal	307.20	117.70	270.00	144.20	651.40	0.74	−0.46
Maize	148.80	59.99	124.40	75.27	333.10	1.27	0.91
Rice_5	354.20	126.90	321.20	163.80	907.00	0.89	0.93
Rice_100	288.20	126.40	232.00	120.80	762.70	0.68	−0.43
Wheat_SRW	178.60	61.97	162.30	85.30	419.60	0.92	0.45
Wheat_HRW	192.50	67.55	172.70	102.20	439.70	1.03	0.35
Banana	0.71	0.28	0.65	0.25	1.30	0.32	−1.18
Orange	0.67	0.23	0.64	0.23	1.43	0.60	−0.26
Beef	3.05	1.08	2.69	1.63	6.18	0.65	−0.65
Chicken	1.60	0.40	1.53	0.88	2.72	0.35	−0.83
Shrimps	12.09	2.14	11.88	7.50	19.25	0.64	0.43
Sugar_EU	0.54	0.12	0.55	0.34	0.78	−0.11	−1.23
Sugar_US	0.53	0.10	0.49	0.38	0.89	1.82	3.09
Sugar_World	0.28	0.11	0.26	0.11	0.65	1.01	0.88
Tobacco	3611.00	812.30	3400.00	2340.00	5118.00	0.31	−1.35

Table A2. Cont.

Variable	Mean	Standard Deviation	Median	Min	Max	Skewness	Kurtosis
Logs_CM	355.50	75.08	344.80	220.50	562.80	0.46	−0.56
Logs_MY	248.30	64.21	251.80	133.30	520.80	0.90	1.53
Sawnwood	694.20	144.20	713.30	374.10	973.60	−0.20	−0.91
Plywood	497.90	93.48	499.90	310.60	751.80	0.09	−0.67
Cotton	1.64	0.49	1.61	0.82	5.06	2.82	14.46
Rubber	1.63	0.98	1.40	0.49	6.26	1.66	3.36
Phosphate_rock	75.95	67.49	44.00	31.00	450.00	3.08	12.69
Dap	284.20	163.50	214.80	112.80	1076.00	1.87	4.93
Tsp	258.60	165.80	198.50	105.10	1132.00	2.36	7.90
Urea	202.80	118.20	185.80	62.75	785.00	1.46	3.31
Potash	204.80	129.90	151.20	83.00	682.50	1.46	1.61
Aluminium	1796.00	442.90	1731.00	1040.00	3578.00	0.82	0.34
Iron	67.32	48.40	37.90	24.30	214.40	1.18	0.31
Copper	4365.00	2472.00	3221.00	1377.00	10,160.00	0.45	−1.25
Lead	1279.00	780.00	935.50	375.70	3720.00	0.47	−1.12
Tin	12,030.00	7412.00	8144.00	3694.00	35,020.00	0.71	−0.64
Nickel	12,830.00	7342.00	11,170.00	3872.00	52,180.00	1.86	5.14
Zinc	1702.00	762.70	1528.00	747.60	4405.00	0.86	0.10
Gold	772.80	506.80	433.90	256.10	1969.00	0.66	−1.06
Platinum	844.10	442.30	809.80	341.20	2052.00	0.69	−0.58
Silver	11.81	8.52	7.03	3.65	42.70	1.15	0.68
dpr	−1.44	0.29	−1.50	−2.03	−0.76	0.34	−0.64
pe	25.58	6.77	25.41	13.32	44.20	0.68	0.22
str	2.85	2.46	2.38	−0.01	8.90	0.41	−1.00
ltr_US	4.58	2.21	4.46	0.62	9.36	0.29	−0.90
ltr_EU	4.83	3.03	4.25	−0.09	11.14	0.39	−0.78
ts	1.73	1.09	1.71	−0.49	3.78	0.03	−1.06
drs	3.08	1.32	3.16	0.24	6.10	−0.03	−1.14
cpi	192.80	42.58	191.60	116.20	273.10	−0.02	−1.24
ppi	157.20	35.17	150.20	104.80	233.40	0.18	−1.55
ip	87.46	14.13	92.35	60.59	104.20	−0.76	−0.90
ee	16.32	4.57	15.84	9.29	26.10	0.21	−1.12
M1	1091.00	1204.00	713.50	614.10	7230.00	4.32	17.86
M2	3680.00	1334.00	3335.00	2305.00	7636.00	1.01	0.25
gea	2.47	59.50	−4.90	−162.40	188.60	0.73	0.86
une	5.88	1.67	5.50	3.50	14.70	1.31	2.36
AUD	0.76	0.12	0.76	0.49	1.10	0.43	0.38
CAD	0.81	0.11	0.79	0.62	1.05	0.42	−0.69
INR	2.55	1.28	2.22	1.32	7.69	2.17	4.62
reer_AUD	95.05	12.43	96.64	71.32	123.90	0.22	−0.60
reer_CAD	106.10	12.08	102.90	85.69	128.40	0.20	−1.32
reer_INR	90.67	12.68	88.83	64.62	128.40	0.32	−0.08
reer_US	97.58	7.57	96.89	83.93	114.60	0.28	−0.99
tb_US	−44,570.00	25,260.00	−48,220.00	−97,680.00	−3492.00	0.10	−1.35

Table A2. Cont.

Variable	Mean	Standard Deviation	Median	Min	Max	Skewness	Kurtosis
GSCI	3525.00	1646.00	2990.00	1086.00	10,560.00	1.13	1.20
oi_USD	296,900,000,000.00	275,400,000,000.00	131,700,000,000.00	27,050,000,000.00	901,500,000,000.00	0.45	−1.43
t_ind	1.11	0.04	1.10	1.04	1.24	0.95	0.99
VXO	19.95	8.21	18.16	7.87	61.38	1.63	3.97
GPR	98.42	48.72	88.57	39.05	512.50	4.58	29.51
stocks_US	1330.00	859.80	1187.00	258.90	4523.00	1.17	1.28
stocks_World	1210.00	566.10	1149.00	423.10	3141.00	0.76	0.33
stocks_G7	1080.00	515.10	1022.00	384.30	2907.00	0.91	0.68
stocks_EU	307.50	116.20	336.30	96.69	582.30	−0.14	−0.91
stocks_EM	659.40	343.40	542.30	109.70	1376.00	0.19	−1.37
stocks_CN	44,140.00	27,480.00	40,970.00	2274.00	141,100.00	0.37	−0.26
ts_BRICS	0.00	0.01	0.00	−0.04	0.05	−0.68	3.74
li_US	99.77	1.33	99.91	92.31	102.20	−1.59	4.55
li_G7	99.86	1.22	100.10	92.26	102.10	−1.83	6.66
li_EU	100.00	1.68	100.10	90.44	103.20	−1.09	2.99
li_CN	99.97	1.41	100.10	85.68	103.10	−2.82	25.36

Table A3. Stationarity tests. (“&gt;” indicates “greater than”, and “&lt;” indicates “smaller than”).

Variable	ADF Stat.	ADF <i>p</i> -Val.	PP Stat.	PP <i>p</i> -Val.	KPSS Stat.	KPSS <i>p</i> -Val.
Brent	−7.9953	<0.01	−253.6471	<0.01	0.0446	>0.10
Dubai	−8.2440	<0.01	−230.5171	<0.01	0.0436	>0.10
WTI	−8.0483	<0.01	−248.0916	<0.01	0.0400	>0.10
Coal_AU	−6.5208	<0.01	−280.4958	<0.01	0.0801	>0.10
Coal_ZA	−6.1623	<0.01	−266.5725	<0.01	0.0538	>0.10
Gas_US	−8.7802	<0.01	−361.7846	<0.01	0.0348	>0.10
Gas_EU	−6.4891	<0.01	−277.7274	<0.01	0.0560	>0.10
Gas_JP	−6.7223	<0.01	−229.5060	<0.01	0.0636	>0.10
Cocoa	−7.3658	<0.01	−321.8202	<0.01	0.0906	>0.10
Coffee_Arabica	−6.9581	<0.01	−314.2563	<0.01	0.0674	>0.10
Coffee_Robusta	−6.2083	<0.01	−298.5168	<0.01	0.1047	>0.10
Tea_Colombo	−8.1779	<0.01	−348.0611	<0.01	0.0361	>0.10
Tea_Kolkata	−12.9817	<0.01	−282.5993	<0.01	0.0150	>0.10
Tea_Mombasa	−7.1632	<0.01	−313.6262	<0.01	0.0259	>0.10
Coconut_oil	−6.1340	<0.01	−323.9682	<0.01	0.0421	>0.10
Groundnuts	−7.0198	<0.01	−284.0919	<0.01	0.0233	>0.10
Fish_meal	−7.7946	<0.01	−275.7443	<0.01	0.0601	>0.10
Palm_oil	−7.1712	<0.01	−269.1340	<0.01	0.0642	>0.10
Soybeans	−7.6204	<0.01	−331.3422	<0.01	0.0596	>0.10
Soybean_oil	−6.7782	<0.01	−263.4829	<0.01	0.0627	>0.10
Soybean_meal	−8.0436	<0.01	−267.6929	<0.01	0.0425	>0.10
Maize	−7.6244	<0.01	−299.4328	<0.01	0.0389	>0.10
Rice_5	−8.7629	<0.01	−233.7177	<0.01	0.0488	>0.10
Rice_100	−7.7580	<0.01	−231.1146	<0.01	0.0521	>0.10
Wheat_SRW	−8.4273	<0.01	−309.6046	<0.01	0.0353	>0.10
Wheat_HRW	−7.7958	<0.01	−304.6156	<0.01	0.0413	>0.10
Banana	−11.8927	<0.01	−388.1070	<0.01	0.0161	>0.10
Orange	−11.8850	<0.01	−268.7407	<0.01	0.0232	>0.10
Beef	−7.7189	<0.01	−228.0717	<0.01	0.1321	>0.10
Chicken	−10.5157	<0.01	−260.6561	<0.01	0.0269	>0.10
Shrimps	−7.3627	<0.01	−196.4938	<0.01	0.0356	>0.10

Table A3. Cont.

Variable	ADF Stat.	ADF <i>p</i> -Val.	PP Stat.	PP <i>p</i> -Val.	KPSS Stat.	KPSS <i>p</i> -Val.
Sugar_EU	−7.4501	<0.01	−315.1929	<0.01	0.1137	>0.10
Sugar_US	−6.3475	<0.01	−282.7717	<0.01	0.0680	>0.10
Sugar_World	−7.3141	<0.01	−288.5904	<0.01	0.0413	>0.10
Tobacco	−5.0524	<0.01	−287.2056	<0.01	0.1256	>0.10
Logs_CM	−7.8160	<0.01	−292.3322	<0.01	0.0386	>0.10
Logs_MY	−7.6421	<0.01	−255.3297	<0.01	0.0395	>0.10
Sawnwood	−6.2682	<0.01	−333.6141	<0.01	0.0931	>0.10
Plywood	−7.3099	<0.01	−281.7184	<0.01	0.0430	>0.10
Cotton	−7.9502	<0.01	−194.0813	<0.01	0.0387	>0.10
Rubber	−6.4845	<0.01	−295.2020	<0.01	0.0616	>0.10
Phosphate_rock	−5.7667	<0.01	−381.9933	<0.01	0.0363	>0.10
Dap	−6.9540	<0.01	−203.3954	<0.01	0.0559	>0.10
Tsp	−7.3063	<0.01	−183.7721	<0.01	0.0515	>0.10
Urea	−8.1881	<0.01	−302.8459	<0.01	0.0324	>0.10
Potash	−6.2091	<0.01	−443.6662	<0.01	0.1158	>0.10
Aluminium	−7.3307	<0.01	−337.5590	<0.01	0.0697	>0.10
Iron	−6.3234	<0.01	−271.0289	<0.01	0.0596	>0.10
Copper	−8.1452	<0.01	−233.0640	<0.01	0.0946	>0.10
Lead	−6.5456	<0.01	−310.2747	<0.01	0.0784	>0.10
Tin	−6.8723	<0.01	−298.2957	<0.01	0.1397	>0.10
Nickel	−6.4269	<0.01	−249.7583	<0.01	0.0438	>0.10
Zinc	−6.3570	<0.01	−293.2702	<0.01	0.0385	>0.10
Gold	−6.6906	<0.01	−331.2990	<0.01	0.5126	0.0388
Platinum	−8.3714	<0.01	−306.8139	<0.01	0.0952	>0.10
Silver	−7.5172	<0.01	−299.9199	<0.01	0.1486	>0.10
dpr	−2.1808	0.5009	−6.7632	0.7318	1.8097	<0.01
pe	−1.8732	0.6309	−5.2990	0.8137	0.9871	<0.01
str	−3.5042	0.0423	−9.1229	0.5999	4.6993	<0.01
ltr_US	−4.0658	<0.01	−31.5982	<0.01	6.2708	<0.01
ltr_EU	−3.1425	0.0979	−13.4612	0.3572	6.0577	<0.01
ts	−3.2958	0.0718	−14.8652	0.2787	0.2490	>0.10
drs	−3.1972	0.0886	−11.4279	0.4710	0.2997	>0.10
cpi	−6.6089	<0.01	−204.7199	<0.01	0.8030	<0.01
ppi	−6.3383	<0.01	−249.5577	<0.01	0.0516	>0.10
ip	−6.5356	<0.01	−298.2689	<0.01	0.2861	>0.10
ee	−4.9675	<0.01	−393.8539	<0.01	0.2569	>0.10
M1	−6.9325	<0.01	−365.9665	<0.01	0.5329	0.0343
M2	−6.0468	<0.01	−156.3054	<0.01	1.3597	<0.01
gea	−2.4424	0.3905	−22.5739	0.0411	0.6736	0.0159
une	−2.5496	0.3451	−21.1600	0.0539	0.3989	0.0776
AUD	−7.0716	<0.01	−393.6667	<0.01	0.0588	>0.10
CAD	−7.0035	<0.01	−413.4559	<0.01	0.0908	>0.10
INR	−6.4130	<0.01	−364.1453	<0.01	0.4552	0.0534
reer_AUD	−7.9739	<0.01	−275.5769	<0.01	0.0580	>0.10
reer_CAD	−7.3475	<0.01	−307.1186	<0.01	0.0951	>0.10
reer_INR	−7.3033	<0.01	−325.3636	<0.01	0.5238	0.0363
reer_US	−7.6200	<0.01	−227.4906	<0.01	0.0742	>0.10
tb_US	−5.7462	<0.01	−103.8847	<0.01	0.1607	>0.10
GSCI	−7.0261	<0.01	−316.9545	<0.01	0.2585	>0.10
oi_USD	−6.8621	<0.01	−330.9185	<0.01	0.0850	>0.10
t_ind	−3.2137	0.0858	−36.2523	<0.01	3.5126	<0.01
VXO	−3.5502	0.0379	−64.6624	<0.01	0.2472	>0.10
GPR	−4.7256	<0.01	−103.1218	<0.01	0.1747	>0.10
stocks_US	−6.1733	<0.01	−394.0153	<0.01	0.1278	>0.10
stocks_World	−6.4726	<0.01	−374.9809	<0.01	0.0608	>0.10
stocks_G7	−6.4349	<0.01	−377.8616	<0.01	0.0720	>0.10
stocks_EU	−6.7591	<0.01	−375.6705	<0.01	0.0936	>0.10
stocks_EM	−7.2645	<0.01	−344.3724	<0.01	0.1365	>0.10

Table A3. Cont.

