

Article

# A Machine-Learning-Based Approach for Natural Gas Futures Curve Modeling

Oleksandr Castello <sup>\*,†</sup>  and Marina Resta <sup>†</sup> 

Department of Economics and Business Studies, School of Social Sciences, University of Genoa,  
16126 Genoa, Italy; resta@economia.unige.it

\* Correspondence: alexander-castello@libero.it; Tel.: +39-345-878-2588

† These authors contributed equally to this work.

**Abstract:** This work studies the term structure dynamics in the natural gas futures market, focusing on the Dutch Title Transfer Facility (TTF) daily futures prices. At first, using the whole dataset, we compared the in-sample fitting performance of three models: the four-factor dynamic Nelson–Siegel–Svensson (4F-DNSS) model, the five-factor dynamic De Rezende–Ferreira (5F-DRF) model, and the B-spline model. Our findings suggest that B-spline is the method that achieves the best in-line fitting results. Then, we turned our attention to forecasting, using data from 20 January 2011 to 13 May 2022 as the training set and the remaining data, from 16 May to 13 June 2022, for day-ahead predictions. In this second part of the work we combined the above mentioned models (4F-DNSS, 5F-DRF and B-spline) with a Nonlinear Autoregressive Neural Network (NAR-NN), asking the NAR-NN to provide parameter tuning. All the models provided accurate out-of-sample prediction; nevertheless, based on extensive statistical tests, we conclude that, as in the previous case, B-spline (combined with an NAR-NN) ensured the best out-of-sample prediction.

**Keywords:** natural gas; futures prices term structure; Nelson–Siegel–Svensson model; De Rezende–Ferreira model; B-spline; artificial neural networks (ANN); Nonlinear Autoregressive Neural Networks (NAR-NNs)



**Citation:** Castello, O.; Resta, M. A Machine-Learning-Based Approach for Natural Gas Futures Curve Modeling. *Energies* **2023**, *16*, 4746. <https://doi.org/10.3390/en16124746>

Academic Editor: Ishaq Bhatti

Received: 15 May 2023

Revised: 8 June 2023

Accepted: 13 June 2023

Published: 15 June 2023



**Copyright:** © 2023 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

During recent decades, the liberalization and financialization wave [1,2] generated a rise in the importance of energy commodities as an alternative asset class within the global market. Futures markets played a fundamental role in the financialization process of energy commodities, indirectly fostered by markets increasing liquidity. In fact, the trading volume of energy futures is constantly expanding, and the increase in the exchange volumes in the Asia–Pacific (APAC) region is the main driver, with a share of 74% [3] of the worldwide trading activity in 2021. Furthermore, increasing returns and inflation have fuelled futures markets' expansion. Consider, for instance, the Euro area: the size of the energy derivatives market increased by 30% in the period January–June 2022, with over 1700 firms involved [4].

In this framework, the availability of proper techniques to model and predict energy futures term structure dynamics is of crucial importance, especially for Western European countries that, in light of unprecedented events such as the COVID-19 pandemic and the more recent Russia–Ukraine war, have to face new challenges to calibrate their policy priorities. As stated in [5], in fact, the European Union is not only one of the major global energy consumers, but it is highly reliant on Russia for imported gas which is, therefore, an important pawn in determining related energy policies. Moreover, as highlighted in [6], the events of the past years have dramatically revealed the many ways in which the energy transition and geopolitics are entangled, and, as predicted in [7] during the previous Russia–Ukraine crisis in 2014, natural gas has assumed a pivotal role in the geopolitics of energy

security in Europe, and it is now the second largest energy commodity behind oil, and the second fastest rising source of energy demand after renewables [8]. Acquiring proper knowledge of the term structure of the natural gas (NG) futures market is therefore helpful in reducing the exposure to price volatility and to assess proper energy policies in light of its key role in the *decarbonization* process and in the transition to sustainable development based on a highly efficient renewable energy system [9,10].

In this perspective, based on the role of energy futures as a hedging tool and indicator of markets trends, we analyzed the term structure of the NG futures market. The scope of this paper is twofold. First, we are interested in testing whether models conventionally employed on the bonds market can be also effective for in-sample fitting in the case of NG futures. Second, moving to the forecasting issue, we investigate the effectiveness of a technique combining yield curve models and machine learning. In detail, we used B-spline [11], four-factor dynamic Nelson–Siegel–Svensson (4F-DNSS) [12], and five-factor dynamic De Rezende–Ferreira (5F-DRF) [13] models for in-sample fitting and a hybrid method that combines the above three techniques to a Nonlinear Autoregressive Neural Network (NAR-NN) for out-of-sample forecasting of NG futures curves. Furthermore, the NAR-NN is also used in the discussion of the results as a benchmark for day-ahead predictions for the futures price time series.

The choice of the fitting models has various motivations: they make a parsimonious use of parameters, being, therefore, easy to handle; in addition, working with fixed-income assets, these models showed a notable ability to replicate the term structure dynamics [12,14–25]. Therefore, provided the similarity with the NG futures market in terms of the varying maturity of the data structure, we tested to what extent the above models can be reliable on the energy markets too. Furthermore, NAR-NNs, with their ease of configuration, gave proof of their ability on both one and multi-step ahead forecasts of time series [26–37], managing highly noisy and volatile data.

So far, the existing literature has mostly focused on the relations between NG and other commodities or securities (see, for instance, [38–47]), as well as on modeling price volatility (e.g., [48–55]), demand and supply (e.g., [56–64]), spot prices (e.g., [65–76]) or futures prices of individual contracts (e.g., [77–79]). Relatively less attention has been paid to NG futures prices term structure modeling and forecasting and only a few studies have partly tackled the issues we are dealing with. For example, Chiarella et al. [80] proposed a two-factor regime-switching volatility model enhanced with the Markov chain Monte Carlo estimation method to model the forward price curve; Almansour [81] studied futures curve dynamics with an extension of the Gibson and Schwartz [82] two-factor model in a regime-switching framework; Leonhardt et al. [83] used a geometric multi-factor model to deal with cointegration of the term structure, regime switching, and seasonality of futures prices. Furthermore, Karstanje et al. [84] studied futures prices comovements of the most traded commodities with a factor approach relying on the Diebold and Li [85] model; Jablonowski and Schicks [86] introduced a three-factor model based on Heath et al. [87] to describe the relationship between gas term structure and temperature forecasts. Finally, Tang et al. [88] developed a predictive method based on artificial neural networks and analyzed the impact of Google search data and internet news sentiment on the model's forecasting ability, Li [89] investigated the abilities of GARCH-type discrete-time models and different Poisson jump-diffusion models to fit NG futures data, while Horváth et al. [90] analyzed the forward curves of 24 different commodities with several polynomial interpolation techniques and provided a comparative study of the predictive abilities of methods based on functional autoregressive processes, Diebold and Li, and naïve approaches.

In light of the above, our study brings to the related literature some contributions that can be summarized as follows. First, we analyzed the stylized facts bridging NG futures and government securities markets to endorse the use of yield curve models in the former. Second, we used parametric yield curve models for in-sample fitting in the NG futures market, and third, we discussed a hybrid scheme with NAR-NNs for day-ahead predictions.

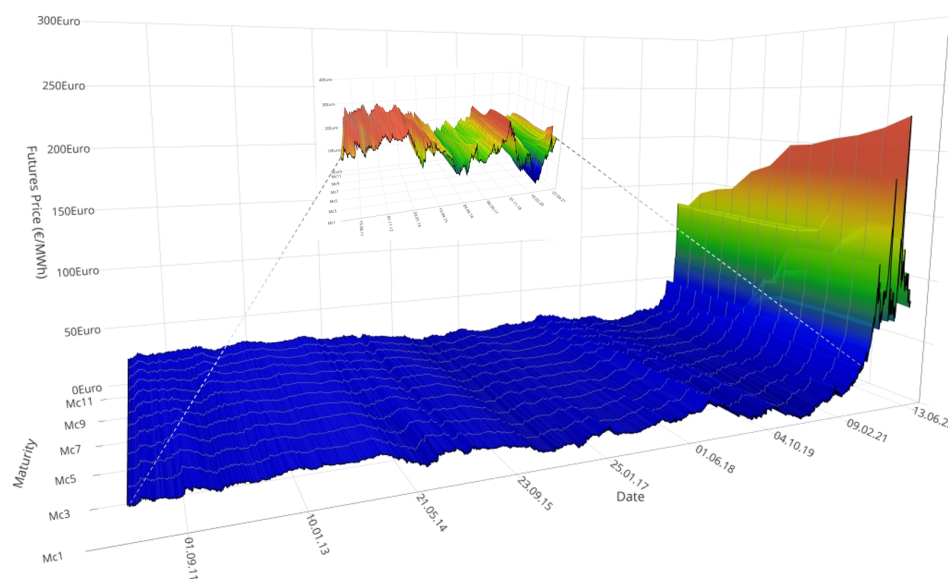
With this aim, we used a dataset of daily prices which spans various market conditions to validate the adequacy of the framework under very different situations.

The remainder of the paper is organized as follows: Section 2 analyzes the features and main stylized facts of the data set; Section 3 presents the methodologies in use for modeling and forecasting, respectively; Section 4 discusses the main results; Section 5 concludes the paper.

## 2. Data

We considered the dataset of daily settlement prices, quoted in €/MWh, of the Mc1–Mc12 natural gas futures contracts expiring in 1 to 12 month(s). The data cover the period from 20 January 2011 to 13 June 2022 for an overall number of 2916 observations. The daily futures prices were obtained from the Dutch Title Transfer Facility (TTF), the virtual trading hub which is the leading European gas trading platform [91], with the highest level of liquidity and highest share of trade. In 2020, the TTF overtook, for the first time, the world’s biggest NG market, Henry Hub, in terms of trading volume and open interest, and reached a new record in 2021, with approximately 1.94 million contracts [92].

Figure 1 displays the three-dimensional surface plot of NG futures curve data for the whole period, while the inset highlights the dynamics of the term structure for the period 20 January 2011–27 April 2021, which is visually flattened because of the severe rise in price level that occurred in 2021–2022. In this temporal frame, states of stability alternated with turbulence; significant upward and downward shifts of the price level at all of the maturities can be observed in various periods, such as between mid-2014 and early 2016, when the global economy faced one of the largest oil price declines due to the global economic slowdown and the surge in production from American shale producers and OPEC members.



**Figure 1.** 3D surface plot of the term structure of natural gas futures prices: the x-axis shows the time expressed in days, the y-axis represents the maturities from 1 month (Mc1) to 12 months (Mc12), and the z-axis the price of the contracts in Euros. The data spans 2916 trading days, from 20 January 2011 to 13 June 2022. The inset shows a zoomed-in area with the markets dynamics in the period 20 January 2011–27 April 2021 otherwise flattened.

A similar situation was replicated twice later: in 2017–2018, when OPEC agreed to cut oil production leading to an increase in the oil price and hence of the natural gas price, and in 2020–2021, with the most significant reduction in the NG price over the whole time span, due to the combined effect of the pandemic and the Russia–Saudi Arabia price war—the oil price dropped down to around 10 \$ per barrel, while WTI oil futures price were traded at

−37.63 \$ per barrel for the first time in history, causing the NG price to slide to 4.5 €/MWh on the European market.

More recently, in the period mid 2021–mid 2022, we observed a turmoil in the NG market caused by several interconnected factors: (i) the surging energy demand driven by the global economic recovery after the pandemic and by the hottest summer of the last century [93]; (ii) low levels of gas storage, with underground storage facilities less than 77% full throughout 2021 and less than 57% in December 2021 [94]; (iii) a shortage of traditional energy resources due to the investment contraction in the hydrocarbon sector and poor renewable performance caused by extreme weather events; (iv) scarce delivery of liquid natural gas (LNG) to the EU market from the Middle East and North America alongside the increase in demand and price in the APAC region; (v) the worsening of the Russia–West relations in connection to sanctions also in the energy sector.