Variable	ADF Stat.	ADF <i>p</i> -Val.	PP Stat.	PP <i>p</i> -Val.	KPSS Stat.	KPSS <i>p</i> -Val.
stocks_CN	−6.5398	<0.01	−415.8805	<0.01	0.1967	>0.10
ts_BRICS	−4.9694	<0.01	−130.1868	<0.01	0.3909	0.0811
li_US	−4.9081	<0.01	−29.9440	<0.01	0.0800	>0.10
li_G7	−5.2672	<0.01	−30.2912	<0.01	0.0946	>0.10
li_EU	−5.4067	<0.01	−25.4211	0.0221	0.1003	>0.10
li_CN	−4.1257	<0.01	−82.3776	<0.01	0.1147	>0.10

Table A4. The Diebold–Mariano test—“best” model vs. ARIMA and vs. NAIVE.

Commodity	Best	Best vs. ARIMA	Best vs. NAIVE
Brent	DMA	0.0004	0.0002
Dubai	DMA	0.0001	0.0000
WTI	BMA	0.0004	0.0004
Coal_AU	B-RIDGRE	0.0729	0.0246
Coal_ZA	BSR av EW rec	0.1906	0.0006
Gas_US	DMA	0.0539	0.2162
Gas_EU	DMA	0.4469	0.0114
Gas_JP	DMA	0.4178	0.0538
Cocoa	ARIMA		0.1483
Coffee_Arabica	DMA	0.4657	0.4603
Coffee_Robusta	ARIMA		0.1928
Tea_Colombo	BMA	0.3237	0.3149
Tea_Kolkata	BMA	0.3972	0.1314
Tea_Mombasa	BMA	0.2629	0.3450
Coconut_oil	BMA	0.1535	0.0087
Groundnuts	DMA	0.2585	0.0245
Fish_meal	B-RIDGRE	0.2576	0.4941
Palm_oil	ARIMA		0.0091
Soybeans	BMA	0.1258	0.0784
Soybean_oil	DMA	0.4197	0.0387
Soybean_meal	ARIMA		0.0007
Maize	BSR av EW rec	0.2956	0.0795
Rice_5	ARIMA		0.4718
Rice_100	NAIVE	0.4002	
Wheat_SRW	ARIMA		0.3001
Wheat_HRW	BMA	0.4565	0.2907
Banana	DMA	0.4266	0.4133
Orange	ARIMA		0.3381
Beef	ARIMA		0.1138
Chicken	ARIMA		0.2135
Shrimps	ARIMA		0.0618
Sugar_EU	RIDGE	0.0884	0.3505
Sugar_US	ARIMA		0.1307
Sugar_World	ARIMA		0.0090
Tobacco	ARIMA		0.2881
Logs_CM	BMA	0.1138	0.0181
Logs_MY	ARIMA		0.0251
Sawnwood	DMA	0.0150	0.0824
Plywood	DMA	0.3120	0.2524
Cotton	ARIMA		0.0036
Rubber	BMA	0.4072	0.2832
Phosphate_rock	GP fix	0.0472	0.1994
Dap	ARIMA		0.1316
Tsp	ARIMA		0.0446
Urea	BSR av MSE rec	0.1435	0.0868
Potash	BMA	0.1253	0.4778
Aluminium	DMA	0.0020	0.0010



**Table A4.** *Cont.*

Commodity	Best	Best vs. ARIMA	Best vs. NAIVE
Iron	RIDGE	0.1258	0.0386
Copper	DMA	0.1459	0.0725
Lead	ARIMA		0.2022
Tin	DMA	0.1651	0.1013
Nickel	ARIMA		0.0693
Zinc	BMS 1V	0.4368	0.0314
Gold	BMA	0.1037	0.1845
Platinum	BMA	0.0013	0.0034
Silver	DMS 1V	0.2438	0.1263

**Table A5.** The Diebold–Mariano test—BSR rec model vs. “best”, vs. ARIMA, and vs. NAIVE.

Commodity	BSR Rec vs. Best	BSR Rec vs. ARIMA	BSR Rec vs. NAIVE
Brent	0.0000	0.0024	0.0352
Dubai	0.0000	0.5241	0.9577
WTI	0.0085	0.0561	0.0825
Coal_AU	0.0167	0.4624	0.1947
Coal_ZA	0.0111	0.0017	0.1732
Gas_US	0.1149	0.5788	0.1229
Gas_EU	0.0151	0.0210	0.2298
Gas_JP	0.0350	0.0696	0.2353
Cocoa	0.0010	0.0010	0.0022
Coffee_Arabica	0.0003	0.0002	0.0000
Coffee_Robusta	0.1578	0.1578	0.1578
Tea_Colombo	0.0118	0.0106	0.0085
Tea_Kolkata	0.1053	0.0368	0.2181
Tea_Mombasa	0.1184	0.1287	0.1229
Coconut_oil	0.0557	0.0747	0.0877
Groundnuts	0.0041	0.0001	0.0091
Fish_meal	0.1584	0.1584	0.1584
Palm_oil	0.0968	0.0968	0.1069
Soybeans	0.0442	0.0516	0.0543
Soybean_oil	0.0605	0.0225	0.8987
Soybean_meal	0.0000	0.0000	0.0000
Maize	0.0567	0.1268	0.1653
Rice_5	0.1174	0.1174	0.0821
Rice_100	0.0008	0.1678	0.0008
Wheat_SRW	0.1148	0.1148	0.0806
Wheat_HRW	0.0360	0.0586	0.0628
Banana	0.0016	0.1621	0.0015
Orange	0.1421	0.1421	0.1424
Beef	0.0594	0.0594	0.1950
Chicken	0.0310	0.0310	0.0527
Shrimps	0.0117	0.0117	0.0087
Sugar_EU	0.0590	0.3408	0.0340
Sugar_US	0.1120	0.1120	0.3439
Sugar_World	0.1591	0.1591	0.1591
Tobacco	0.1131	0.1131	0.1203
Logs_CM	0.0006	0.0232	0.0948
Logs_MY	0.0000	0.0000	0.0001
Sawnwood	0.0002	0.0432	0.0610
Plywood	0.0056	0.0069	0.0070
Cotton	0.0005	0.0005	0.0164
Rubber	0.0003	0.0182	0.0360
Phosphate_rock	0.0155	0.9250	0.0061
Dap	0.1408	0.1408	0.3374

**Table A5.** *Cont.*

Commodity	BSR Rec vs. Best	BSR Rec vs. ARIMA	BSR Rec vs. NAIVE
Tsp	0.1142	0.1142	0.8300
Urea	0.0474	0.0868	0.2359
Potash	0.0875	0.5945	0.1037
Aluminium	0.0000	0.0030	0.0073
Iron	0.2240	0.7191	0.9356
Copper	0.0158	0.4215	0.6329
Lead	0.0206	0.0206	0.1275
Tin	0.1509	0.6005	0.8598
Nickel	0.1016	0.1016	0.4385
Zinc	0.4096	0.5121	0.8845
Gold	0.1634	0.6838	0.6001
Platinum	0.0319	0.0679	0.0675
Silver	0.0982	0.3400	0.3480

**Table A6.** The Diebold–Mariano test—GP rec model vs. “best”, vs. ARIMA, and vs. NAIVE.

Commodity	GP Rec vs. Best	GP Rec vs. ARIMA	GP Rec vs. NAIVE
Brent	0.0000	0.0001	0.0038
Dubai	0.0000	0.0001	0.0005
WTI	0.0000	0.0092	0.0405
Coal_AU	0.1305	0.7823	0.9500
Coal_ZA	0.0107	0.2715	0.7661
Gas_US	0.1499	0.1547	0.1528
Gas_EU	0.0007	0.0018	0.0083
Gas_JP	0.0363	0.0599	0.2664
Cocoa	0.0000	0.0000	0.0000
Coffee_Arabica	0.0004	0.0004	0.0005
Coffee_Robusta	0.0301	0.0301	0.0301
Tea_Colombo	0.0000	0.0000	0.0000
Tea_Kolkata	0.0001	0.0000	0.0000
Tea_Mombasa	0.0000	0.0000	0.0000
Coconut_oil	0.0000	0.0000	0.0001
Groundnuts	0.0000	0.0000	0.0000
Fish_meal	0.1590	0.1590	0.1590
Palm_oil	0.0009	0.0009	0.0209
Soybeans	0.0125	0.0125	0.0125
Soybean_oil	0.0011	0.0012	0.0514
Soybean_meal	0.0000	0.0000	0.0002
Maize	0.0004	0.0005	0.0005
Rice_5	0.1026	0.1026	0.1026
Rice_100	0.0011	0.2012	0.0011
Wheat_SRW	0.0000	0.0000	0.0000
Wheat_HRW	0.0300	0.0448	0.0586
Banana	0.0000	0.0000	0.0000
Orange	0.1269	0.1269	0.1279
Beef	0.0327	0.0327	0.0381
Chicken	0.0762	0.0762	0.0588
Shrimps	0.0000	0.0000	0.0000
Sugar_EU	0.0012	0.0014	0.0001
Sugar_US	0.0746	0.0746	0.1168
Sugar_World	0.0000	0.0000	0.0000
Tobacco	0.0000	0.0000	0.0000
Logs_CM	0.0000	0.0003	0.0019
Logs_MY	0.0000	0.0000	0.0000
Sawnwood	0.0028	0.0029	0.0029
Plywood	0.0000	0.0000	0.0000
Cotton	0.0003	0.0003	0.0007

Table A6. Cont.

Commodity	GP Rec vs. Best	GP Rec vs. ARIMA	GP Rec vs. NAIVE
Rubber	0.0062	0.0493	0.0585
Phosphate_rock	0.2654	0.9534	0.4070
Dap	0.0434	0.0434	0.0293
Tsp	0.0209	0.0209	0.4023
Urea	0.0000	0.0002	0.0005
Potash	0.0187	0.4791	0.0157
Aluminium	0.0000	0.0000	0.0001
Iron	0.1423	0.4858	0.7309
Copper	0.0004	0.0596	0.0722
Lead	0.0081	0.0081	0.0111
Tin	0.0566	0.1823	0.3743
Nickel	0.0022	0.0022	0.0062
Zinc	0.0001	0.0097	0.0310
Gold	0.1449	0.4744	0.3852
Platinum	0.0161	0.8765	0.8434
Silver	0.4522	0.6101	0.6420

Table A7. The Diebold–Mariano test—“rec” vs. “fix”.

Commodity	BSR	BSR av MSE	BSR av EW	GP
Brent	0.9997	0.0000	0.7999	0.9982
Dubai	1.0000	0.0000	0.0044	0.9995
WTI	0.1591	0.8409	0.0008	0.0005
Coal_AU	0.3120	0.8751	0.0000	0.0000
Coal_ZA	0.9830	0.0417	0.0003	0.6665
Gas_US	0.1963	0.1555	0.0013	0.1225
Gas_EU	0.1298	0.8661	0.1405	0.0000
Gas_JP	0.1591	0.8409	0.0416	0.0314
Cocoa	0.9930	0.0000	0.0124	0.0000
Coffee_Arabica	0.9999	0.0004	0.0010	0.9980
Coffee_Robusta	0.1591	0.8409	0.0000	0.9686
Tea_Colombo	0.7948	0.0000	0.0000	1.0000
Tea_Kolkata	0.7471	0.0000	0.8468	1.0000
Tea_Mombasa	0.8812	0.0000	0.0012	0.0000
Coconut_oil	0.9214	0.0000	0.0003	0.9972
Groundnuts	0.9960	0.0000	0.0029	0.0000
Fish_meal	0.8416	0.1590	0.0301	0.8410
Palm_oil	0.1325	0.8768	0.0227	0.0000
Soybeans	0.1585	0.0125	0.8363	0.9875
Soybean_oil	0.3132	0.0314	0.0712	0.0000
Soybean_meal	1.0000	0.0001	0.0011	0.9940
Maize	0.9349	0.0004	0.0005	0.0000
Rice_5	0.8793	0.1027	0.0449	0.1594
Rice_100	0.9973	0.0269	0.0904	0.9792
Wheat_SRW	0.9144	0.0000	0.0135	0.9999
Wheat_HRW	0.9400	0.0628	0.0000	0.0411
Banana	0.8607	0.0000	0.2094	0.0001
Orange	0.1591	0.8409	0.0000	0.1117
Beef	0.8903	0.0337	0.0658	0.0000
Chicken	0.9754	0.1210	0.6837	0.8680
Shrimps	0.9739	0.0000	0.0016	0.0001
Sugar_EU	0.9703	0.0002	0.1612	0.0083
Sugar_US	0.6617	0.1312	0.5337	0.6361
Sugar_World	0.8409	0.8727	0.0491	0.0379
Tobacco	0.8706	0.0002	0.0403	0.9175
Logs_CM	0.9990	0.0002	0.0040	0.0001
Logs_MY	0.0227	0.9773	0.9212	1.0000

**Table A7.** *Cont.*

Commodity	BSR	BSR av MSE	BSR av EW	GP
Sawnwood	0.6473	0.0030	0.0906	0.0408
Plywood	0.1648	0.8005	0.1591	0.7624
Cotton	0.2596	0.7893	0.0004	0.0002
Rubber	0.9974	0.0262	0.0000	0.9615
Phosphate_rock	0.9965	0.3159	0.1899	0.7346
Dap	0.0788	0.9212	0.0012	0.0044
Tsp	0.5278	0.0060	0.0758	0.6297
Urea	0.9526	0.0000	0.0085	1.0000
Potash	0.8955	0.0181	0.1150	0.9811
Aluminium	0.9999	0.0000	0.0000	0.9998
Iron	0.5166	0.3003	0.0374	0.0418
Copper	0.8625	0.0029	0.0000	0.1166
Lead	0.9535	0.0013	0.0008	0.9830
Tin	0.4807	0.1016	0.0806	0.0004
Nickel	0.6666	0.0004	0.0024	0.0010
Zinc	0.2026	0.0004	0.0055	0.1246
Gold	0.3883	0.3878	0.0105	0.0000
Platinum	0.9602	0.2326	0.0006	0.1839
Silver	0.1561	0.8442	0.0000	0.0000

**Table A8.** The Diebold–Mariano test—BSR vs. GP.

Commodity	BSR Rec vs. GP Rec	BSR Fix vs. GP Fix
Brent	0.0655	0.8724
Dubai	0.0002	1.0000
WTI	0.7229	0.0188
Coal_AU	0.9406	0.0000
Coal_ZA	0.8899	0.9977
Gas_US	0.1541	0.0816
Gas_EU	0.3310	1.0000
Gas_JP	0.6818	0.9523
Cocoa	0.0009	0.0000
Coffee_Arabica	0.2536	1.0000
Coffee_Robusta	0.7545	0.9212
Tea_Colombo	0.0000	0.9933
Tea_Kolkata	0.0000	0.9952
Tea_Mombasa	0.7340	0.0000
Coconut_oil	0.3622	0.9736
Groundnuts	0.0001	0.0000
Fish_meal	0.5007	0.9910
Palm_oil	0.8764	1.0000
Soybeans	0.0125	0.9222
Soybean_oil	0.0269	0.0000
Soybean_meal	0.0347	1.0000
Maize	0.0007	0.0000
Rice_5	0.1030	0.1589
Rice_100	0.4754	0.9999
Wheat_SRW	0.0000	0.8917
Wheat_HRW	0.4674	0.7429
Banana	0.0000	0.0000
Orange	0.7725	0.1117
Beef	0.2953	0.7829
Chicken	0.5324	0.8318
Shrimps	0.0000	0.0000
Sugar_EU	0.0031	0.0063
Sugar_US	0.3293	0.3542
Sugar_World	0.8409	0.8409

Table A8. Cont.