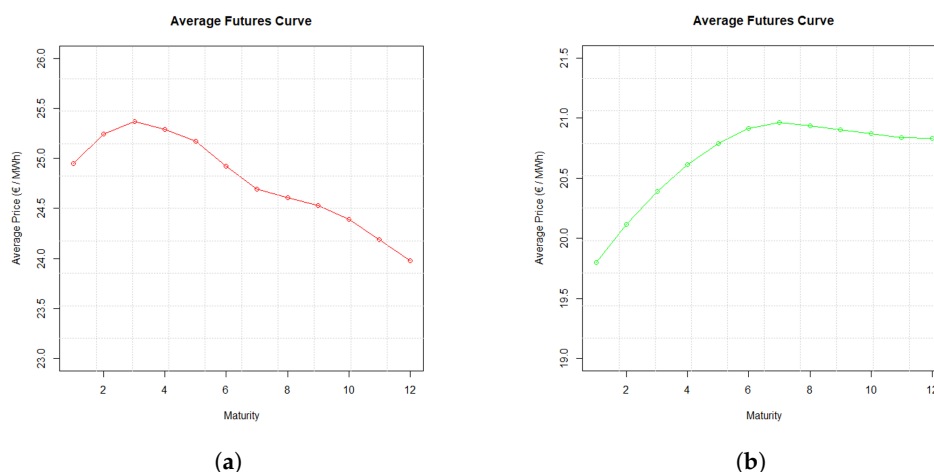
In summary, the observation period poses a rich set of dynamics in the NG term structure, with futures curves assuming a great variety of shapes: upward sloping (contango), downward sloping (backwardation), as well as inverted or humped/multi-humped, that is, conditions all observable in the government bonds market, thus justifying the extension to the NG market of the framework and the methods discussed in Diebold and Li [85]. We, therefore, argue that stylized facts in the bonds and in the NG futures market are similar. To examine these properties, in Table 1 we present the main descriptive statistics of futures prices and of the daily volatility ( $\sigma_{daily}$ ) of futures prices series for each NG futures contract calculated as the absolute value of price returns, following [95].

**Table 1.** Descriptive statistics of prices and daily volatility for each natural gas futures contract. For the price, we reported the mean, the standard deviation (SD), the minimum (Min), and the maximum (Max) values, while for the daily volatility ( $\sigma_{daily}$ ) we examined the mean and the median.

Maturity	Price				$\sigma_{daily}$	
	Mean	SD	Min	Max	Mean	Median
Mc1	24.951	20.928	3.509	227.201	2.031	1.167
Mc2	25.244	20.764	4.058	217.293	1.870	1.088
Mc3	25.369	20.170	4.618	210.804	1.743	1.000
Mc4	25.288	19.095	5.406	206.905	1.688	0.942
Mc5	25.170	18.327	7.082	200.902	1.611	0.923
Mc6	24.923	17.376	7.921	199.052	1.557	0.913
Mc7	24.690	16.645	9.194	179.233	1.507	0.906
Mc8	24.605	16.410	10.692	171.752	1.449	0.894
Mc9	24.527	16.078	11.130	154.291	1.405	0.848
Mc10	24.394	15.544	10.828	149.990	1.368	0.806
Mc11	24.190	14.733	10.801	143.515	1.329	0.797
Mc12	23.979	13.750	10.739	130.742	1.299	0.771

The results suggest that the average futures curve has a downward sloping trend and a slight hump in the short term (see Figure 2a for the visual inspection), with average prices ranging between a maximum of 25.37 €/MWh at maturity Mc3 and a minimum of 23.98 €/MWh for Mc12. At first glance, the empirical evidence seem to be in contrast to the feature of the increasing average yield curve characterizing the bond market. Nevertheless, if we consider the period from January 2011 to September 2021, excluding the most acute phase of the 2021–2022 downturn, and we plot again the corresponding average futures curve in that resized time frame (Figure 2b), then the shape is consistent with the increasing and concave curve which is typical of bond markets.

We can, therefore, argue that the discrepancy originally highlighted in Figure 2a, considering the whole observation period, is probably due to record-breaking fluctuations of the TTF futures prices between December 2021 and May 2022, with price peaks reaching over 200 €/MWh for short and mid-term maturities: the all-time record of 227 €/MWh was reached on 7 March 2022, that means an average increase of 945% over the same period of 2021 and of 127% over the previous month.



**Figure 2.** Plot of the behavior of the average futures curves in the period January 2011–June 2022 (a) and in the resized period January 2011–September 2021 (b).

It is then reasonable to assume that this record-breaking trend is temporary, reflecting market players’ concerns in the short to medium term. As soon as the crisis is overcome and the NG sector restores a secure and stable supply chain and storage, the curve presumably should turn to contango again, in analogy to the average spot yield curve behavior.

Furthermore, we can observe a trade-off between the volatility and the maturity: price volatility (daily volatility) is higher for contracts at shorter maturities and decreases for contracts with longer expiration dates; in fact, the standard deviation given in column three of Table 1 (mean given in column six) spans from a maximum value of 20.93 (2.03) at maturity Mc1 to a minimum value of 13.75 (1.29) for the maturity Mc12, with an average decrease of 4.0%. This is consistent with the phenomenon known as the Samuelson hypothesis [96], observed on the bond market as well, i.e., futures price volatility is a decreasing function of the time to maturity. To prove this assertion, we followed [97,98] and ran the Jonckheere–Terpstra test (JT test) [99,100] for ordered differences among classes to verify whether the medians of the time series of daily volatilities across maturity are decreasingly ordered. At first, we verified whether the  $\sigma_{daily}$  series at each maturity presents homogeneous statistical features by computing the Jarque–Bera test for normality, the Ljung–Box test for autocorrelation, as well as the augmented Dickey–Fuller test for stationarity. The results, summarized in Table 2, indicate that the  $\sigma_{daily}$  time series are not normally distributed, they are autocorrelated, and they do not contain unit roots.

We then ran the JT test with the null hypothesis  $H_0$  that the median values of the volatility series are the same at all maturities, against the alternative  $H_1$ , with at least one strictly decreasing inequality:

$$\begin{cases} H_0 : \tilde{\sigma}_{12} = \tilde{\sigma}_{11} = \dots = \tilde{\sigma}_1 \\ H_1 : \tilde{\sigma}_{12} \leq \tilde{\sigma}_{11} \leq \dots \leq \tilde{\sigma}_1 \end{cases} \tag{1}$$

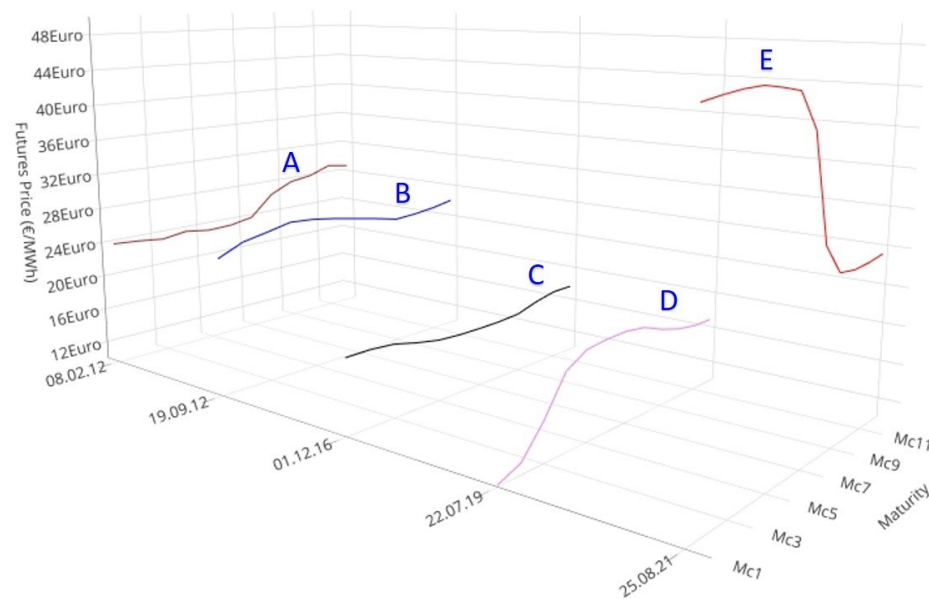
where  $\tilde{\sigma}_i, i = 1 \dots 12$ , represents the median of the daily volatility time series at maturity  $i$ . The result of the JT test, with the Z statistic equal to  $2.59 \times 10^8$  and a  $p$ -value of  $2 \times 10^{-16}$ , allows us to reject  $H_0$  at the 1% significance level, and thus supports the Samuelson hypothesis.

Another stylized fact shared with the yield curve in the government bond market is the great variety of shapes exhibited by the NG futures curves. Figure 3 shows five slices extracted from the 3D surface plot, representing the main trends observed on the NG market in different periods.



**Table 2.** Results of the Jarque–Bera test for normality (JB test), the Ljung–Box test for autocorrelation (LB test), and the augmented Dickey–Fuller test for stationarity (ADF test) computed on the daily volatility ( $\sigma_{daily}$ ). The symbol \* is used to denote the rejection of the null hypothesis  $H_0$  (data are normally distributed in the JB test; the series exhibits no autocorrelation in the LB tests; data series are not stationary in the ADF test) at the 1% significance level.

Maturity	$\sigma_{daily}$		
	JB Test	LB Test	ADF Test
Mc1	$1.87 \times 10^5 *$	$4.07 \times 10^3 *$	$-9.628 *$
Mc2	$2.35 \times 10^5 *$	$4.23 \times 10^3 *$	$-9.376 *$
Mc3	$3.65 \times 10^5 *$	$3.98 \times 10^3 *$	$-9.131 *$
Mc4	$8.67 \times 10^5 *$	$3.53 \times 10^3 *$	$-8.068 *$
Mc5	$7.79 \times 10^5 *$	$3.91 \times 10^3 *$	$-8.485 *$
Mc6	$8.12 \times 10^5 *$	$4.21 \times 10^3 *$	$-8.728 *$
Mc7	$6.19 \times 10^5 *$	$4.48 \times 10^3 *$	$-8.144 *$
Mc8	$7.13 \times 10^5 *$	$4.78 \times 10^3 *$	$-7.911 *$
Mc9	$7.75 \times 10^5 *$	$5.24 \times 10^3 *$	$-8.204 *$
Mc10	$7.59 \times 10^5 *$	$5.11 \times 10^3 *$	$-8.088 *$
Mc11	$8.23 \times 10^5 *$	$4.92 \times 10^3 *$	$-7.952 *$
Mc12	$1.49 \times 10^6 *$	$4.51 \times 10^3 *$	$-7.238 *$



**Figure 3.** Plot of the main futures curves shapes observed on the market at various times  $t = 8$  February 2012, 19 September 2012, 1 December 2016, 22 July 2019, and 25 August 2021. Time is represented on the x-axis, while maturities and prices (€/MWh) are on the y-axis and z-axis, respectively.

On 8 February 2012, for instance, curve **A** was normal, i.e., slightly increasing for longer maturities; on 19 September 2012 and 1 December 2016, the term structure (see **B** and **C**) was almost flat; on 22 July 2019, the increasing and slightly humped curve **D** is associated with the market in contango; on 25 August 2021, the market is in backwardation, as testified by the decreasing futures curve behavior at longer maturities (see **E**).

Overall, we can preliminarily conclude that the information-rich content and the similarities highlighted in the previous sections make the NG futures market a fruitful ground for testing models usually employed to fit and forecast the behavior in the government bond market.

### 3. Modeling Approach

#### 3.1. Parametric Factor Models

The four-factor dynamic Nelson–Siegel–Svensson (4F-DNSS) and the five-factor dynamic De Rezende–Ferreira (5F-DRF) are the most flexible exponential parametric models in the so-called Nelson–Siegel class. They are characterized by an improved fitting ability with respect to the three-factor Nelson–Siegel [101] model in which they have their roots, that makes them suitable to describe and replicate the overwhelming majority of yield curves trends and dynamics, including humps/basins, in the range of short and long-term maturities.