Commodity	BSR Rec vs. GP Rec	BSR Fix vs. GP Fix
Tobacco	0.7458	1.0000
Logs_CM	0.0801	0.0005
Logs_MY	0.0001	0.0266
Sawnwood	0.0030	0.0000
Plywood	0.0001	0.6267
Cotton	0.8532	0.0008
Rubber	0.1652	1.0000
Phosphate_rock	0.9117	0.7826
Dap	0.4567	0.8416
Tsp	0.0556	0.9980
Urea	0.0006	1.0000
Potash	0.3157	0.9868
Aluminium	0.0436	1.0000
Iron	0.3080	0.1071
Copper	0.0158	0.1789
Lead	0.0969	1.0000
Tin	0.0972	0.0006
Nickel	0.0027	0.0004
Zinc	0.0008	0.2556
Gold	0.3184	0.0001
Platinum	0.9505	1.0000
Silver	0.7129	0.0091

## Appendix B

Below is a description of the time-varying importance of price predictors, based on the DMA model according to the procedure described in Section 5.5.

In particular, the dividend to price ratio had a mostly positive impact, especially for Coal\_ZA, Phosphate\_rock, and Dap up to 2000; Palm\_oil, Cotton, Rubber, Phosphate\_rock, Potash, and Tin around 2011–2015. This impact was negative for Coal\_AU and Coal\_ZA between 2010 and 2012; Phosphate\_rock between 2003 and 2006; and Gold since 2018.

The price earnings ratio also had a mostly negative impact, especially on the following: Coal\_ZA between 2001 and 2011; Coffee\_Robusta between 2006 and 2020; Coconut\_oil, Fish\_meal, Palm\_oil, Soybean\_oil, Soybean\_meal, and Maize around 2015–2018; Rice\_5 and Rice\_100 since 2020; Tobacco between 2009 and 2015; Potash and Iron between 2005 and 2009 and since 2017; Tin and Zinc between 2016 and 2018; and Gold since 2019. It was positive for Gas\_EU between 2000 and 2003; Gas\_JP since 2018; Phosphate\_rock between 1999 and 2001 and since 2010; Potash between 2009 and 2014; and Tin between 2011 and 2015.

The short-term interest rate had a mostly negative impact, especially on Coal\_Za before 2011 and Coal\_AU between 2006 and 2011, and both of these commodities since 2019; Gas\_JP and Groundnuts since 2019; Fish\_meal before 2002; Rice\_5 and Rice\_100 between 2008 and 2010; Wheat\_HRW between 1997 and 2001; Beef and Chicken between 2007 and 2009; Sugar\_EU between 2009 and 2011; Rubber before 2011; Phosphate\_rock between 2010 and 2015 and between 2018 and 2020; and Dap and Tsp between 2004 and 2011. It was also negative for various metals (i.e., Lead, Tin, Nickel, Zinc, Gold, Platinum and Silver) between 2000 and 2009 (but in different sub-periods for different metals). For Gold, this negative impact was also present between 2012 and 2015. There was a positive impact on Soybean\_meal between 2014 and 2016; Tobacco before 2002; Rubber between 2012 and 2014; Phosphate\_rock between 2004 and 2007 and between 2016 and 2018; and Iron between 1997 and 2001, between 2005 and 2009, and between 2014 and 2020.

The long-term interest rate for the U.S. had a mostly negative impact. For many commodities, this was especially seen around 2007–2011 (i.e., Dubai, WTI, Coconut\_oil, Palm\_oil, Soybean\_oil, Maize, Wheat\_HRW, Beef, Sugar\_US, Sugar\_World, Tobacco, Rubber, Phosphate\_rock, Dap, Tsp, Potash, Copper, Tin, Gold, Platinum, and Silver). Some



negative impact was also present before 1999 for Fish\_meal, Wheat\_SRW, Wheat\_HRW, Dap, Lead, and Silver. For Phosphate\_rock, the negative impact was present between 2007 and 2020. A positive impact was present for Coal\_AU between 1998 and 2000, Rubber between 2002 and 2003, Gas\_JP between 2009 and 2010, Groundnuts since 2019, and Iron between 2006 and 2018.

On the other hand, the long-term interest rate for the Euro area had mixed impacts. It was negative for Brent and Dubai around 2011 and 2012, Coal\_AU between 2003 and 2006 and between 2011 and 2019, Coffee\_Robusta between 2001 and 2004 and between 2006 and 2009, Palm\_oil between 2011 and 2013, Shrimps between 2016 and 2020, Sugar\_World between 2016 and 2017, Tobacco before 2002 and between 2012 and 2015, Cotton between 2013 and 2014, Rubber between 2011 and 2014, Tsp since 2011, Iron since 2005, Copper between 2011 and 2013, Tin between 2009 and 2016, and Platinum between 2011 and 2016. A positive impact was present for Coal\_ZA before 2001, Gas\_EU between 1999 and 2004, Gas\_JP between 2015 and 2016, Soybean\_meal between 2015 and 2017, Logs\_MY and Plywood before 1997, Phosphate\_rock in several periods between 2008 and 2020, Tsp between 2007 and 2011, and Potash between 2009 and 2012 and between 2016 and 2020.

The term spread had a mixed impact. The impact was negative on all energy commodities in some periods around 2016 and 2020; Soybeans, Maize, Wheat\_SRW and Wheat\_HRW in some periods before 2002; Tea\_Mombasa, Fish\_meal, and Soybean\_meal between 2009 and 2016; Tobacco between 2008 and 2012; Phosphate\_rock since 2005; Potash between 2006 and 2009 and between 2015 and 2020. It was also negative for some metals (i.e., Iron, Copper, Lead, Tin, and Zinc) around 2015–2019. On the other hand, it was positive on Coffee\_Arabica between 2014 and 2019, Coffee\_Robusta between 1999 and 2005, Groundnuts since 2014, Fish\_meal and Palm\_oil between 1999 and 2003, Logs\_CM between 2004 and 2007, Logs\_MY between 2011 and 2016, Cotton and Rubber between 2002 and 2004, Iron between 2006 and 2014, and Gold after 2014.

The default return spread had a slightly more positive (than negative) impact on commodity prices. The positive impact was on WTI, Coal\_AU, and Coal\_ZA around 2019 and 2020; Coal\_AU around 2008; Coal\_ZA before 2002; Gas\_EU since 2019; Gas\_JP between 2011 and 2018; Coffee\_Arabica and Coffee\_Robusta between 1999 and 2002; Palm\_oil between 2002 and 2004; Soybeans between 1997 and 2000; Soybean\_oil between 2001 and 2006; wheat prices around 1999 and 2001; Sugar\_World between 2002 and 2006; Logs\_MY and Plywood between 1999 and 2001; Rubber between 2000 and 2002 and between 2012 and 2014; Phosphate\_rock between 2006 and 2020; and Silver before 1997. The negative impact was on Tea\_Mombasa between 2015 and 2018, Groundnuts since 2014, Fish\_meal for several periods between 1999 and 2013, wheat prices since 2019, Tobacco before 2009, Logs\_MY between 2012 and 2016, Dap and Tsp around 2007 and 2012, Iron between 2005 and 2017, Copper and Lead between 2015 and 2018, Zinc for several periods between 2006 and 2018, and Gold since 2013.

The U.S. Consumer Price Index for all urban consumers had a mostly negative impact. In particular, this was on most crude oil prices in various periods between 2000 and 2009 and since 2016; Coal\_AU before 2002 and since 2016; Cocoa between 2002 and 2005 and between 2010 and 2020; Tea\_Kolkata between 2008 and 2020; Coconut\_oil since 2012; Fish\_meal between 2000 and 2009; Soybean\_meal between 2013 and 2016; Maize between 2009 and 2016; Shrimps between 2011 and 2016; Phosphate\_rock before 1999; Tsp between 2012 and 2020; Potash between 2005 and 2020; Aluminium, Iron, Tin and Zinc around 2006–2009; Platinum between 1997 and 1999 and between 2011 and 2016. It had a positive impact on Gas\_JP between 1999 and 2004, Palm\_oil and Soybean\_oil around 2001–2003, rice prices around 2010–2012, Wheat\_HRW between 1997 and 2003, Tobacco between 2015 and 2020, Phosphate\_rock between 2010 and 2018, Dap between 2009 and 2012, and Urea since 2014.

The U.S. Producer Price Index had a mostly negative impact. In particular, this was on oil prices, coal prices and Gas\_US around 2004–2010; Cocoa between 2002 and 2007; Tea\_Mombasa between 1997 and 2000; Coconut\_oil between 2004 and 2006 and between

2010 and 2012; Palm\_oil, Soybeans, Soybean\_oil, Soybean\_meal, and Maize in several periods between 2001 and 2019; wheat prices before 2016; Phosphate\_rock between 2009 and 2018; Potash since 2005; Iron between 2005 and 2011; Zinc between 2011 and 2016; Gold between 2003 and 2014; and Platinum and Silver around 2008–2012. It had a positive impact on coffee prices since 2015, Tea\_Kolkata since 2008, rice prices since 2020, Beef since 2015, Chicken before 2000, Shrimps between 2011 and 2015, Sugar\_EU and Sugar\_World since 2020, Tobacco between 2012 and 2017, Plywood between 2015 and 2020, Cotton since 2018, Dap since 2016, Tsp since 2007, and Urea between 2008 and 2014.

U.S. industrial production had a mostly positive impact, especially since 2008 on Coffee\_Robusta, Soybeans, Sugar\_EU, Potash, Aluminium, Iron, Tin, Nickel, and Zinc. Furthermore, it had a positive impact in shorter periods on Brent between 2003 and 2005, WTI between 2001 and 2006, all oil prices around 2008–2009, Gas\_EU and Gas\_JP between 2015 and 2020, Fish\_meal between 2013 and 2020, Beef between 2015 and 2019, Sugar\_US between 2011 and 2017, Tobacco between 2006 and 2009, Plywood between 2011 and 2016, Rubber between 2008 and 2017, Phosphate\_rock before 2005, Platinum between 2016 and 2020, and Silver since 2020. It had a negative impact on Gas\_EU and Soybeans around 2008, Orange and Sugar\_US between 2000 and 2008, Logs\_MY between 2012 and 2016, Phosphate\_rock between 2012 and 2019, Gold between 2000 and 2008, and Platinum between 2000 and 2002.

U.S. average hourly earnings of production and nonsupervisory employees had a mixed impact on less commodities than the previously reported variables. In particular, it had a negative impact on oil prices before 2015, Tea\_Mombasa between 2015 and 2019, Fish\_meal between 2011 and 2014, Soybean\_meal between 2004 and 2013, Tsp between 2002 and 2007, and Potash and Iron between 2005 and 2020. It had a positive impact on Gas\_JP between 2004 and 2014; coffee prices between 2014 and 2020; Sugar\_EU between 2009 and 2013; Cotton between 2010 and 2017; Phosphate\_rock before 2003 and since 2009; and Aluminium, Copper, Zinc, Gold and Platinum around 2006–2009.

The M1 money stock for the U.S. and M2 money stock for the U.S. had sometimes contradictory impacts, whereas M2 seemed to have more impact overall than M1. In particular, M1 had a negative impact on Cocoa between 2004 and 2007 and between 2018 and 2020, Palm\_oil between 2001 and 2004, Orange between 2001 and 2008, Sugar\_EU between 2008 and 2012, Sugar\_World between 2001 and 2005, logs prices before 1998, Sawnwood and Plywood around 2009–2011, Phosphate\_rock between 2008 and 2020, Dap between 2009 and 2012, Potash between 2005 and 2017, and Iron between 2015 and 2018. M1 had a positive impact on coal prices around 2001–2008, Gas\_EU and Gas\_JP around 2018–2020, Maize before 2002, Tsp and Urea around 2012–2015, Iron between 2009 and 2013, and Silver between 2001 and 2008. M2 had a negative impact on coal prices and gas prices around 2015–2019; Tea\_Kolkata between 2015 and 2020; Coconut\_oil between 2011 and 2017; Palm\_oil, Soybeans, and Soybean\_oil around 2003–2012; Sugar\_EU before 2002; Sugar\_US between 2008 and 2020; logs prices around 2000; Cotton before 2008 and between 2014 and 2020; Rubber and Phosphate\_rock before 1999; Dap and Tsp between 2007 and 2017; and Lead between 2008 and 2018. M2 had a positive impact on Groundnuts and Fish\_meal since 2013, rice prices between 2010 and 2013, wheat prices between 2001 and 2004, Chicken and Tobacco between 2014 and 2019, Phosphate\_rock since 2008, Potash between 2009 and 2017, Gold and Platinum between 2014 and 2020, and Silver between 2018 and 2020.

Killian's Index of Global Real Economic Activity had a mostly positive impact. In particular, this was on coal prices for almost all of the analysed period; Gas\_EU and Gas\_JP around 2008–2011 and since 2015; Tea\_Mombasa and Coconut\_oil since 2020; Soybeans, Soybean\_meal, Maize, rice prices, and wheat prices before 2003; Soybeans, Soybean\_oil, and Soybean\_meal around 2008; Sugar\_US between 2011 and 2018; Sugar\_World between 2000 and 2003; Tobacco between 2014 and 2018; Sawnwood since 2019; Phosphate\_rock since 2009; Dap since 2011; Potash since 2016; Iron since 2004; Lead and Zinc around 2004; and Silver between 1998 and 1999. However, it was negative on Dubai and WTI between 2001

and 2004, Palm\_oil and rice prices around 2008–2011, Beef before 1998, Chicken between 2015 and 2020, Potash between 2009 and 2013, and Platinum between 2008 and 2011.

The U.S. unemployment rate had a rather negative impact, in particular on oil prices and Gas\_EU around 2000, Coal\_ZA before 2010, Groundnuts since 2020, Fish\_meal between 1997 and 1998, rice prices around 2008 and since 2020, Sugar\_US between 2009 and 2015, Phosphate\_rock between 2000 and 2003 and between 2006 and 2017, Dap before 2000, Tsp and Urea before 1997, Potash since 2006, and Iron between 2004 and 2009. However, it had a positive impact on Gas\_EU and Gas\_JP around 2010; Coffee\_Robusta between 2005 and 2006; Wheat\_HRW between 1997 and 2003; Sugar\_World between 2002 and 2005; Tobacco before 2002; and Iron, Copper, Lead, Tin, Nickle, and Platinum around 2009.

U.S. dollar exchange rates had a mostly positive impact. Mostly, this was indicated by the Australian dollar to U.S. dollar exchange rate, and, secondly, by the Indian rupee to U.S. dollar exchange rate. The impact from the Canadian dollar to U.S. dollar exchange rate was least often observed.

In particular, the Australian dollar to U.S. dollar exchange rate had a positive impact on Coal\_AU between 1997 and 2002, between 2008 and 2011, and since 2017; Coal\_ZA between 2001 and 2005 and between 2008 and 2015; Gas\_EU in similar periods as for Coal\_ZA, Cocoa, and coffee prices around 2009; Tea\_Mombasa between 2006 and 2013; Groundnuts between 2011 and 2019; Palm\_oil between 2007 and 2013; Soybeans and Soybean\_meal between 2004 and 2016; Soybean\_oil between 2009 and 2012 and between 2015 and 2017; Rice\_100 and wheat prices between 2012 and 2016; wheat prices since 2019; Beef since 2016; Logs\_CM since 2003; Sawnwood between 1998 and 2004; Plywood between 2005 and 2008; Cotton between 2007 and 2011; Rubber between 1998 and 2012; Tsp since 2008; and various metal prices around 2005, 2009, and 2020. In the case of metals, those mostly impacted were Aluminium, Copper, Lead, Nickel, Platinum, and Silver. However, the impact was negative on oil prices since 2016, Phosphate\_rock between 2008 and 2016 and since 2020, and Potash between 2009 and 2016.

The Canadian dollar to U.S. dollar exchange rate had a positive impact on Gas\_EU between 2008 and 2016; Gas\_JP between 1998 and 2009; Cocoa between 2012 and 2016; Tea\_Kolkata between 2001 and 2008; Coconut\_oil, Soybeans, Wheat\_HRW, Phosphate\_rock, Tsp, and Gold around 2009; Tobacco between 2012 and 2013; Iron, Gold, Platinum, and Silver since 2016; and Gold in several periods since 2000. It had a negative impact on oil prices around 2004–2007 and 2015; Coal\_AU between 2008 and 2010 and around 2015; Gas\_US between 1998 and 2001 and around 2003, 2009, and 2017; Coffee\_Arabica between 2006 and 2009; Sawnwood between 1998 and 2008; and Phosphate\_rock in several periods before 2006.

The Indian rupee to U.S. dollar exchange rate had a positive impact on Coal\_ZA since 2011, Coffee\_Robusta between 2009 and 2020, Maize in several periods since 2006, Rubber between 2008 and 2019, Urea before 2004, Potash between 2008 and 2015, Iron between 2005 and 2019, Platinum in several periods between 1997 and 2019, and Silver around 2009 and between 2011 and 2014. It also had a positive impact on several other commodities (i.e., Dubai, Coffee\_Arabica, Tea\_Kolkata, Tea\_Mombasa, Coconut\_oil, Fish\_meal, Palm\_oil, Soybean\_oil, Soybean\_meal, and wheat prices) around 2009. It had a negative impact on Coal\_AU and Gas\_US before 1998, Gas\_US between 2004 and 2007, Groundnuts since 2014, Phosphate\_rock since 2008, and Potash since 2016.

In the case of real effective exchange rates, the one based on the manufacturing Consumer Price Index for Australia and the one based on the manufacturing Consumer Price Index for India had mostly positive impacts, whereas the one based on the manufacturing Consumer Price Index for Canada and the one based on the manufacturing Consumer Price Index for the U.S. had mostly negative impacts.

In particular, the real effective exchange rate based on the manufacturing Consumer Price Index for Australia had a positive impact on Brent, Dubai, coal prices, Palm\_oil, Soybean\_oil, Sugar\_EU, Phosphate\_rock, and Urea in various periods since 2011; Coffee\_Robusta between 2005 and 2011; Beef between 2009 and 2015; Cotton between 2002

and 2013; Urea before 2005; Potash between 2007 and 2017; and Iron between 1999 and 2003. It had a negative impact on Gas\_EU since 2008, Sugar\_US between 2001 and 2014, and Phosphate\_rock since 2010.

Real effective exchange rates based on the manufacturing Consumer Price Index for Canada had a negative impact on Coal\_AU, Gas\_US, Coffee\_Arabica, Coconut\_oil, Palm\_oil, rice prices, Beef, and Tin in several periods after 2009; Coffee\_Robusta between 1997 and 2016; Tea\_Colombo before 2003; Soybeans, Soybean\_meal, sugar prices, logs prices, Phosphate\_rock, Tsp, and Copper around 2009–2013; Sawnwood before 1999; Platinum between 2008 and 2014; and Coffee\_Arabica around 1997–2002. It had a positive impact on Gas\_EU and Tea\_Kolkata since 2017, Phosphate\_rock between 2000 and 2004, and Potash since 2005.

Real effective exchange rates based on the manufacturing Consumer Price Index for India had a positive impact on Coal\_AU in several sub-periods over the whole analysed time period, Tean\_Mombasa between 2001 and 2020, Fish\_meal between 2009 and 2020, rice prices around 2020, Wheat\_HRW between 2003 and 2008, Chicken since 2013, Sawnwood between 1998 and 2005, Plywood since 2017, Tsp between 2012 and 2016, Iron between 2009 and 2013, and Tin around 2009. They had a negative impact on oil prices around 1999, coffee prices since 2018, Tea\_Colombo since 2013, Tea\_Kolkata between 2009 and 2013, Plywood between 2003 and 2008, Potash since 2005, Copper around 2007 and since 2016, Nickel since 2012, Zinc since 2019, and Silver between 2008 and 2009.

Real effective exchange rates based on the manufacturing Consumer Price Index for the U.S. had a positive impact on Gas\_US and Gas\_JP since 2020; Tea\_Kolkata since 2013; Tea\_Mombasa between 1998 and 2002; Orange in several sub-periods since 2008; Beef before 1998 and between 2005 and 2009; Shrimps and Sugar\_EU around 2014–2018; Cotton before 2002; Copper, Tin, and Silver around 2007; and Copper Nickel and Silver since 2020. It had a negative impact on Coal\_AU and Cola\_ZA around 2015–2017, Cocoa since 2017, Tea\_Mombasa between 2010 and 2015, Fish\_meal, Palm\_oil and Soybeans since 2019, Soybean\_meal and rice prices since 2009, Shrimps before 2002, Sugar\_EU between 2003 and 2012, Logs\_CM over the whole analysed period, Sawnwood between 1999 and 2006 and since 2017, Rubber between 2013 and 2016, Phosphate\_rock between 2003 and 2006, Urea between 2015 and 2018, Iron since 2005, and Lead and Zinc around 2005.

The U.S. trade balance had a mixed impact. It was positive mostly after 2006, in particular, on Coal\_AU since 2005; Coal\_ZA since 2018; Tea\_Colombo since 2017; Iron between 2013 and 2019; and Coconut\_oil, Soybeans, Soybean\_oil, Soybean\_meal, Maize, Rice\_100, Sugar\_EU, Sugar\_US, Sugar\_World, Phosphate\_rock, Dap, and Tsp around 2008. It was negative on oil prices between 2000 and 2007, Gas\_EU and Gas\_JP between 2003 and 2011, Groundnuts between 2002 and 2007, Shrimps in several periods since 2003, Logs\_CM between 2002 and 2006, Logs\_MY and Plywood since 2015, Aluminium between 2000 and 2009, Iron between 2003 and 2009, and Copper around 2000.

The S&P GSCI Commodity Total Return Index had a positive impact on the majority of commodities and over multiple periods. In particular, this was on oil prices during the whole analysed period; Coal\_AU since 2002; Coal\_ZA since 2009; Gas\_US, Tea\_Mombasa, and Tin over almost whole analysed period; Gas\_US since 2009; Gas\_EU since 2017; Gas\_JP before 2015; Tea\_Kolkata before 2004; Soybean\_meal between 1997 and 2005; Wheat\_HRW between 1998 and 2003; Beef between 2008 and 2015; Rubber between 2008 and 2019; Dap between 2009 and 2014; Urea before 1999; Potash since 2005; Aluminium since 2009; Copper since 2007; Zinc between 2017 and 2019; Gold between 2001 and 2015; Platinum between 2013 and 2015; and Silver between 2011 and 2015. However, it had a negative impact on Groundnuts since 2014, Soybean\_oil between 2003 and 2007, and Phosphate\_rock since 2016.

The dollar open interest had a mostly positive impact, in particular, on Cocoa, coffee prices, Palm\_oil, Soybean\_meal, and Cotton in various periods after 2010; Coconut\_oil in various sub-periods of the whole analysed period; Fish\_meal between 1999 and 2006; Maize between 2005 and 2017; rice prices and Sugar\_US between 2009 and 2018; Sugar\_World

between 2010 and 2020; Phosphate\_rock between 1999 and 2001 and between 2009 and 2013; Dap between 2014 and 2017; and Platinum between 2006 and 2020. It had a negative impact on oil prices between 1997 and 2003 and around 2020; Gas\_EU before 2004; Gas\_JP since 2003; Tea\_Kolkata between 2014 and 2019; Tea\_Mombasa between 2006 and 2018; rice prices, Orange, and Chicken around 2001; and Potash between 2005 and 2010 and since 2014.

Working's dollar T-index had a mostly negative impact, in particular, on oil prices around 2016; Coal\_AU since 2005; Coal\_ZA between 2001 and 2015; Gas\_JP between 2002 and 2008 and since 2017; Coconut\_oil, Groundnuts, Palm\_oil, Soybean\_oil, Maize, Wheat\_HRW, and Sugar\_EU around 2008–2014; Chicken and Shrimps since 2004; Rubber between 2012 and 2014; Phosphate\_rock since 2013; Dap between 2009 and 2015; Tsp since 2010; Urea between 2012 and 2014; Tin between 2007 and 2015; and Platinum between 2008 and 2015. It had a positive impact on Coal\_ZA and Dap before 1999, Gas\_EU between 2010 and 2012, Coffee\_Robusta between 2005 and 2007, Fish\_meal before 2002, and Potash since 2005.

The VXO index of implied volatility had a mixed impact. It was negative on: oil prices around 2008 and 2020–2021, Coal\_ZA and Gas\_US before 1998, Coal\_AU and Gas\_US between 2018 and 2020, Chicken between 2001 and 2005, Rubber between 2008 and 2014, Dap since 2010, Urea between 2011 and 2015, Potash between 2005 and 2020, Lead between 1997 and 2001 and between 2008 and 2013, and Gold in various periods before 2014. It was positive on Coal\_AU between 2005 and 2009; Gas\_EU between 2010 and 2014; Coconut\_oil between 2012 and 2019; Groundnuts since 2017; Palm\_oil, Soybeans, Soybean\_oil, Soybean\_meal, Maize, and rice prices around 2020; wheat prices between 2015 and 2019; Cotton between 2012 and 2019; Phosphate\_rock since 2009; and Tin around 2006–2007.

The Global Geopolitical Risk Index had a limited impact. It was negative on Dubai between 2004 and 2014; Coal\_AU between 2004 and 2009; Coal\_ZA before 2013; Soybeans, Soybean\_oil, Soybean\_meal, Maize, and Rice\_5 since 2020; Tobacco since 2015; Rubber, Gold, and Silver since 2018; Phosphate\_rock between 2009 and 2017; and Potash since 2006. It was positive on Rice\_100 before 2002, Sawnwood between 1998 and 2001, Dap and Tsp since 2009, Aluminium and Lead since 2017, Iron between 2008 and 2017, and Tin between 2001 and 2017.

Stock prices had a mixed impact. The S&P 500 Index and the MSCI G7 index had mostly negative impacts, whereas the MSCI WORLD for developed markets index, the MSCI EU index, and the MSCI EM for emerging markets index had mostly positive impacts.

In particular, the S&P 500 Index had a negative impact on oil prices before 2001 and since 2015, coal prices between 2002 and 2005 and since 2017, Cocoa in various periods since 2003, Tea\_Colombo between 2008 and 2016, Tea\_Mombasa between 2015 and 2020, wheat prices around 2010–2015 and since 2019, Orange between 2009 and 2014, Sugar\_EU over the whole analysed period, Sugar\_World between 2013 and 2017, logs prices around 2006–2008, Logs\_CM between 2010 and 2014, Logs\_MY since 2014, Sawnwood since 2000, Urea between 2001 and 2008, Iron since 2008, Copper since 2016, Lead before 1999, Tin since 2014, Nickel between 2011 and 2016, Zinc since 2019, Gold and Platinum over the whole analysed period, and Silver since 2006. It was positive on rice prices around 2000–2001 and between 2008 and 2015, Phosphate\_rock since 2009, Dap between 2009 and 2013, and Potash between 2009 and 2013.

In the case of the MSCI G7 index since 2015, the impact was positive (opposite than that of the S&P 500 Index) on oil prices. It was also positive on Groundnuts since 2013, Shrimps since 2011, Cotton between 2011 and 2019, Phosphate\_rock between 2010 and 2016, Gold in various periods between 1997 and 2016, Platinum in various periods before 2008, and Silver between 1998 and 2002. It was negative on Coal\_ZA between 2001 and 2010 and since 2017; Cocoa between 2001 and 2010; Tea\_Colombo and Tea\_Kolkata around 2010–2013; Fish\_meal, Palm\_oil and Soybeans around 1998–2001; rice prices in various sub-period of the whole analysed period; wheat prices since 2009; Logs\_CM since 2002;



Dap since 2009; Tsp before 2006; Potash since 2013; and Copper, Lead, Tin, Nickel, and Zinc between 2007 and 2016.

The MSCI WORLD for developed markets index had a positive impact on Brent since 2015; Coal\_ZA and Gas\_US around 2007–2010 and 2018; Cocoa between 2003 and 2014; Tea\_Colombo and Tea\_Mombasa around 2011–2015; Rice\_5 before 2006 and both rice prices around 2012; wheat prices between 2013–2018; and Iron, Copper, Lead, Tin, Nickel, and Zinc around 2010. In the case of individual metals, the positive impact was observed in some other periods before or after 2010. In the cases of Gold and Platinum, it was positive over almost the whole analysed period. The MSCI WORLD for developed markets index had a negative impact on Tea\_Mombasa and Groundnuts since 2014, Shrimps since 2010, Logs\_CM in various period since 2003, Rubber between 2003 and 2006, Phosphate\_rock since 2009, Dap between 2008 and 2016, Tsp in various periods since 2014, Potash between 2006 and 2014, and Silver between 1998 and 2002.

The MSCI EU index had a positive impact on oil prices around 2009; Tea\_Mombasa and Coconut\_oil between 2017 and 2020; Palm\_oil before 2001; Rice\_5 between 2004 and 2012; wheat prices since 2018; Sugar\_EU in various sub-periods of the whole analysed period; Logs\_CM since 2002; Logs\_MY, Sawnwood, and Plywood between 2015 and 2018; and Tin between 2010 and 2017. It had a negative impact on Coal\_ZA before 1997 and between 2001 and 2010, Gas\_US since 2017, Rice\_5 and Beef between 2014 and 2017, and Phosphate\_rock and Dap around 2012–2014.

The MSCI EM for the emerging markets index had a positive impact on the majority of commodities, and it was mostly after 2015. In particular, this was on oil prices, coal prices, Coconut\_oil, Groundnuts, Palm\_oil, Soybeans, Soybean\_oil, Shrimps, Sugar\_US, Potash, and metals. In the case of Gold, Platinum, and Silver, the impact was present even since 2010. Maize, rice prices, wheat prices, Banana, and Orange were impacted positively around 2005–2015. The impact was negative on Gas\_US between 2000 and 2010; Cocoa before 2001; Logs\_CM and Phosphate\_rock over almost the whole analysed period; Dap, Tsp, and Urea before 1998; Plywood before 1999; and Tsp since 2011.

The index representing the Chinese stock market had a slightly more negative than positive impact. It became an important price predictor in around 2008. In particular, it had a negative impact on oil prices around 2009 and since 2015; coal prices since 2017; Gas\_EU and Gas\_JP since 2010; Rice\_5 between 2008 and 2014; both rice prices around 2009 and around 2011; Beef since 2015; Sawnwood since 2006; Gold since 2016; and Copper, Lead, and Tin around 2016. It had a positive impact on Tea\_Mombasa since 2016, Fish\_meal since 2013, Maize between 2005 and 2008, Chicken since 2016, Cotton and Rubber around 2020, Phosphate\_rock between 2009 and 2015, Iron since 2009, Tin in various periods before 2008, and Nickel between 2009 and 2015.