Let us indicate by  $p(t)$  the  $N \times 1$  vector of observed gas futures prices available at maturity  $m \in M = (m_1, m_2, \dots, m_N)'$ , where  $N$  is the maximum maturity length in months, and consider a time horizon of length  $t, t = 1, \dots, T$  expressed in days. The price dynamic is described by:

$$p(t) = F(t)\beta + \eta(t). \tag{2}$$

The variables in (2) deserve some further explanation. We start with  $F$ , which is an  $(N \times T) \times k$  matrix of factor loadings, with  $k = 4$  or  $k = 5$ , depending on the model, that is, 4F-DNSS or 5F-DRF. The generic  $m$ -th row is either in the form:

$$F_m^{DNSS}(t) = \left[ 1 \quad \tau_1 \frac{1-e^{-m/\tau_1}}{m} \quad \tau_1 \frac{1-e^{-m/\tau_1}}{m} - e^{-m/\tau_1} \quad \tau_2 \frac{1-e^{-m/\tau_2}}{m} - e^{-m/\tau_2} \right], \tag{3}$$

in the 4F-DNSS model, i.e., when  $k = 4$ , or:

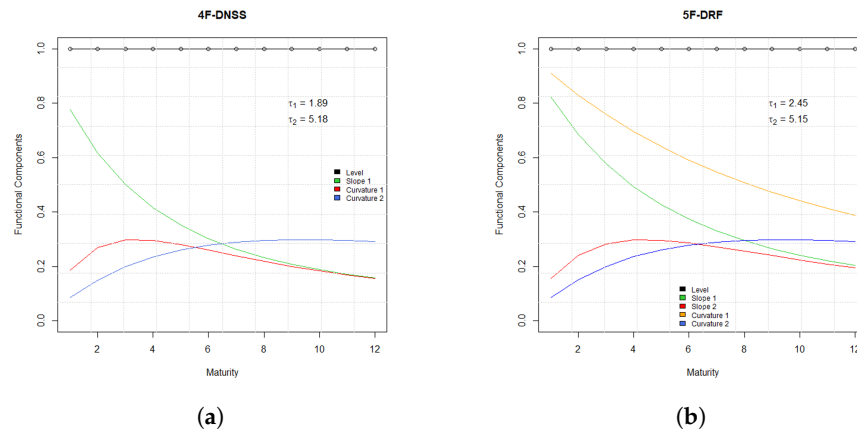
$$F_m^{DRF}(t) = \left[ 1 \quad \tau_1 \frac{1-e^{-m/\tau_1}}{m} \quad \tau_2 \frac{1-e^{-m/\tau_2}}{m} \quad \tau_1 \frac{1-e^{-m/\tau_1}}{m} - e^{-m/\tau_1} \quad \tau_2 \frac{1-e^{-m/\tau_2}}{m} - e^{-m/\tau_2} \right], \tag{4}$$

in the 5F-DRF model, i.e., when  $k = 5$ ; here,  $\tau_1$  and  $\tau_2$  are the decay terms which regulate the exponential components' decaying speed.

Three elements characterize the  $F$  matrix, that is, the factor loadings, i.e., the building blocks of futures curves. The first element, and also the first component, in both (3) and (4) is the level that represents the long-term component, constant for every maturity. The second element, occupying position 2 in (3) and positions 2 and 3 in (4) is  $\tau_i(1 - e^{-m/\tau_i})/m, i = 1, 2$ , is the slope of the futures curve. It is a proxy of the short-term component, as it starts at 1 and quickly converges monotonically to zero. Finally, the third element is  $[\tau_j(1 - e^{-m/\tau_j})/m] - e^{-m/\tau_j}, j = 1, 2$ , we find it in positions 3 and 4 in (3) and in positions 4 and 5 in (4), and it represents the curvature of the futures curve. It is a proxy of the medium-term component of the futures curve as it begins at zero, reaches the maximum value at medium-term maturities, and monotonically returns to zero at long-term maturities. The models presented in (3) and (4) consider different combinations of the above elements: the 4F-DNSS uses a single slope and two curvature components, while the 5F-DRF introduces an additional slope element. Figure 4 shows the behavior of the factor loadings in the case of the 4F-DNSS (a) and 5F-DRF (b) models.

Turning to  $\beta$ , we have  $\beta = [\beta_0 \beta_1 \beta_2 \beta_3]'$  in the 4F-DNSS model, and  $\beta = [\beta_0 \beta_1 \beta_2 \beta_3 \beta_4]'$  in the 5F-DRF model. Each element represents the weight associated with the corresponding factor loading, hence, changes in the vector  $\beta$  components impact the level, slope, and curvature of the NG futures function and, thus, its shape. As a result, all the futures curve shapes can be replicated by a proper weights calibration and combination with the corresponding loadings. Finally,  $\eta(t) \sim \mathcal{N}(0, \Sigma)$  represents the error terms vector, assumed to be normally distributed with a zero mean vector and a variance–covariance matrix  $\Sigma$ .

Concerning the estimation process of  $\beta$ , we applied an approach organized into two stages. At first, following [102], we identified for each time  $t$  the optimal combination  $[\hat{\tau}_1(t), \hat{\tau}_2(t)]$ , and hence  $\hat{\beta}(t)$ , as the weights vector associated with the lowest root mean square error (RMSE). Then, we determined the average values of  $\hat{\tau}_j (j = 1, 2)$  to derive the optimal estimate of  $\hat{\beta}^*(t)$  for each available day. In this way the model maintains a high adaptive capability and gains in stably estimated parameters.



**Figure 4.** 4F-DNSS (a) and 5F-DRF (b) factor loadings at different times to maturity. In (a,b) we indicate the average values of  $\tau_1$  and  $\tau_2$  determined during the estimation process and used for the daily fits.

For an easier understanding we are going to provide a more detailed description of the main steps of the procedure for both the 4F-DNSS and 5F-DRF models in the following section.

1. Define the set  $\Phi_j = \{m_{j,k}\}_{k=1,\dots,N_j}$  of maturities  $m_{j,k}$  with  $j = 1, 2$  and  $N_j$  equal to the set's cardinality;  $m_{1,1}$  represents the lower bound of  $\Phi_1$  ( $m_{1,L}$ ) and corresponds to the first available maturity of the NG futures market, while the upper bound  $m_{1,U}$  is, at the same time, the lower bound of  $\Phi_2$  ( $m_{2,L}$ ) and the straddling maturity between the short and medium-term period. The upper bound of  $\Phi_2$  ( $m_{2,U}$ ) is the longest observed maturity. Values in  $\Phi_1$  and  $\Phi_2$  range between the corresponding lower/upper values by way of the proper step sizes  $\Delta_1$  and  $\Delta_2$ . Given the absence of a closed form expression for  $\Delta_1$  and  $\Delta_2$ , and given the trade-off between the step sizes discretization level and the speed of the estimation procedure, we conducted different simulations testing  $\Delta_1$  and  $\Delta_2$  in the range  $[0.25, 1]$  and  $[0.25, 1.5]$ , respectively. In this way, we selected  $\Delta_1 = 0.75$  and  $\Delta_2 = 1$  for both the 4F-DNSS and the 5F-DRF models.
2. For each maturity  $m_{j,k}$  in the sets  $\Phi_1$  and  $\Phi_2$ , estimate the vectors  $\hat{\tau}_1$  and  $\hat{\tau}_2$  that maximize the medium-term component:

$$\frac{1 - e^{-m_{j,k}/\tau_j}}{m_{j,k}/\tau_j} - e^{-m_{j,k}/\tau_j}, \quad k = 1, \dots, N_j; j = 1, 2.$$

3. For every  $t = 1, \dots, T$ :
  - (a) for each component of  $\hat{\tau}_1$ , vary the components of  $\hat{\tau}_2$  to estimate by OLS regression different array sets  $\hat{\beta}(t)$ , choosing the one with the lowest sum of squared residuals (SSR), computed as the squared magnitude of the difference between the vectors of the observed,  $p(t)$ , and estimated,  $\hat{p}(t)$ , prices:

$$SSR(t) = \left\| p(t) - \hat{p}(t) \right\|^2.$$

Clearly there are as many sets of optimal parameters as the number of  $\hat{\tau}_1$  values;

- (b) choose the optimal  $\hat{\beta}$ , associated with the lowest SSR.

$$SSR(t) = \left\| [p(t) - \hat{p}(t, \hat{\beta}(t), \hat{\tau}_1(t), \hat{\tau}_2(t))] \right\|^2.$$

4. Repeat step 3 for each time  $t$  ( $t = 1, \dots, T$ ) to obtain the time series parameters of both  $\hat{\tau}_1(t)$  and  $\hat{\tau}_2(t)$ .
5. Compute the average values  $\bar{\tau}_1(t)$  and  $\bar{\tau}_2(t)$  of the decay terms times series of step 4, then estimate again the set  $\hat{\beta}^*(t)$  for each time  $t$  ( $t = 1, \dots, T$ ) to obtain the final set of  $T$  estimated futures curves.



### 3.2. B-Spline Interpolation Method

B-spline [11,103,104] is a powerful modeling tool to fit observable data without strong functional form assumptions. The B-spline function is:

$$f(x) = \sum_{i=1}^{k+d-2} \pi_i B_{i,d}(x), \tag{5}$$

where  $\pi_i$  ( $i = 1, \dots, k + d - 2$ ) are the spline coefficients and  $B_{i,d}(x)$ ,  $d \geq 1$  are B-spline basis functions. Those, in turn, are fully determined once the order  $d \geq 1$  and the sequence of nondecreasing real values  $\xi_1 \leq \xi_2 \leq \dots \leq \xi_k$  is set, acting as control points or knots. To have a well-defined B-spline of order  $d$  and degree  $d - 1$  covering the whole span of knots, the sequence of knots is extended as following:

$$\underbrace{\xi_1, \dots, \xi_1}_{k-1 \text{ times}}, \xi_1, \xi_2, \dots, \xi_k, \underbrace{\xi_k, \dots, \xi_k}_{k-1 \text{ times}}$$

The  $i$ -th B-spline basis of order  $d$  is then recursively defined for  $d > 1$  as:

$$B_{i,d}(x) = \delta_{i,d}(x)B_{i,d-1}(x) + [1 - \delta_{i+1,d}(x)]B_{i+1,d-1}(x), \tag{6}$$

with

$$B_{i,1}(x) = \begin{cases} 1, & \text{if } \xi_i \leq x \leq \xi_{i+1} \\ 0, & \text{otherwise} \end{cases}, \tag{7}$$

and

$$\delta_{i,d}(x) = \begin{cases} \frac{x - \xi_i}{\xi_{i+d-1} - \xi_i}, & \xi_i \neq \xi_{i+d-1} \\ 0, & \text{otherwise} \end{cases}. \tag{8}$$

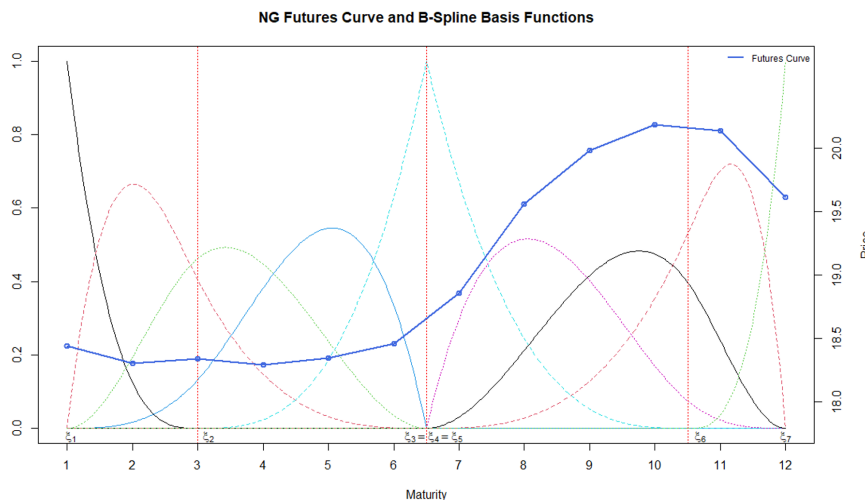
In practical applications, the choice of the number  $k$  of knots is of paramount importance: too many (too few) knots, in fact, can result in overfitting (underfitting) issues. In general, the problem is addressed by the use of priors, enforcing smoothness across the coefficients  $\pi_i$ : in general, the closer the consecutive  $\pi_i$  are to each other, the smoother the resulting B-spline is, with lower local variability.