The share of BRIC countries' trade in the total global trade had a mixed impact. In particular, it was positive on Gas\_EU between 2014 and 2017, coffee prices before 1999 and between 2010 and 2014, Fish\_meal between 2008 and 2020, Rice\_100 around 1997, between 2004 and 2007 and between 2014 and 2017, wheat prices between 2017 and 2020, Shrimps before 2007 and between 2010 and 2013, Logs\_MY and Plywood between 2009 and 2011, Cotton between 2010 and 2015, Dap and Tsp before 1998, Potash between 2009 and 2018, and Copper before 1998 and between 2006 and 2009. It was negative on oil prices around 2018; coal prices since 2018; Coffee\_Robusta, Tea\_Colombo and Tea\_Kolkata around 2019; Coconut\_oil and Groundnuts since 2018; sugar prices around 2016–2017; Sugar\_World before 2004; Tobacco between 2001 and 2016; Sawnwood before 2006; Phosphate\_rock since 2009; Dap, Tsp, and Urea around 2010–2014; Iron between 2009 and 2013; Aluminium, Lead, Nickel, and Silver around 2019; Zinc between 2016 and 2019; Gold between 1996 and 2004, between 2011 and 2014, and since 2019; and Platinum between 1999 and 2001 and between 2009 and 2019.

In the case of leading indicators, those for the U.S. had a rather negative impact, whereas those for G7 countries, the Euro area, and China had a mostly positive impact.

Similarly, as before, the variables linked with the Chinese market and the Euro area became important mostly after 2008.

In particular, the leading indicator for the U.S. had a negative impact on oil and coal prices since 2020, Gas\_EU between 2010 and 2013, Gas\_JP around 2010 and since 2016, Groundnuts between 2013 and 2017, Fish\_meal before 2013, Palm\_oil and Soybean\_oil around 2006, Soybean\_meal between 1997 and 200, Sugar\_US between 2009 and 2020, Tobacco between 1998 and 2009, Phosphate\_rock between 2001 and 2007, Potash between 2009 and 2013, Iron since 2001, Zinc and Platinum around 2008, and Gold between 2015 and 2019. It had a positive impact on coffee prices before 1997, logs prices before 1998, Plywood and Cotton around 1998; Phosphate\_rock since 2009; Aluminium and Copper around 1997; Tin, Platinum and Silver around 2008; and, moreover, Platinum between 2000 and 2002.

The leading indicator for the G7 countries had a positive impact on oil prices around 2009, Coal\_AU and Gas\_JP in various sub-periods of the whole analysed period; Gas\_EU between 2002 and 2004 and between 2010 and 2013; Coconut\_oil between 2015 and 2019; Fish\_meal between 1998 and 2009; wheat prices around 2016–2017; Tobacco between 1998 and 2009; Logs\_CM between 2008 and 2018; Cotton between 2013 and 2018; Dap and Tsp around 2020; Ureal between 2016 and 2020; Iron between 2009 and 2017; Copper before 2009; Tin, Nickel and Zinc around 2008; and Platinum around 2000, between 2002 and 2004, and between 2008 and 2011. It had a negative impact on oil prices since 2020, Coconut\_oil between 2001 and 2003, Groundnuts between 2013 and 2018, wheat prices around 1997, Logs\_MY, Sawnwood and Plywood before 2000, Phosphate\_rock since 2009, Potash since 2014, Gold since 2013, and Silver since 2020.

The leading indicator for the Euro area had a positive impact on oil prices around 2009 and since 2020, coal prices since 2008, Groundnuts since 2013, Palm\_oil, Soybean\_oil and Maize between 2008 and 2019, rice prices around 2008–2009, wheat prices before 1998 and between 2007 and 2014, logs prices between 2016 and 2020, Rubber between 1997 and 2005, Phosphate\_rock, Dap, Tsp and Potash since 2008, and Gold since 2013. It had a negative impact on Gas\_EU between 2010 and 2014, Gas\_JP between 2001 and 2007, Coffee\_Arabica between 1997 and 2000 and between 2002 and 2005, Tobacco between 1998 and 2009 and between 2016 and 2017, Iron between 2005 and 2008, Copper around 2008 and since 2020, and Zinc between 2006 and 2013.

The leading indicator for China had a positive impact on oil prices since 2020; coal prices between 2009 and 2011; rice prices between 2009 and 2017; wheat prices before 1999; Chicken between 2008 and 2016; Phosphate\_rock between 2004 and 2008; Dap between 2008 and 2011 and between 2014 and 2020; Tsp between 2008 and 2020, Potash between 2011 and 2015; and Tea\_Colombo, Palm\_oil, Soybean\_oil, Sugar\_World, Sawnwood, Cotton, Copper, Tin, Zinc, Platinum, and Silver since 2020. It had a negative impact on oil prices around 2006, Gas\_JP between 2001 and 2008, between 2011 and 2015 and since 2017, Logs\_CM before 2002, Cotton and Rubber between 2013 and 2019, Phosphate\_rock between 2009 and 2016, Iron between 2005 and 2013 and between 2017 and 2020, Copper around 2014–2018, Nickel and Zinc between 2008 and 2019, and Gold between 2006 and 2007.

## References

- Abd Elaziz, Mohamed, Ahmed A. Ewees, and Zakaria Alameer. 2020. Improving adaptive neuro-fuzzy inference system based on a modified salp swarm algorithm using genetic algorithm to forecast crude oil price. *Natural Resources Research* 29: 2671–86. [\[CrossRef\]](#)
- Aguilar-Rivera, Rubén, Manuel Valenzuela-Rendon, and J. Rodriguez-Ortiz. 2015. Genetic algorithms and Darwinian approaches in financial applications: A survey. *Expert Systems with Applications* 42: 7684–97. [\[CrossRef\]](#)
- Ahumada, Hildegart, and Magdalena Cornejo. 2015. Explaining commodity prices by a cointegrated time series-cross section model. *Empirical Economics* 48: 1667–90. [\[CrossRef\]](#)
- Akram, Qaisar Farooq. 2009. Commodity prices, interest rates and the dollar. *Energy Economics* 31: 838–51. [\[CrossRef\]](#)
- Alam, Md Rafayet, and Scott Gilbert. 2017. Monetary policy shocks and the dynamics of agricultural commodity prices: Evidence from structural and factor-augmented VAR analyses. *Agricultural Economics* 48: 15–27. [\[CrossRef\]](#)

- Alameer, Zakaria, Ahmed Fathalla, Kenli Li, Haiwang Ye, and Jianhua Zhang. 2020. Multistep-ahead forecasting of coal prices using a hybrid deep learning model. *Resources Policy* 65: 101588. [CrossRef]
- Alameer, Zakaria, Mohamed Abd Elaziz, Ahmed A. Ewees, Haiwang Ye, and Jianhua Zhang. 2019a. Forecasting copper prices using hybrid adaptive neuro-fuzzy inference system and genetic algorithms. *Natural Resources Research* 28: 1385–401. [CrossRef]
- Alameer, Zakaria, Mohamed Abd Elaziz, Ahmed A. Ewees, Haiwang Ye, and Jianhua Zhang. 2019b. Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm. *Resources Policy* 61: 250–60. [CrossRef]
- Algieri, Bernardina, Matthias Kalkuhl, and Nicolas Koch. 2017. A tale of two tails: Explaining extreme events in financialized agricultural markets. *Food Policy* 69: 256–69. [CrossRef]
- Aloui, Riadh, Mohamed Safouane Ben Aissa, and Duc Khuong Nguyen. 2013a. Conditional dependence structure between oil prices and exchange rates: A copula-GARCH approach. *Journal of International Money and Finance* 32: 719–38. [CrossRef]
- Aloui, Riadh, Shawkat Hammoudeh, and Duc Khuong Nguyen. 2013b. A time-varying copula approach to oil and stock market dependence: The case of transition economies. *Energy Economics* 39: 208–21. [CrossRef]
- Al-Qudsi, Sulayman. 2010. Oil and commodity price volatility: Origins and impact on the Arab economy and capital markets. *Geopolitics of Energy* 32: 3–24.
- Alquist, Ron, Lutz Kilian, and Robert Vigfusson. 2013. Forecasting the price of oil. In *Handbook of Economic Forecasting* 2. Edited by Graham Elliott, C. Granger and Allan Timmermann. Amsterdam: Elsevier, pp. 427–507.
- Andreasson, Pierre, Stelios Bekiros, Duc Khuong Nguyen, and Gazi Salah Uddin. 2016. Impact of speculation and economic uncertainty on commodity markets. *International Review of Financial Analysis* 43: 115–27. [CrossRef]
- Apergis, Nicholas, and James E. Payne. 2010. Renewable energy consumption and economic growth: Evidence from a panel of OECD countries. *Energy Policy* 38: 656–60. [CrossRef]
- Arango, Luis, Fernando Arias, and Adriana Florez. 2012. Determinants of commodity prices. *Applied Economics* 44: 135–45. [CrossRef]
- Arora, Vipin, and Matthew Tanner. 2013. Do oil prices respond to real interest rates? *Energy Economics* 36: 546–55. [CrossRef]
- Arouri, Mohamed El Hedi, Jamel Jouini, and Duc Khuong Nguyen. 2011. Volatility spillovers between oil prices and stock sector returns: Implications for portfolio management. *Journal of International Money and Finance* 30: 1387–405. [CrossRef]
- Arouri, Mohamed El Hedi, Thanh Huong Dinh, and Duc Khuong Nguyen. 2010. Time-varying predictability in crude-oil markets: The case of GCC countries. *Energy Policy* 38: 4371–80. [CrossRef]
- Arslan-Ayaydin, Özgür, and Inna Khagleeva. 2013. The dynamics of crude oil spot and futures markets. In *Energy Economics and Financial Markets*. Edited by André Dorsman, John L. Simpson and Wim Westerman. Berlin: Springer, pp. 159–73.
- Atil, Ahmed, Amine Lahiani, and Duc Khuong Nguyen. 2014. Asymmetric and nonlinear pass-through of crude oil prices to gasoline and natural gas prices. *Energy Policy* 65: 567–73. [CrossRef]
- Ayres, Joao, Constantino Hevia, and Juan Pablo Nicolini. 2020. Real exchange rates and primary commodity prices. *Journal of International Economics* 122: 103261. [CrossRef]
- Bal, Debi Prasad, and Badri Narayan Rath. 2015. Nonlinear causality between crude oil price and exchange rate: A comparative study of China and India. *Energy Economics* 51: 149–56.
- Banerjee, Debanjan, Arijit Ghosal, and Imon Mukherjee. 2019. Prediction of gold price movement using geopolitical risk as a factor. *Advances in Intelligent Systems and Computing* 814: 879–86.
- Banner, Katharine M., and Megan D. Higgs. 2016. Considerations for assessing model averaging of regression coefficients. *Ecological Applications* 27: 78–93. [CrossRef] [PubMed]
- Barbieri, Maria, and James Berger. 2004. Optimal predictive model selection. *The Annals of Statistics* 32: 870–97. [CrossRef]
- Basher, Syed Abul, Alfred A. Haug, and Perry Sadorsky. 2012. Oil prices, exchange rates and emerging stock markets. *Energy Economics* 34: 227–40. [CrossRef]
- Baumeister, Christiane, and Lutz Kilian. 2015. Forecasting the real price of oil in a changing world: A forecast combination approach. *Journal of Business and Economic Statistics* 33: 338–51. [CrossRef]
- Bekiros, Stelios, Rangan Gupta, and Alessia Paccagnini. 2015. Oil price forecastability and economic uncertainty. *Economics Letters* 132: 125–28. [CrossRef]
- Belmonte, Miguel, and Gary Koop. 2014. Model switching and model averaging in time-varying parameter regression models. *Advances in Econometrics* 34: 45–69.
- Benmoussa, Amor Aniss, Reinhard Ellwanger, and Stephen Snudden. 2020. *The New Benchmark for Forecasts of the Real Price of Crude Oil*. Working Papers of Bank of Canada 39. Ottawa: Bank of Canada.
- Bernabe, Araceli, Esteban Martina, Jose Alvarez-Ramirez, and Carlos Ibarra-Valdez. 2004. A multi-model approach for describing crude oil price dynamics. *Physica A: Statistical Mechanics and its Applications* 338: 567–84. [CrossRef]
- Bernardi, Mauro, and Leopoldo Catania. 2018. The model confidence set package for R. *International Journal of Computational Economics and Econometrics* 8: 144–58. [CrossRef]
- Bhattacharya, Maumita, Rafiqul Islam, and Jemal Abawajy. 2016. Evolutionary optimization: A big data perspective. *Journal of Network and Computer Applications* 59: 416–26. [CrossRef]
- Bistline, John E. 2014. Natural gas, uncertainty, and climate policy in the US electric power sector. *Energy Policy* 74: 433–42. [CrossRef]
- Bloom, Nicholas. 2009. The impact of uncertainty shocks. *Econometrica* 77: 623–85.
- Bloomberg. 2022. S&P GSCI Commodity Total Return Index. Available online: <https://www.bloomberg.com/quote/SPGSCITR:IND> (accessed on 1 December 2022).