In our study, we selected seven knot points, with three overlapping knots, and we partitioned the maturity domain [1, 12], that is, from 1 to 12 months, into four sub-periods, that is, from 1 to 3, 3 to 6.5, 6.5 to 10.5, and 10.5 to 12. This implies that the fitted futures curves are divided into four segments, each approximated by a set of piecewise basis functions of the same degree, as shown in Figure 5 where we overlay an NG futures curve taken as an example from the dataset and the corresponding basis functions used to approximate it. Cubic B-spline functions were used and the conditions were such as to assure continuity of the slope (curve segments have the same first derivative at joint, i.e., the corresponding function is of class  $\mathcal{C}^1$ ) and curvature (curve segments with same second derivative at joint, i.e., functions belonging to  $\mathcal{C}^2$ ) were applied at each knot, except to those overlapping; here, the B-spline is of class  $\mathcal{C}^0$ , that is, curve segments are connected at the joint.

The estimation of the vector of parameters  $\pi(t)$  was performed for each time  $t$ ,  $t = 1, \dots, T$ , with the least squares method minimizing the sum of the weighted squared residuals (WSSE):

$$WSSE(\pi(t)) = \sum_{j=1}^N \omega(m_j, t) [p(m_j, t) - f(m_j, t)]^2. \tag{9}$$

where  $\omega(m_j, t)$  is the error weight at maturity  $m_j$  and time  $t$ ;  $p(m_j, t)$  is the NG futures price observed at maturity  $m_j$  and time  $t$ ; while  $f(m_j, t)$  represents the point on the B-spline curve at maturity  $m_j$  and time  $t$ .



**Figure 5.** Example of an NG futures curve (blue line), the corresponding 4-order basis polynomial functions (dashed lines), and the  $\xi_1, \dots, \xi_7$  knots.

### 3.3. Nonlinear Autoregressive Neural Network (NAR-NN)

An artificial neural network is a system aimed at simulating the human nervous system. It is characterized by a computational scheme which is not programmed but trained by a machine learning algorithm. Thanks to its ability in identifying nonlinear relationships in the data, it can approximate any differentiable function [105] and it is recognized as a universal approximator [106].

In this work, we explored the potential of Nonlinear Autoregressive Neural Networks (NAR-NNs), which belong to the class of dynamic recurrent neural networks. The model portrays a nonlinear relationship between the current  $x_{m,t}$  value of the observed univariate time series, that is, NG futures prices at maturity  $m \in M = (m_1, m_2, \dots, m_N)'$  at time  $t$ , and  $h$  past values or feedback delays for each  $t = 1 + h, \dots, T$ , capturing the autoregressive properties:

$$x_{m,t} = g(x_{m,t-1}, x_{m,t-2}, \dots, x_{m,t-h}) + \epsilon_{m,t}, \tag{10}$$

where  $g(\cdot)$  represents an unknown nonlinear transfer function that the network tries to approximate, while  $\epsilon_{m,t}$  stands for the approximation error at maturity  $m$  and time  $t$ .

The NAR-NN is made by interconnected processing units, called nodes or neurons [107], arranged in sequential and fully connected layers: the input layer, one or more hidden layers, and the output layer.

The network operates in two phases: the open loop phase, during which the network is created and trained, and the closed loop phase, during which predictions are made. In the open loop, the network aims at identifying the appropriate transfer function to form a correct mapping between inputs and target values even in the presence of nonlinear dynamics, based on a pure feed-forward architecture. In particular, given the  $\{x_{m,t}\}_{t=1}^T$  time series, the network creates a vector of  $T - h$  historical target values, each of which is associated with the vector (defined input pattern) of  $h$  previous target values  $x_{m,t-j}$ ,  $j = 1, \dots, h$ . Each input pattern is used as the input to the network. Its elements are multiplied by an assigned weight  $\theta_{i,j}$  ( $j = 1, \dots, h$ ) and sent to the closest hidden layer. Then, each hidden node  $i$  sums the incoming weighted signals with a bias value  $\theta_{i,0}$ , according to:

$$\varphi_{i,t} = \theta_{i,0} + \sum_{j=1}^h \theta_{i,j} x_{m,t-j}. \tag{11}$$

The resulting value  $\varphi_{i,t}$  is processed via the activation function  $\Lambda_{\chi}(\cdot)$  that applies a transformation (either sigmoid or linear) and activates (deactivates) the network hidden neurons. If the value  $\varphi_{i,t}$  exceeds a given threshold  $\chi$ , then the hidden node generates a

response signal which is broadcast either to the nodes of the next hidden layer(s), if there are any, or to the node of the output layer, where it is processed through an activation function  $\psi(\cdot)$  (usually linear), generating the networks final response  $\hat{x}_{m,t}$ :

$$\hat{x}_{m,t} = \psi \left( \gamma_0 + \sum_{i=1}^n \gamma_i \Lambda_{\chi}(\varphi_{i,t}) \right), \quad (12)$$

where  $\gamma_i$  is the weight assigned to the connection between the hidden unit  $i$  and the output unit; and  $\gamma_0$  represents the bias used to optimize the working point of the neuron in the output unit.

During the training process, to improve the performance and obtain the closest response to the target values the network determines the best vector  $\mathbf{v}^*$  of weights and bias by means of a learning algorithm that minimizes the error function:

$$Err(\mathbf{v}) = \sum_{t=1+h}^T (x_{m,t} - \hat{x}_{m,t})^2, \quad (13)$$

After the training phase, the neural network is converted into a closed loop network and, for  $t > T$ , the  $(x_{m,t-1}, x_{m,t-2}, \dots, x_{m,t-h+1}, x_{m,t-h})$  original lagged values are used to generate the first prediction  $\hat{x}_{m,t}$ . The forecasted value is fed back to the tap delay line in the input layer and added to form the new set  $(\hat{x}_{m,t}, x_{m,t-1}, x_{m,t-2}, \dots, x_{m,t-h+1})$ , which produces the next forecast  $\hat{x}_{m,t+1}$ . Based on such a recursive approach, new forecasted values update the previous set of lagged values to make new predictions  $(\hat{x}_{m,t+q})$  in the next step  $q$ .

We examined various network architecture layouts and we chose the optimal one according to a trial and error approach; in our case, the best solution turned out to be an NAR-NN made by one hidden layer. We set the number of hidden nodes and feedback delays in the ranges of [8, 10] and [3, 9], respectively, depending on the 4F-DNSS or 5F-DRF model and the parameter's time series. We used the logistic sigmoid activation function for the hidden nodes:

$$\Lambda_{\chi}(\varphi_{i,t}) = \frac{1}{1 + e^{-\varphi_{i,t}}}, \quad (14)$$

and a linear activation function for the output nodes.

The training and the learning of the network generally uses the available input data partitioned into training (70%), validation (15%), and testing (15%) sets. The supervised learning process was carried out implementing the Levenberg–Marquardt back propagation learning algorithm (LMBP) [108,109] with the weights update rule:

$$\Delta \mathbf{v}_k = - \left( \mathbf{J}^T(\mathbf{v}_{k-1}) \mathbf{J}(\mathbf{v}_{k-1}) + \mu \mathbf{I} \right)^{-1} \mathbf{J}^T(\mathbf{v}_{k-1}) \mathbf{e}_{k-1}. \quad (15)$$

where  $\mathbf{J}$  is the Jacobian matrix of the network errors with respect to the weights and biases;  $\mu$  represents the damping factor;  $\mathbf{I}$  is the Identity matrix; while  $\mathbf{e}_{k-1}$  represents the vector of the training errors at step  $k - 1$ . In the initial phase, the algorithm initializes random weights and calculates the value of the error function; then, the LMBP sets a large  $\mu$  and updates weights moving in the steepest-descent direction. If the update fails to reduce the error, then  $\mu$  is raised; otherwise, if the error decreases, the damping factor is reduced. Generally, the training process stops when either the maximum number of training cycles or the maximum training time is reached, or when a specific level of accuracy is attained.

#### 4. Empirical Study

We present and discuss the results of the estimation of the NG futures prices carried out with the 4F-DNSS, 5F-DRF, and B-spline models, and then we evaluate the term structure forecasts obtained with the hybrid scheme, combining the above models with the NAR-NN.

#### 4.1. Goodness-of-Fit

In this stage we tested the fitting abilities of the above mentioned methods using the whole dataset. The observation period spans from 20 January 2011 to 13 June 2022.

Table 3 lists the average descriptive statistics of the fitted prices for each available maturity. The interpretation is twofold: a first reading concerns the models’ performances, while another relates to the models’ ability to replicate stylized facts.

**Table 3.** Descriptive statistics of daily fitted futures prices at different maturities obtained with the 4F-DNSS, 5F-DRF, and B-spline models. For each model we report the mean, the standard deviation (SD), the MSE, and the RMSE at every maturity.

Maturity	4F-DNSS				5F-DRF				B-Spline			
	Mean	SD	MSE	RMSE	Mean	SD	MSE	RMSE	Mean	SD	MSE	RMSE
Mc1	24.944	21.003	0.416	0.645	24.930	20.848	$7.6 \times 10^{-2}$	$2.7 \times 10^{-1}$	24.950	20.923	$7.5 \times 10^{-4}$	$2.7 \times 10^{-2}$
Mc2	25.278	20.780	1.579	1.256	25.304	21.033	$6.5 \times 10^{-1}$	$8.0 \times 10^{-1}$	25.252	20.813	$4.8 \times 10^{-2}$	$2.2 \times 10^{-1}$
Mc3	25.341	19.975	1.156	1.075	25.352	20.060	$5.3 \times 10^{-1}$	$7.3 \times 10^{-1}$	25.349	20.066	$3.1 \times 10^{-1}$	$5.6 \times 10^{-1}$
Mc4	25.267	19.032	1.893	1.376	25.257	18.950	1.4681	1.2117	25.310	19.169	$4.0 \times 10^{-1}$	$6.3 \times 10^{-1}$
Mc5	25.128	18.199	3.446	1.856	25.107	18.035	1.5893	1.2607	25.162	18.282	$2.1 \times 10^{-1}$	$4.6 \times 10^{-1}$
Mc6	24.963	17.506	2.151	1.467	24.946	17.354	1.6048	1.2668	24.918	17.331	$2.4 \times 10^{-1}$	$4.9 \times 10^{-1}$
Mc7	24.791	16.893	1.166	1.080	24.787	16.858	1.2568	1.1211	24.697	16.703	$2.2 \times 10^{-1}$	$4.7 \times 10^{-1}$
Mc8	24.623	16.302	0.865	0.930	24.634	16.450	$3.8 \times 10^{-1}$	$6.2 \times 10^{-1}$	24.603	16.385	$8.5 \times 10^{-2}$	$2.9 \times 10^{-1}$
Mc9	24.464	15.712	0.956	0.978	24.484	16.011	$3.5 \times 10^{-1}$	$5.9 \times 10^{-1}$	24.523	16.066	$7.0 \times 10^{-2}$	$2.6 \times 10^{-1}$
Mc10	24.314	15.139	0.906	0.952	24.334	15.435	$6.3 \times 10^{-1}$	$7.9 \times 10^{-1}$	24.400	15.564	$8.0 \times 10^{-2}$	$2.8 \times 10^{-1}$
Mc11	24.174	14.621	0.542	0.736	24.179	14.686	$5.8 \times 10^{-1}$	$7.6 \times 10^{-1}$	24.188	14.726	$9.3 \times 10^{-3}$	$9.6 \times 10^{-2}$
Mc12	24.044	14.210	1.227	1.107	24.018	13.834	$4.8 \times 10^{-1}$	$7.0 \times 10^{-1}$	23.979	13.750	$1.3 \times 10^{-5}$	$3.6 \times 10^{-3}$