- Borychowski, Michał, and Andrzej Czyzewski. 2015. Determinants of prices increase of agricultural commodities in a global context. *Management* 19: 152–67. [CrossRef]
- Brabazon, Anthony, Michael Kampouridis, and Michael O'Neill. 2020. Applications of genetic programming to finance and economics: Past, present, future. *Genetic Programming and Evolvable Machines* 21: 33–53. [CrossRef]
- Brown, Pablo Pincheira, and Nicolás Hardy. 2019. Forecasting base metal prices with the Chilean exchange rate. *Resources Policy* 62: 256–81. [CrossRef]
- Brown, Stephen P. A., and Mine K. Yucel. 2008. What drives natural gas prices? *The Energy Journal* 29: 45–60. [CrossRef]
- Buncic, Daniel, and Carlo Moretto. 2015. Forecasting copper prices with dynamic averaging and selection models. *The North American Journal of Economics and Finance* 33: 1–38. [CrossRef]
- Burnham, Kenneth, and David R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information*. Berlin: Springer.
- Buyuksahin, Bahattin, and Michel A. Robe. 2014. Speculators, commodities and cross-market linkages. *Journal of International Money and Finance* 42: 38–70. [CrossRef]
- Byrne, Joseph, Giorgio Fazio, and Norbert Fiess. 2013. Primary commodity prices: Co-movements, common factors and fundamentals. *Journal of Development Economics* 101: 16–26. [CrossRef]
- Byun, Sung. 2017. Speculation in commodity futures markets, inventories and the price of crude oil. *Energy Journal* 38: 93–113. [CrossRef]
- Cade, Brian S. 2015. Model averaging and muddled multimodel inferences. *Ecology* 96: 2370–82. [CrossRef]
- Caginalp, Gunduz, and Mark DeSantis. 2011. Nonlinearity in the dynamics of financial markets. *Nonlinear Analysis: Real World Applications* 12: 1140–51. [CrossRef]
- Caldara, Dario, and Matteo Iacoviello. 2022a. Measuring geopolitical risk. *American Economic Review* 112: 1194–225. [CrossRef]
- Caldara, Dario, and Matteo Iacoviello. 2022b. Measuring Geopolitical Risk. Available online: <https://matteoiacoviello.com/gpr.htm> (accessed on 1 December 2022).
- Carmona, René. 2015. Financialization of the commodities markets: A non-technical introduction. In *Commodities, Energy and Environmental Finance*. Edited by R. Aid, M. Ludkovski and R. Sircar. New York: Springer, pp. 3–37.
- Cashin, Paul, Luis F. Cespedes, and Ratna Sahay. 2004. Commodity currencies and the real exchange rate. *Journal of Development Economics* 75: 239–68. [CrossRef]
- CBOE. 2022. VIX Historical Price Data. Available online: [https://www.cboe.com/tradable\\_products/vix/vix\\_historical\\_data](https://www.cboe.com/tradable_products/vix/vix_historical_data) (accessed on 1 December 2022).
- Ceperic, Vladimir, Niko Bako, and A. Baric. 2014. A symbolic regression-based modelling strategy of AC/DC rectifiers for RFID applications. *Expert Systems with Applications* 41: 7061–67. [CrossRef]
- Chai, Jian, Quanying Lu, Yi Hu, Shouyang Wang, Kin Lai, and Hongtao Liu. 2018. Analysis and Bayes statistical probability inference of crude oil price change point. *Technological Forecasting and Social Change* 126: 271–83. [CrossRef]
- Chen, Pei-Fen, Chien-Chiang Lee, and Jhih-Hong Zeng. 2014. The relationship between spot and futures oil prices: Do structural breaks matter? *Energy Economics* 43: 206–17. [CrossRef]
- Chen, Peng. 2015. Global oil prices, macroeconomic fundamentals and China's commodity sector comovements. *Energy Policy* 87: 284–94. [CrossRef]
- Chen, Qi, Bing Xue, Lin Shang, and Mengjie Zhang. 2016. Improving generalisation of genetic programming for symbolic regression with structural risk minimisation. In *Proceedings of the Genetic and Evolutionary Computation Conference GECCO 2016*, Denver, CO, USA, July 20–24; Edited by Tobias Friedrich and Frank Neumann. New York: Association for Computing Machinery, pp. 709–16.
- Chen, Shiu-Sheng. 2016. Commodity prices and related equity prices. *Canadian Journal of Economics* 49: 949–67. [CrossRef]
- Chen, Shiu-Sheng, and Hung-Chyn Chen. 2007. Oil prices and real exchange rates. *Energy Economics* 29: 390–404. [CrossRef]
- Chen, Yu-Chin, Kenneth Rogoff, and Barbara Rossi. 2010. Can exchange rates forecast commodity prices? *The Quarterly Journal of Economics* 125: 1145–94. [CrossRef]
- Chen, Yu-Chin, Kenneth Rogoff, and Barbara Rossi. 2012. Predicting agri-commodity prices: An asset pricing approach. In *Global Uncertainty and the Volatility of Agricultural Commodities Prices*. Edited by Bertrand Munier. Amsterdam: IOS Press, pp. 45–71.
- Chiou-Wei, Song-Zan, Sheng-Hung Chen, and Zhen Zhu. 2020. Natural gas price, market fundamentals and hedging effectiveness. *The Quarterly Review of Economics and Finance* 78: 321–37. [CrossRef]
- Chipman, Hugh A., Edward I. George, and Robert E. McCulloch. 1998a. BART: Bayesian additive regression trees. *The Annals of Applied Statistics* 4: 266–98. [CrossRef]
- Chipman, Hugh A., Edward I. George, and Robert E. McCulloch. 1998b. Bayesian CART model search. *Journal of the American Statistical Association* 93: 935–48. [CrossRef]
- Ciner, Cetin. 2017. Predicting white metal prices by a commodity sensitive exchange rate. *International Review of Financial Analysis* 52: 309–15. [CrossRef]
- Clark, Todd E., and Michael W. McCracken. 2009. Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review* 50: 363–95. [CrossRef]
- Claveria, Oscar, Enric Monte, and Salvador Torra. 2016. Quantification of survey expectations by means of symbolic regression via genetic programming to estimate economic growth in Central and Eastern European economies. *Eastern European Economics* 54: 171–89. [CrossRef]



- Claveria, Oscar, Enric Monte, and Salvador Torra. 2017. Evolutionary computation for macroeconomic forecasting. *Computational Economics* 51: 1–17. [CrossRef]
- Claveria, Oscar, Enric Monte, and Salvador Torra. 2022. A genetic programming approach for economic forecasting with survey expectations. *Applied Sciences* 12: 6661. [CrossRef]
- Clements, Kenneth W., and Renée Fry. 2008. Commodity currencies and currency commodities. *Resources Policy* 33: 55–73. [CrossRef]
- Commodity Futures Trading Commission. 2022. Historical Compressed. Available online: <https://www.cftc.gov/MarketReports/CommitmentsofTraders/HistoricalCompressed/index.htm> (accessed on 1 December 2022).
- Cornelius, Peter, and Jonathan Story. 2007. China and global energy markets. *Orbis* 51: 5–20. [CrossRef]
- Coulombe, P. G., Maxime Leroux, Dalibor Stevanovic, and Stéphane Surprenant. 2021. Macroeconomic data transformations matter. *International Journal of Forecasting* 37: 1338–54. [CrossRef]
- Cross, Jamie, and Bao Nguyen. 2017. The relationship between global oil price shocks and China's output: A time-varying analysis. *Energy Economics* 62: 79–91. [CrossRef]
- Cuaresma, Jesus Crespo, Jaroslava Hlouskova, and Michael Obsersteiner. 2018. Fundamentals, speculation or macroeconomic conditions? Modelling and forecasting Arabica coffee prices. *European Review of Agricultural Economics* 45: 583–615. [CrossRef]
- Cuaresma, Jesus Crespo, Jaroslava Hlouskova, and Michael Obsersteiner. 2021. Agricultural commodity price dynamics and their determinants: A comprehensive econometric approach. *Journal of Forecasting* 40: 1245–73. [CrossRef]
- Diaz-Rainey, Ivan, Helen Roberts, and David H. Lont. 2017. Crude inventory accounting and speculation in the physical oil market. *Energy Economics* 66: 508–22. [CrossRef]
- Diebold, Francis X., and Robert S. Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13: 253–63. [CrossRef]
- Dimoulkas, Ilias, Lars Herre, Dina Khastieva, Elis Nycander, Mikael Amelin, and Peyman Mazidi. 2018. A hybrid model based on symbolic regression and neural networks for electricity load forecasting. Paper presented at the 2018 15th International Conference on the European Energy Market (EEM), Lodz, Poland, June 27–29; New York: IEEE, pp. 1–5.
- Dogan, Eyup. 2016. The relationship between economic growth, energy consumption and trade. *Bulletin of Energy Economics* 4: 70–80.
- Dong, Baomin, Xuefeng Li, and Boqiang Lin. 2010. Forecasting long-run coal price in China: A shifting trend time-series approach. *Review of Development Economics* 14: 499–519. [CrossRef]
- Downes, John, and Jordan Elliot Goodman. 2018. *Dictionary of Finance and Investment Terms*. Hauppauge: Barron's Educational Series, Inc.
- Drachal, Krzysztof. 2016. Forecasting spot oil price in a dynamic model averaging framework—Have the determinants changed over time? *Energy Economics* 60: 35–46. [CrossRef]
- Drachal, Krzysztof. 2018a. Determining time-varying drivers of spot oil price in a Dynamic Model Averaging framework. *Energies* 11: 1207. [CrossRef]
- Drachal, Krzysztof. 2018b. Some novel Bayesian model combination schemes: An application to commodities prices. *Sustainability* 10: 2801. [CrossRef]
- Drachal, Krzysztof. 2020. Dynamic Model Averaging in economics and finance with fDMA: A package for R. *Signals* 1: 47–99. [CrossRef]
- Du, Limin, and Yanan He. 2015. Extreme risk spillovers between crude oil and stock markets. *Energy Economics* 51: 455–65. [CrossRef]
- Duc Huynh, Toan Luu Duc, Tobias Burggraf, and Muhammad Ali Nasir. 2020. Financialisation of natural resources & instability caused by risk transfer in commodity markets. *Resources Policy* 66: 101620. [CrossRef]
- EIA. 2020. China's Crude Oil Imports Surpassed 10 Million Barrels per Day in 2019. Available online: <https://www.eia.gov/todayinenergy/detail.php?id=43216> (accessed on 1 December 2022).
- EIA. 2022. U.S. Energy Information Administration. Available online: <https://www.eia.gov> (accessed on 1 December 2022).
- Eiben, A., and Jim Smith. 2015. *Introduction to Evolutionary Computing*. Berlin: Springer.
- Etienne, Xiaoli L., Scott H. Irwin, and Philip Garcia. 2018. Speculation and corn prices. *Applied Economics* 50: 4724–44. [CrossRef]
- Ewees, Ahmed A., Mohamed Abd Elaziz, Zakaria Alameer, Haiwang Ye, and Jianhua Zhang. 2020. Improving multilayer perceptron neural network using chaotic grasshopper optimization algorithm to forecast iron ore price volatility. *Resources Policy* 65: 101555. [CrossRef]
- Fattouh, Bassam, and Pasquale Scaramozzino. 2011. Uncertainty, expectations, and fundamentals: Whatever happened to long-term oil prices? *Oxford Review of Economic Policy* 27: 186–206. [CrossRef]
- Fattouh, Bassam, Lutz Kilian, and Lavan Mahadeva. 2013. The role of speculation in oil markets: What have we learned so far? *The Energy Journal* 34: 20–30. [CrossRef]
- Fernandez-Diaz, Jose M., and Bruce Morley. 2019. Interdependence among agricultural commodity markets, macroeconomic factors, crude oil and commodity index. *Research in International Business and Finance* 47: 174–94. [CrossRef]
- Fishe, Raymond P.H., and Aaron Smith. 2019. Do speculators drive commodity prices away from supply and demand fundamentals? *Journal of Commodity Markets* 15: 100078. [CrossRef]
- Frankel, Jeffrey, and Andrew K. Rose. 2010. *Determinants of Agricultural and Mineral Commodity Prices*. Cambridge: Harvard University, John F. Kennedy School of Government.
- FRED. 2015. FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics* 34: 574–89. [CrossRef]

- FRED. 2022. Economic Data. Available online: <https://fred.stlouisfed.org> (accessed on 1 December 2022).
- Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33: 1–22. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fuad, Mohd Nazri Mohd, and Mohd Azlan Hussain. 2015. Systematic design of chemical reactors with multiple stages via multi-objective optimization approach. In *Computer Aided Chemical Engineering*. Edited by K. V. Gernaey, J. K. Huusom and R. Gani. Oxford: Elsevier, vol. 37, pp. 869–74.
- Funashima, Yoshito. 2020. Global economic activity indexes revisited. *Economics Letters* 193: 109269. [\[CrossRef\]](#)
- Galipaud, Matthias, Mark A. F. Gillingham, Morgan David, and François-Xavier Dechaume-Moncharmont. 2014. Ecologists overestimate the importance of predictor variables in model averaging: A plea for cautious interpretations. *Methods in Ecology and Evolution* 5: 983–91. [\[CrossRef\]](#)
- Gangopadhyay, Kausik, Abhishek Jangir, and Rudra Sensarma. 2016. Forecasting the price of gold: An error correction approach. *IIMB Management Review* 28: 6–12. [\[CrossRef\]](#)
- Garcia, Diego, and Werner Kristjanpoller. 2019. An adaptive forecasting approach for copper price volatility through hybrid and non-hybrid models. *Applied Soft Computing Journal* 74: 466–78. [\[CrossRef\]](#)
- Gargano, Antonio, and Allan Timmermann. 2014. Forecasting commodity price indexes using macroeconomic and financial predictors. *International Journal of Forecasting* 30: 825–43. [\[CrossRef\]](#)
- Geman, Hélyette, and William Smith. 2013. Theory of storage, inventory and volatility in the LME base metals. *Resources Policy* 38: 18–28. [\[CrossRef\]](#)
- Ghalayini, Latife. 2017. Modeling and forecasting spot oil price. *Eurasian Business Review* 7: 355–73. [\[CrossRef\]](#)
- Ghoshray, Atanu, and Madhavi Pundit. 2021. Economic growth in China and its impact on international commodity prices. *International Journal of Finance and Economics* 26: 2776–89. [\[CrossRef\]](#)
- Giacomini, Raffaella, and Barbara Rossi. 2010. Forecast comparisons in unstable environments. *Journal of Applied Econometrics* 25: 595–620. [\[CrossRef\]](#)
- Golafshani, Emadaldin Mohammadi, and Ashraf Ashour. 2016. Prediction of self-compacting concrete elastic modulus using two symbolic regression techniques. *Automation in Construction* 64: 7–19. [\[CrossRef\]](#)
- Gramacy, Robert B. 2019. Monomvn: Estimation for MVN and Student-t Data with Monotone Missingness. Available online: <https://CRAN.R-project.org/package=monomvn> (accessed on 1 December 2022).
- Green, Peter J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82: 711–32. [\[CrossRef\]](#)
- Guidolin, Massimo, and Manuela Pedio. 2021. Forecasting commodity futures returns with stepwise regressions: Do commodity-specific factors help? *Annals of Operations Research* 299: 1317–56. [\[CrossRef\]](#)
- Guimera, Roger, Ignasi Reichardt, Antoni Aguilar-Mogas, Francesco A. Massucci, Manuel Miranda, Jordi Pallares, and Marta Sales-Prado. 2020. A Bayesian machine scientist to aid in the solution of challenging scientific problems. *Science Advances* 6: eaav6971. [\[CrossRef\]](#)
- Guzman, Juan Ignacio, and Enrique Silva. 2018. Copper price determination: Fundamentals versus non-fundamentals. *Mineral Economics* 31: 283–300. [\[CrossRef\]](#)
- Haeri, Maryam Amir, Mohammad Mehdi Ebadzadeha, and Gianluigi Folino. 2017. Statistical genetic programming for symbolic regression. *Applied Soft Computing* 60: 447–69. [\[CrossRef\]](#)
- Haider, C., F. O. de Franca, B. Burlacu, and G. Kronberger. 2023. Shape-constrained multi-objective genetic programming for symbolic regression. *Applied Soft Computing* 132: 109855. [\[CrossRef\]](#)
- Haigh, Michael S. 2018. Fundamentals and commodity prices. In *Commodities: Markets, Performance, and Strategies*. Edited by Kent H. Baker, Greg Filbeck and Jeffrey H. Harris. Oxford: Oxford University Press, pp. 90–108.
- Hamid, Mohd Fahmi Abdul, and Ani Shabri. 2017. Palm oil price forecasting model: An autoregressive distributed lag (ARDL) approach. *AIP Conference Proceedings* 1842: 030026. [\[CrossRef\]](#)
- Hamilton, James D. 2009. Causes and consequences of the oil shock of 2007–2008. *Brookings Papers on Economic Activity* 40: 215–59. [\[CrossRef\]](#)
- Hansen, Peter R., Asger Lunde, and James Nason. 2011. The model confidence set. *Econometrica* 79: 453–97. [\[CrossRef\]](#)
- Hara, Akira, Jun-ichi Kushida, and Tetsuyuki Takahama. 2019. Time series prediction using deterministic geometric semantic genetic programming. Paper presented at the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, October 6–9; New York: IEEE, pp. 1945–49.
- Harris, Charles R., K. Jarrod Millman, Stéfan J. Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, and et al. 2020. Array programming with NumPy. *Nature* 585: 357–62. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hartley, Peter, and Kenneth Barry Medlock. 2014. The relationship between crude oil and natural gas prices: The role of the exchange rate. *The Energy Journal* 35: 25–44. [\[CrossRef\]](#)
- Harvey, David I., Neil M. Kellard, Jakob B. Madsen, and Mark E. Wohar. 2018. The resource curse, commodity prices and economic growth. In *Global Commodity Markets and Development Economics*. Edited by Stephan Pfaffenweller. London: Routledge, pp. 16–49.
- Harvey, David, Stephen Leybourne, and Paul Newbold. 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13: 281–91. [\[CrossRef\]](#)