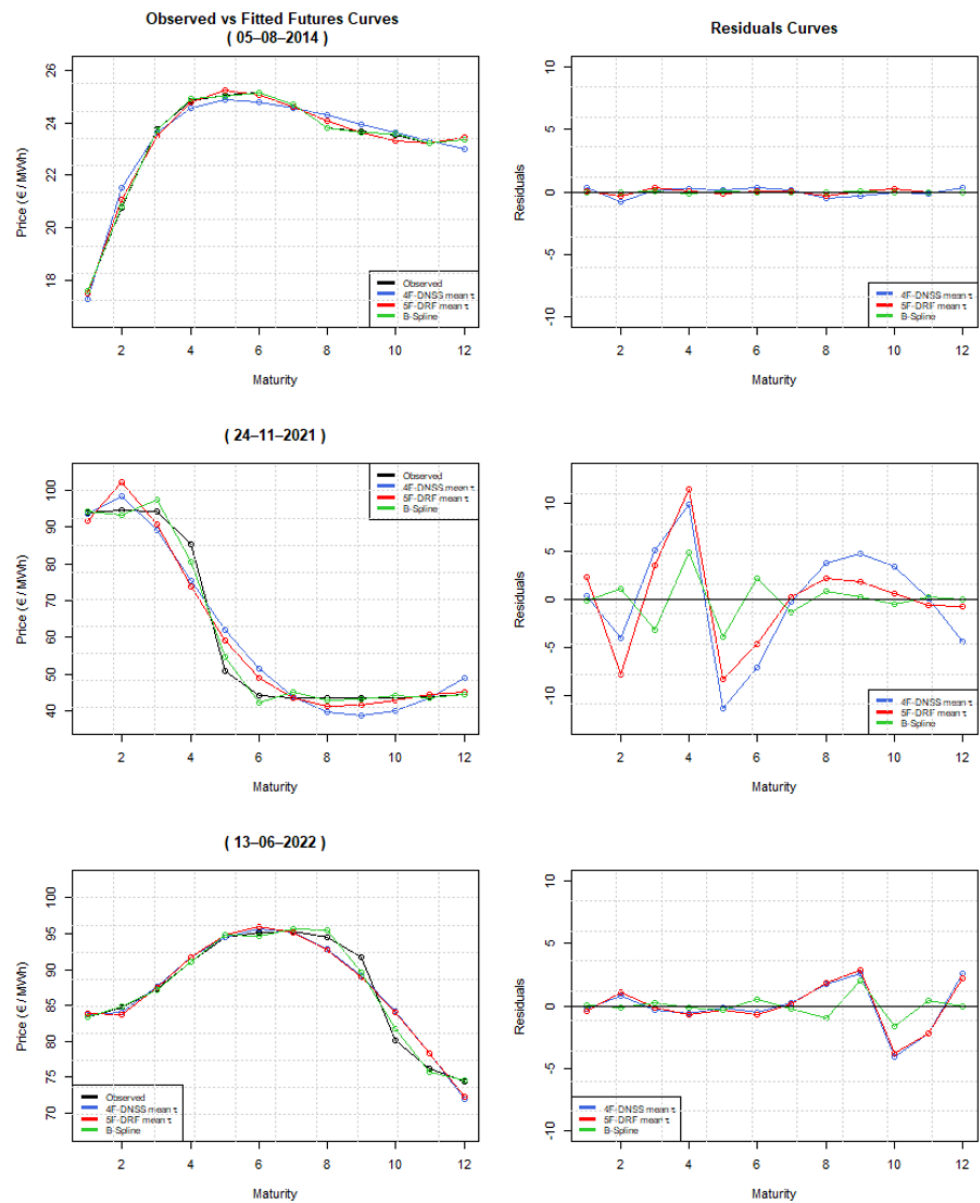
By comparison with the descriptive statistics summarized in Table 1, all the models achieve similar outcomes, faithfully mimicking the observed prices with negligible differences at each maturity. Nevertheless, the B-spline model performed better than the other methods. The B-spline model was 90.41% (83.61%) more effective than the 4F-DNSS (5F-DRF) model, in terms of the average mean squared error (MSE) performance metric.

Furthermore, the analysis revealed that all the methods were able to replicate the main stylized facts of the price series: the average curves are humped and slightly decreasing like the observable one, the volatility decreases at longer maturities, the curve at shorter maturities is more volatile than at medium and long ones. Additionally, the MSE and RMSE metrics are very low, confirming the models’ ability to replicate very accurately, on average, prices’ time series at each maturity. To support this argument, Figure 6 displays the observed and fitted futures curves and the residuals for three different times: 5 August 2014, characterized by an upward sloping curve; 24 November 2021, featuring an inverted S-shaped trend; and 13 June 2022, where a humped curve was observed. The above dates were chosen as representative of the most difficult curve shapes. At a visual inspection, we note a high degree of accuracy for all the shapes. Moreover, despite the use of constant decaying parameters, both the 4F-DNSS and 5F-DRF models seem to be flexible enough to deal with challenging curve dynamics. Nevertheless, the B-spline generated better approximations, which overlap the observed trends in every case and outperform the parametric models, which encountered some difficulties, especially in fitting curves with multiple inflection points.

For an in-depth view of the in-sample outcomes, Table 4 shows the performance metrics computed on the models’ residuals generated during the estimation process, while Figure 7 presents a 2D visualization of the mean absolute error (MAE) metric generated by the 4F-DNSS, 5F-DRF, and B-spline models.

The B-spline model gives overall superior results, with the lowest errors in the whole fitting horizon and peaks of limited magnitude in the range [0.0038, 2.8780]. Clusters of peaks, i.e., departures from the observed values, can be observed at the extremes of the

time series. These clusters reflect pronounced market volatility associated with specific historical events.

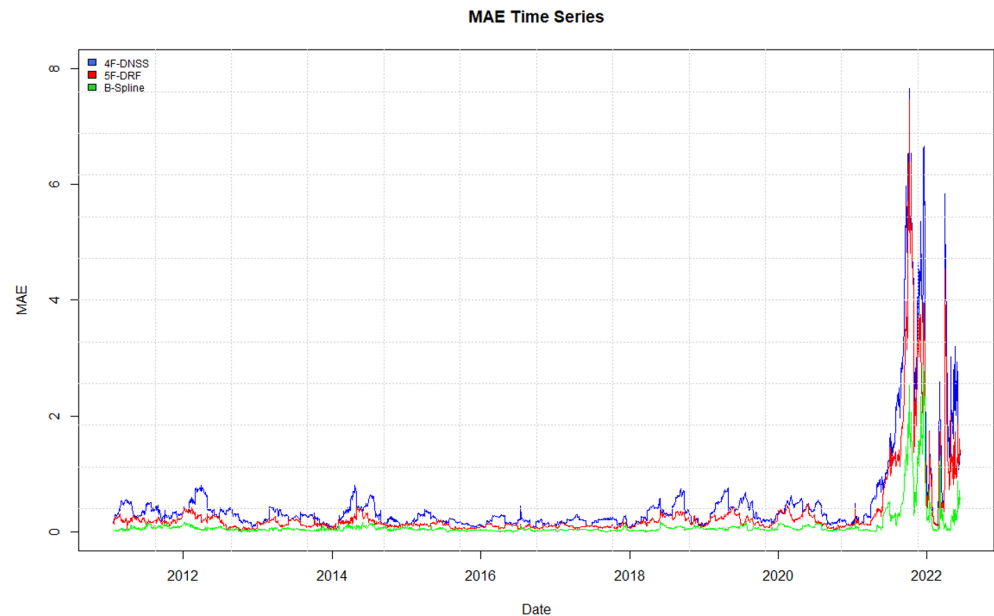


**Figure 6.** The first column shows the observed and fitted futures curves with the 4F-DNSS (blue), 5F-DRF (red), and B-spline (green) models; the related residuals curves are shown on the right-hand side. The days chosen are representative of the most difficult dynamics of the futures curves.

**Table 4.** Main MSE and RMSE statistics for the 4F-DNSS, 5F-DRF, and B-spline models. For each metric and model we report the mean, the standard deviation (SD), the minimum (Min), and the maximum (Max) values.

	MSE			RMSE		
	4F-DNSS	5F-DRF	B-Spline	4F-DNSS	5F-DRF	B-Spline
<i>Mean</i>	1.3586	$8.0050 \times 10^{-1}$	$1.3978 \times 10^{-1}$	$5.8370 \times 10^{-1}$	$3.9148 \times 10^{-1}$	$1.4050 \times 10^{-1}$
<i>SD</i>	6.5667	4.5393	$9.5190 \times 10^{-1}$	1.0091	$8.0460 \times 10^{-1}$	$3.4653 \times 10^{-1}$
<i>Min</i>	$3.9908 \times 10^{-3}$	$3.2177 \times 10^{-4}$	$2.5177 \times 10^{-5}$	$6.3173 \times 10^{-2}$	$1.7938 \times 10^{-2}$	$5.0177 \times 10^{-3}$
<i>Max</i>	98.9464	83.6844	17.4452	9.9472	9.14792	4.17675





**Figure 7.** MAE time series obtained with the 4F-DNSS (blue), 5F-DRF (red), and B-spline (green) models in the natural gas futures market.

In summary, the results confirm the adequacy of the techniques to model NG futures curves, as they are able to effectively replicate all the features and dynamics of the market.

#### 4.2. Out-of-Sample Forecasting

We applied NAR-NNs to perform both direct and indirect out-of-sample forecasts on the NG futures price term structure. In the former, the neural network was used to predict prices time series; in the latter, in the way shown by Diebold and Li [85], based on the data from 20 January 2011 to 13 May 2022, we used the neural network to predict the vector of parameters of the 4F-DNSS, 5F-DRF, and B-spline models, deriving futures prices at each forecasting horizon in a second time. The implementation was carried out with the *narnet* function of the deep learning toolbox of MATLAB R2022a. The forecasts cover 21 working days: the period 16 May 2022–13 June 2022. The forecasting period is short but very turbulent, with a variety of behaviors of the future curves.

Finally, we evaluated the effectiveness of the proposed approach using the mean absolute percentage error (MAPE)

$$MAPE = \frac{100}{T} \sum_{t=1}^T \frac{1}{M} \sum_{m=1}^M \left| \frac{p_{t+h,m} - \hat{p}_{t+h,m}}{p_{t+h,m}} \right|, \quad (16)$$

and the mean squared forecast error (MSFE)

$$MSFE = \frac{1}{T} \sum_{t=1}^T \frac{1}{M} \sum_{m=1}^M (p_{t+h,m} - \hat{p}_{t+h,m})^2, \quad (17)$$

as performance metrics. In order to increase the robustness of the analysis and to assess the accuracy of the results, we also computed the Theil's inequality coefficient [110], which returned the level of accuracy of the forecasting methods:

$$U_2 = \frac{\sqrt{\sum_{t=1}^{T-1} \left( \frac{\hat{p}_{t+h,m} - p_{t+h,m}}{p_{t,m}} \right)^2}}{\sqrt{\sum_{t=1}^{T-1} \left( \frac{p_{t+h,m} - p_{t,m}}{p_{t,m}} \right)^2}} \tag{18}$$

where  $\hat{p}_{t+h,m}$  is the forecast at time  $t + h$ , for the maturity  $m$ . In detail, we employed the so-called Theil’s  $U_2$  statistic [111,112]. Here, perfect forecasting accuracy, that is,  $\hat{p}_{t+h,m} = p_{t+h,m}$  for every  $t + h$  and  $m$  receives a zero value for the statistics; the value one, on the other hand, is a kind of threshold of the forecasting accuracy relative to the naïve approach. In practice,  $U_2$  values above (below) 1 indicate a forecasting performance which is less (more) accurate than the naïve approach. The results of the performance metrics are reported in Table 5.

**Table 5.** Comparison of the models’ forecasting performances using different error metrics. Columns two and three, respectively, show the average MAPE (%) and MSFE values, while column four reports the Theil’s  $U_2$  test score.

Forecasting Model	MAPE	MSFE	Theil’s $U_2$ Score
4F-DNSS/NAR-NN	6.1309	43.4550	0.0644
5F-DRF/NAR-NN	7.6939	85.4639	0.0670
B-spline/NAR-NN	2.7562	9.2254	0.0268
NAR-NN	2.7619	9.4914	0.0280

Based on Table 5, all of the candidate methods achieved satisfying results. In particular, looking at the MAPE values in column two, we observe that the prediction accuracy ranges on average between 92.3% and 97.2%. Furthermore, it is possible to rank the models depending on MAPE: the 5F-DRF/NAR-NN is in fourth place, with the highest MAPE, the 4F-DNSS/NAR-NN is in 3rd place, the NAR-NNs gains second place, while the top position is taken by the B-spline/NAR-NN. Column four, which reports the Theil’s  $U_2$  scores, confirms the ranking: the 5F-DRF/NAR-NN, in fact, although showing a notable predictive accuracy, with a score of 0.0670, is in fourth place; the score of the 4F-DNSS/NAR-NN is slightly better (0.0644); and B-spline/NAR-NN achieves the best overall accuracy score.

Indeed we moved one step further, performing additional tests to validate the forecasting power of all of the competing models. With the adjusted Diebold–Mariano (ADM) test [113] we pairwise tested the difference in the forecast errors of the prediction models and we discussed whether those are significantly different from zero. When comparing two forecasting models, say model 1 and model 2, the null hypothesis  $H_0$  is that forecasts of model 2 are more accurate than those of model 1 at the 95% significance level. The results are reported in Table 6.

**Table 6.** Results of the adjusted Diebold–Mariano test. The  $p$ -value is accompanied by the \* label when the null hypothesis  $H_0$  is accepted at the 95% significance level.

Model 1	Model 2	ADM Statistics	$p$ -Value
4F-DNSS/NAR-NN	B-spline/NAR-NN	4.2532	0.9998 *
5F-DRF/NAR-NN	B-spline/NAR-NN	3.3970	0.9986 *
NAR-NN	B-spline/NAR-NN	1.9269	0.9658 *
5F-DRF/NAR-NN	4F-DNSS/NAR-NN	−0.1332	0.4477 *
5F-DRF/NAR-NN	NAR-NN	3.1252	0.9973 *
4F-DNSS/NAR-NN	NAR-NN	4.1745	0.9998 *

The scores of the ADM statistic and the related  $p$ -values (columns 3 and 4 of Table 6) are notably higher than the critical value at the 95% significance level  $t(0.05, n - 1) = -1.72$ . This, in turn, implies that in all of the examined comparisons, we have fallen within the acceptance region, thus, strongly accepting  $H_0$ , that is, that forecasts of the model indicated

in column 2 (Model 2) are more accurate than those of the model indicated in column 1 (Model 1), In detail, the test results over the six different comparisons show that:

- B-spline/NAR-NN is more accurate than 4F-DNSS/NAR-NN;
- B-spline/NAR-NN is more accurate than 5F-DRF/NAR-NN;
- B-spline/NAR-NN is more accurate than NAR-NN;
- NAR-NN is more accurate than 4F-DNSS/NAR-NN;
- NAR-NN is more accurate than 5F-DRF/NAR-NN;
- 4F-DNSS/NAR-NN is more accurate than 5F-DRF/NAR-NN.

We can therefore conclude that B-spline/NAR-NN supersedes both 4F-DNSS/NAR-NN and 5F-DRF/NAR-NN and, as already testified by the results discussed in the previous sections, 4F-DNSS/NAR-NN works better than 5F-DRF/NAR-NN. Furthermore, as we compared the performance of B-spline/NAR-NN to that of the NAR-NN *as it is*, used likewise as a benchmark, also in that case, according to the test score, the results highlight that the combination of B-spline to the neural engine provides better results.

An additional validation of the results has then come from the model confidence set (MCS) procedure [114]: we constructed a superior set model (SSM) for the dataset under examination, for which the null hypothesis of equal predictive ability (EPA) is not rejected at the 95% confidence level, with the punctual mean forecast as loss function, as in [115]. The results are summarized in Table 7.

**Table 7.** Results of the MCS test at the 95% confidence level.

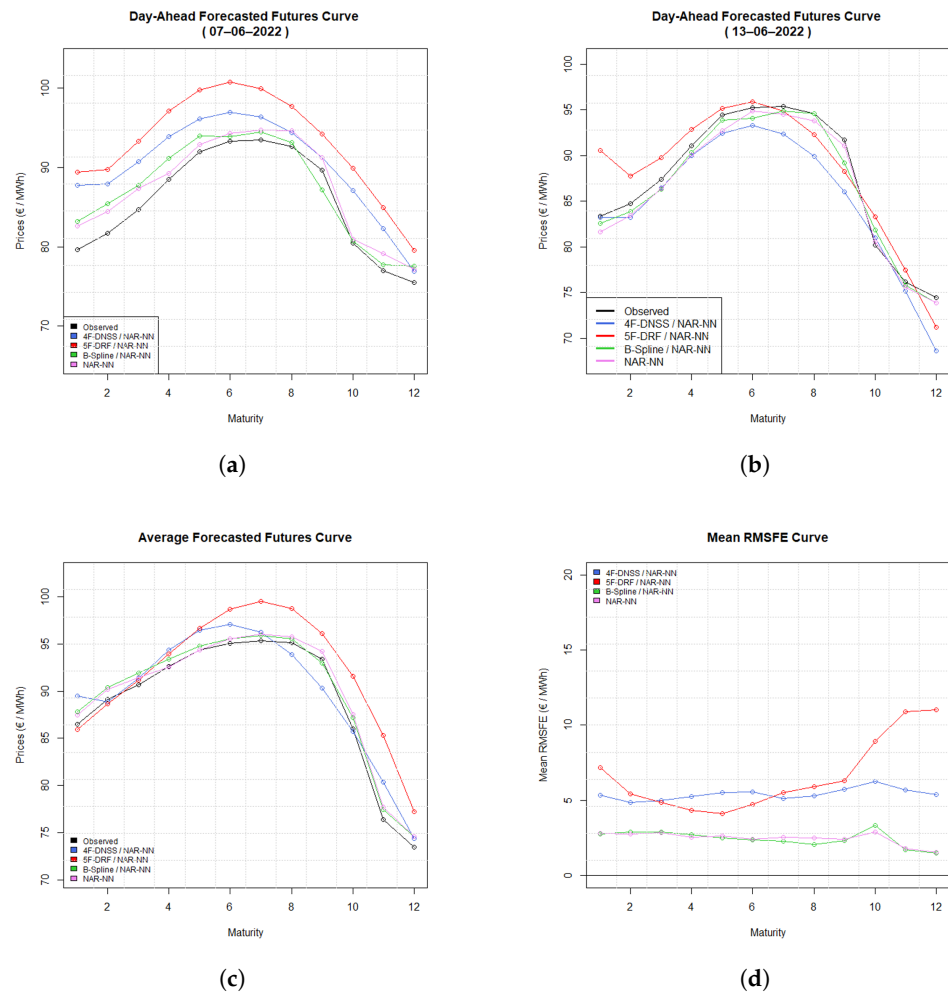
SSM	Rank	Loss
B-spline/NAR-NN	1	5.6868
NAR-NN	2	6.2330

The MCS must be interpreted as a confidence interval: in our case, for instance, as we have set  $\alpha = 0.05$ , the  $MCS_{1-\alpha}$  contains the best model, with  $100(1 - \alpha)\%$  confidence. Table 7 simply suggests that, within the examined methods, B-spline/NAR-NN and NAR-NN are those providing the most reliable forecasts at the 95% confidence level. Nevertheless, looking at the loss values in column 3 of Table 7, the technique achieving the best results is the one combining B-spline with NAR-NN, as it is associated with a lower loss function value.

All in all, a better understanding of the results can be gained with an example: Figure 8a,b compare futures curves observed on two days taken at random from our sample (7 and 13 June 2022) with those predicted according to the proposed methodology. Figure 8c,d show the average forecasted curves along with the average RMSFE generated by the models for each maturity, respectively. At first sight, the approaches generated accurate parameters and curves forecasts, with predicted futures prices being close enough to the observed ones for each available maturity, without notable spikes or outliers.

Nevertheless, it is worth highlighting some differences among the models in terms of smoothness and accuracy. In fact, factor models combined with neural networks produced smooth curves which, however, struggle to follow the shapes of the observations; on the contrary, those of the B-spline/NAR-NN and NAR-NNs appear more rough, although they follow the original futures curves trends more faithfully.

The curves predicted by means of B-spline/NAR-NN and NAR-NN alone are systematically closer to the observable ones at all the maturities than those obtained with either the 4F-DNSS/NAR-NN or 5F-DRF/NAR-NN models. These models, in fact, are characterized by higher error rates for every maturity, as testified by the RMSFE, with futures prices significantly under/overpredicted. Furthermore, although the 5F-DRF model exhibited superior in-sample fitting performance compared to that of 4F-DNSS, however, the same did not happen with the forecast. In fact, the 4F-DNSS/NAR-NN combination showed a consistent improvement of predictive performance (+20%) compared to the 5F-DRF/NAR-NN, whose predictions were characterized by larger errors across all the maturity spectrum.



**Figure 8.** Comparison of day-ahead forecasts in some worst cases (a,b). Average predictions (c) and average RMSFE (d) generated by the 4F-DNSS (blue), 5F-DRF (red), B-spline (green), and NAR-NNs (violet) models.

Overall, the reason for the weaker performances of both the factor models can be probably found in the volatility of the  $\beta$  observed along the forecasting period, since those models carry out a single approximation of the whole futures curve, which is therefore very challenging to manage and predict even with the aid of a flexible tool such as the neural network. On the contrary, with the B-spline/NAR-NN combination it is possible to capture the dynamics of futures curves, thanks to the local piecewise approximation.

To conclude, empirical evidence proved the effectiveness of the proposed framework for both modeling and predictive analysis, delivering, in the end, results which are very close to the true values, even under extreme conditions such as those affecting the NG futures market in the period 2021–2022. Among all the models, the B-spline model emerged as the best model for in-sample fitting. Furthermore, its joint use with NAR-NNs made this the best model also for out-of-sample day-ahead predictions within the NG futures market.

## 5. Concluding Remarks

In this study we addressed the problem of modeling and predicting futures prices term structure in the natural gas (NG) market. With this aim, we proposed a framework based on the use of interest rate models, given the similarities between the NG futures and fixed-income markets, and machine learning techniques as well. In particular, we used two models in the Nelson–Siegel family for fitting purposes, i.e., the four-factor dynamic Nelson–Siegel–Svensson (4F-DNSS) and the five-factor dynamic De Rezende–Ferreira (5F-DRF),

as well as B-spline, investigating their ability to replicate trends and dynamics of the NG futures market. Moreover, for the estimation procedure of the factor models we discussed a methodology based on time-varying parameters and fixed decay terms to ensure both high interpolating performances and parameter stability for the predictive process. Relative to the out-of-sample forecasting process, we conducted day-ahead predictions by means of Nonlinear Autoregressive Neural Networks (NAR-NNs) used in two ways: in the first case, we used them to predict model parameters' time series and then derive the prices, while in the second case we used NAR-NNs to directly predict prices' time series.

We provided empirical evidence of the ability of the suggested framework to achieve very satisfying results for both in-sample fitting and forecasting purposes within the NG futures market, even under extreme conditions, such as geopolitical turmoils and huge price jumps. The 4F-DNSS, 5F-DRF, and B-spline models demonstrated high levels of flexibility as well as adaptability to a wide variety of dynamics and trends characterizing the NG futures term structure, which resulted in very small magnitudes of the error metrics. Nevertheless, the B-spline model performed substantially better than the parametric models, especially in representing the medial and final parts of futures curves, properly replicating all the observed curve shapes, even the ones with multiple inflection points. Furthermore, the predictive performance clearly demonstrated the consistency of the implemented forecasting strategy. In this way, our results highlight that the hybrid B-spline/NAR-NN method is the preferable approach for day-ahead forecasting as it provides the lowest errors, outperforming both the 4F-DNSS/NAR-NN and 5F-DRF/NAR-NN combinations as well as the NAR-NN directly employed on the data.

The satisfying results obtained herein lay the groundwork for further experimental investigations oriented to the use of more sophisticated machine learning techniques, as well as for further research to enhance fitting and prediction of NG futures prices. In fact, all of these topics represent a part of our ongoing research.