- Hasheminia, Hamed, and Seyed Taghi Akhavan Niaki. 2006. A genetic algorithm approach to find the best regression/econometric model among the candidates. *Applied Mathematics and Computation* 183: 337–49. [CrossRef]
- Hassanat, Ahmad, Khalid Almohammadi, Esra'a Alkafaween, Eman Abunawas, Awni Hammouri, and V. B. Surya Prasath. 2019. Choosing mutation and crossover ratios for genetic algorithms—A review with a new dynamic approach. *Information* 10: 390. [CrossRef]
- Hastie, Trevor, and Brad Efron. 2013. lars: Least Angle Regression, Lasso and Forward Stagewise. Available online: <https://CRAN.R-project.org/package=lars> (accessed on 1 December 2022).
- Hastie, Trevor, and Robert Tibshirani. 2000. Bayesian backlifting. *Statistical Science* 15: 196–213.
- Hastings, W. Keith. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57: 97–109. [CrossRef]
- Hatzenbuehler, Patrick L., Philip C. Abbott, and Kenneth A. Foster. 2016. Agricultural commodity prices and exchange rates under structural change. *Journal of Agricultural and Resource Economics* 41: 204–24.
- Herrera, Gabriel Paes, Michel Constantino, Benjamin Miranda Tabak, Hemerson Pistori, Jen-Je Su, and Athula Naranpanawa. 2019. Long-term forecast of energy commodities price using machine learning. *Energy* 179: 214–21. [CrossRef]
- Hong, Harrison, and Motohiro Yogo. 2012. What does futures market interest tell us about the macroeconomy and asset prices? *Journal of Financial Economics* 105: 473–90. [CrossRef]
- Hotelling, Harold. 1931. The economics of exhaustible resources. *Journal of Political Economy* 39: 137–75. [CrossRef]
- Huang, Jianbai, Yingli Li, Hongwei Zhang, and Jinyu Chen. 2021. The effects of uncertainty measures on commodity prices from a time-varying perspective. *International Review of Economics and Finance* 71: 100–14. [CrossRef]
- Huang, Zhixing, Jinghui Zhong, Liang Feng, Yi Mei, and Wentong Cai. 2020. A fast parallel genetic programming framework with adaptively weighted primitives for symbolic regression. *Soft Computing* 24: 7523–39. [CrossRef]
- Hyndman, Rob J., and Anne B. Koehler. 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22: 679–88. [CrossRef]
- Hyndman, Rob J., and Yeasmin Khandakar. 2008. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software* 26: 1–22.
- Irwin, Scott H., Dwight R. Sanders, and Robert P. Merrin. 2009. Devil or angel? The role of speculation in the recent commodity price boom (and bust). *Journal of Agricultural and Applied Economics* 41: 377–91. [CrossRef]
- Irz, Xavier, Jyrki Niemi, and Xing Liu. 2013. Determinants of food price inflation in Finland—The role of energy. *Energy Policy* 63: 656–63. [CrossRef]
- Jacks, David S., and Martin Stuermer. 2020. What drives commodity price booms and busts? *Energy Economics* 85: 104035. [CrossRef]
- Ji, Qiang, Jiang-Bo Geng, and Ying Fan. 2014. Separated influence of crude oil prices on regional natural gas import prices. *Energy Policy* 70: 96–105. [CrossRef]
- Jin, Ying. 2021. A Bayesian MCMC Based Symbolic Regression Algorithm. Available online: <https://github.com/ying531/MCMC-SymReg> (accessed on 1 December 2022).
- Jin, Ying, Weilin Fu, Jian Kang, Jiadong Guo, and Jian Guo. 2019. Bayesian symbolic regression. *arXiv arXiv:1910.08892*.
- Juvenal, Luciana, and Ivan Petrella. 2014. Speculation in the oil market. *Journal of Applied Econometrics* 30: 621–49. [CrossRef]
- Kagraoka, Yusho. 2016. Common dynamic factors in driving commodity prices: Implications of a generalized dynamic factor model. *Economic Modelling* 52: 609–17. [CrossRef]
- Kaufmann, Robert K. 2011. The role of market fundamentals and speculation in recent price changes for crude oil. *Energy Policy* 39: 105–15. [CrossRef]
- Kaur, Gursimran, and Babli Dhiman. 2017. Dynamic linkage between Indian stock market and commodity market. *International Journal of Applied Business and Economic Research* 15: 401–11.
- Kaya, H. 2016. Forecasting the price of crude oil with multiple predictors. *Siyasal Bilgiler Fakultesi Dergisi (ISMUS)* 1: 133–51.
- Keijzer, Maarten. 2004. Scaled symbolic regression. *Genetic Programming and Evolvable Machines* 5: 259–69. [CrossRef]
- Kilian, Lutz. 2009. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review* 99: 1053–69. [CrossRef]
- Kilian, Lutz. 2019. Measuring global real economic activity: Do recent critiques hold up to scrutiny? *Economic Letters* 178: 106–10. [CrossRef]
- Kilian, Lutz, and Dan Murphy. 2014. The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics* 29: 454–78. [CrossRef]
- Kilian, Lutz, and Xiaoqing Zhou. 2018. Modeling fluctuations in the global demand for commodities. *Journal of International Money and Finance* 88: 54–78. [CrossRef]
- Killian, Lutz, and Bruce Hicks. 2013. Did unexpectedly strong economic growth cause the oil price shock of 2003–2008? *Journal of Forecasting* 32: 385–94. [CrossRef]
- Kim, Soohyeon, J. Baek, and Eunbyeong Heo. 2017. Buffer vs. speculation: A review on the role of crude oil inventory. *IAEE Energy Forum* 26: 13–14.
- Klotz, D., M. Herrnegger, and K. Schulz. 2017. Symbolic regression for the estimation of transfer functions of hydrological models. *Water Resources Research* 53: 9402–23. [CrossRef]
- Koop, G. 2017. Bayesian methods for empirical macroeconomics with big data. *Review of Economic Analysis* 9: 33–56. [CrossRef]



- Koop, Gary, and Dimitris Korobilis. 2011. UK macroeconomic forecasting with many predictors: Which models forecast best and when do they do so? *Economic Modelling* 28: 2307–18. [CrossRef]
- Koop, Gary, and Dimitris Korobilis. 2012. Forecasting inflation using Dynamic Model Averaging. *International Economic Review* 53: 867–86. [CrossRef]
- Koop, Gary, and Dimitris Korobilis. 2013. Large time-varying parameter VARs. *Journal of Econometrics* 177: 185–98. [CrossRef]
- Korns, Michael F. 2011. Accuracy in symbolic regression. In *Genetic Programming Theory and Practice IX*. Edited by Rick Riolo, Ekaterina Vladislavleva and Jason H. Moore. New York: Springer, pp. 129–51.
- Koza, J. 1998. *Genetic Programming*. Cambridge: MIT Press.
- Kronberger, Gabriel, Stefan Fink, Michael Kommenda, and Michael Affenzeller. 2011. Macro-economic time series modeling and interaction networks. In *Applications of Evolutionary Computation*. Edited by C. Chio, A. Brabazon, G. A. Caro, R. Drechsler, M. Farooq, J. Grahl, G. Greenfield, C. Prins, J. Romero, G. Squillero and et al. Berlin: Springer, pp. 101–10.
- Kubalik, Jiří, Erik Derner, and Robert Babuska. 2020. Symbolic regression driven by training data and prior knowledge. In *GECCO '20: Proceedings of the 2020 Genetic and Evolutionary Computation Conference*. New York: Association for Computing Machinery, pp. 958–66.
- La Cava, William, Patryk Orzechowski, Bogdan Burlacu, de F. Franca, Marco Virgolin, Ying Jin, Michael Kommenda, and Jason Moore. 2021. Contemporary symbolic regression methods and their relative performance. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*; Edited by J. Vanschoren and S. Yeung. Available online: [https://datasets-benchmarks-proceedings.neurips.cc/paper\\_files/paper/2021/file/c0c7c76d30bd3dcaefc96f40275bdc0a-Paper-round1.pdf](https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/file/c0c7c76d30bd3dcaefc96f40275bdc0a-Paper-round1.pdf) (accessed on 1 December 2022).
- Labys, Walter C. 2006. *Modeling and Forecasting Primary Commodity Prices*. London: Routledge.
- Lan, Gongjin, Jakub M. Tomczak, Diederik M. Roijers, and A. E. Eiben. 2022. Time efficiency in optimization with a Bayesian-evolutionary algorithm. *Swarm and Evolutionary Computation* 69: 100970. [CrossRef]
- Landajuela, Mikel, Chak Lee, Jiachen Yang, Ruben Glatt, Claudio P. Santiago, Ignacio Aravena, Terrell Mundhenk, Garrett Mulcahy, and Brenden K. Petersen. 2022. A unified framework for deep symbolic regression. *Advances in Neural Information Processing Systems* 35: 33985–98.
- LaRose, A. 2014. *Global Natural Gas Markets Overview*. Washington, DC: U.S. Energy Information Administration.
- Lee, Geum Yong. 1999. Genetic recursive regression for modeling and forecasting real-world chaotic time series. In *Advances in Genetic Programming* 3. Edited by L. Spector, W. B. Langdon, U.-M. O'Reilly and P. J. Angeline. Cambridge, MA: MIT Press, pp. 401–23.
- Li, Raymond, and Guy C. K. Leung. 2011. The integration of China into the world crude oil market since 1998. *Energy Policy* 39: 5159–166. [CrossRef]
- Li, Raymond, Roselyne Joyeux, and Ronald D. Ripple. 2014. International natural gas market integration. *The Energy Journal* 35: 159–79. [CrossRef]
- Linn, Scott C., and Zhen Zhu. 2004. Natural gas prices and the gas storage report: Public news and volatility in energy futures markets. *Journal of Futures Markets* 24: 283–313. [CrossRef]
- Liu, Li, Yudong Wang, Chongfeng Wu, and Wenfeng Wu. 2016. Disentangling the determinants of real oil prices. *Energy Economics* 56: 363–73. [CrossRef]
- Liu, Yunling, and Yansong Lv. 2020. Commodity price evaluation based on improved data mining methods. Paper presented at the 2020 International Conference on E-Commerce and Internet Technology (ECIT), Zhangjiajie, China, April 22–24; New York: IEEE, pp. 145–48.
- Lubbers, Johannes, and Peter Posch. 2016. Commodities' common factor: An empirical assessment of the markets' drivers. *Journal of Commodity Markets* 4: 28–40. [CrossRef]
- Mayer, Herbert, Andreas Rathgeber, and Markus Wanner. 2019. Financialization of metal markets: Does futures trading influence spot prices and volatility? *Resources Policy* 53: 300–16. [CrossRef]
- McKinney, Wes. 2010. Data structures for statistical computing in Python. Paper presented at the 9th Python in Science Conference, Austin, TX, USA, June 28–July 3; vol. 445, pp. 56–61.
- Medeiros, Marcelo C., Gabriel F. R. Vasconcelos, Álvaro Veiga, and Eduardo Zilberman. 2019. Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics* 39: 98–119.
- Mensi, Walid, Makram Beljid, Adel Boubaker, and Shunsuke Managi. 2013. Correlations and volatility spillovers across commodity and stock markets: Linking energies, food, and gold. *Economic Modelling* 32: 15–22. [CrossRef]
- Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21: 1087–92. [CrossRef]
- Mohammadi, Hassan. 2011. Long-run relations and short-run dynamics among coal, natural gas and oil prices. *Applied Economics* 43: 129–37. [CrossRef]
- Moody's. 2022. Home. Available online: <https://www.moodys.com> (accessed on 1 December 2022).
- Mostafa, Mohamed M., and Ahmed A. El-Masry. 2016. Oil price forecasting using gene expression programming and artificial neural networks. *Economic Modelling* 54: 40–53. [CrossRef]
- MSCI. 2022. End of Day Index Data Search. Available online: <https://www.msci.com/end-of-day-data-search> (accessed on 1 December 2022).
- Mu, Xiaoyi. 2007. Weather, storage, and natural gas price dynamics: Fundamentals and volatility. *Energy Economics* 29: 46–63. [CrossRef]

- Narotam, Pradeep K., John F. Morrison, Michael D. Schmidt, and Narendra Nathoo. 2014. Physiological complexity of acute traumatic brain injury in patients treated with a brain oxygen protocol: Utility of symbolic regression in predictive modeling of a dynamical system. *Journal of Neurotrauma* 31: 630–41. [CrossRef]
- Nazlioglu, Saban, and Ugur Soytas. 2012. Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis. *Energy Economics* 34: 1098–104. [CrossRef]
- Nick, Sebastian, and Stefan Thoenes. 2014. What drives natural gas prices?—A structural VAR approach. *Energy Economics* 45: 517–27. [CrossRef]
- Nicolau, Miguel, and Alexandros Agapitos. 2021. Choosing function sets with better generalisation performance for symbolic regression models. *Genetic Programming and Evolvable Machines* 22: 73–100. [CrossRef]
- Nonejad, Nima. 2019. Crude oil price volatility dynamics and the great recession. *Applied Economics Letters* 26: 622–27. [CrossRef]
- Nonejad, Nima. 2020. A detailed look at crude oil price volatility prediction using macroeconomic variables. *Journal of Forecasting* 39: 1119–141. [CrossRef]
- Nurmakhanova, Mira. 2020. Oil and growth challenge in Kazakhstan. *International Journal of Economics and Business Research* 20: 100–16. [CrossRef]
- Obadi, Saleh Mothana, and Matej Korcek. 2020. Driving fundamentals of natural gas price in Europe. *International Journal of Energy Economics and Policy* 10: 318–24. [CrossRef]
- OECD. 2022. Main Economic Indicators. Available online: <https://doi.org/10.1787/data-00052-en> (accessed on 1 December 2022).
- Olsen, Kyle, James Mjelde, and David Bessler. 2015. Price formulation and the law of one price in internationally linked markets: An examination of the natural gas markets in the USA and Canada. *The Annals of Regional Science* 54: 117–42. [CrossRef]
- Onorante, Luca, and Adrian Raftery. 2016. Dynamic Model Averaging in large model spaces using dynamic Occam's window. *European Economic Review* 81: 2–14. [CrossRef] [PubMed]
- Orzechowski, Patryk, William La Cava, and Jason H. Moore. 2018. Where are we now?: A large benchmark study of recent symbolic regression methods. In *GECCO '18: Proceedings of the Genetic and Evolutionary Computation Conference*. Edited by Hernan Aguirre. New York: Association for Computing Machinery, pp. 1183–90.
- Osathanunkul, Rossarin, Chatchai Khiewngamdee, Woraphon Yamaka, and Songsak Sriboonchitta. 2018. The role of oil price in the forecasts of agricultural commodity prices. In *Predictive Econometrics and Big Data*. Edited by V. Kreinovich, S. Sriboonchitta and N. Chakpitak. Berlin: Springer, pp. 422–29.
- Ouyang, Ruolan, and Xuan Zhang. 2020. Financialization of agricultural commodities: Evidence from China. *Economic Modelling* 85: 381–89. [CrossRef]
- Pincheira, Pablo, and Nicolás Hardy. 2021. Forecasting aluminum prices with commodity currencies. *Resources Policy* 73: 102066. [CrossRef]
- Pincheira-Brown, Pablo, Andrea Bentancor, Nicolás Hardy, and Nabil Jarsun. 2022. Forecasting fuel prices with the Chilean exchange rate: Going beyond the commodity currency hypothesis. *Energy Economics* 106: 105802. [CrossRef]
- Prates, D. 2007. The recent rise of commodities prices. *Revista de Economía Política* 27: 323–44.
- R Core Team. 2018. *R: A Language and Environment for Statistical Computing*; Vienna: R Foundation for Statistical Computing. Available online: <https://www.R-project.org> (accessed on 1 December 2022).
- Raftery, Adrian E., Miroslav Kárný, and Pavel Ettler. 2010. Online prediction under model uncertainty via Dynamic Model Averaging: Application to a cold rolling mill. *Technometrics* 52: 52–66. [CrossRef]
- Reboredo, Juan. 2012. Modelling oil price and exchange rate co-movements. *Journal of Policy Modeling* 34: 419–40. [CrossRef]
- Reboredo, Juan Carlos, Miguel A. Rivera-Castro, and Gilney F. Zebende. 2014. Oil and US dollar exchange rate dependence: A detrended cross-correlation approach. *Energy Economics* 42: 132–39. [CrossRef]
- Regnier, Eva. 2007. Oil and energy price volatility. *Energy Economics* 29: 405–27. [CrossRef]
- Regolin, Evandro Nunes, and Aurora Trinidad Ramirez Pozo. 2005. Bayesian automatic programming. In *Genetic Programming*. Edited by M. Keijzer, A. Tettamanzi, P. Collet, J. van Hemert and M. Tomassini. Berlin: Springer, pp. 38–49.
- Rezitis, Anthony N., and Maria Sassi. 2013. Commodity food prices: Review and empirics. *Economics Research International* 2013: 694507. [CrossRef]
- Ribeiro, Celma O., and Sydney M. Oliveira. 2011. A hybrid commodity price-forecasting model applied to the sugar-alcohol sector. *Australian Journal of Agricultural and Resource Economics* 55: 180–98. [CrossRef]
- Riggi, Marianna, and Fabrizio Venditti. 2015. The time varying effect of oil price shocks on euro-area exports. *Journal of Economic Dynamics and Control* 59: 75–94. [CrossRef]
- Rubaszek, Michał, and Gazi Salah Uddin. 2020. The role of underground storage in the dynamics of the US natural gas market: A threshold model analysis. *Energy Economics* 87: 104713. [CrossRef]
- Rueda, R., M. P. Cuellar, M. C. Pegalajar, and M. Delgado. 2019a. Straight line programs for energy consumption modelling. *Applied Soft Computing Journal* 80: 310–28. [CrossRef]
- Rueda, R., M. P. Cuellar, M. Molina-Solana, Y. Guo, and M. C. Pegalajar. 2019b. Generalised regression hypothesis induction for energy consumption forecasting. *Energies* 12: 1069. [CrossRef]
- Salisu, Afees A., Kazeem O. Isah, and Ibrahim D. Raheem. 2019. Testing the predictability of commodity prices in stock returns of G7 countries: Evidence from a new approach. *Resources Policy* 64: 101520. [CrossRef]