**Author Contributions:** Conceptualization, O.C. and M.R.; methodology, O.C. and M.R.; software, O.C.; validation, M.R.; formal analysis, O.C. and M.R.; investigation, O.C. and M.R.; data curation, O.C. and M.R.; writing—original draft preparation, O.C. and M.R.; writing—review and editing, O.C. and M.R.; visualization, O.C. and M.R.; supervision, M.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** The authors are thankful to the anonymous referees for their kind comments and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

NG	Natural gas
TTF	Title Transfer Facility
WTI	West Texas Intermediate
NYMEX	New York Mercantile Exchange
4F-DNSS	Dynamic Nelson–Siegel–Svensson four-factor model
5F-DRF	Dynamic De Rezende–Ferreira five-factor Model
NAR-NN	Nonlinear Autoregressive Neural Network
LMBP	Levenberg–Marquardt back propagation learning algorithm



## References

1. Cheng, I.; Xiong, W. *The Financialization of Commodity Markets*; Working Paper 19642; National Bureau of Economic Research: Cambridge, MA, USA, 2014. [\[CrossRef\]](#)
2. Creti, A.; Nguyen, D. Energy markets financialization, risk spillovers, and pricing models. *Energy Policy* **2015**, *82*, 260–263. [\[CrossRef\]](#)
3. Futures Industry Association (FIA). *2021 Annual ETD Volume Review*; FIA: Washington, DC, USA, 2021.
4. Furtuna, O.; Grassi, A.; Ianiro, A.; Kallage, K.; Koci, R.; Lenoci, F.D.; Sowinski, A.; Vacirca, F. *Financial Stability Risks from Energy Derivatives Markets*; Techreport; European Central Bank: Frankfurt am Main, Germany, 2022.
5. Pascual, C.; Zambetakis, E. *The Geopolitics of Energy: From Security to Survival*; Technical Report 07; Brookings Institution: Washington, DC, USA, 2016.
6. Bordoff, J.; O’Sullivan, M. The Age of Energy Insecurity. How the Fight for Resources Is Upending Geopolitics. *Foreign Affairs* **2023**, *102*, 104–119.
7. Bartuska, V.; Lang, P.; Nosko, A. *The Geopolitics of Energy Security in Europe*; Carnegie Reports; Carnegie Europe: Brussels, Belgium, 2019.
8. Snam. *Global Gas Report 2020*; Snam: Milano, Italy, 2020.
9. Chen, B.; Xiong, R.; Li, H.; Sun, Q.; Yang, J. Pathways for sustainable energy transition. *J. Clean. Prod.* **2019**, *228*, 1564–1571. . [\[CrossRef\]](#)
10. Guidolin, M.; Alpcan, T. Transition to sustainable energy generation in Australia: Interplay between coal, gas and renewables. *Renew. Energy* **2019**, *139*, 359–367. [\[CrossRef\]](#)
11. Schoenberg, I. Contributions to the Problem of Approximation of Equidistant Data by Analytic Functions: Part A—On the Problem of Smoothing or Graduation. A First Class of Analytic Approximation Formulae. *Q. Appl. Math.* **1946**, *4*, 45–99. [\[CrossRef\]](#)
12. Svensson, L. *Estimating and Interpreting forward Interest Rates: Sweden 1992–1994*; NBER Working Papers 4871; National Bureau of Economic Research, Inc.: Cambridge, MA, USA, 1994. [\[CrossRef\]](#)
13. De Rezende, R.; Ferreira, M. *Modeling and Forecasting the Brazilian Term Structure of Interest Rates by an Extended Nelson-Siegel Class of Models: A Quantile Autoregression Approach*; Resreport; Escola Brasileira de Economia e Finanças: Rio de Janeiro, Brazil, 2008.
14. Lin, B. Fitting term structure of interest rates using B-splines: The case of Taiwanese Government bonds. *Appl. Financ. Econ.* **2002**, *12*, 57–75. [\[CrossRef\]](#)
15. Koopman, S.; Mallee, M.; Van der Wel, M. *Analyzing the Term Structure of Interest Rates Using the Dynamic Nelson-Siegel Model with Time-Varying Parameters*; Tinbergen Institute Discussion Papers 07-095/4; Tinbergen Institute: Amsterdam, The Netherlands, 2007.
16. De Pooter, M. *Examining the Nelson-Siegel Class of Term Structure Models—In-Sample Fit versus Out-of-Sample Forecasting Performance*; Resreport TI 2007-043/4; Tinbergen Institute: Amsterdam, The Netherlands, 2007.
17. Caldeira, J.; Moura, G.; Portugal, M. Efficient Yield Curve Estimation and Forecasting in Brazil. *Rev. Econ.* **2010**, *11*, 27–51.
18. De Rezende, R.; Ferreira, M. Modeling and Forecasting the Yield Curve by an Extended Nelson-Siegel Class of Models: A Quantile Autoregression Approach. *J. Forecast.* **2013**, *32*, 111–123. [\[CrossRef\]](#)
19. Muthoni, L.; Onyango, S.; Ongati, O. Extraction of Zero-Coupon Yield Curve for Nairobi Securities Exchange: Finding the Best Parametric Model for East African Securities Markets. *J. Math. Stat. Sci.* **2015**, *2015*, 51–74.
20. Chouikh, A.; Yousfi, R.; Chehibi, C. Yield Curve Estimation: An Empirical Evidence from the Tunisian Bond Market. *J. Financ. Econ.* **2017**, *5*, 300–309. [\[CrossRef\]](#)
21. Ullah, W.; Bari, K. The Term Structure of Government Bond Yields in an Emerging Market. *Rom. J. Econ. Forecast.* **2018**, *21*, 5–28.
22. Faria, A.; Almeida, C. A hybrid spline-based parametric model for the yield curve. *J. Econ. Dyn. Control* **2018**, *86*, 72–94. [\[CrossRef\]](#)
23. Nunes, M.; Gerding, E.; McGroarty, F.; Niranjana, M. A comparison of multitask and single task learning with artificial neural networks for yield curve forecasting. *Expert Syst. Appl.* **2019**, *119*, 362–375. [\[CrossRef\]](#)
24. Nagy, K. Term structure estimation with missing data: Application for emerging markets. *Q. Rev. Econ. Financ.* **2020**, *75*, 347–360. [\[CrossRef\]](#)
25. Mineo, E.; Alencar, A.; Moura, M.; Fabris, A. Forecasting the Term Structure of Interest Rates with Dynamic Constrained Smoothing B-Splines. *J. Risk Financ. Manag.* **2020**, *13*, 65. [\[CrossRef\]](#)
26. Guo, W.; Xue, H. Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models. *Math. Probl. Eng.* **2014**, *2014*, 857865. [\[CrossRef\]](#)
27. Barunik, J.; Malinská, B. Forecasting the term structure of crude oil futures prices with neural networks. *Appl. Energy* **2016**, *164*, 366–379. [\[CrossRef\]](#)
28. Ruiz, L.; Cuéllar, M.; Calvo-Flores, M.; Jiménez, M. An Application of Non-Linear Autoregressive Neural Networks to Predict Energy Consumption in Public Buildings. *Energies* **2016**, *9*, 684. [\[CrossRef\]](#)
29. Zhou, L.; Zhao, P.; Wu, D.; Cheng, C.; Huang, H. Time series model for forecasting the number of new admission inpatients. *BMC Med. Inform. Decis. Mak.* **2018**, *18*, 39. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Aliberti, A.; Pupillo, I.; Terna, S.; Macii, E.; Di Cataldo, S.; Patti, E.; Acquaviva, A. A Multi-Patient Data-Driven Approach to Blood Glucose Prediction. *IEEE Access* **2019**, *7*, 69311–69325. [\[CrossRef\]](#)

31. Benrhmach, G.; Namir, K.; Namir, A.; Bouyaghroumni, J. Nonlinear Autoregressive Neural Network and Extended Kalman Filters for Prediction of Financial Time Series. *J. Appl. Math.* **2020**, *2020*, 5057801. [[CrossRef](#)]
32. Di Franco, G.; Santurro, M. Machine learning, artificial neural networks and social research. *Qual. Quant.* **2020**, *55*, 1007–1025. [[CrossRef](#)]
33. Butler, S.; Kokoszka, P.; Miao, H.; Shang, H. Neural network prediction of crude oil futures using B-splines. *Energy Econ.* **2021**, *94*, 105080. [[CrossRef](#)]
34. Chi, Y. Time Series Forecasting of Global Price of Soybeans using a Hybrid SARIMA and NARNN Model: Time Series Forecasting of Global Price of Soybeans. *Data Sci. J. Comput. Appl. Inform.* **2021**, *5*, 85–101. [[CrossRef](#)]
35. Xu, X.; Zhang, Y. Corn cash price forecasting with neural networks. *Comput. Electron. Agric.* **2021**, *184*, 106120. [[CrossRef](#)]
36. Blanchard, T.; Samanta, B. Wind speed forecasting using neural networks. *Wind Eng.* **2020**, *44*, 33–48. [[CrossRef](#)]
37. Xu, X.; Zhang, Y. Soybean and Soybean Oil Price Forecasting through the Nonlinear Autoregressive Neural Network (NARNN) and NARNN with Exogenous Inputs (NARNN-X). *Intell. Syst. Appl.* **2022**, *13*, 200061. [[CrossRef](#)]
38. Emery, G.; Liu, Q. An analysis of the relationship between electricity and natural-gas futures prices. *J. Futures Mark.* **2002**, *22*, 95–122. [[CrossRef](#)]
39. Brown, S.; Yücel, M. What Drives Natural Gas Prices? *Energy J.* **2008**, *29*, 45–60. [[CrossRef](#)]
40. Brigida, M. The switching relationship between natural gas and crude oil prices. *Energy Econ.* **2014**, *43*, 48–55. [[CrossRef](#)]
41. Etienne, X.; Trujillo-Barrera, A.; Wiggins, S. Price and volatility transmissions between natural gas, fertilizer, and corn markets. *Agric. Financ. Rev.* **2016**, *76*, 151–171. [[CrossRef](#)]
42. Gatfaoui, H. Linking the gas and oil markets with the stock market: Investigating the U.S. relationship. *Energy Econ.* **2016**, *53*, 5–16. [[CrossRef](#)]
43. Zhang, Y.; Chevallerier, J.; Guesmi, K. “De-financialization” of commodities? Evidence from stock, crude oil and natural gas markets. *Energy Econ.* **2017**, *68*, 228–239. [[CrossRef](#)]
44. Li, H.; Chen, L.; Wang, D.; Zhang, H. Analysis of the Price Correlation between the International Natural Gas and Coal. *Energy Procedia* **2017**, *142*, 3141–3146. [[CrossRef](#)]
45. Ji, Q.; Zhang, H.; Geng, J. What drives natural gas prices in the United States?—A directed acyclic graph approach. *Energy Econ.* **2018**, *69*, 79–88. [[CrossRef](#)]
46. Behmiri, N.; Manera, M.; Nicolini, M. Understanding Dynamic Conditional Correlations between Oil, Natural Gas and Non-Energy Commodity Futures Markets. *Energy J.* **2019**, *40*, 55–76. [[CrossRef](#)]
47. Tiwari, A.; Mukherjee, Z.; Gupta, R.; Balcilar, M. A wavelet analysis of the relationship between oil and natural gas prices. *Resour. Policy* **2019**, *60*, 118–124. [[CrossRef](#)]
48. Suenaga, H.; Smith, A.; Williams, J. Volatility dynamics of NYMEX natural gas futures prices. *J. Futures Mark.* **2008**, *28*, 438–463. [[CrossRef](#)]
49. Lv, X.; Shan, X. Modeling natural gas market volatility using GARCH with different distributions. *Phys. A Stat. Mech. Appl.* **2013**, *392*, 5685–5699. [[CrossRef](#)]
50. van Goor, H.; Scholtens, B. Modeling natural gas price volatility: The case of the UK gas market. *Energy* **2014**, *72*, 126–134. [[CrossRef](#)]
51. Saltik, O.; Degirmen, S.; Ural, M. Volatility Modelling in Crude Oil and Natural Gas Prices. *Procedia Econ. Fin.* **2016**, *38*, 476–491. [[CrossRef](#)]
52. Fałdziński, M.; Fiszeder, P.; Orzeszko, W. Forecasting Volatility of Energy Commodities: Comparison of GARCH Models with Support Vector Regression. *Energies* **2021**, *14*, 6. [[CrossRef](#)]
53. Lu, F.; Ma, F.; Li, P.; Huang, D. Natural gas volatility predictability in a data-rich world. *Int. Rev. Financ. Anal.* **2022**, *83*, 102218. [[CrossRef](#)]
54. Liang, C.; Xia, Z.; Lai, X.; Wang, L. Natural gas volatility prediction: Fresh evidence from extreme weather and extended GARCH-MIDAS-ES model. *Energy Econ.* **2022**, *116*, 106437. [[CrossRef](#)]
55. Guo, K.; Liu, F.; Sun, X.; Zhang, D.; Ji, Q. Predicting natural gas futures’ volatility using climate risks. *Fin. Res. Lett.* **2023**, *55*, 103915. [[CrossRef](#)]
56. Szoplik, J. Forecasting of natural gas consumption with artificial neural networks. *Energy* **2015**, *85*, 208–220. [[CrossRef](#)]
57. Khan, M.A. Modelling and forecasting the demand for natural gas in Pakistan. *Renew. Sustain. Energy Rev.* **2015**, *49*, 1145–1159. [[CrossRef](#)]
58. Shaikh, F.; Ji, Q. Forecasting natural gas demand in China: Logistic modelling analysis. *Int. J. Electr. Power Energy Syst.* **2016**, *77*, 25–32. [[CrossRef](#)]
59. Panapakidis, I.; Dagoumas, A. Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. *Energy* **2017**, *118*, 231–245. [[CrossRef](#)]
60. Shaikh, F.; Ji, Q.; Shaikh, P.H.; Mirjat, N.H.; Uqaili, M.A. Forecasting China’s natural gas demand based on optimised nonlinear grey models. *Energy* **2017**, *140*, 941–951. [[CrossRef](#)]
61. Chen, Y.; Chua, W.; Koch, T. Forecasting day-ahead high-resolution natural-gas demand and supply in Germany. *Appl. Energy* **2018**, *228*, 1091–1110. [[CrossRef](#)]
62. Özmen, A.; Yılmaz, Y.; Weber, G. Natural gas consumption forecast with MARS and CMARS models for residential users. *Energy Econ.* **2018**, *70*, 357–381. [[CrossRef](#)]
63. Hribar, R.; Potočník, P.; Šilc, J.; Papa, G. A comparison of models for forecasting the residential natural gas demand of an urban area. *Energy* **2019**, *167*, 511–522. [[CrossRef](#)]