- Sarradj, Ennes, and Thomas Geyer. 2014. Symbolic regression modeling of noise generation at porous airfoils. *Journal of Sound and Vibration* 333: 3189–3202. [CrossRef]
- Schewe, Jacob, Christian Otto, and Katja Frieler. 2017. The role of storage dynamics in annual wheat prices. *Environmental Research Letters* 12: 054005. [CrossRef]
- Schiller, R. 2000. *Irrational Exuberance*. Princeton, NJ: Princeton University Press.
- Schiller, R. 2022. Online Data. Available online: <http://www.econ.yale.edu/~shiller/data.htm> (accessed on 1 December 2022).
- Senkerik, Roman, Adam Viktorin, Michal Pluhacek, Tomas Kadavy, and Ivan Zelinka. 2017a. Differential evolution driven analytic programming for prediction. In *Artificial Intelligence and Soft Computing*. Edited by L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh and J. Zurada. Cham: Springer, pp. 676–87.
- Senkerik, Roman, Adam Viktorin, Michal Pluhacek, Tomas Kadavy, and Ivan Zelinka. 2017b. Hybridization of analytic programming and differential evolution for time series prediction. In *Hybrid Artificial Intelligent Systems*. Edited by F. Martinez de Pison, R. Urraca, H. Quintian and E. Corchado. Cham: Springer, pp. 686–98.
- Sermpinis, Georgios, Charalampos Stasinakis, Konstantinos Theofilatos, and Andreas Karathanasopoulos. 2015. Modeling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms: Support vector regression forecast combinations. *European Journal of Operational Research* 247: 831–46. [CrossRef]
- Sheta, Alaa, Hossam Faris, and Mouhammd Alkasassbeh. 2013. A genetic programming model for S&P 500 stock market prediction. *International Journal of Control and Automation* 6: 303–14.
- Shilling, Henry. 1996. *The International Guide to Securities Market Indices*. Chicago: Routledge.
- Sinha, Ankur, Pekka Malo, and Timo Kuosmanen. 2015. A multiobjective exploratory procedure for regression model selection. *Journal of Computational and Graphical Statistics* 24: 154–82. [CrossRef]
- Smiech, Sławomir, and Monika Papież. 2013. Fossil fuel prices, exchange rate, and stockmarket: A dynamic causality analysis on the European market. *Economics Letters* 118: 199–202. [CrossRef]
- Smits, Guido F., and Mark Kotanchek. 2005. Pareto-front exploitation in symbolic regression. In *Genetic Programming Theory and Practice II*. Edited by Una-May O'Reilly, Tina Yu, Rick Riolo and Bill Worzel. Boston: Springer, pp. 283–99.
- Souza, Rodrigo da Silva, Leonardo B. de Mattos, and João E. de Lima. 2021. Commodity prices and the Brazilian real exchange rate. *International Journal of Finance and Economics* 26: 3152–72. [CrossRef]
- Steel, Mark F. 2020. Model averaging and its use in economics. *Journal of Economic Literature* 58: 644–719. [CrossRef]
- Steermmer, Martin. 2018. 150 years of boom and bust: What drives mineral commodity prices? *Macroeconomic Dynamics* 22: 702–17. [CrossRef]
- Stephens, T. 2021. Genetic Programming in Python, With a Scikit-Learn Inspired API: Gplearn. Available online: <https://github.com/trevorstephens/gplearn> (accessed on 1 December 2022).
- Stock, James H., and Mark W. Watson. 2004. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* 23: 405–30. [CrossRef]
- Stooq. 2022. Quotes. Available online: <https://stooq.com> (accessed on 1 December 2022).
- Sukcharoen, Kunlapath, and David Leatham. 2018. Analyzing extreme comovements in agricultural and energy commodity markets using a regular vine copula method. *International Journal of Energy Economics and Policy* 8: 193–201.
- Tan, Xiaofen, and Yongjiao Ma. 2017. The impact of macroeconomic uncertainty on international commodity prices: Empirical analysis based on TVAR model. *China Finance Review International* 7: 163–84. [CrossRef]
- Tapia Cortez, C. A., S. Saydam, J. Coulton, and C. Sammut. 2018. Alternative techniques for forecasting mineral commodity prices. *International Journal of Mining Science and Technology* 28: 309–22. [CrossRef]
- Tashman, Len. 2000. Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting* 16: 437–50. [CrossRef]
- The Pandas Development Team. 2020. pandas-dev/pandas: Pandas. Available online: <https://doi.org/10.5281/zenodo.3509134> (accessed on 1 December 2022).
- The World Bank. 2022. Commodities Markets. Available online: <https://www.worldbank.org/en/research/commodity-markets> (accessed on 1 December 2022).
- Thiyagarajan, Saloni, G. Naresh, and S. Mahalakshmi. 2015. Forecasting volatility in Indian agri-commodities market. *Global Business and Finance Review* 20: 95–104. [CrossRef]
- Tiwari, Aviral Kumar, Arif Billah Dar, and Niyati Bhanja. 2013. Oil price and exchange rates: A wavelet based analysis for India. *Economic Modelling* 31: 414–22. [CrossRef]
- Uddin, Gazi Salah, Aviral Kumar Tiwari, Mohamed Aroui, and Frédéric Teulon. 2013. On the relationship between oil price and exchange rates: A wavelet analysis. *Economic Modelling* 35: 502–7. [CrossRef]
- UNCTAD. 2012. Don't Blame the Physical Markets: Financialization Is the Root Cause of Oil and Commodity Price Volatility, Policy Brief 25. Available online: [https://unctad.org/system/files/official-document/presspb2012d1\\_en.pdf](https://unctad.org/system/files/official-document/presspb2012d1_en.pdf) (accessed on 1 December 2022).
- United Nations Statistics Division. 2022. Monthly Bulletin of Statistics. Available online: [http://comtrade.un.org/api/getmbsdata?series\\_type=T35.M.V.\\$&fmt=CSV](http://comtrade.un.org/api/getmbsdata?series_type=T35.M.V.$&fmt=CSV) (accessed on 1 December 2022).
- Van Rossum, G., and F. L. Drake, Jr. 1995. *Python Reference Manual*. Amsterdam: Centrum voor Wiskunde en Informatica.



- Vazquez, Daniel, Roger Guimera, Marta Sales-Prado, and Gonzalo Guillen-Gosalbez. 2022. Automatic modeling of socioeconomic drivers of energy consumption and pollution using Bayesian symbolic regression. *Sustainable Production and Consumption* 30: 596–607. [CrossRef]
- Wagner, Neal, Zbigniew Michalewicz, Moutaz Khouja, and Rob Roy McGregor. 2007. Time series forecasting for dynamic environments: The DyFor genetic program model. *IEEE Transactions on Evolutionary Computation* 11: 433–52. [CrossRef]
- Wang, Qiang, Shuyu Liu, and Rongrong Li. 2018. China's dependency on foreign oil will exceed 80% by 2030: Developing a novel NMGM-ARIMA to forecast China's foreign oil dependence from two dimensions. *Energy* 163: 151–67. [CrossRef]
- Wang, Qingfeng, and Xu Sun. 2017. Crude oil price: Demand, supply, economic activity, economic policy uncertainty and wars—From the perspective of Structural Equation Modelling (SEM). *Energy* 133: 483–90. [CrossRef]
- Wang, Shouyang, Lean Yu, and K. K. Lai. 2004. A novel hybrid AI system framework for crude oil price forecasting. *Lecture Notes in Computer Science* 3327: 233–42.
- Wang, Yudong, Bing Zhang, Xundi Diao, and Chongfeng Wu. 2015a. Commodity price changes and the predictability of economic policy uncertainty. *Economics Letters* 127: 39–42. [CrossRef]
- Wang, Yudong, Li Liu, and Chongfeng Wu. 2017. Forecasting the real prices of crude oil using forecast combinations over time-varying parameter models. *Energy Economics* 66: 337–48. [CrossRef]
- Wang, Yudong, Li Liu, Xundi Diao, and Chongfeng Wu. 2015b. Forecasting the real prices of crude oil under economic and statistical constraints. *Energy Economics* 51: 599–608. [CrossRef]
- Weiss, Mark Allen. 2014. *Data Structures and Algorithm Analysis in C++*. Upper Saddle River: Pearson Education, Inc.
- Weng, Futian, Muzhou Hou, Tianle Zhang, Yunlei Yang, Zheng Wang, Hongli Sun, Hao Zhu, and Jianshu Luo. 2018. Application of regularized extreme learning machine based on BIC criterion and genetic algorithm in iron ore price forecasting. In *Proceedings of the 2018 3rd International Conference on Modelling, Simulation and Applied Mathematics*. Edited by A. Luevanos Rojas, G. Ilewicz, D. J. Jakobczak and K. Weller. Paris: Atlantis Press, pp. 212–17.
- Winkler, Stephan M., Gabriel Kronberger, Michael Kommenda, Stefan Fink, and Michael Affenzeller. 2015a. Dynamics of predictability and variable influences identified in financial data using sliding window machine learning. In *Computer Aided Systems Theory—EUROCAST 2015*. Edited by Roberto Moreno-Díaz, Franz Pichler and Alexis Quesada-Arencibia. Cham: Springer, pp. 326–33.
- Winkler, Stephan M., Michael Affenzeller, Gabriel Kronberger, Michael Kommenda, Bogdan Burlacu, and Stefan Wagner. 2015b. Sliding window symbolic regression for detecting changes of system dynamics. In *Genetic Programming Theory and Practice XII*. Edited by Rick Riolo, William P. Worzel and Mark Kotanchek. Cham: Springer, pp. 91–107.
- Working, Holbrook. 1960. Speculation on hedging markets. *Food Research Institute Studies* 1: 185–220.
- WTO. 2022. Statistics on Merchandise Trade. Available online: <https://data.wto.org> (accessed on 1 December 2022).
- Xu, Yang, Liyan Han, Li Wan, and Libo Yin. 2019. Dynamic link between oil prices and exchange rates: A non-linear approach. *Energy Economics* 84: 104488. [CrossRef]
- Yan, Y., and L. Yuan. 2019. Commodity market financialisation, information transmission and driving factors: Research based on environmental intraday high frequency data of China. *Journal of Environmental Protection and Ecology* 20: 2047–58.
- Yang, Chi-Jen, Xiaowei Xuan, and Robert B. Jackson. 2012. China's coal price disturbances: Observations, explanations, and implications for global energy economies. *Energy Policy* 51: 720–27. [CrossRef]
- Yang, Guangfei, Tao Sun, Jianliang Wang, and Xianneng Li. 2015a. Modeling the nexus between carbon dioxide emissions and economic growth. *Energy Policy* 86: 104–17. [CrossRef]
- Yang, Guangfei, Xianneng Li, Jianliang Wang, Lian Lian, and Tiejun Ma. 2015b. Modeling oil production based on symbolic regression. *Energy Policy* 82: 48–61. [CrossRef]
- Yin, Libo, and Qingyuan Yang. 2016. Predicting the oil prices: Do technical indicators help? *Energy Economics* 56: 338–50. [CrossRef]
- Yin, Xuluo, Jiangang Peng, and Tian Tang. 2018. Improving the forecasting accuracy of crude oil prices. *Sustainability* 10: 454. [CrossRef]
- Yousefi, Ayoub, and Tony S. Wirjanto. 2004. The empirical role of the exchange rate on the crude-oil price information. *Energy Economics* 26: 783–99. [CrossRef]
- Yuan, Fong-Ching, Chao-Hui Lee, and Chaochang Chiu. 2020. Using market sentiment analysis and genetic algorithm-based least squares support vector regression to predict gold prices. *International Journal of Computational Intelligence Systems* 13: 234–46. [CrossRef]
- Zegklitz, J., and P. Posik. 2021. Benchmarking state-of-the-art symbolic regression algorithms. *Genetic Programming and Evolvable Machines* 22: 5–33. [CrossRef]
- Zhang, Byong-Tak. 2000. Bayesian methods for efficient genetic programming. *Genetic Programming and Evolvable Machines* 1: 217–42. [CrossRef]
- Zhang, Hui Jun, Jean-Marie Dufour, and John W. Galbraith. 2016. Exchange rates and commodity prices: Measuring causality at multiple horizons. *Journal of Empirical Finance* 36: 100–20. [CrossRef]
- Zhang, Yaojie, M. I. M. Wahab, and Yudong Wang. 2023. Forecasting crude oil market volatility using variable selection and common factor. *International Journal of Forecasting* 39: 486–502. [CrossRef]

- Zhao, Yang, Jianping Li, and Lean Yu. 2017. A deep learning ensemble approach for crude oil price forecasting. *Energy Economics* 66: 9–16. [[CrossRef](#)]
- Zhao, Yuan, Weiguo Zhang, Xue Gong, and Chao Wang. 2021. A novel method for online real-time forecasting of crude oil price. *Applied Energy* 303: 117588. [[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.