64. Su, H.; Zio, E.; Zhang, J.; Xu, M.; Li, X.; Zhang, Z. A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model. *Energy* **2019**, *178*, 585–597. [[CrossRef](#)]
65. Mu, X. Weather, storage, and natural gas price dynamics: Fundamentals and volatility. *Energy Econ.* **2007**, *29*, 46–63. [[CrossRef](#)]
66. Panella, M.; Barcellona, F.; D'Ecclesia, R. Forecasting Energy Commodity Prices Using Neural Networks. *Adv. Decis. Sci.* **2012**, *2012*, 289810. [[CrossRef](#)]
67. Salehnia, N.; Falahi, M.; Seifi, A.; Mahdavi Adeli, M. Forecasting natural gas spot prices with nonlinear modeling using Gamma test analysis. *J. Nat. Gas Sci. Eng.* **2013**, *14*, 238–249. [[CrossRef](#)]
68. Mason, C.; Wilmot, N.A. Jump processes in natural gas markets. *Energy Econ.* **2014**, *46*, S69–S79. [[CrossRef](#)]
69. Jin, J.; Kim, J. Forecasting Natural Gas Prices Using Wavelets, Time Series, and Artificial Neural Networks. *PLoS ONE* **2015**, *10*, e0142064. [[CrossRef](#)]
70. Čeperić, E.; Žiković, S.; Čeperić, V. Short-term forecasting of natural gas prices using machine learning and feature selection algorithms. *Energy* **2017**, *140*, 893–900. [[CrossRef](#)]
71. Dolatabadi, S.; Narayan, P.; Nielsen, M.; Xu, K. Economic significance of commodity return forecasts from the fractionally cointegrated VAR model. *J. Futures Mark.* **2018**, *38*, 219–242. [[CrossRef](#)]
72. Berrisch, J.; Ziel, F. Distributional modeling and forecasting of natural gas prices. *J. Forecast.* **2022**, *41*, 1065–1086. [[CrossRef](#)]
73. Kwas, M.; Rubaszek, M. Forecasting Commodity Prices: Looking for a Benchmark. *Forecasting* **2021**, *3*, 447–459. [[CrossRef](#)]
74. Li, J.; Wu, Q.; Tian, Y.; Fan, L. Monthly Henry Hub natural gas spot prices forecasting using variational mode decomposition and deep belief network. *Energy* **2021**, *227*, 120478. [[CrossRef](#)]
75. Wang, J.; Cao, J.; Yuan, S.; Cheng, M. Short-term forecasting of natural gas prices by using a novel hybrid method based on a combination of the CEEMDAN-SE-and the PSO-ALS-optimized GRU network. *Energy* **2021**, *233*, 121082. [[CrossRef](#)]
76. Pei, Y.; Huang, C.; Shen, Y.; Wang, M. A Novel Model for Spot Price Forecast of Natural Gas Based on Temporal Convolutional Network. *Energies* **2023**, *16*, 2321. [[CrossRef](#)]
77. Borovkova, S.; Mahakena, D. News, volatility and jumps: The case of natural gas futures. *Quant. Financ.* **2015**, *15*, 1217–1242. [[CrossRef](#)]
78. Jana, R.; Ghosh, I. A residual driven ensemble machine learning approach for forecasting natural gas prices: Analyses for pre-and during-COVID-19 phases. *Ann. Oper. Res.* **2022**. [[CrossRef](#)]
79. Li, R.; Song, X. A multi-scale model with feature recognition for the use of energy futures price forecasting. *Expert Syst. Appl.* **2023**, *211*, 118622. [[CrossRef](#)]
80. Chiarella, C.; Clewlow, L.; Kang, B. *Modelling and Estimating the Forward Price Curve in the Energy Market*; Technical Report 260; University of Technology Sydney: Sydney, Australia, 2009.
81. Almansour, A. Convenience yield in commodity price modeling: A regime switching approach. *Energy Econ.* **2016**, *53*, 238–247. [[CrossRef](#)]
82. Gibson, R.; Schwartz, E. Stochastic Convenience Yield and the Pricing of Oil Contingent Claims. *J. Financ.* **1990**, *45*, 959–976. [[CrossRef](#)]
83. Leonhardt, D.; Ware, A.; Zagst, R. A Cointegrated Regime-Switching Model Approach with Jumps Applied to Natural Gas Futures Prices. *Risks* **2017**, *5*, 48. [[CrossRef](#)]
84. Karstanje, D.; van der Wel, M.; van Dijk, D. Common Factors in Commodity Futures Curves. January, 2023. Available online: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2558014](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2558014) (accessed on 14 May 2023).
85. Diebold, F.; Li, C. Forecasting the term structure of government bond yields. *J. Econom.* **2006**, *130*, 337–364. [[CrossRef](#)]
86. Jablonowski, C.; Schicks, M. A Three-Factor Model on the Natural Gas Forward Curve Including Temperature Forecasts. *J. Energy Mark.* **2017**, *10*, 87–105. [[CrossRef](#)]
87. Heath, D.; Jarrow, R.; Morton, A. Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica* **1992**, *60*, 77–105. [[CrossRef](#)]
88. Tang, Y.; Wang, Q.; Xu, W.; Wang, M.; Wang, Z. Natural Gas Price Prediction with Big Data. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 5326–5330. [[CrossRef](#)]
89. Li, B. Pricing dynamics of natural gas futures. *Energy Econ.* **2019**, *78*, 91–108. [[CrossRef](#)]
90. Horváth, L.; Liu, Z.; Rice, G.; Wang, S. A functional time series analysis of forward curves derived from commodity futures. *Int. J. Forecast.* **2020**, *36*, 646–665. [[CrossRef](#)]
91. Heather, P. *European Traded Gas Hubs: German Hubs about to Merge*; Techreport; The Oxford Institute for Energy Studies: Oxford, UK, 2021.
92. Intercontinental Exchange. *ICE Announces Record Activity in TTF and JKM Gas Complexes as They Evolve into Global Benchmarks*; Techreport; Intercontinental Exchange (ICE): Atlanta, GA, USA, 2021.
93. Copernicus Climate Change Service (C3S). *Summer 2022 Europe's Hottest on Record*; Copernicus Climate Change Service (C3S): Brussels, Belgium, 2022.
94. Gas Infrastructure Europe (GIE). *Gas Storage Inventory*; Gas Infrastructure Europe (GIE): Brussels, Belgium, 2022.
95. Bessembinder, H.; Coughenour, J.; Seguin, P.; Smoller, M. Is There a Term Structure of Futures Volatilities? Reevaluating the Samuelson Hypothesis. *J. Deriv.* **1996**, *4*, 45–58. [[CrossRef](#)]
96. Samuelson, P.A. Proof That Properly Discounted Present Values of Assets Vibrate Randomly. *Bell J. Econ. Manag. Sci.* **1973**, *4*, 369–374. [[CrossRef](#)]

97. Duong, H.; Kalev, P. The Samuelson hypothesis in futures markets: An analysis using intraday data. *J. Bank. Financ.* **2008**, *32*, 489–500. [[CrossRef](#)]
98. Jaeck, E.; Lautier, D. Samuelson hypothesis and electricity derivative markets. In Proceedings of the 31st International French Finance Association Conference, AFFI 2014, Aix-en-Provence, France, 20–21 May 2014; p. 24.
99. Jonckheere, A. A Distribution-Free k-Sample Test Against Ordered Alternatives. *Biometrika* **1954**, *41*, 133–145. [[CrossRef](#)]
100. Terpstra, T. The asymptotic normality and consistency of kendall’s test against trend, when ties are present in one ranking. *Indag. Math.* **1952**, *55*, 327–333. [[CrossRef](#)]
101. Nelson, C.; Siegel, A. Parsimonious Modeling of Yield Curves. *J. Bus.* **1987**, *60*, 473–489. [[CrossRef](#)]
102. Castello, O.; Resta, M. Modeling the Yield Curve of BRICS Countries: Parametric vs. Machine Learning Techniques. *Risks* **2022**, *10*, 36. [[CrossRef](#)]
103. Curry, H.; Schoenberg, I. On spline distributions and their limits: The Polya Distribution Functions. *Bull. Am. Math. Soc.* **1947**, *4*, 109.
104. Curry, H.; Schoenberg, I. On Pólya frequency functions IV: The fundamental spline functions and their limits. *J. D’Anal. Math.* **1966**, *17*, 71–107. [[CrossRef](#)]
105. Reed, R.; Marks, R. *Neural Smoothing: Supervised Learning in Feedforward Artificial Neural Networks*; The MIT Press: Cambridge, CA, USA, 1999.
106. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [[CrossRef](#)]
107. Rosenblatt, F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychol. Rev.* **1958**, *65*, 386–408. [[CrossRef](#)]
108. Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **1944**, *2*, 164–168. [[CrossRef](#)]
109. Marquardt, D.W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441. [[CrossRef](#)]
110. Theil, H. *Economic Forecasts and Policy*; Technical Report; Netherland School of Economics: Rotterdam, The Netherlands, 1958.
111. Bliemel, F. Theil’s Forecast Accuracy Coefficient: A Clarification. *J. Mark. Res.* **1973**, *10*, 444–446. [[CrossRef](#)]
112. Granger, C.; Newbold, P. Some comments on the evaluation of economic forecasts. *Appl. Econ.* **1973**, *5*, 35–47. [[CrossRef](#)]
113. Harvey, D.; Leybourne, S.; Newbold, P. Testing the equality of prediction mean squared errors. *Int. J. Forecast.* **1997**, *13*, 281–291. . [[CrossRef](#)]
114. Hansen, P.; Lunde, A.; Nason, J. The model confidence set. *Econometrica* **2011**, *79*, 453–497. [[CrossRef](#)]
115. Hansen, P.; Lunde, A. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *J. Appl. Econom.* **2005**, *20*, 873–889. [[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.